ENAC - Faculté Environnement naturel architectural et construit
INTER- Institut de l'urbain et des territoires
BPE – Bioénergie et Planification énergétique

EPFL ENAC GR-GN, Bât. GC A3
Station 18
CH - 1015 LAUSANNE
tél: 021 - 693 06 27 fax: 021 - 693 28 63
E-mail: edgard.gnansounou@epfl.ch http://bpe.epfl.ch/

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Energy Doctoral Programme

Energy Planning

Modeling, simulation and interpretation

Selected papers

Edgard Gnansounou

Introduction to Energy Economics and Modeling

APPLIED ECONOMIC- ENVIRONMENT- ENERGY MODELING FOR QUANTITATIVE IMPACT ASSESSMENT

Claudia KEMFERT*
In: Amelung, Rotmans, Valkering: Integrated Assessment for Policy Modelling, 2003

Abstract

Applied economic- environment- energy modeling is becoming increasingly important in assessing diverse policy strategy quantitative impacts and in anticipating and simulating potential future economic performances under environmental and energy constraints. The aim of this paper is to shed light on different modeling approaches, compare methodologies and theories and classify existing modeling approaches. The creation of an economic- environment- energy modeling approach is primarily influenced by the purpose of modeling and merit of explanation. The quality of a model can be accurately evaluated only when both targets are explicated and unambiguously reached.
Key words: Economic- environment- energy modeling
JEL- classification: C0, O3, Q0

Content

* Claudia Kemfert is head of section ”SPEED” (Scientific Pool of Environmental Economic Disciplines) at the Dept. of Economics , University of Oldenburg. Address for correspondence: Oldenburg University; SPEED, D-26111 Oldenburg; E- mail: kemfert@uni-oldenburg.de; Tel: +49 441 798 4106, Fax: +49 441 798 4101

1 Introduction

Due to the complexity and solidity of economic interactions, modelers try to portray a simplified picture of reality: they create a model using methods of reduced abstraction through an improved description of sincerity. The quality of a model is predominantly appraised through distinguished explanation value. The demand for economic modeling is derived from three central economic targets. The *explanation target* is defined as the aim of economists to understand and illuminate economic interrelations such as economic consequences of individual behavior, direct and feedback effects of key economic interactions, and allocation and distribution of scarce goods and resources. Through knowledge of verified economic interrelations and connections the *prediction target*'s goal is anticipating future economic performances. Furthermore, in assessing economic impacts of alternative policy strategies and games or distinct institutional implementations the *simulation target* recently revealed itself as the most prominent factor in economic modeling.

Concern about fossil fuel resource depletion has led to an increase in applied pure economic models with a detailed representation of the energy market. Pioneering energy – economic modeling efforts focused primarily on the representation of scarce resources like oil and its impact on world economies. More recently, not only the scarcity of energy resources, but also other natural resources and the environment and climate played a major role in economic-environment- energy modeling. The complexity of models has increased considerably, notably regarding global environmental issues such as acid rain, ozone depletion and climate change. Climatic changes induced by human activities (e.g. increased greenhouse gas emissions) are largely caused by energy related activities such as fossil fuel consumption. One important source of climate change is anthropogenic greenhouse gas emissions. Increasing atmospheric concentrations of greenhouse gases have a substantial impact on global temperature and sea level which creates extensive economic, ecological and climatic impacts. The threat of climate change makes designing economic development strategies and energy and environmental policies increasingly important. Recent developments in sophisticated economic- environment-energy modeling intend to integrate climate, ecosystem and economic impacts within a so-called Integrated Assessment Modeling approach (IAM).[1]

Existing literature focuses mainly on a comparison of modeling results, see Weyant (1999), Bosello et. al (1998), Springer (2003) and Hourcade et al. (2001). Grubb (1993) and Hourcade et. al. (1996) summarize a representation of some modeling approaches and classifications.

The aim of this paper is providing an overview of different approaches of economic-environment-energy modeling, as well as their methodologies, theories and general interrelations. The first part of the paper explains the theories and general methodologies of economic models. It summarizes the different approaches of economic-environment-energy modeling, differentiates their purposes, scales and mathematical approaches and describes corresponding data requirements. Furthermore, the substitutability and the role of endogenisation as major driving factors in economic-environment-energy modeling is briefly discussed; some examples of these modeling approaches are subsequently compared. The last chapter concludes.

2 Methodologies- Some General Remarks

The procedure for gaining scientific findings is determined by the principal methodology and concepts. Scientific theory necessitates methodologies analyzing techniques and contemplation

[1] Edmonds (1998) gives an overview of the newest modeling approaches. Previous overviews can be found in Dowlatabadi (1993), Dowlatabadi, and Rotmans (1998) and Toth (1995).

modes. The concept of equilibria, static or dynamic investigation, and partial and total analyses remain the most important base perceptions of modeling.

2.1 Stability of Equilibria

The determination of an equilibrium is crucial in economic modeling. However, the concept of finding equilibria is not standardized in scientific theories. The methodological term of an equilibrium is more often used in natural sciences as a continual steady state over time. Therefore, an economic system is situated in an equilibrium if all endogenous variables do not change over time and all other exogenous variables are constant. Because of the universal character of the above equilibrium determination, this concept can be applied to different economic areas such as market, balance of payment or growth equilibrium. The growth equilibrium can also be classified as dynamic equilibrium because all equilibrium values change over time. Nevertheless, a dynamic equilibrium is a temporary steady state because the rate of changes are stable.

The equilibrium concept in economic theory is usually meant as market equilibrium or market clearance by the balance of supply and demand. An equilibrium is *ex definitione* stable over time, whereas disequilibria are only temporarily constant. Therefore, economies tend towards a equalized state (which acts as reference position).

A stable equilibrium is characterized by an obligatory return to the base equilibrium after external shocks. If this cannot be reached, the equilibrium is indifferent or unstable if the reaction induces an equilibrium position remote from the reference and base equilibrium situation. In economic theory only stable equilibria are considered. All equilibria analysis need to incorporate a stability investigation to determine the conditions of a stable equilibrium.

2.2 Static, Comparative Static and Dynamic

In economic modeling, time can be treated as a constant (static), a parameter (comparative static) or a variable (dynamic). In static analysis, all variables are related to the same time reference location. An example for a static investigation is determining a market clearance price by the balance of supply and demand at a specific time position. The comparative static analysis investigates and contrasts the variables of different time scales. A comparative static analysis does not explain adjustment processes like direct economic reactions and feedback effects . The dynamic analysis does include these effects; the adaptation processes can be simulated and evaluated. The stability conditions determine the convergence towards an equilibrium.

2.3 Partial Versus Total Analyses

In order to separate particular economic relations and reactions, not all markets and interrelations are modeled simultaneously. Instead, specific sections of the economy are detached and reflected. The *total analysis* investigates all economic reactions and interrelations whereas the research of economy sections is named *partial analysis*. Sensitivity analyses use mostly the so-called *ceteris paribus* stipulation: in order to assess the impacts of one or several influence factors, only one is modified whereas all others remain fixed. Both partial or total analyses use the *ceteris paribus* clause to investigate the sensitivity of a model through assumption variations.

3 Economic Theories

Economists usually distinguish between two major economic theories: *neo classical* and the *neo Keynesian* theory. The neo classical economic theory covers the microeconomic decisions of individuals and investigates the distribution and allocation of scarce resources towards alternative utility purposes in order to reach market clearance situations. Consumers maximize their utility under budget constraints, firms maximize their profits under costs constraints. Optimally, the marginal utility and product of input factors is equal to its relative price. Substitution processes are induced by an change in relative price. Market clearance of all markets is reached by the adjustment of market prices, known as general equilibrium theory. Therefore, the general equilibrium microeconomic theory is primarily concerned with the allocation and distribution of scarce goods and resources against the background of alternative usability potentials. From a more macroeconomic perspective, the neoclassical model covers four markets which ought to be cleared: goods, labor, capital and money. Both labor supply and demand are influenced by real wages; full employment is determined by the balance of supply and demand. Capital demand is revealed by the investment decisions of firms, supply by the savings of households. The market clearance price on the capital market is represented by the market interest rate. The equilibrium of money supply and demand is reached by the market clearance price in which money supply is assumed to be exogenous. The general equilibrium is always reached on all markets without any policy intervention due to the so-called “invisible hand of the markets” and Says´ law stating that each supply creates the demand and leads necessarily to the equalization of demand and supply.
The *Neo- Keynesianian* macroeconomic theory also includes the equilibrium concept to determine the market clearance of all markets. However, on the labor market this equilibrium does not necessarily cause full employment. In contrast to the neoclassical theory, nominal wages, not real wages determine labor supply which are not as flexible as assumed; they are rigid and on the down scale. Due to this, a market equilibrium where unemployment occurs is feasible. Furthermore, investment decisions are not only determined by the interest rate but also by expectations and uncertainties. Private savings are not only influenced by the market interest rate but also by real income changes. Therefore, consumption augments with real income growth. The main economic driving factor is the effective demand which increases the real GNP to a large extent. For that reason, the Says´ law is reversed and the market cannot clear itself autonomously. It can be said that neoclassical economic theory focuses on allocation whereas the neo Keynesian economic theory draws its attention principally to the employment theme to investigate the facility degree of non employed production factors. Further developments of economic theories include ideas of both concepts, like the Monetarian economic theory which principally intends to explain the driving factors of inflation or the theory of rational expectations which concentrates on the employment and inflation processes.
Economic modeling approaches can be based on neoclassical theories (e.g. general equilibrium models), neo Keynesian theoretic approaches (e.g. most of the econometric modeling efforts) or a combination of both. More frequently, neoclassical modeling methods start from the benchmark of optimizing individual behavior in perfectly competitive markets. Optimal growth models under conditions of certainty and under the assumption of infinite consumers as homogenous agents, along with competitive markets and constant returns to scale in production typically imply that the allocation of resources achieved by a decentralized economy will be the same as that chosen by a central planer maximizing the utility of the representative economic agent. This modeling concept is known as the Ramsey infinite horizon optimization model which derives the intertemporal conditions needed to satisfy the optimal growth path chosen by a central planner. Applied general equilibrium models include microeconomic and individual based optimization behavior and a macroeconomic optimal growth Ramsey agent to determine

intertemporal equilibrium conditions. In contrast to the infinite horizon maximizers of the Ramsey model, overlapping generation models incorporate individuals of different generations. These various generations can trade with one another: each generation trades with different generations in distinct periods of time, and generations not yet born with unknown preferences and unregistered in current market transactions exist as well. The model provides an example of an economy where competitive equilibrium is not necessarily the same that would have been chosen by a central planner. Therefore, it is not automatically optimally Pareto. Life cycle savers may overaccumulate capital, leading to a situation in which everyone could be better off by consuming part of the capital stock.

Most simulation models start with the neoclassical economic theory benchmark being enlarged by assumptions of incomplete markets and imperfect competition or non-neoclassical constructs as bounded rationality and interdependent utility functions in order to understand important aspects of financial and labor markets. Applied general equilibrium models can also be used to investigate impacts of negotiation games. General equilibrium models are suited for game theoretic investigations due to equivalent phenomena of the correspondence of the Walras' equilibrium and the core of expected negotiation results. More generally, game theory studies decision processes of several competitive individuals intending to reach the same target. It focuses therefore on conflict and cooperation solutions. The game theory investigates strategic actions, coalition formation and market power of individuals and coalitions by the determination of stable, fair, optimal or equitable distribution of utility. A Nash equilibrium determines a set of strategies of all players which is stable and cannot be improved by payoffs.

Economic modeling concepts are based on different theoretical concepts for the purpose of explanation, prediction or simulation within a time, geographic or sectoral scale. The following chapter explains the different purposes of modeling and their approaches.

4 Economic – Environment- Energy Modeling

To avoid significant misinterpretations of modeling results by incorrect applications, models must be classified and evaluated against the background of the purpose they were designed for. Pure scientific objectives may focus on either an improvement in real economic theory descriptions or on a model application evaluating or simulating economic market reactions and assessing policy strategies. If the main purpose of a model creation is to use it as a decision support tool, two main issues need to be clarified: a. which policy is to be investigated: environmental, energy or climate policy and which geographical and time scale should be covered? b. Is the main target the *evaluation* or *optimization* of policy strategies?

For example, detailed information regarding the costs and benefits of regional and local environmental programs must be studied by highly disaggregated cost-benefit environmental assessment models. In contrast, impacts of global climate change policy strategies require an investigation by aggregated worldwide economic simulation models covering the main interlinkages of economic actors. Moreover, the time horizon plays a crucial role through which a specific policy option is explored. The purpose, time and geographical scale determine the nature of each economic modeling instrument.

4.1 Purpose of Modeling

The purpose of modeling is principally determined by the intention to *forecast* and anticipate future economic performances or to *simulate* economic reactions of scenario analysis. Hourcade et al. add a third purpose of modeling: the "backcasting" target that attempts to look back from the future to the present in order to construct visions of desired futures by interviewing experts. Forecasting models must be based on historical time data to extrapolate potential future trends.

Therefore, they are usually applied only for analyzing short to medium term impacts of action. This approach requires an endogenous representation of economic behavior and general growth patterns and is principally found in short-term, econometrically driven economic models. Simulation models are widely used to study "if-then" questions by scenario analysis in which a specific number of policy or intervention scenarios is compared against a so-called "business as usual" reference scenario. Depending on the time scale, simulation purposes principally aim at an exploration of the future by scenario analysis. Because of that, main economic driving factors must be given exogenously so that general assumptions must be made about population growth, economic behavior, physical resources, substitution options and ordinary technical progress. Specific purposes arise from the area of interests. Energy models could focus on concrete notions of energy demand, supply, specific impacts or evaluation. Environmental and climate models could also focus on specific areas or on the overall phenomenon in general. More frequently, modelers integrate subjects of different disciplines as a means to combine several specific purposes. In order to investigate economic-environment-energy interactions, an integrative modeling approach is indispensable.

4.2 Time Dimension and Geographical Scale

Both time horizon and the geographical scale within applied modelling concepts crucially depend on investigation purpose. If long term global climatic impacts are to be evaluated and simulated, the time horizon of the model needs to cover at least the next 50 years. Grubb et al. (1993) classify time scales into a short term period of five years or less, a medium term period covering between three and fifteen years and a long term time period of ten years or more. Economic forecast purposes are based primarily on an extrapolation of past reactions into the short to medium future time period.
The geographical coverage reflects the level at which the analysis takes place. Global models comprise international regions and nations and simulate economic relations and reactions on a highly aggregated level. Regional analysis focus primarily on a specific region like Europe or Asia, while local models replicate specific interrelations of market actors and systems within a region of a country. National models cover only one country and represent economic behaviour within this country. Depending on the size and market power of respective countries, national models consider the impact of trade relations on other countries in a different manner. Usually, small countries cannot influence the market price for goods or energy services. Larger countries however could have this potential and shape market feedbacks and prices by changing trade reactions. Applied models may concentrate their analysis primarily on one sector within an economy or could cover all relevant sectors (multi sector models). Single-sector models only provide information regarding a particular sector and do not include economic relations and interlinkages with other economic sectors. Multi-sector models focus primarily on a national level but can also be applied to regional or international investigations.

4.3 Data Requirements

Model results crucially depend on data quality. Forecasting models must be based on long term time series data, the longer and disaggregated the available data, the more precise and detailed future predictions can be. Data reliability plays a fundamental role. Some countries do not provide the necessary data or otherwise use different statistical accounting methods that are not comparable to other techniques and data. On a global level, highly aggregated time series data for the main economic sizes and driving factors are obligatory. However, no common database exists for a consistent and comprehensive collection of time series data for all world regions. Recent developments and cooperation of global economy and trade modelers established a common data base primarily covering the major input-output tables of world regions and their bilateral trade flows.[2] This data is based mainly on the input-output tables provided by the statistical offices of each nation and collects data from UN and OECD statistics. On a very detailed level of a country or region, efforts have already been undertaken to harvest and establish common databases; depending on the modeling purpose, some data still must be imposed.

5 Different Modeling Approaches

A very common and accepted classification of energy – economy models is the distinction between top-down and bottom-up approaches. Top-down modeling approaches tend to cover economic relations from a more (as the name implies) "top down" economic perspective whereas bottom-up modeling approaches focus more on the description of individual energy systems and technologies.
According to Hourcade et.al. (1996) differences in outcomes of both types of models stem from the distinct treatment of how technology is adopted. The following table illustrates decision making behavior of economic agents and how markets and economic institutions actually operate over a given period of time.

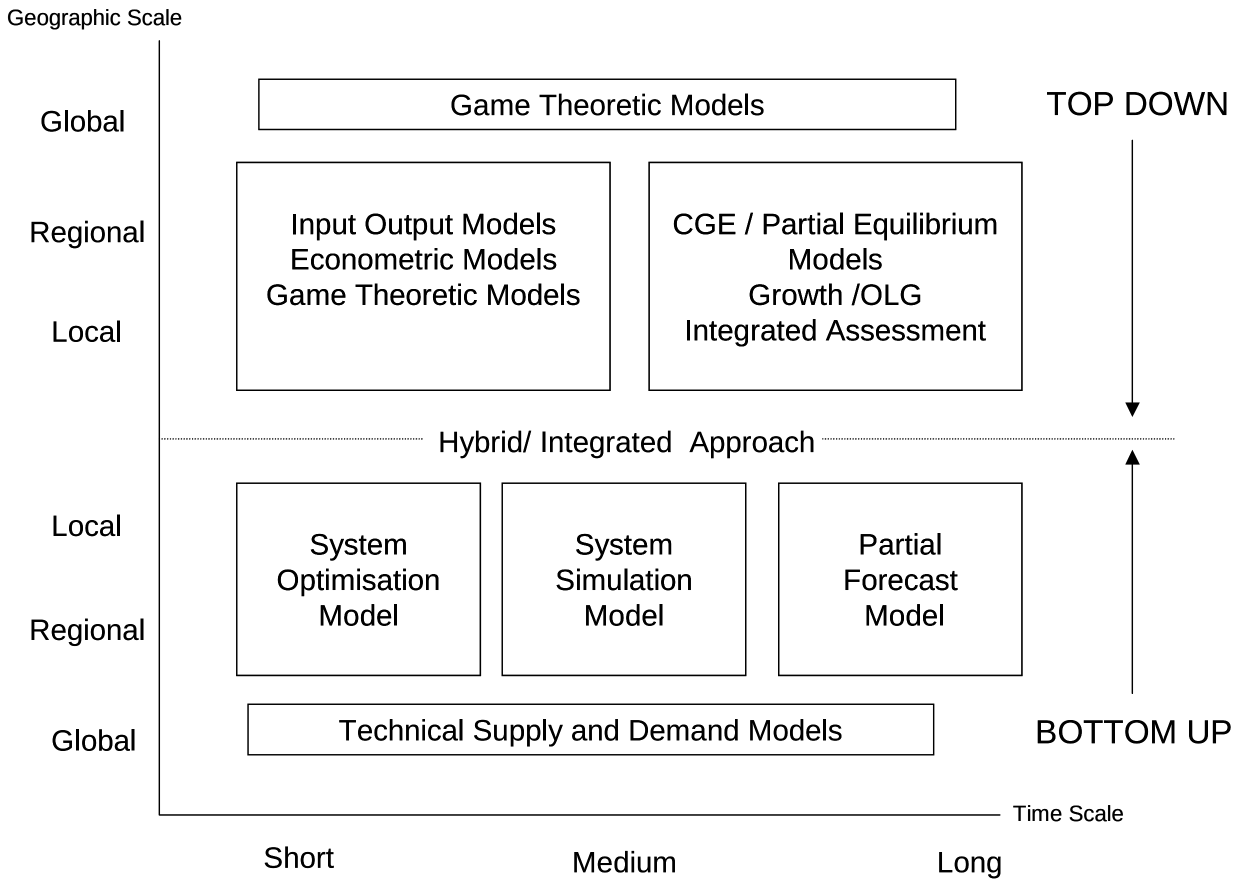


Figure 1: Classification of Model Types

Grubb et al. (1993) refers to the very different paradigms both approaches follow, the top down approach is associated with the so called "pessimistic" economic paradigm, while the bottom up approach is associated with "optimistic" engineering paradigm.

[2] See GTAP , Hertel (1998)

5.1 Bottom-Up Models

Bottom-up models provide detailed information about the techniques, structural effects and sectoral behavior of an economy. These partial equilibrium economic models describe the economy from the bottom side and emphasize the “green” potential of available technologies. They can be classified into technical supply and demand models, linear and non-linear system optimization models, or system simulation models and partial forecast models. As an outcome, bottom-up models determine marginal abatement costs and so called “no regret” possibilities, i.e. energy saving potentials at a low cost. According to Hourcade et.al. (1996), bottom-up models can be classified into descriptive and prescriptive models. Descriptive models attempt to provide practical estimates of the technology mix resulting from actual decisions based on assumptions such as complex preferences, intangible costs, capital constraints, risk treatment, uncertainty and market barriers. Alternatively, prescriptive models present an estimate of technological potentials by evaluating the most efficient existing technologies found through cost minimization. Consequently, prescriptive studies provide a more optimistic estimation of technological potentials when compared to descriptive studies.

The limitations of bottom-up modeling approaches are that they generally neglect feedback effects to and from the economy as well as the rebound effects on international energy markets. Moreover, they do not account for uncertainties related to many environmental and climate phenomena. The diffusion process of new technologies and market thresholds such as hidden costs and market constraints tend to be neglected, resulting in an overestimation of efficiency improvement potentials and environmental policy effects. Furthermore, bottom-up models hardly assess the costs of reducing greenhouse gases on a global scale.[3]

5.2 Top-Down Models

Top-down modeling approaches focus primarily on the economic description of interactions and relations and are classified within this context into IO, macroeconometric and integrated assessment models. Input Output (IO) models are largely applied to investigate direct and indirect economic and sectoral effects of demand-driven policies. Based on national input output tables, dynamic IO model, in contrast to static IO models, determine intertemporal and dynamic effects of investment changes. In order to account for intrasectoral substitution opportunities of factor inputs, the traditional limited Leontief production functions must be replaced by Constant Elasticity of Substitution (CES) production functions that ideally also cover energy as a further input factor. Technological progress is taken into account by an autonomous energy efficiency improvement factor (AEEI). International trade relations can be simulated by including flexible exchange rates. The major limits of IO models are that they can only be used on a national level; in the original format IO models do not cover substitution and feedback effects and neglect intersectoral substitution effects. National IO models are widely used to estimate direct and indirect sectoral effects of diverse policy strategies, primarily in assessing potential employment effects.

Most macro econometric models applied for economic-environment-energy studies are based on the above described neo Keynesian theory. Because macro econometric modeling approaches are based on national input-output data tables and econometrically estimated sets of equations, they are widely used for national, short-term economic forecasting. Since most of the macro econometric modeling approaches are based on the neo Keynesian theory, they allow for a

[3] Conversely, energy system models are widely used to assess global costs of GHG mitigation, See Kram (1998), Criqui, Klaassen et al. (2000) and Capros and Mantzos (2000).

determination of equilibria and structural unemployment short term. Because of data constraints, macro econometric models are mainly applied for national policy evaluation and economic forecast issues rather as long term and global future prediction tools. Macro econometric models were initially developed as pure economic models; as energy and environmental concerns become more serious, they are adapted to investigate these issues by introducing energy as a traditional production factor and by representing an energy market and interaction between both components. In most cases, macro econometric models must be based on long-run time series data which are not always available for specific environmental goods and services, nor for broad trade flows on a global scale.

Computational general equilibrium Models (CGE) are primarily based on the above described neoclassical theoretical background of equalizing supply and demand in all markets by means of market clearance prices. Profit maximization under perfect competition and free market entrance guarantee zero profits and optimal allocation and distribution of resources. The dynamics of CGE is produced by capital accumulation and/or by the exogenous growth of production factors and productivity. Recursive dynamic models determine temporary equilibria under myopic expectations, intertemporal CGEs consider capital and investment changes over time. In order to guarantee the consistency of all exogenous given parameter, CGEs are frequently benchmarked on a given base year. CGE models produce (as a response to external shocks) a general equilibrium of all markets according to economic behavior of individual agents. Mainly CGEs use non-linear substitution-based production and utility functions of the CES type to describe production and utility behavior. Energy can be considered a further input production factor; to describe technical progress, autonomous energy efficiency factors are considered. Recent model improvements include an endogenous representation of technological progress. The main shortcomings of CGE are their reliance on an assumption of perfect market equilibrium; they do not allow for structural unemployment in the long run. Recent model development take this into account by allowing imperfect markets and incomplete information. CGEs are simulation tools that assess “if-then” questions in a medium to long term time horizon and can hardly be applied for short and long term prediction tasks.

In order to take into account not only economic but environmental, energy and climatic impacts, existing economic modeling approaches must be integrated into ecological, ecosystem, and climatic models. The economic assessment of climate change is based on pure economic models focusing on economic relations and interlinkages, economic models enlarged by stylized climatic interrelations, or submodels usually known as integrated assessment (IAM) models. Costs and benefits of climate change are predominantly assessed by integrated assessment models (IAM) incorporating physical relations of climate change and the economic effects of damage functions. Integrated assessment models are characterized by combining multidisciplinary approaches to thoroughly evaluate climate change impacts.

IAMs are primarily applied in global economic and climate simulations, should incorporate uncertainty and risk analysis, social and economic organization of developing economies and an endogenous representation of technological change.[4] Weyant et al. (1996) distinguish two primary IAMs: policy optimization and policy evaluation models. Policy optimization models optimise key policy control variables whereas policy evaluation models assess the impacts of specific policies. Policy optimization models can be classified into a. cost–benefit models evaluating climate policies, b. target-based models optimizing reactions under exogenous climate targets, c. uncertainty-based models dealing with decision making processes under uncertainty. Policy evaluation models can be separated into two types: a. deterministic projection models in which each input and output takes on a single value, b. stochastic projection models treating some input and outputs stochastically.

[4] See Bosello et al. (1998).

Hybrid modelling approaches combine bottom up modeling tools with top down models by either a so called "soft link" of an integrated and internal modeling approach or a "hard link" that externally exchange key parameters and variables of both models.[5] Most applied general equilibrium models and overlapping generation models use as input data a national input-output table or social accounting matrix (SAM). Econometric input-output models apply solution methods of static or dynamic IO analysis but merge it with econometrically estimated parameters and equations.[6]

5.3 Mathematical Approach and Solution Methods

Linear and non linear optimization models maximize or minimize target functions under specific side constraints often formulated under disequilibria conditions. Most general equilibrium modeling systems and integrated assessment models are formulated by an optimization approach where the optimization conditions are equalized to market clearance conditions. Non integratable general equilibrium problems can be formulated as a sequence of mathematical optimization problems and solved by an interative process. The method of sequencial joint maximization finds an equilibrium through determination of Negishi welfare weights. Computational modeling approaches benefit from modern algorithms techniques that improve and accelerate solution procedures. Whereas linear optimization models are solved mostly by general Simplex algorithms, non linear optimization models are principally solved by a general version of the Newton algorithms or variants of them.[7]

In order to resolve a system of generalized equations, computational general equilibrium models (CGE) use an equalization format (Walras law) by explaining a simultaneous system of nonlinear equations. The primal approach derives supply and demand functions from the production function and determines a market clearance price. Though the dual approach ascertains a system of marginal cost functions as inverse supply functions to determine cost prices, price dependent demand is derived from Sheppards Lemma which reveals supply. The difficulties arise from the fact that no limits on prices and activities (inequalities) occurs so that very often an optimal solution cannot be found (Manne 1985). As a consequence, iterative equalization methods are preferred.

Every optimization problem under unequal side constraints can be formulated by a complementary approach. Generally, the complementary format assures that each non negative variable must be zero, or the corresponding inequality condition must be equalized. The mixed complementary format (MCP) represents the first order conditions of a non linear optimization problem (Karush Kuhn Tucker conditions).[8] In contrast to original non linear optimization problems the complementarity format exposes the advantage of allowing simultaneous restrictions on prices and activities. MCPs arise in many application areas including applied economics, game theory, structural engineering and chemical engineering.

Econometric regression methods are based mostly on the ordinary least squares technique (OLS); a linear system of simultaneous equations is solved by the method of three stage ordinary least squares. Non linear econometric equation systems can be solved by maximum likelihood estimator processes computerized by the Gauss method and Gauss Newton algorithm, by the

---

[5] See Böhringer (1998) for an integrated bottom up and top down approach see the next chapter for applied modeling concepts.

[6] Like Panta Rhei, see next chapter.

[7] Like the Scarf algorithm, for more information and algorithms see Shoven and Whalley (1992).

[8] The programming language GAMS offers two different solution algorithms, MILES and Path, see (Ferris and Sinaoiromsaran 1998), MILES see (Rutherford 1993).

method of nonlinear two stage least squares, nonlinear three stage squares or generalized method of moments estimation.[9]

5.4 Substitutability

Linear optimization models constructed to depict production behavior can only represent substitution processes between input factors through a change of production technologies. Non linear substitution processes between input factors may only be modeled through a non linearisation of supply side production practices.
Most macroeconomic modeling approaches include CES (constant elasticity of substitution) function types to characterize substitution procedures. Computational general equilibrium models (CGE) mainly consider CES production and utility functions to describe production and consumption behavior. Depending on substitution elasticities, CES functions represent a substitution of complementary relationships between goods and factors. So called nested CES production functions include more than two input factors, whereas a conglomerate of two or more input factors can be substituted against another input factor. For example:

$$Y = e^{mt}\left[a(bK^{-\alpha} + (1-b)E^{-\alpha})^{\frac{\beta}{\alpha}} + (1-\alpha)L^{-\beta}\right]^{-\frac{1}{\beta}}$$

demonstrates a production function with three input factors, capital K, energy E and labor L. Capital and energy can be substituted against each other, just as a nest can be substituted against labor. Substitution elasticities can be derived from

$\sigma_{\alpha_i} = \dfrac{1}{1+\alpha_i}$ and $\sigma_{\beta_i} = \dfrac{1}{1+\beta_i}$ with α and β as substitution parameter α, β> -1, *a* and *b*

distribution parameter *a*, *b* >0. If the estimation of β is below –1, the substitution elasticity will become negative, a theoretical impossibility. If the substitution elasticity is negative, the two input factors can be interpreted as complements. If the substitution elasticities are positive, the input factors are substitutes.[10] A substitution elasticity between zero and one indicates that both input factors are incomplete substitutes, whereas a substitution elasticity higher than one full substitution opportunity is feasible. To represent trade relations, goods and services are widely traded and must be treated as incomplete substitutes, determined by the Armington trade function:

$Y_i = \left\{ d_i^D D^{\frac{\sigma_i^{IM-1}}{\sigma_i^{DM}}} + d_i^{IM} IM^{\frac{\sigma_i^{IM}-1}{\sigma_i^{IM}}} \right\}^{\frac{\sigma_i^{IM}}{\sigma_i^{IM}-1}}$ with D as domestic goods and IM imports, σ represents the

substitution elasticity between domestic and imported goods.

---

[9] See Kemfert (1998).
[10] See Kemfert (1998).

Supply and demand substitution processes are induced by relative price changes. If CGE models are based on an input output table such as data input, intra and intersectoral substitution processes can be covered. Input output models can only represent factor substitution processes if the above mentioned CES functions are implemented, because IO models include limited production functions with fixed coefficients of the Leontief type. Linear programming models do not allow for an inclusion of CES production functions with variable factor coefficients.

5.5 Role of Technological Change

The majority of bottom up models use a linear programming or optimization framework. Due to this, technological change is represented as a shift towards a new and more efficient technology. Linear programming models are widely applied in firms to detect optimal planning tracks. A chosen technology is optimal if the gross yield is maximized under the condition that each resource is restricted. Each linear equation system is solved by the Simplex algorithm to find a maximum value by jumping from one upper extremity to another. The dual optimization minimizes costs and determine the corresponding shadow prices of scarce resources. Because no transactions costs, information costs, market reactions, uncertainties about discount rates are considered, bottom up models tend to generate overly optimistic cost estimates.

In top down models, technological change is either represented exogenously by a so-called autonomous energy efficiency (AEEI) factor or endogenously by explicit investment decisions in R&D which increases innovation and new technology learning due to decreasing costs. Technological change is usually included within the production function representing technological changes resulting from price substitutions. Mainly, technological changes influence the productivity of input factors (labor, capital, energy) within the production process. Technological changes within the production process of the economy and especially the interdependencies and interrelations of driving forces to describe endogenous technological progress sufficiently have been studied by various authors within a theoretic or applied modeling framework. Traditional neoclassical growth models like (Solow 1956) and (Swan 1956) gave the first interpretations and modeling frameworks of standard growth models, while (Arrow 1962) found the Solow-Swan model's ability to demonstrate per-capita growth driven by exogenous technological change but not by the endogenous accumulation of inputs. Following (Barro and Sala-i-Martin 1992) endogenous growth and technological change models can be classified into three main categories: a. Models based on human capital accumulation, b. Schumpeterian models based on improvements in the quality of products and c. models based on the enlargement of product variety. (Lucas 1988) proposed the first endogenous growth model including human capital mainly based on the idea that production of the final good can be reached by the input of physical capital, specific human capital and the average level of human capital. (Aghion and Howitt 1992) elaborated an endogenous growth model including creative destruction in the sense that each new product takes the place of the older one in the Schumpterian tradition; (Grossman and Helpman 1991) introduced the same idea. Within this model, three sectors are distinguished: the intermediate good, the final good, and research. During final good production the productivity of the intermediate good can increase as a means of technological progress. Technological advances and innovation are considered uncertain by a stochastic process characterized by a Poisson process.

(Romer 1990) and (Grossmann and Helpman 1991) elaborated an endogenous growth theory by considering innovations expanding the variety of available goods. R&D activities are treated as other production activities converting primary inputs like capital and labor into knowledge. The model of (Romer 1996) and (Aghion and Howitt 1992) consider two sectors. The first produces the final good while the second is the R&D sector aiming to increase the level of technology.

(Goulder and Mathai 1999) reflect on induced technological progress both by R&D expenditures and learning by doing. Their main findings indicate that if knowledge comes from increasing R&D expenditures, carbon abatement can be moved to later time periods, but if knowledge is gained through learning by doing, the cost effective option could be accelerating the timing of carbon emission abatement.

5.6 Discounting

The technique of discounting effects is applied by economists to compare immediate positive or negative policy strategy impacts with those occurring in the distant future.
Influenced by a long discussion regarding the "right" discount rate in economic-environment-energy modeling, modelers tend to be as pragmatic as possible. Two schools of thought have been developed based on both prescriptives and descriptive approaches. The former leads to a selection of the discount rate by "ethical principles" or rules relating to the way benefits and costs of different generations ought to be weighted. This means that the choice of a discount rate is based on an observation of the rates of return to capital invested in a variety of alternative assets. The first approach mostly leads to relatively low discount rate values (close to zero or even negative) whereas the second approach would mean a higher discount rate (five to 20 percent). However, as these long discussions about discount rates reveal, nearly all scientists agreed that it is appropriate to discount future benefits and costs at some positive rates, see Arrow (1996), Weitzman (1999) or Manne (1999).
The discounting technique used in most Ramsey type or OLG models are typically very similar, i.e. each generation maximizes the present value of lifetime utilities. Therefore, future utility units are discounted to the beginning of their respective lives. Welfare present values are calculated by discounting generation-specific utilities to the beginning of the planning horizon using the "social discount rate". Generation-specific myopia equals the myopic attitude of a central planner summing up all generation-specific present values. Kemfert and Bayer (2002) refrain from this very strong assumption because the assumed discounting technique biases in favor of current living generations and discriminates against future ones and is therefore not "neutral" within an intergenerational framework. Tol *(Tol* 1999*)* as well as Bayer (*Bayer* 2000) analyzed different kinds of discounting measures and their impacts on climate change and economic reactions. They both found that the discounting method has substantial impacts on long-term emissions control and short run emissions abatement. Howarth (*Howarth* 1998) shows that welfare statements depend heavily on transfer assumptions between different generations. Distributional aspects are focal points within investigations by Stephan et al. (*Stephan/Müller-Fürstenberger* 1998, *Stephan et al.* 1997), Manne (*Manne* 1999), Nordhaus (*Nordhaus* 1994, who argues in a Ramsey-type-model) as well as the more qualitative paper by Schelling (*Schelling* 1995). In general, they assert that distributional reasons are the most important arguments for not abating at all. If today's living generations would heavily abate, future generations will not only be wealthier due to conventional capital accumulation but also due to returns induced by GHG-abatement. On the other hand, renunciation of GHG-abatement leads to a more equal distributional effect. Conventional capital formation is used more intensely, leading to increasing consumption and investment possibilities for future generations while simultaneously decreasing "green capital," resulting in welfare losses for future generations.

6 Examples of Economic-Environment-Energy Models

6.1 Overview

This short chapter provides a classification of applied economic-environment-energy models according to their geographical scale and bottom up (LP, NLP) or top down (IO, CGE, IAM) nature. Recent advances in economic-environment-energy modeling have moved towards global, regionalized IAMs based mostly on a microeconomic general equilibrium framework.[11]

	National	EU	Global	Global-Regionalized
Input Output Models	MIS (D) MEPA			
LP/NLP Models	MARKAL ETSAP MESSAGE III	HERMES-MIDAS MARKAL		IEA MARKAL POLES PRIMES CERT
IA Models		ESCAPE	DICE, R&DICE, PRICE SLICE CETA	AIM IMAGE RICE FUND PAGE MERGE IIAM ICAM MINICAM OXFORD SGM
CGE /AGE	Conrad (D) Bovemberg- Goulder (USA) Jorgeson- Wilcoxen (USA) NEWAGE-D	GEM-E3 NEWAGE-EU LEAN ETAS		ERM EPPA SGM MS-MRT G-TEM GREEN C-Cubed WAGEM Wordscan
Econometric Models	MDM (UK) Panta Rhei (D)	Quest WARM E3 ME		Panta Rhei (World)

Table 1: Classification of Economic-Environment-Energy Models[12]

Empirically oriented and applied theoretic models incorporate endogenous or induced technological changes using different approaches. Macroeconometric models like E3ME (Lee, Pesaran et al. 1990) or WARM (Carraro and Galeotti 1997) include simple approaches of endogenous technological changes. Integrated assessment models like ICAM3 (Dowlabadi

[11] See Bosello et al. (1998).

[12] See Kemfert and Kuckshinrichs (1995):MIS; Commission (1993): HERMES-MIDAS; Bahn, Barreto et al. (1997), Gielen and Kram (1998) and Manne and Wene (1994): MARKAL- MACRO; Gielen and Kram (1998; Kram (1998; Kram (1998): ETSAP; Messner and Strubegger (1994): MESSAGE; Kouvaritakis, Soria et al. (2000): POLES; Capros (1996): PRIMES; Nordhaus (1993): DICE; Peck and Teisberg (1992): CETA; Batjes and Goldewijk (1994): IMAGE; Nordhaus and Yang (1996): RICE; Tol (1999): FUND; Manne, Mendelsohn et al. (1995):MERGE; Rutherford (1992): IIAM; Conrad (1993); Goulder (1995) and Goulder and Mathai (1999); Jorgenson and Wilcoxen (1993); Capros, Georgakopoulos et al. (1995): GEM E3; Böhringer (1997): NEWAGE; Welsch and Hoster (1995): LEAN; Babiker, Reilly et al. (2001): EPPA; Edmonds (1998): Minicam; Burniaux, Nicoletti et al. (1992): GREEN; McKibbin and Wilcoxen (1999): C-Cubed; Kemfert (2001): WAGEM; Bollen, Gielen et al. (1999):Worldscan; Barker and Zagame (1995): E3ME; MacCracken, Edmonds et al. (1999): SGM; Carraro and Galeotti (1996): WARM; Meyer (1998): Panta Rhei.

1998) uses more sophisticated approaches of modeling endogenous technological changes. Macroeconomic general equilibrium models like DICE (R&DICE) (Nordhaus 1999) encompass induced innovations; the use of carbon energy is controlled by induced technological change. Energy system models like MESSAGE (Grübler and Messner 1998) include learning by doing in special functions within an energy system framework. New versions of POLES (Kouvaritakis, Soria et al. 2000) or MARKAL (Barreto and Kypreos 2000) contain approaches to endogenously determine technological changes in their approaches. (Goulder and Schneider 1999) investigate the implications of incorporating induced technological progress through increased R&D effort with the result that amplified environmentally related R&D efforts might crowd out other non-environmentally related R&D efforts and lead to a decrease in output and gross world product.

7 Conclusion

This paper gives an overview of applied economy- energy- environmental modelling for quantitative impact assessment. The aim of this paper is to shed light on different modeling approaches, compare methodologies and theories and classify existing modeling approaches. The creation of an economic- environment- energy modeling approach is primarily influenced by the purpose of modeling and merit of explanation. The quality of a model can be accurately evaluated only when both targets are explicated and unambiguously reached.

8 References

Aghion, P. and P. Howitt (1992). Endogenous Growth Theory.

Arrow, K. (1962). "The Economic Implications of Learning by Doing." Review of Economic Studies **29**: 155-173.

Arrow, K. (1962). Economic Welfare and Allocation of Resources for Invention. Princeton, NBER.

Arrow, K.J. et al.: 1996, 'Intertemporal Equity, Discounting, and the Economic Efficiency', in Bruce, J.P. et al. (eds.), *Climate Change 1995. Economic and Social Dimensions of Climate Change*, Cambridge et al., Cambridge University Press, 125-144.

Bahn, O., L. Barreto, et al. (1997). A Multi-Regional MARKAL-MACRO Model to Study an International Market of CO2 Emission Permits. Villingen / CH, Paul Scherrer Institut.

Barker, T. and P. Zagame (1995). E3ME: An Energy-Environment-Economy Model for Europe. Brüssel, European Commission.

Barreto, L. and S. Kypreos (2000). Multi-Regional Technological Learning: Some Experiences with the MARKEL Model. Villigen, Energy Modelling Group.

Barro, R. J. and X. Sala-i-Martin (1992). Economic Growth, McGraw-Hill.

Batjes, J. J. and C. G. M. Goldewijk (1994). The IMAGE 2 Hundred Year (1890-1990) Database of the Global Environment (HYDE). Bilthoven, RIVM.

Bayer, S. and Cansier, D.: 1999, 'Intergenerational Discounting: A New Approach', *International Journal of Development Planning Literature*, **14**, 301-325.

Bayer, S.: 2000, *Intergenerationelle Diskontierung am Beispiel des Klimaschutzes*, Marburg, Metropolis.

Bayer, S.: 2002, 'Generation Adjusted Discounting in Long-term Decision-Making', *International Journal of Sustainable Development*, (forthcoming).

Bayer, S., Kemfert, C.: 2002, Reaching National Kyoto Targets in Germany and Sustainable Development, in: Environment, Development and Sustainability, 4 (4), 371-390

Blanchard, O.J. and Fischer, S.: 1989, *Lectures on Macroeconomics*, Cambridge, Mass., London, MIT Press.

Böhringer, C. (1997). NEWAGE - Modellinstrumentarium zur gesamtwirtschaftlichen Analyse von Energie- und Umweltpolitiken. Energiemodelle in der Bundesrepublik Deutschland - Stand der Entwicklung. S. Molt and U. Fahl. Jülich**:** S. 99-122.

Böhringer, C. (1998). The Synthesis of Bottom-up and Top-Down in Energy Policy Modeling. Energy Models for Decision Support- New Challenges and Possible Solutions. E. S. Läge, P. Stuttgart**:** 113-131.

Bollen, J., A. Gielen, et al. (1999). "Clubs, Ceilings and CDM: Macroeconomics of Compliance with the Kyoto Protocol." The Energy Journal - A Special Issue: The Costs of the Kyoto Protocol: A Multi-Model Evaluation: 177-206.

Bosello, F., Carraro, C. and Kemfert, C.: Advances of Climate Modelling for Policy Analysis, Nota di Lavoro, 82/98

Burniaux, J.-M., G. Nicoletti, et al. (1992). "GREEN: A Global Model for Quantifiying the Costs of Policies to Curb CO_2 - Emissions." OECD-Studies: The Economic Costs of

Reducing Emissions **No.19**.

Cansier, D.: 1996, *Umweltökonomie*, 2., neubearbeitete Auflage, Stuttgart, Lucius & Lucius.

Capros, P. (1996). PRIMES and GEM-E3 Contribution to Climate Change Policy Debate. Athens, National Technical University of Athens.

Capros, P. and L. Mantzos (2000). The Economic Effects of EU-Wide Industry Level Emission Trading to Reduce Greenhouse Gases.

Capros, P., T. Georgakopoulos, et al. (1995). GEM-E3 Computable General Equilibrium Model for Studying Economy - Energy - Environment Interactions. Bruxelles, European Commission.

Carraro, C. and M. Galeotti (1996). WARM: A European Model for Energy and Environmental Analysis. Milan, Fondazione Eni Enrico Mattei.

Carraro, C. and M. Galeotti (1997). “Economic Growth, International Competetiveness and Environmental Protection,R&D and Innovation Strategies with the WARM Model.” Energy Economics **19**(1): 2-28.

Commission, E. (1993). HERMES, Harmonised Econometric Research for Modeling Economic Systems. Amsterdam.

Conrad, K. (1993). Problems of Modeling CO_2 Reductions with GE Models: Applied General Equlibrium Modeling for Environmental Policy Analysis. Economics of the Greenhouse Effect. W. Kuckhinrichs, W. Pfaffenberger and W. Ströbele. Jülich. **13:** 33 - 88.

Criqui, P., G. Klaassen, et al. (2000). The Efficiency of Energy R&D Expenditures.

Dowlabadi, H. (1998). “Sensitivity of Climate Change Mitigation Estimates to Assumptions about Technical Change.” Energy Economics **20**(5-6): 473-493.

Dowlatabadi, H., and M. Granger Morgan (1993) "Integrated Assessment of Climate Change." *Science* 259 (26 March): 1813, 1932.

Edmonds, J (1998): Climate Change Economic Modelling: Background Analysis for the Kyoto Protocol, OECD Workshop Paris

Edmonds, J. (1998). Climate Change Economic Modelling: Background Analysis for the Kyoto Protocol. Paris.

Ferris, M. and J. Pang (1995). “Engineering and Economic Applications of Complementarity Problems.” .

Gerlagh, R. and van der Zwaan, B.: 2001, ‘The Effects of Aging and an Environmental Trust Fund in an Overlapping Generations Model on Carbon Emission Reductions’, *Ecological Economics*, **36**, 311-326.

Gerlagh, R.: 1999, *The Efficient and Sustainable Use of Environmental Resource Systems*, Amsterdam, Thela Thesis.

Gielen, D. and T. Kram (1998). The MATTER Projekt on Integrated Energy/Materials Strategies for Western Europe. IEA-ETSAP Workshop, Berlin.

Gielen, R. and T. Kram (1998). The Role of Kyoto Mechanisms: Result from MARKAL Analysis, ECN-Policy Studies.

Goulder, L. and S. Schneider (1999). “Induced Technological Change and the Attractiveness of CO2 Abatement Policies.” Resource and Energy Economics **21**: 211-253.

Goulder, L. H. (1995). "Effects of Carbon Taxes in a Economy with Prior tax Distortions: An Intertemporal General Equilibrium Analysis." Journal of Environmental Economics and Management **29**(3): 271-297.

Goulder, L. H. and K. Mathai (1999). "Optimal CO2 Abatement in the Presence of Induced Technological Change." Journal of Environmental Economics and Management **2000**(38): 1-38.

Goulder, L. H. and K. Mathai (1999). "Optimal CO2 Abatement in the Presence of Induced Technological Change." Journal of Environmental Economics and Management **2000**(38): 1-38.

Grossman, G. and E. Helpman (1991). "Quality Ladders in the Theory of Growth." Review of Economic Studies **58**: 43-61.

Grossmann, G. M. and E. Helpman (1991). Innovation and Growth in the Global Economy. Cambridge (USA), MIT Press.

Grubb et al. (1993): The Costs of Limiting Fossil Fuel CO2 Emissions: A Survey and Analysis. in : R.H. Socolow et. Al.(eds.) *Annual Review of Energy and the Environment*, Vol. 18, 1993.

Grübler, A. and S. Messner (1998). "Technological Change and the Timing of Mitigation Measures." Energy Economics **20**: 495-512.

Hertel, T. (1998): Global Trade Analysis, GTAP

Hourcade et. Al. (2001): The Economics of a Lost Deal.

Hourcade et.al. (1996): Estimating the Costs of Mitigating Greenhouse Gases. In: J.P. Bruce, H. Lee and E. Haites (eds.) Climate Change 1995: Economic and Social Dimensions of Climate Change. Contribution of Working Group III to the Second Assessment Report of the IPCC. University Press, pp. 263- 296.

Howarth, R.B.: 1998, 'An Overlapping Generations Model of Climate-Economy Interactions', *Scandinavian Journal of Economics*, **100**, 575-591.

Jorgenson, D. W. and P. J. Wilcoxen (1993). "Reducing US Carbon Emissions: An Econometric General Equilibrium Assessment." Resource and Energy Economics **15**: 7-25.

Kemfert, C. (1998). "Estimated Substitution Elasticities of a Nested CES."

Kemfert, C. and W. Kuckshinrichs (1995). MIS- A Model - Based Macroeconomic Information System for Energy Analysis in Germany. Oldenburg.

Kemfert, C.: 1999, *Economic Implications of the Kyoto Protocol, Perspectives of Newest Climate Change Policy Options*, Working Paper ESA-AP-99-2, Stuttgart, Stuttgart University.

Kemfert, C.: 2000, *Emissions Trading and its Impacts on World Economies - Contemplation of Baseline Emissions Paths and a Ceiling on Emissions Trading*, Nota Di Lavoro 01/2000, Fondazione ENI Enrico Mattei (FEEM), Milano, Italy

Kouvaritakis, N., A. Soria, et al. (2000). Endogenizing Technology Change in Energy Models: The POLES Experience.

Kram, T. (1998). History, the ETSAP Kyoto Statement and Post-Kyoto Analysis. FEES/ETSAP Workshop: Energy Models for Decision Support - New Challenges and Possible Solutions, Berlin, ECN Policy Studies.

Kram, T. (1998). The Energy Technology Systems Analysis Programme: History, the ETSAP Kyoto Statement and Post-Kyoto Analysis. OECD Workshop: The Economic Modelling of Climate Change: Backround Analysis for the Kyoto Protocol, Paris, ECN Policy Studies.

Lee, K., M. H. Pesaran, et al. (1990). Aggragation Bias in Labour Demand Equations. Disaggregation in Econometric Modelling. T. Barker and M. H. Pesaran. London.

Lucas, R. E. J. (1988). “On the Mechanics of Economic Development.” Journal of Monetary Economics **22**: 3-42.

MacCracken, C. N., J. A. Edmonds, et al. (1999). “The Economics of the Kyoto Protocol.” The Energy Journal - A Special Issue: The Costs of the Kyoto Protocol: A Multi-Model Evaluation: 25-72.

Manne (1985), : On the Formulation and Solution of Economic Equilibrium Models, in: A.S. Manne (1985): Economic Equilibrium- Model Formulation and Solution, Mathematical Programming Study 23, 1-22

Manne, A. S. and C. O. Wene (1994). MARKAL/MACRO: A Linked Model for Energy-Economy Analysis. Advances in Systems Analysis: Modelling Energy-Related Emissions on a National and Global Level. J. F. Hake, M. Kleemann, W. Kuckshinrichs, D. Martinsen and M. Walbeck. Jülich. **15:** 153-191.

Manne, A. S., R. Mendelsohn, et al. (1995). “MERGE - A Model for Evaluating Regional and Global Effects of GHG Reduction Policies.” Energy Policy **23**(1): 17-34.

Manne, A.S.: 1999, ‘Equity, Efficiency, and Discounting’, in P.R. Portney and J.P. Weyant (eds.), *Discounting and Intergenerational Equity*, Washington D. C., Resources for the Future, 111-129.

McKibbin, J. W. and J. P. Wilcoxen (1999). Permit Trading Under the Kyoto Protocol and Beyond. EMF/IEA/IEW Workshop, International Energy Agency, Paris.

Messner, S. and M. Strubegger (1994). The Energy Model MESSAGE III. Advances in Systems Analysis: Modelling Energy-Related Emissions on a National and Global Level. J. F. Hake, M. Kleemann, W. Kuckshinrichs, D. Martinsen and M. Walbeck. Jülich. **15:** 29-49.

Meyer, B. (1998). Panta Rhei: Econometric E3 Modelling for Germany. Energy Models for Decision Support- New Challenges and Possible Solutions. E. S. Läge, P **.:** 155-167.

Nordhaus, W. (1999). Modelling Induced Innovation in Climate Change Policy. Laxenburg.

Nordhaus, W. D. (1993). “Rolling the 'DICE': An Optimal Transition Path for Controlling Greenhouse Gases.” Resource and Energy Economics **15**: 27-50.

Nordhaus, W. D. and Z. Yang (1996). “RICE: A Regional Dynamic General Equilibrium Model of Optimal Climate-Change Policy.” American Economic Review **86**(4): 741-765.

Nordhaus, W.D.: 1994, *Managing the Global Commons*, Cambridge, Mass., MIT Press.

Pahlke, A.: 2000, *Allokative und intergenerationale Effekte einer ökologischen Steuerreform*, Josef Eul Verlag, Lohmar, Germany.

Peck, S. C. and T. J. Teisberg (1992). “CETA: A Model for Carbon Emission Trajectory

Assessment." The Energy Journal **13**(1): 55-77.

"Production Function Approach for Germany." Energy Economics(20): 249 - 264.

Romer, D. (1996). Advances Macroeconomics, MCGraw-Hill.

Romer, P. (1990). "Endogenous Technological Change." Journal of Political Economy 98(71-102).

Rotmans, J., and Dowlatabadi, H. (1998), 'Integrated Assessment of Climate Change: Evaluation of Models and Other Methods', in Rayner, S. and Malone, E.(eds.), Human Choice and Climate Change: An International Social Science Assessment, Batelle Press, USA.

Rutherford, T. (1992). The Welfare Effects of Fossil Carbon Restrictions: Results from a Recursive Dynamic Trade Model. Paris, OECD.

Rutherford, T. (1993). MILES: A Mixed Inequality and Non-Linear Equation Solver.

Schelling, T.C.: 1995, 'Intergenerational Discounting', *Energy Policy*, **23**, 395-401.

Shoven, J. B. and J. Whalley (1992). Applying General Equilibrium; Cambridge Surveys of Economic Literature. Cambridge.

Solow, R. (1956). "A Contribution to the Theory of Economic Growth." Quarterly Journal of Economics **70**: 65-94.

Springer, Urs (2003): The Market for GHG Permits Under the Kyoto Protocol: A Survey of Model Studies, in Enegry Economics, forthcoming

Stephan, G. and Müller-Fürstenberger, G.: 1998, 'Discounting and the Economic Costs of Altruism in Greenhouse Gas Abatement', *Kyklos*, **51**, 321-338.

Stephan, G., Müller-Fürstenberger, G., and Previdoli, P.: 1997, 'Overlapping Generations or Infinitely-Lived Agents. Intergenerational Altruism and the Economics of Global Warming', *Environmental and Resource Economics*, **10**, 27-40.

Swan, T. (1956). "Economic Growth and Capital Accumulation." Economic Record **32**: 334-361.

Tol, R. S. J. (1999). "Kyoto, Efficieny, and Cost-Effectiveness: Applications of FUND." The Energy Journal - A Special Issue: The Costs of the Kyoto Protocol: A Multi-Model Evaluation: 131-156.

Tol, R.S.J.: 1999, 'Time Discounting and Optimal Emission Reduction: An Application of FUND', *Climatic Change*, **41**, 351-362.

Toth, F.(1996): Practice and Progress in Integrated Assessment of Climate Change, in: Energy Policy, Vol. 23, No.4/5, pp. 253-267Babiker, M. H., J. M. Reilly, et al. (2001). The MIT Emissons Prediction and Policy Analysis (EPPA) Model: Revisions, Sensitivities, and Comparisons of Results, MIT Joint Programm on the Science and Policy of Global Change.

Weitzman, M. (1999). Just Keep Discounting, But... Discounting and Intergenerational Equity. P. Portney and J. Weyant**:** 23-31.

Welsch, H. and F. Hoster (1995). "A General Equilibrium Analysis of European Carbon/ Energy Taxation: Model Structure and Macroeconomic Results." Zeitschrift für Wirtschafts- und Sozialwissenschaften **115**(2): 275 - 303.

2 The theory of energy economics: an overview

Thomas Weyman-Jones

1 Introduction

In reality there is no such subject as energy economics, because energy, although a meaningful concept in the physics or engineering sense, is not a commodity that can be bought and sold in the marketplace. However, individual fuels (primary and secondary electricity, natural gas, oil, coal) can be bought and sold; in this context, primary electricity includes renewable sources and nuclear power. Therefore 'energy economics' is really the economics of fuel markets, and the phrase: energy economics is used for convenience to represent all the useful economic concepts which arise in studying different fuels. The energy industries are organised in different ways in different countries; many are investor owned, especially in the USA and the UK, but state ownership is also common. Many are characterised by economies of scale and hence have considerable market power, which usually leads to them being regulated. Fuels are widely traded in solid, liquid and gaseous form, and are transported all over the world in tankers, pipes and wires.

In some of these fuel markets we can see that it is cheaper to have one company do all the business rather than many. Examples are the national power and gas grid companies engaged in the activity of bulk transmission of electric power and natural gas. Such companies are traditionally referred to as public utilities (although there is no presumption that they are or should be owned by the state). Because these companies are believed to operate most cheaply or efficiently when there is only one of them in each market we call them 'natural monopolies' (that is, the traditional public utilities: water, gas, electricity, telecommunications, have the characteristics known as natural monopoly even when they are not statutory monopolies). Consequently there is a wide public interest in the possibility of regulating their behaviour, and the economics of regulation becomes an intrinsic part of energy economics.[1]

The format of this chapter follows from these fundamental ideas. It begins by looking at the basic economic ideas of resource allocation in capital-intensive fuel industries with emphasis on the nature of cost–benefit analysis of fuel investment decisions, and the consequent implications for efficient market pricing. The topics covered here include the nature of short- and long-run marginal cost of energy supply, the process of investment decision making, and the design of efficient price mechanisms in industries where storage of the product is very costly, and in industries where delivery of the product through a grid differs from the economic activity of creating the product. Both of these features are critical characteristics of the energy industries. When such characteristics stem from the fact that the industry delivers its output through a network of wires or pipes, analysts often use the alternative description: network industries. This is followed by a discussion of the market conditions that are frequently found in the fuel industries.

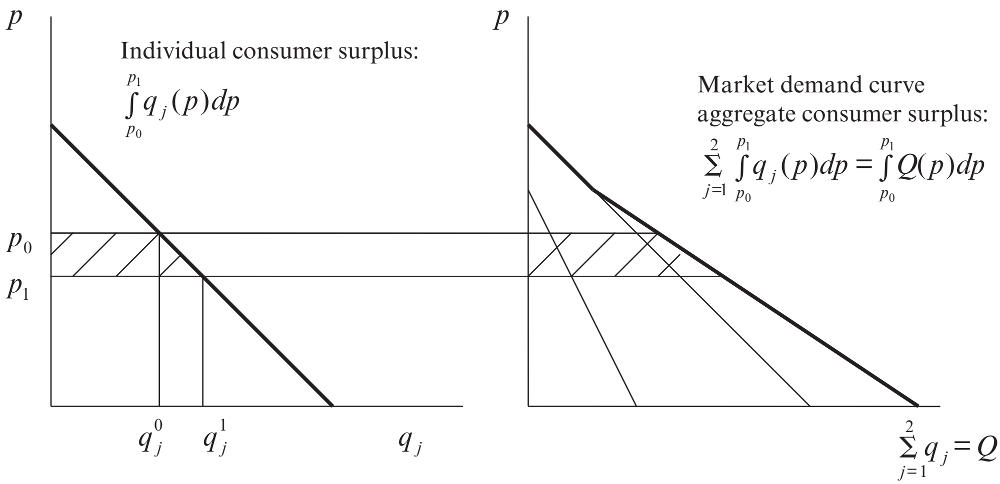


Figure 2.1 Individual and aggregate consumer surplus

2 Cost–Benefit Analysis and Market Structure

A considerable part of energy economics and policy is concerned with optimal resource allocation which is normative rather than positive economics. However, a normative economics approach can be useful to understand market outcomes. This is because a competitive market will mimic the allocation of resources that is achieved in a welfare-maximising model. For that reason, a useful way to simulate the behaviour of a competitive market equilibrium is to characterise the equilibrium through welfare analysis[2] (Mas-Colell et al. 1995, pp. 630–31). Therefore, cost–benefit analysis is a useful building block because it conveniently describes a route to an optimal allocation of resources. In fact cost–benefit analysis has a stronger property as well: the conventional economics approach to efficient resource allocation, the Pareto criterion,[3] is unable to offer policy recommendations when there are losers as well as winners from a policy change. A fundamental tool of cost–benefit analysis is the individual consumer's demand curve which expresses the quantity demanded of any commodity (good or service) as a negative function of its price:

$$q = q(p); \quad q'(p) < 0.$$

This is illustrated in the left-hand panel of Figure 2.1, for one consumer labelled: j. In the figure, the price has fallen from p_0 to p_1 and quantity demanded has risen as a result from q_j^0 to q_j^1. The demand curve expresses the consumer's willingness to pay for different units of a commodity, with the marginal willingness to pay for additional units falling as more units are consumed. The area left of the demand curve but above the price actually being charged at present is called the 'consumer surplus', and it is the willingness to pay for so many units of a commodity minus the amount actually paid for those units, using the traditional Marshallian definition.

When the price of a commodity falls, the consumer obtains additional consumer surplus. In Figure 2.1 (left-hand panel):

$$CS_j = \int_{p''}^{p'} q_j(p)\,dp.$$

Note that this is measurable as an amount of money, and can be measured from an empirically estimated demand function. If the compensated demand function has been measured, that is, the demand function based only on the substitution effect of a price change after compensating for the income effect, an alternative definition is: consumer surplus is the amount of real income a consumer would pay to be as well off after a fall in price as he/she would be if the price had not fallen; this is Hicks's compensating variation definition of consumer surplus.

To arrive at the market demand curve for a commodity, horizontal summation of the individual demand curves of different persons or households, (j) is used:

$$Q(p) = \sum_{1}^{J} q_j(p).$$

Horizontal summation is illustrated in the right-hand panel of Figure 2.1, and is required when the consumption of the good in question by person 1 reduces the amount available for person 2. Such goods (the majority) are called 'private goods'.

The area left of the market demand curve and above the price charged is then the aggregate consumer surplus from consumption of the commodity at the prevailing market price, p^*:

$$CS = \int_{p^*}^{\infty} Q(p)\,dp.$$

This is interpreted as one part of the gross benefit from supply of the commodity at the price p^* and is the economist's universal measure of aggregate consumer welfare. It represents the sum of all persons' compensating variation measures of consumer surplus.

The supplier's revenue is: pQ, and the cost of supplying a commodity is given by the *cost function*:

$$C = C(Q); \quad C'(Q) \equiv \text{Marginal Cost } (MC) > 0.$$

Marginal cost is a forward-looking measure, and represents the change in total cost that would be observed if the level of output were to change by one unit. Aggregate producer surplus is the other part of the gross benefit from supply of the commodity, and this is the area left of the supply curve and below the price charged for the product. The supply curve of a product to a market is the horizontal summation of the marginal cost curves of the individual firms so that producer surplus,[4] written π is:

$$\pi = pQ(p) - C[Q(p)].$$

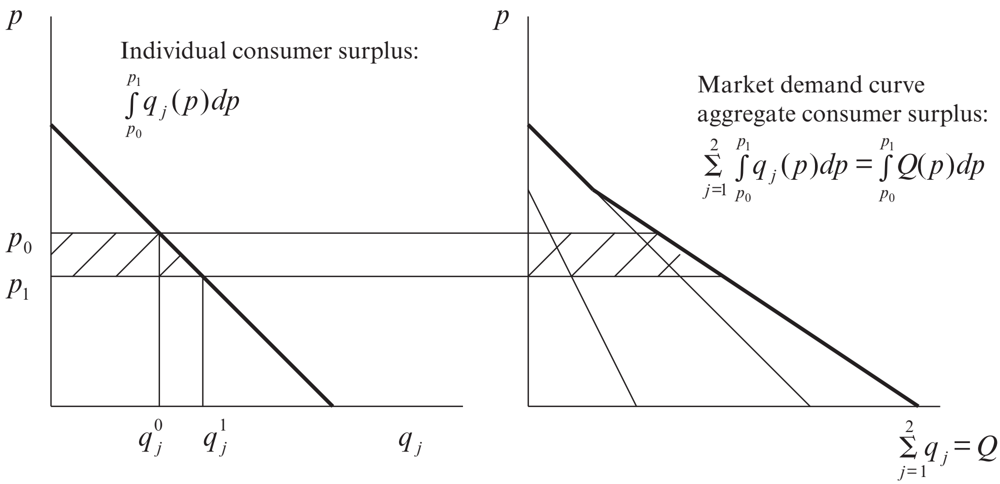


Figure 2.1 Individual and aggregate consumer surplus

2 Cost–Benefit Analysis and Market Structure

A considerable part of energy economics and policy is concerned with optimal resource allocation which is normative rather than positive economics. However, a normative economics approach can be useful to understand market outcomes. This is because a competitive market will mimic the allocation of resources that is achieved in a welfare-maximising model. For that reason, a useful way to simulate the behaviour of a competitive market equilibrium is to characterise the equilibrium through welfare analysis[2] (Mas-Colell et al. 1995, pp. 630–31). Therefore, cost–benefit analysis is a useful building block because it conveniently describes a route to an optimal allocation of resources. In fact cost–benefit analysis has a stronger property as well: the conventional economics approach to efficient resource allocation, the Pareto criterion,[3] is unable to offer policy recommendations when there are losers as well as winners from a policy change. A fundamental tool of cost–benefit analysis is the individual consumer's demand curve which expresses the quantity demanded of any commodity (good or service) as a negative function of its price:

$$q = q(p); \quad q'(p) < 0.$$

This is illustrated in the left-hand panel of Figure 2.1, for one consumer labelled: *j*. In the figure, the price has fallen from p_0 to p_1 and quantity demanded has risen as a result from q_j^0 to q_j^1. The demand curve expresses the consumer's willingness to pay for different units of a commodity, with the marginal willingness to pay for additional units falling as more units are consumed. The area left of the demand curve but above the price actually being charged at present is called the 'consumer surplus', and it is the willingness to pay for so many units of a commodity minus the amount actually paid for those units, using the traditional Marshallian definition.

When the price of a commodity falls, the consumer obtains additional consumer surplus. In Figure 2.1 (left-hand panel):

$$CS_j = \int_{p''}^{p'} q_j(p)\,dp.$$

Note that this is measurable as an amount of money, and can be measured from an empirically estimated demand function. If the compensated demand function has been measured, that is, the demand function based only on the substitution effect of a price change after compensating for the income effect, an alternative definition is: consumer surplus is the amount of real income a consumer would pay to be as well off after a fall in price as he/she would be if the price had not fallen; this is Hicks's compensating variation definition of consumer surplus.

To arrive at the market demand curve for a commodity, horizontal summation of the individual demand curves of different persons or households, (j) is used:

$$Q(p) = \sum_{1}^{J} q_j(p).$$

Horizontal summation is illustrated in the right-hand panel of Figure 2.1, and is required when the consumption of the good in question by person 1 reduces the amount available for person 2. Such goods (the majority) are called 'private goods'.

The area left of the market demand curve and above the price charged is then the aggregate consumer surplus from consumption of the commodity at the prevailing market price, p^*:

$$CS = \int_{p^*}^{\infty} Q(p)\,dp.$$

This is interpreted as one part of the gross benefit from supply of the commodity at the price p^* and is the economist's universal measure of aggregate consumer welfare. It represents the sum of all persons' compensating variation measures of consumer surplus.

The supplier's revenue is: pQ, and the cost of supplying a commodity is given by the *cost function*:

$$C = C(Q); \quad C'(Q) \equiv \text{Marginal Cost } (MC) > 0.$$

Marginal cost is a forward-looking measure, and represents the change in total cost that would be observed if the level of output were to change by one unit. Aggregate producer surplus is the other part of the gross benefit from supply of the commodity, and this is the area left of the supply curve and below the price charged for the product. The supply curve of a product to a market is the horizontal summation of the marginal cost curves of the individual firms so that producer surplus,[4] written π is:

$$\pi = pQ(p) - C[Q(p)].$$

Then the net economic welfare, $W(p^*)$ from supplying the commodity at a price of p^* is the taken to be the unweighted sum of aggregate consumer surplus (CS) plus aggregate producer surplus π, that is, total revenue minus the cost of supply, $C(Q)$:

$$W(p) = CS + \pi = \left[\int_{p^*}^{\infty} Q(p)\,dp\right] + \{p^*Q(p) - C[Q(p)]\}.$$

The cost–benefit analysis of microeconomic economic policy therefore requires the choice of p^* to maximise this objective with first-order condition depending on the slope[5] of the aggregate market demand curve, $dQ/dp = Q'(p)$:

$$\frac{dW}{dp} = \left(\frac{dCS}{dp}\right) + \left(\frac{d\pi}{dp}\right) = [-Q(p^*)] + \{Q(p^*) + [p^* - C'(Q)]Q'(p^*)\} = 0,$$

and simplifying:

$$\frac{dW}{dp} = \left(p - \frac{dC}{dQ}\right)\frac{dQ}{dp} = (p - MC)\frac{dQ}{dp} = 0,$$

which requires that price should equal marginal cost: $p^* = C'(Q)$. This coincides with the condition for a Pareto optimum, but it is derived by allowing for winners and losers, with the winners gaining enough to sufficiently compensate the losers,[6] and hence is consistent only with the potential Pareto criterion; this is the basis of cost–benefit analysis. In turn, this leads to the prediction that a sufficiently competitive market will choose the socially optimal behaviour of marginal cost pricing. The problem of economic regulation is whether a given market can be expected to be sufficiently competitive. As shown above, the standard social welfare function adopted for policy choices in energy economics is based on unweighted consumer and producer surplus. For energy policy that leads to discrete changes a useful approximation to the consequent welfare change is:

$$\Delta W = \tfrac{1}{2}(p - MC)\Delta Q.$$

It is immediately clear that a necessary condition for the policy to be desirable according to the potential Pareto criterion is that after the policy change there are no further welfare gains, $\Delta W = 0$, in other words, price equals marginal cost. But who gets what when there is a policy change? Conventional cost–benefit does not weight these gains differently, but different weights to reflect social preferences for one group in society *vis-à-vis* another is always a possibility.

What happens when there are large fixed costs to setting up an energy company, for example, the installation of a distribution network: total cost is $C = F + cQ$? This is illustrated in Figure 2.2, where average cost lies above marginal cost because the role of fixed costs is never entirely absent irrespective of the volume of output. Marginal cost pricing at the optimal output Q^* leads to losses, and consequently no firm will enter the industry to supply the commodity, despite the fact that at every output below Q^*, willingness to pay for the product exceeds the total cost, including fixed cost, of supplying it.

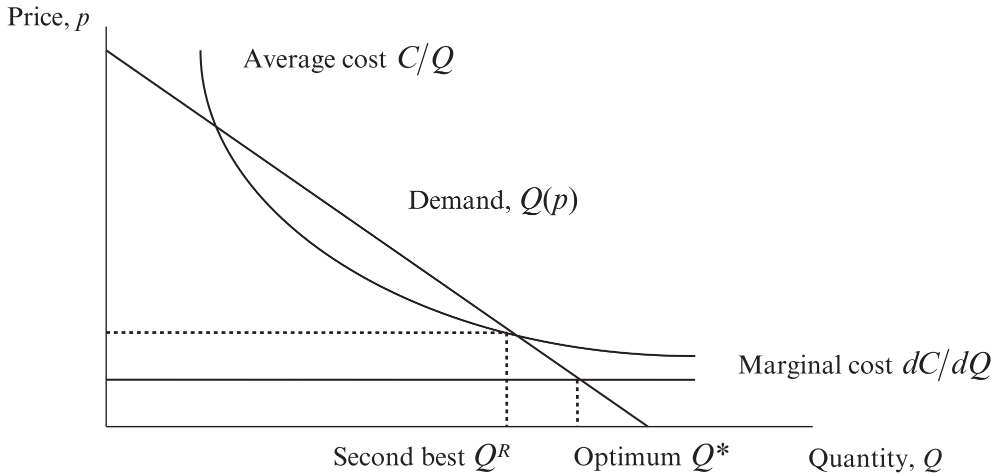


Figure 2.2 First- and second-best allocations for natural monopoly

Try average cost pricing, in this case a specific example of the more general idea of Ramsey pricing. The second-best outcome is at $Q^R(p^R)$ which is the solution to the problem:

$$\max W(p) = CS + \pi = \left[\int_{p^*}^{\infty} Q(p)\,dp\right] + \{p^*Q(p) - C[Q(p)]\},$$

such that:

$$p * Q(p) - C[Q(p)] \geq 0.$$

Since there are equal weights on consumer and producer surplus, social welfare improves for every fall in price that gives a monetary transfer from producer to consumer until the constraint is just satisfied. Therefore lower price with an implied welfare gain of $\{p - C'[Q(p)]\}Q'(p)$ until $p^R = C[Q(p^R)]/Q(p^R)$.

3 The Social Discount Rate in Cost–Benefit Analysis

The passage of time is regarded as one of the most important issues in an economic decision since it affects the delay with which benefits arrive and costs can be postponed. The discount rate, i, measures the loss of interest on cash flows that arrive one year from now and so cannot be invested until then. The procedure of discounted cash flow analysis states that the standard formula for net present value (NPV) (including both negative and positive cash flows, where each cash flow is assumed to occur at the *beginning* of the year) is:

$$NPV = x_0 + \frac{x_1}{1+i} + \frac{x_2}{(1+i)^2} + \ldots + \frac{x_t}{(1+i)^t} + \ldots + \frac{x_T}{(1+i)^T} = \sum_{t=0}^{t=T} \frac{x_t}{(1+i)^t}.$$

Projects with a positive net present value are worth doing. A useful version of the present value formula occurs when the cash flow is expected to be the same in every year: $(x/i)[1 - (1 + i)^{-T}]$. This is the net present value of an annuity.

What is the appropriate choice for the discount rate i in cost–benefit analysis? There are two suggested solutions for the choice of social discount rate (SDR): the social time preference rate (STP), and the social opportunity cost of capital (SOC).

Begin with a social welfare function that depends on the level of consumption in different periods: $W = \phi(C_0, C_1, \ldots)$. This weights the levels of total consumption for society in each period (t) (including the distribution among individuals, j). One example of this social welfare function is:

$$W = \sum_t \sum_j \delta(t)\, U_j(C_{jt}),$$

where:

$$U_j(C_{jt}) = \frac{1}{1 - \eta} C_{jt}^{1-\eta}.$$

If society consisted of a single individual, $j = 1$, who is assumed to have diminishing marginal utility, then a specific example of the social welfare function could be:

$$W = 2\sqrt{C_0} + 2\sqrt{C_1}.$$

This example is a special case corresponding to $\delta(t) = 1$ and $\eta = \frac{1}{2}$. More generally, this is an example where the present and future generations are weighted exactly equally:

$$\delta(0) = \delta(1) = \cdots = \delta(t) = \cdots = 1.$$

Note that this example has the property that when present and future consumption is the same, the marginal social welfare of consumption is the same, so that the marginal rate of substitution between present and future consumption is unity. Consequently in this case the generational weights will not affect the fundamental choice of the social discount rate. Figure 2.3 illustrates this example by using the property of 45° lines, and it can be seen that society's preference for present over future consumption is represented by the slope of the welfare contour:

$$dC_1/dC_0 = -[(\partial W/\partial C_0)/(\partial W/\partial C_1)] = -1 = -C_1/C_0 \Leftrightarrow C_1 = C_0.$$

Another special example of the social welfare function corresponds to $\eta = 1$. It plays a major role in the UK government report on the economics of climate change (Stern 2006).

$$W = \sum_t \sum_j \delta(t) \ln C_{jt}.$$

In Figure 2.4, the consumption possibility frontier represents the rate at which present consumption can be turned into future consumption in the economy's

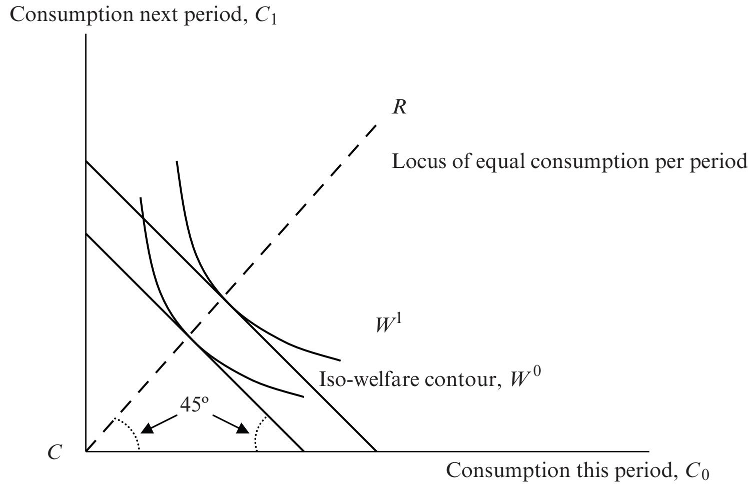


Figure 2.3 Equal welfare weights for current and future consumption

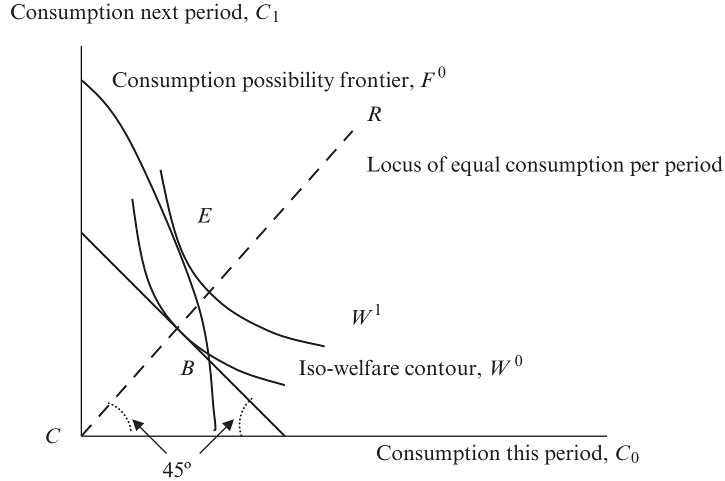


Figure 2.4 Positive marginal social return to capital requires lower consumption in the current period even when there are equal welfare weights for current and future consumption

production of real national income, that is, the rate of return to saving and investment. Consider a simple derivation of this frontier for two periods: this period is $t = 0$ and next period is $t = 1$. Suppose the economy starts with a capital stock of K_0 and that the maximum output available for consumption is $f(K)$. The fundamental constraint limits the sum of total consumption over the two periods to the total output available from the capital stock. Period 1's capital, K_1, is equal to the initial stock plus any saving (that is, output – consumption) done in period 0. Assume for the present that capital does not wear out.

The constrained optimisation model for choosing the social discount rate is shown below:

$$\max W = \phi(C_0, C_1)$$

$$\text{s.t. } C_0 + C_1 = f(K_0) + f[K_0 + f(K_0) - C_0].$$

The Lagrangean function is:

$$L = \phi(C_0, C_1) + \lambda(C_0 + C_1 - \{f(K_0) + f[K_0 + f(K_0) - C_0]\}),$$

with first-order conditions:

$$\partial L/\partial C_0 = \partial\phi/\partial C_0 + \lambda[1 + f'(K_1)] = 0,$$

$$\partial L/\partial C_1 = \partial\phi/\partial C_1 + \lambda = 0,$$

$$\partial L/\partial\lambda = (C_0 + C_1 - \{f(K_0) + f[K_0 + f(K_0) - C_0]\}) = 0.$$

Eliminating λ yields the tangency condition:

$$\frac{\partial\phi/\partial C_0}{\partial\phi/\partial C_1} = [1 + f'(K_1)].$$

This is shown at point E in Figure 2.4, where:

$$dC_1/dC_0 = -[(\partial W/\partial C_0)/(\partial W/\partial C_1)] = -[1 + f'(K)] \Leftrightarrow C_1 > C_0.$$

In general, therefore, the social discount rate should be different from zero, because otherwise the marginal product of capital is treated as zero. The implication of discounting the future to reflect that positive return to capital is that society should refrain from consumption today to build up capital for the future.

The equilibrium condition can be rearranged to give:

$$\frac{\partial\phi/\partial C_0}{\partial\phi/\partial C_1} = 1 + \left(\frac{\partial\phi/\partial C_0 - \partial\phi/\partial C_1}{\partial\phi/\partial C_1}\right) = [1 + f'(K_1)].$$

The left-hand side can be expanded further:

$$1 + \left(\frac{\partial\phi/\partial C_0 - \partial\phi/\partial C_1}{\partial\phi/\partial C_1}\right) = 1 + \left(\frac{\partial\phi/\partial C_0 - \partial\phi/\partial C_1}{dC} \frac{C}{\partial\phi/\partial C_1} \frac{dC}{C}\right).$$

If it is assumed that the weights on intergenerational consumption are constant at $\delta(t) = 1$, then this expression representing the left-hand side of the equilibrium condition can be interpreted as:

$$1 + \left(\frac{dMU}{dC}\frac{C}{MU}\frac{dC}{C}\right) = 1 + (\eta \Delta \log C),$$

where η is the elasticity of the marginal utility of consumption:

$$\eta = (dMU/dC)(C/MU),$$

and $\Delta \log C$ is the growth rate of consumption.

However, for reasons to be explored later in the context of Stern (2006), economists sometimes do assume that generations are weighted differently, that is, that there is a positive rate of pure time preference resulting in the discounting of the welfare of a future population, that is,

$$\delta(t) = 1/(1 + \delta)^t.$$

In this case the social welfare function would be written in the form:

$$W = U(C_0) + [U(C_1)/(1 + \delta)].$$

The slope of the welfare contour must take account of this intergenerational rate of pure time preference, so that the social time preference rate becomes:

$$-\frac{dC_1}{dC_0} = (1 + \eta \Delta \log C)(1 + \delta) \approx (1 + \delta + \eta \Delta \log C).$$

The right-hand side of the equilibrium condition can also be expanded:

$$1 + f'(K)$$

$$= 1 + \{f'(K)[f(K)/K]K/f(K)\} = 1 + (\Delta \log Y/\Delta \log K)(Y/K) = 1 + (\alpha Y/K).$$

In this expression, $f(K) \equiv Y$ is the real income producible by the capital stock, and α is the elasticity of real national income with respect to capital. The common tangent slope at E in Figure 2.4 is the discount factor to be applied to socially desirable investments:

$$1 + SDR = 1 + [\delta + \eta(\Delta \log C)] = 1 + (\alpha Y/K),$$

that is,

$$SDR = STP = SOC.$$

The left-hand side of the basic equilibrium condition is the rate of social time preference, *STP*, while the right-hand side is the rate of social opportunity cost of capital, *SOC*. Note that neither side allows for risk, because each individual social investment project is assumed to have returns per head of the population that are small relative to and uncorrelated with national income.

Estimating this discount factor is problematic. Suppose, which can usually be expected to be the case, that the economy is not at an efficient equilibrium, but is at a point such as B in Figure 2.4, where the economy is underinvesting (that is, overconsuming) for next year compared with point E. Here the *STP* rate, the left-hand side of the equilibrium condition, is given by the flatter slope of the welfare contour compared with the *SOC* rate, the right-hand side of the equilibrium condition, which is given by the steeper slope of the production possibility frontier. Using either of these two rates to compute the social discount rate will result in an error: when there is underinvestment: $STP < SDR < SOC$.

Now that the essential building blocks of cost–benefit analysis have been established, the optimal allocation of resources in energy economics can be investigated.

4 Marginal Cost and Investment Decisions in Energy Supply

The application of cost–benefit analysis in energy economics was pioneered at Électricité de France in the 1950s (see Boiteux 1960). It came into English economics through the work of Turvey (1967, 1971) at the UK Electricity Council and subsequently spread worldwide through the work of Turvey and Anderson (1977), and Rees (1984). Other important theoretical contributions have been made by Crew and Kleindorfer (1979) (uncertainty and pricing), Littlechild (1970) (non-linear programming models), Wenders (1976) (tariff schedule implications) and Bohn et al. (1983) (spot and real-time pricing), among others. The textbook model needs to be amended to take account of capital-intensive energy production, transmission and distribution (Berrie, 1983; Stoft 2002). A principal distinction is between output and capacity to produce output. Both are measured in the same units: electricity = kilowatt-hours per hour (= kilowatts); gas: therms per day; oil: barrels per day or tonnes per year; coal: tonnes per year; renewables: tonnes of oil equivalent per year (that is, the amount of heat generated that is the same as the amount generated by burning 1 tonne of oil).

Assume that one unit of plant and equipment is used to produce one unit of output, and that it costs £c per period to hire this plant. Alternatively it costs £c per period to repay with interest the loan used to buy the plant. Once installed, it costs £r per period to operate 1 unit of plant to produce 1 unit of output. Note that r is the running or operating cost of 1 unit of power production; c is capacity cost of 1 unit of power production. Operating cost is constant up to the level of capacity installed, then it is infinite because no more capacity is available. Figure 2.5 illustrates this.

In this model:

$$SRMC = \begin{cases} r\text{: demand} \leq \text{capacity} \\ \infty\text{: demand} > \text{capacity} \end{cases}$$

$$LRMC = r + c.$$

The *SRMC* (short-run marginal cost) curve shifts to the right whenever more capacity is installed, and it always intersects *LRMC* (long-run marginal cost) from below, as shown in Figure 2.5.

Optimal resource allocation using cost–benefit analysis therefore requires:

1. Set price, $p = SRMC$ to ration demand to capacity, or to make maximum use of spare capacity: $p = m$, where m is whatever level of *SRMC* intersects the demand curve.

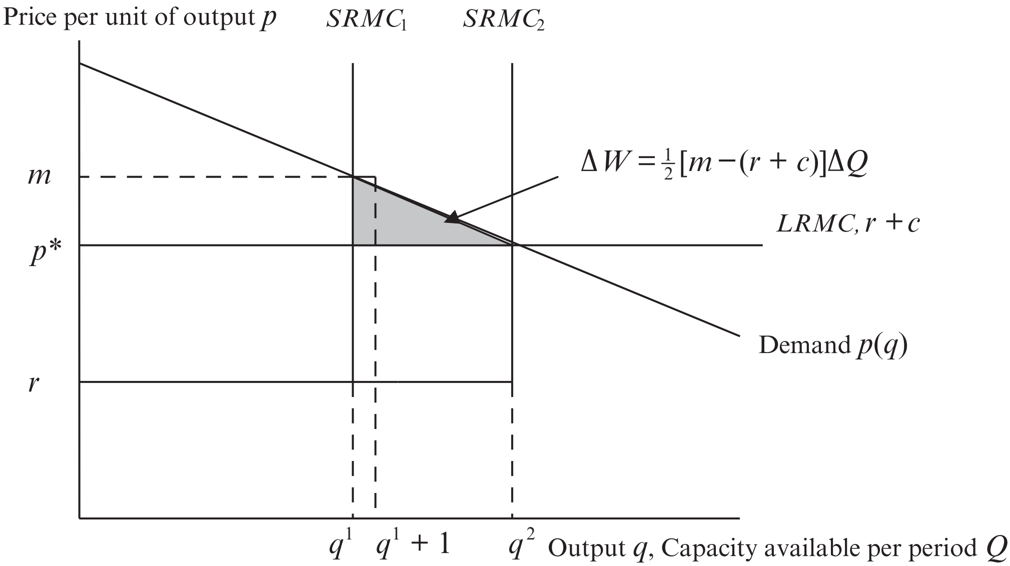


Figure 2.5 *Single-period marginal net benefit of increasing capacity by 1 unit and by* $\Delta Q = q^2 - q^1$ *units*

2. Compute the net benefit of changing capacity, and invest in or scrap capacity until the net benefit has been used up:

$$\Delta W = \tfrac{1}{2}(p - LRMC)\Delta q = \tfrac{1}{2}[m - (r + c)]\Delta q = 0.$$

This net benefit for a discrete change in capacity is shown as the shaded triangle of consumer and producer surplus[7] in Figure 2.5. At this point:

$$p^* = SRMC = LRMC.$$

It is often convenient to work in terms of a single unit change in capacity: $\Delta q = 1$ and in this case the net benefit is illustrated in Figure 2.5 by the rectangular sliver with base equal to $q^1 + 1 - q^1$. The marginal net benefit of 1 unit of capacity is:

$$dW/dp = (p - LRMC)(dq/dp) = [m - (r + c)](dq/dp).$$

Now think of a single unit of capacity and suppose it lasts for T years. The net present value of installing that unit over its life is:

$$NPV = \sum_{t=0}^{t=T} \frac{[m - (r + c)]}{(1 + i)^t},$$

and the optimal decision is to invest if $NPV > 0$. An alternative expression uses the total cost of installing 1 unit of capacity instead of the periodic repayment:

$$NPV = -C + \sum_{t=0}^{t=T} \frac{(m - r)}{(1 + i)^t}.$$

In the energy industry it is customary to write this in the reverse as net effective cost of capacity (*NEC*):

$$NEC = C - \sum_{t=0}^{t=T} \frac{(m - r)}{(1 + i)^t},$$

and invest if $NEC < 0$.

The *NEC* is the cost of installing 1 unit of capacity less the lifetime opportunity cost savings of having that unit and therefore not having to ration demand. Note the ingredients required in this recipe: (i) forecast of the market-clearing price of energy up to T years ahead (m), (ii) choice of discount rate, (i), and (iii) forecast of the technically efficient operating cost of capacity up to T years ahead.

This has led to a well-known controversy. If demand fluctuates or is uncertain, then *SRMC* pricing may become very volatile and scrapping and investment policy may show many changes of direction. Some economists have suggested setting price = *LRMC* all the time, and using non-price rationing, or maintaining surplus capacity to match demand with supply. This was the UK Treasury view in the 1970s–1980s for the electricity supply industry. The two opposing viewpoints are represented by Munasinghe and Warford (1982) and Newbery (1985). The analysis just completed sets out the essence of the merchant investment model of electricity and gas production. It proceeds as if each capacity investment decision is taken separately by a different competitive firm or merchant. This is the model that lies at the core of many major studies of power plant investment such as MIT (2004).

Much policy analysis of individual power plant and renewable technology decisions takes the merchant investor approach but it is not clear how to compare different technology choices in this model. The net present value criterion applies to a single plant type but different plant types may have different lifetime durations. One solution for comparing different plant types uses a system-based approach discussed later in this chapter. Another solution to this comparison problem which can be applied to the merchant investor problem is to use the annuitised *NEC*s for different technologies (Rees 1973).[8] Imagine $s = 1 \ldots S$ different technologies, with different lives: $T(s)$. Compute the annuity factor for each, that is, the annual constant sum for which the present value is equal to the *NEC* (or *NPV*) of the corresponding technology:

$$A\left\{C^s - \sum_{t=0}^{t=T(s)} [(m_t^s - r_t^s)/(1 + i)^t]\right\} = (i \times NEC^s)/[1 - (1 + i)^{-T(s)}].$$

Note the appropriate value for m_t^s varies with the type of capacity being evaluated. Choose the technology with the lowest annuitised *NEC* or highest annuitised *NPV*.

Another approximation used in many studies of energy investments, is based on levelised discounted cost. The purpose is to obtain an equivalent energy price (expressed in terms of gas or electricity or oil and so on) for each technology. This ignores system implications, and in effect treats each separate capacity investment as a mini-supply industry of its own. It asks what constant price through time, $\bar{p}$, would allow a plant operating independently to break even?

$$\sum_{t=1}^{t=T(s)} [\bar{p}^s q_t^s/(1+i)^t] = C^s + \sum_{t=1}^{t=T(s)} [r_t^s q_t^s/(1+i)^t],$$

so that the levelised discounted cost, *LDC* is:

$$\bar{p}^s = \frac{C^s + \sum_{t=1}^{t=T(s)} [r_t^s q_t^s/(1+i)^t]}{\sum_{t=1}^{t=T(s)} [q_t^s/(1+i)^t]},$$

that is, the present value of lifetime costs relative to the present value of lifetime energy delivered.

Figure 2.5 has become the most widely used investment tool by energy regulators, and governments, although not necessarily by energy utilities. Many widely publicised studies of electricity generation costs, for example, calculate *LDC* for different plant types and then recommended on the basis of lowest *LDC*.

There are several objections to this method of cost evaluation, although its ease of use and apparent financial soundness makes it very popular:

1. The forecast of energy refers to that generated by the plant, not the demand on the system so it assumes that the plant will largely maintain its position in the merit order of relative operating costs.
2. The calculation takes no account of the mix of other plant types on the system, and does not calculate running cost savings relative to these other plant types.
3. The calculation directly compares plants with different lives.

All of these factors mean that *LDC* expresses what the average discounted price of electricity would be in a hypothetical situation in which all of a utility's generating system was converted to the plant in question. *LDC* is logically coherent as an accounting calculation, but whether it is economically relevant to cost-minimising plant choice is another question.

5 Peak-load Pricing

The analysis can be extended to cover several periods of demand when energy cannot be stored from one period to the next. The critical idea is that a period – day, week, month, year – is composed of a cycle of subperiods each with its own demand schedule. For example, in electricity supply a 24-hour day consists of two demand subperiods: daytime peak demand and night-time off-peak demand. In telecoms we might distinguish weekday from weekend calls in a 7-day cycle of subperiods. Gas demand fluctuates between summer and winter. Figure 2.6 assumes two subperiods of equal length in each cycle for convenience, labelled with superscript 1 for off-peak demand and superscript 2 for peak demand. The lower demand curve ($p^1(q)$ corresponding to the prices labelled p_0^1 and p_1^1) represents off-peak demand, and it lies entirely below the upper demand curve ($p^2(q)$ corresponding to the prices labelled p_0^2 and p_1^2) which represents peak demand).

The cost assumptions are a development of those used earlier. The critical aspect of

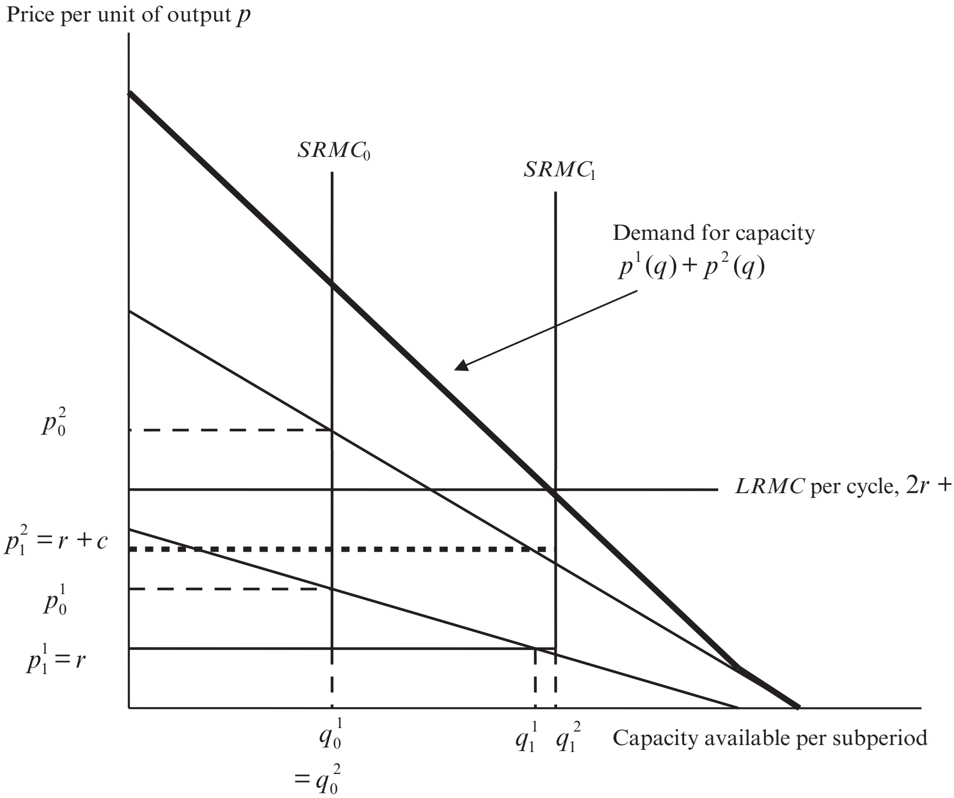


Figure 2.6 *Two-period peak and off-peak pricing when capacity is below optimum:* $p_0^1 + p_0^2 > 2r + c$*, and at the optimum:* $p_1^1 + p_1^2 = 2r + c$

capacity is that it is available for both peak and off-peak demand. These demands are not rivals for the same capacity. This is called the 'public good nature of peak demand' and allows us to sum the demand curves for each subperiod vertically to obtain a demand for capacity curve for the cycle of subperiods.

Once installed, it costs £r per subperiod to operate 1 unit of plant for 1 subperiod to produce 1 unit of output per subperiod. Capacity that produces 1 unit of output for the whole cycle incurs £$2r$ operating cost. The investment rule requires that price = *LRMC* = $2r + c$, but price for the cycle is a hypothetical concept constructed by summing the peak and off-peak demand curves vertically to represent the demand for capacity curve: $p^1(q) + p^2(q)$. Now the pricing rule requires that demand is rationed to capacity in each subperiod, by charging a price equal to or greater than operating cost, r:

$$p^1 = r + k_1,$$

$$p^2 = r + k_2.$$

The investment rule requires that the vertical summation of the peak and off-peak prices should cover $LRMC = 2r + c$ for the cycle of subperiods:

$$p^1 + p^2 = (r + k_1) + (r + k_2) = 2r + c \Rightarrow k_1 + k_2 = c.$$

The general properties of the solution are clear from Figure 2.6. Two different positions are illustrated. With the upper limit to capacity given by the $SRMC_0$ curve, the prices which ration demand to capacity are p_0^1 in the off-peak period, and p_0^2 in the peak period. At this point both prices exceed operating cost, and each subperiod's demand makes a contribution to capacity cost, the capacity payment (k_1 or k_2, with $k_2 > k_1$). The distances $p_0^1 - r$ and $p_0^2 - r$ represent these capacity payments in Figure 2.6 when capacity is limited along $SRMC_0$. However, $SRMC_0$ is not an equilibrium outcome; there is a positive net benefit to increasing capacity, represented in Figure 2.6 by the fact that the vertically summed demand for capacity curve $p^1(q) + p^2(q)$ intersects $LRMC$ further to the right at a capacity level represented by $SRMC_1$. The willingness to pay for an extra unit of capacity at the margin exceeds the marginal cost of another unit of capacity. This net benefit is captured by expanding capacity until $p^1(q) + p^2(q) = 2r + c$, and at this level the prices which ration demand to capacity are p_1^1 in the off-peak period, and p_1^2 in the peak period. Figure 2.6 illustrates two different possible shapes for the demand profile and the distribution of capacity payments across periods. At the initial capacity level represented by $SRMC_0$, both off-peak and peak prices exceed the operating cost in order to ration demand to capacity. This has the effect of removing the actual peak in demand and the resulting load profile is flat with the same power consumption in both subperiods: $q_0^1 = q_0^2$. However, in this example, it is the strength of peak demand that generates most of the positive net benefit of expanding capacity. When this has occurred, the optimal prices are such that all of the capacity cost is recovered from the peak period: $p_1^2 = r + c$ while the off-peak demand covers operating cost only: $p_1^1 = r$. A consequence of this is that the load profile is no longer flat and an actual peak in consumption has occurred: $q_1^1 < q_1^2$.

Another useful way of thinking of the issue is this. If the only way of meeting peak demand is to build more capacity, the difference between the peak and off-peak prices must equal the willingness to pay for more capacity at the peak less the willingness to pay for more capacity in the off-peak period: $k_2 - k_1 \leq c$.

6 A Simplified Spot Pricing Model with and without Random Demand

An important model of energy markets such as gas and electricity is the competitive spot pricing equilibrium where the corresponding welfare-maximising equilibrium is analysed using Kuhn–Tucker nonlinear programming analysis (similar to classical Lagrangean optimisation) to construct a simple model. A much more detailed review of this topic is contained in the masterly survey paper by Crew et al. (1995), who, in particular, discuss the issue of modelling actual rather than planned consumer surplus.

$B(y_t)$ is the aggregate benefit function, associated with demand of y in period t.

Assume that the marginal benefit of electricity at a given level of consumption is its market price, $B'(y_t^*) = p_t$. The aggregate benefit could be the consumer surplus plus the revenue component of producer surplus:

$$B(y_t) = \int_0^{y_t^*} p_t(y_t)\,dy_t,$$

that is, the area under the inverse demand function $p_t(y_t)$. Net welfare benefit is then $B(y_t) - Cost$. Assume that there is a finite value for the aggregate benefit of the first unit of consumption: $B(0) = V^*$. This is taken as the willingness to pay to avoid loss of consumption, and in energy market terms is the *value of lost load.*

x_t is the load produced in period t which may differ from the demand y_t,
q is the capacity installed for all periods $t = 1 \ldots T$,
e_t is the excess of demand over output load available in period t, so that $e_t \equiv y_t - x_t$; therefore e_t is the random variable in the model when uncertainty of demand is permitted,
r_t is the operating cost of output per unit in period t, and
β is the unit cost of new capacity installed; installed capacity is $q^* = q/a$ where a is availability of capacity.

When there is no uncertainty, the standard model for one plant and many equal length subperiods is:

$$\max W = \sum_{t=1}^{t=T} B(y_t) - \sum_{t=1}^{t=T} r_t x_t - \beta q,$$

subject to the demand constraints: $x_t \geq y_t$, $t = 1 \ldots T$ with dual variables: m_t and the capacity constraints: $x_t \leq q$, $t = 1 \ldots T$ with dual variables: k_t. The Lagrangean is:

$$L = \sum_{t=1}^{t=T} B(y_t) - \sum_{t=1}^{t=T} r_t x_t - \beta q + \sum_{t=1}^{t=T} m_t(x_t - y_t) + \sum_{t=1}^{t=T} k_t(q - x_t)$$

The firm chooses to maximise net economic benefit with respect to y_t, x_t, q, because it chooses capacity, price and output simultaneously, but not independently. The necessary conditions are:

$$\partial L/\partial y_t = p(y_t) - m_t \leq 0, \quad y_t(\partial L/\partial y_t) = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial x_t = -r_t + m_t - k_t \leq 0, \quad x_t(\partial L/\partial x_t) = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial q = \beta - \sum_t k_t \leq 0, \quad q(\partial L/\partial q) = 0,$$

$$\partial L/\partial m_t = x_t - y_t \geq 0, \quad m_t(\partial L/\partial m_t) = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial k_t = q - x_t \geq 0, \quad k_t(\partial L/\partial k_t) = 0, \quad t = 1 \ldots T.$$

Assume an interior optimum: $y_t, x_t, q > 0$, then these conditions are written:

$p_t = m_t$: price equals marginal cost on the system,
$m_t = r_t + k_t$: system marginal cost equals operating cost plus capacity payment, and
$\sum_t k_t = \beta$: sum of periodic capacity payments equals the cost of capacity.

These conditions apply to a span of time periods that could cover one day or a cycle of subperiods, but they generalise to optimisation over many years with the addition of a discount factor; for example, to make an investment decision compare the present value of lifetime capacity payments to the lifetime capacity cost:

$$\sum_t [k_t/(1 + i)^t] = \beta.$$

The conditions also generalise to many different types of capacity: $s = 1 \ldots S$ with the addition of an appropriate subscript, and an aggregated form of the demand constraint:

$$\sum_{s+1}^{s=S} x_{st} - y_t \geq 0.$$

Then, for example, the marginal cost calculation is:

$$m_{1t} = r_{1t} + k_{1t} = \ldots = m_{st} = r_{st} + k_{st} = m_{St} = r_{St} + k_{St}.$$

This last result is illustrated in Figure 2.7, which is based on Turvey (1971).

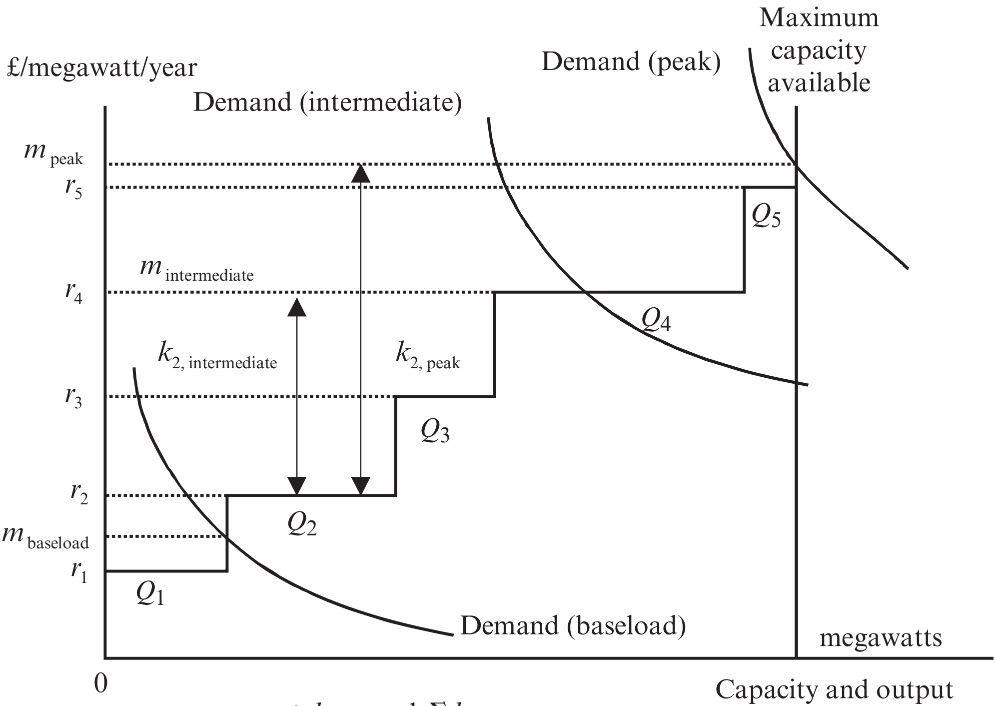
In the figure, five different types (or vintages) of capacity are shown with installed values of $Q_1 \ldots Q_5$. They are arrayed in ascending order of operating cost to represent the idea of the merit order of plant dispatch. Critically the long-run marginal cost is no longer immediately obvious. Since the optimisation solves the pricing and investment model simultaneously, the system marginal cost is a measure of both short- and long-run marginal cost.

With uncertainty, model specification is particularly important. The basic idea in this model is to penalise the proximity of load to available capacity, and this can be demonstrated in a very simple setting. Load shedding or the use of unsatisfied demand is now included in the model. In particular, it is necessary to distinguish between potential demand associated with the current price and the actual load which can be delivered. This simple model is based on Stoft (2002, p.136).[9] It uses the concept of lost load, served load and states that the sum of the two is defined as load: $y_t \equiv x_t + e_t$. The difference between potential demand and actual load can now be positive: $e_t \equiv y_t - x_t$, and this random variable has a known probability density function, $f(e_t)$. The cumulative distribution function defines the probability of any given size of outage:

$$F(e_t^*) = \int_{-\infty}^{e_t^*} f(e_t)\, de_t = \text{prob}(e_t \leq e_t^*),$$

and two values are of interest: the probability of non-positive outage (no load shedding), $F(0)$, and the probability of positive outage: $1 - F(0)$.

The cost of load which is shed is: V^* per unit of $y_t - x_t = e_t$, that is, the value of lost



$m_{st} = r_s + k_{st}$ and $\Sigma k_{st} = c_s$

e.g.

$m_{\text{intermediate}} = r_1 + k_{1,\text{ intermediate}} = r_2 + k_{2,\text{ intermediate}} = r_3 + k_{3,\text{ intermediate}} = r_4$

Source: Based on Turvey (1971).

Figure 2.7 Multiplant and multiperiod marginal cost of energy generation

load. The demand constraints of the certainty model remain: $x_t - y_t \geq 0$; however, they will have shadow prices that include the probability that the constraint is binding. The constraint may be violated if load shedding is allowed, and then this is penalised by an additional term in costs that again reflects the probability of this occurring, that is, the probability of lost load: $1 - F(0)$. The problem has the Lagrangean function:

$$L = \sum_{t=1}^{t=T} B(y_t) - \sum_{t=1}^{t=T} r_t x_t - \beta q - \sum_{t=1}^{t=T} [1 - F(0)] V^*(y_t - x_t)$$

$$+ \sum_{t=1}^{t=T} [m_t F(0)](x_t - y_t) + \sum_{t=1}^{t=T} k_t (q - x_t).$$

Note how the demand constraints have been replaced by an expression composed of two terms: the first records positive outages: $y_t - x_t \equiv e_t > 0$ which are associated with an expected monetary cost, $[1 - F(0)]V^*$, and the second records non-positive outages with an expected shadow cost: $F(0)m_t$. The necessary conditions for this simplified statement of the problem read:

$$\partial L/\partial y_t = p(y_t) - [1 - F(0)]V^* - F(0)m_t \leq 0, \quad y_t(\partial L/\partial y_t) = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial x_t = -r_t + [1 - F(0)]V^* + F(0)m_t - k_t \leq 0, \quad x_t(\partial L/\partial x_t) = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial q = \beta - \sum_t k_t \leq 0, \quad q(\partial L/\partial q) = 0,$$

$$\partial L/\partial [F(0)m_t] = x_t - y_t \geq 0, \quad F(0)m_t\{\partial L/\partial [F(0)m_t]\} = 0, \quad t = 1 \ldots T,$$

$$\partial L/\partial k_t = q - x_t \geq 0, \quad k_t(\partial L/\partial k_t) = 0, \quad t = 1 \ldots T.$$

Assume an interior optimum: $y_t, x_t, q > 0$, then these conditions can now be interpreted in a simple way. Refer to the probability of positive outage as loss of load probability, *LOLP*:

$$LOLP \equiv 1 - F(0),$$

and refer to short-run marginal cost as system marginal price, *SMP*:

$$SMP \equiv m_t.$$

Then:

$$p_t = [1 - F(0)](V^* - m_t) + m_t = LOLP(V^* - SMP) + SMP$$

$$= LOLP \times V^* + (1 - LOLP) \times SMP.$$

That is, the spot price equals the loss of load probability times the value of lost load plus the probability of maintaining load times the system marginal price. System marginal price is the cost of the marginal production unit and equals operating cost plus capacity payment. The capacity payments sum to the cost of capacity:

$\sum_t k_t = \beta$: sum of periodic capacity payments equals the cost of capacity,

but each now has two components which depend on the loss of load probability:

$$k_t = [1 - F(0)](V^* - r)_t + F(0)(m_t - r_t).$$

These are the standard results in the spot pricing literature: in each half hour the efficient spot price equals:

marginal generation cost + *marginal capital cost . . . no uncertainty case*,
weighted average of marginal generation and outage costs . . . uncertainty case.

Outages are modelled as output from non-existent capacity which has zero capacity cost but a very high operating (outage) cost. An important qualification remains, however. The model with certainty has a set of *ex ante* price relationships that will automatically be realised in practice because uncertainty is absent. This is not the case in the model with uncertainty; the *ex ante* relationships are based on the maximisation of expected net welfare benefit, but the actual *ex post* realisation will be different. To handle the divergences between expected values of the variables and their realisations there should also exist an *ex post* balancing market. Thus the uncertainty model outlines the equilibrium before trading, but the real-time spot market must allow instantaneous adjustment of *ex ante* values to realised outcomes.

7 Energy Market Architecture

The analysis to this point gives an insight into the price relationships at the efficient allocation of resources. However, the welfare maximisation model has been used only as a means of simulating the competitive outcome. The mechanism for achieving this outcome still relies on competitive markets rather than centralised regulation, as the issue of capacity payments highlights. In the uncertainty model the capacity payment which covers the cost of building new capacity depends on the strength of demand at the peak relative to the system marginal price. In early applications of spot pricing with investor-owned producing firms, many market designs arranged for separate capacity payments with regulator-determined value of lost load in addition to system marginal price recovery.[10] However, such a regulated market architecture is open to abuse of market power if a producer with sufficient capacity can increase loss of load probability by withdrawing nominated capacity availability at the last moment. Consequently, during the years after 2000, many spot markets such as that of the UK moved away from a pool with separate capacity payments. The efficient spot market outcome was left to competitive entrants to make offers and bids to supply through individual negotiated contracts with a balancing market to adjust realised values to *ex ante* planned supply and demand. The mechanism for spreading the risk of faulty contracting is an active market in financial options related to spot and forward electricity and gas contracts (Stoft 2002; Wolak 2006). In the UK case, it is arguable that the stimulus to a more efficient wholesale market after the disappearance of the capacity payments system owes as much if not more to competitive entry by new generators as it did to the evolution of new trading and contracting arrangements (Evans and Green 2005). Consequently, it is important to keep in mind that the structure of the welfare maximisation model is not a guide to market architecture; it simulates the competitive outcome, but it is still the mechanism of free entry and exit in response to profit incentives that implements the spot pricing equilibrium.

Joskow (2006) has suggested some practical critical ingredients for liberalised electricity markets on the basis of several years of international experience. In Table 2.1, Joskow's architecture for energy market reform to replace public or state-owned utilities (POUs) with investor-owned utilities (IOUs) is summarised. Several of the ideas raised in Joskow's table are considered below, including the rate of entry into energy markets, and access to networks. Wholesale market spot prices can even be signalled to retail consumers with the option of a fixed price tariff instead.

8 Competition in Wholesale Energy Markets

In UK energy markets a classic case study of the competition in wholesale power markets concerns the trading arrangements for electricity in England and Wales. The analysis of

Table 2.1 Architecture for energy market reform

Component		Policy	Objectives
a.	IOUs	Privatise state-owned utilities	High-powered incentives, non-political objectives, hard budget constraints
b.	Separation	Vertical separation of generation, transmission, distribution and supply	Barriers to cross-subsidisation, and discrimination against access
c.	Demerger	Horizontal demerger of generation	Wholesale market competition
d.	Integration	Horizontal integration of transmission	Single *independent* system operator for system reliability and economic standards
e.	Wholesale market	Voluntary public wholesale spot energy and operating reserve markets	Support for real-time supply–demand balancing, economic trading, quick response to outages
f.	Demand-side response	Develop active demand-side institutions	Consumer demand-side response to wholesale prices
g.	Access	Promote efficient access to transmission network	Efficient competitive production and exchange, and allocates scarce transmission capacity among competing users
h.	Unbundling	Unbundle retail tariffs into retail power supply and delivery charges	Competition in supply separate from regulated (natural monopoly) distribution and transmission
i.	Economic procurement	Benchmark supply costs for small consumers	Yardstick for supply by distribution company where small consumers not open to competition
j.	Independent regulation	Independent regulatory authority with expert staff	Performance-based regulation using good information to regulate for distribution and transmission, e.g., yardstick competition
k.	Transition	Transition from POUs to IOUs	Mechanisms compatible with competitive markets

Source: Joskow (2006).

Green and Newbery (1992) suggested that the small number and concentrated size of the original market participants led to a Nash equilibrium in supply schedules (offer curves) that produced large efficiency losses. They suggested that firms used market power to manipulate the availability of capacity in order to push up capacity payments, and increase the marginal price of electricity. This produces the *policy implication* that divestment of plant and enhanced competitive entry is required to improve competition in electricity generation, but a difficulty with the analysis of markets with a finite number of firms is to determine the optimal number in the market. Many oligopoly models use the Nash equilibrium for a Cournot game in which firms choose output quantities to maximise profit, taking the quantity of output from rivals as given. Powell (1993) used such a model to show that forward contract commitments would reduce the ability of firms to exercise spot market power. In power markets, however, it is often more interesting to focus on price-setting behaviour. In a two-player Bertrand game each duopolist chooses his/her price, taking the other's price as given: for example, for duopolist 1, where π, p, c, q are respectively profit, price, marginal cost and output, the model states:

$$\max_{\{P_1:\text{ given } p_2\}} \pi_1 = (p_1 - c)q_1.$$

Here, a pure strategy Nash equilibrium has five properties (Rasmusen 1994), where market demand is written $p = a - bQ$, and c is marginal cost (the same for each firm):

if $p_1 < p_2$, then $q_1 = Q = (a - p_1)/b$ and $q_2 = 0$;
if $p_2 < p_1$, then $q_2 = Q = (a - p_2)/b$ and $q_1 = 0$;
if $p_1 = p_2 = p$, then $q_1 + q_2 = Q = (a - p)/b$;
neither deviates and the unique equilibrium is where $p_1 = p_2 = c$.

The essence is that the lower-price duopolist captures the whole market. Prices cannot differ because the higher-price duopolist can respond by shaving price sufficiently to capture the other's market share. This stops when each has shaved price to marginal cost. Any division of the market is then a Nash equilibrium because each just breaks even while any deviation of price from marginal cost will mean zero or negative profits.

In a classic paper, Klemperer and Meyer (1989) described a way of extending the Cournot and Bertrand models. Instead of saying that players must choose either quantity or price as the strategic variable, they argued that each firm would look for its profit-maximising supply *curve* relating quantity to price. Hence this is called a 'supply function' model. Here the Nash equilibrium strategies consist not of a set of outputs or a set of prices but a set of supply functions stating how much each firm will supply for any given market price: $q_i = q_i(p)$. There have been several applications of this model, particularly to markets where a small number of firms participate in auctions to supply a product and each firm's bid consists of both a nominated supply quantity and a price that is required for that supply to be available. This is very relevant to the nature of spot energy markets where energy producers bid in supplies and prices to a daily market organised by an independent system operator, as envisaged in the Joskow architecture. Note that there is an additional problem with auctions because the difficulty of monitoring the firms' signals to each other and the transparency of any firm cheating tend to encourage cartel bidding. The version of the Klemperer–Meyer model used here is that of Green (1996) and Green and Newbery (1992). Green's (1996) model is restricted to linear supply functions.

Market demand is $D = q_1 + q_2$ and market demand is a function of market price: $D = D(P)$. Each firm thinks of the market price as its strategic variable and recognises that its share of market demand is the difference between total demand and the other firm's share: $q_1 \equiv D - q_2$ and $q_2 \equiv D - q_1$. It takes as given the other firm's supply function: $q(P)$. In the derivation, note that this identity holds:

$$\frac{dq_1}{dP} \equiv \frac{dD(P)}{dP} - \frac{dq_2}{dP}.$$

Firm 1:

$$\max_{P,\, q_2(P)\, given} \pi_1 = Pq_1 - TC(q_1) = P(D - q_2) - TC(q_1)$$

$$\text{so } \frac{d\pi_1}{dP} = (D - q_2) + (P - MC_1)\left(\frac{dD}{dP} - \frac{dq_2}{dP}\right) = 0$$

$$\text{solve: } q_1 = (P - MC_1)\left(-\frac{dD}{dP} + \frac{dq_2}{dP}\right),$$

and Firm 2:

$$\max_{P,\, q_1(P)\, given} \pi_2 = Pq_2 - TC(q_2) = P(D - q_1) - TC(q_2)$$

$$\text{so } \frac{d\pi_2}{dP} = (D - q_1) + (P - MC_2)\left(\frac{dD}{dP} - \frac{dq_1}{dP}\right) = 0$$

$$\text{solve: } q_2 = (P - MC_2)\left(-\frac{dD}{dP} + \frac{dq_1}{dP}\right).$$

These are a pair of simultaneous differential equations, so the solutions take the form of equations: $q = q(P)$ rather than numbers. Now we restrict our search for the solutions to linear supply curves of the form: $q = \beta P$. We assume linear marginal cost curves: $MC_i = c_i q_i$ and a linear market demand curve:

$$D = a - bP \Rightarrow \frac{dD(P)}{dP} = -b.$$

Our differential equation response functions are:

$$\beta_1 P = (P - c_1\beta_1 P)(b + \beta_2)$$

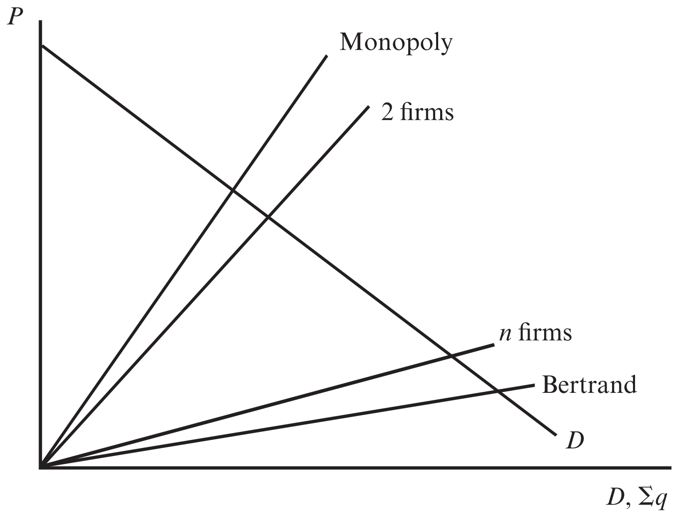
$$\beta_2 P = (P - c_2\beta_2 P)(b + \beta_1),$$

but P can be cancelled:

$$\beta_1 = (1 - c_1\beta_1)(b + \beta_2)$$

$$\beta_2 = (1 - c_2\beta_2)(b + \beta_1).$$

The problem now is to solve for β_1 and β_2, the slope of each firm's best response supply function with respect to the P axis. Market supply is then the horizontal summation of the individual supply curves: $\beta_1 P + \beta_2 P = (\beta_1 + \beta_2)P$. The critical question asked by Green and Newbery is this: will the market supply curve approximate the competitive industry marginal cost curve (what they call the 'Bertrand curve') or will it approximate the cartel monopoly bidding curve? Even more interesting is this question: how many



Nash equilibrium supply functions

Source: Green and Newbery (1992).

Figure 2.8 Green and Newbery power spot market model

new entrants are needed to ensure that the market supply curve is close to the Bertrand curve? The effect of adding more firms is to make the market supply curve steeper with respect to the price axis, that is, flatter with respect to the quantity axis:

$$\beta_1 P + \cdots + \beta_n P = \left(\sum_{i=1}^{i=n} \beta_i\right) P,$$

but it also increases the number of equations to be solved simultaneously:

$$\beta_i = (1 - c_i\beta_i)\left(b + \sum_{j \neq i}^{n} \beta_j\right), \ i = 1 \ldots n.$$

The effect is shown in Figure 2.8. The steepest supply function is the bidding curve for a monopoly firm, and the least steep supply schedule is that corresponding to a number of firms which behave as if they comprise a Bertrand–Nash equilibrium in supply schedules. As the number of firms entering the market increases, each maximising profit while taking the bidding supply schedule of the others as given, the aggregate of the supply schedules moves closer to the efficient Bertrand equilibrium schedule.

How many firms are needed for efficient resource allocation? Green and Newbery (1992) simulated the UK spot electricity market shortly after privatisation in 1990, and argued that with five or more players of equal scale, the aggregate Nash equilibrium supply schedule hardly differed from the Bertrand equilibrium supply schedule in terms of the estimated deadweight welfare loss. Consequently, although efficiency of outcome required more than simply splitting the incumbent monopolist into two separate players,

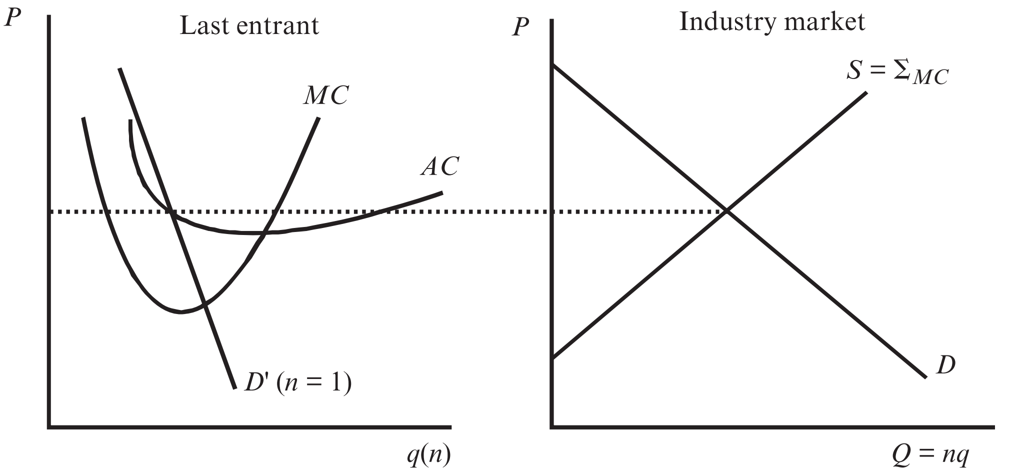


Figure 2.9 Incentive to new entry

nevertheless a feasible and finite number of entrants would deliver an outcome relatively close to a competitive equilibrium.

9 What Determines the Optimal Degree of Entry to Market?

A simple model of the optimal number of firms in an industry is given by Armstrong et al. (1994, p. 107). This is explained below.

The aggregate output of an industry is $Q = nq$ where n is the number of firms and q is the average output per firm. Generally we expect the average output per firm to fall as the number of firms rises. Trading off producer surplus against consumer surplus suggests that the net economic benefit of an extra entrant to an industry is:

Profit of last entrant + effect of last entrant in lowering price towards marginal cost.

The first term accrues to the producer while the second accrues to the consumers

The net benefit can be zero for two reasons. First, if both terms are zero and n is very large, that is, if there are constant or decreasing returns to scale then the gain from more entry is zero when both the last entrant's profit is zero, and price equals marginal cost. Second, if the two terms are non-zero but cancel out when n is small. If there are increasing returns to scale (or fixed costs are important), then when entry has pushed profit down to zero, this implies $P = AC > MC$. The second term is negative and entry is excessive. The more important the fixed costs, or increasing returns to scale, the lower should be n. Figure 2.9 illustrates. First assume that demand is large relative to the output of a single firm. This means start with the diagram on the right and ignore the demand curve labelled D' in the left diagram. If demand is large enough (D) then free entry leads to $P = MC > AC$ and further entry incentives exist until $P = AC = MC$. Now instead assume that the output of one firm is large relative to the market demand. Ignore the diagram on the right and assume that market demand is: D'. If minimum efficient scale is high relative to market demand (D') then entry should stop when $P = AC > MC$.

The formal analysis of the effect on a firm's output as the number of firms rises is given by:

$$\frac{dq(n)}{dn} = q'(n) < 0,$$

$$\frac{dW}{dn} = \pi_n + (P - MC)nq'(n)\;(>0)\;+\;[(>0)\;\times\;(<0)].$$

Here π_n is the profit of the last entrant, and the second term is the effect of an extra firm in reducing the $P - MC$ gap. The optimal number of firms is:

$$n = -\frac{q(P_n - AC_n)}{q'(n)(P_n - MC_n)}.$$

10 The Access Pricing Problem

The Joskow (2006) architecture argues for vertical de-integration of different aspects of power supply, but there is a link between vertical integration and the important topic of access pricing. In vertical integration the key question is the determination of the price of input charged to a downstream firm by the upstream firm. In access pricing, one firm owns the network for distributing the commodity to the final consumers. It could but need not be vertically integrated. However, there is now another firm, the third party, which wishes to supply the commodity to some of these final customers. It can obtain the commodity as input (the third party might be an upstream firm) but must use the available network owned by the downstream firm. The downstream firm can charge for access to this network. The access charge must cover the network costs associated with the customers which the third-party firm detaches from the network owner. These costs may be very difficult to measure separately. What is the marginal opportunity cost of access that will form the basis of an efficient access price? Baumol and Sidak's (1994) efficient component pricing rule argues that the marginal opportunity cost of access is the profit forgone by the network owner in permitting the third party to detach some customers that the downstream firm would otherwise supply. In principle, profit per customer is calculable, but it will be difficult to distinguish the fraction which covers network costs from the fraction which reflects the downstream firm's market power. The access pricing problem is discussed in detail by Armstrong et al. (1994).

Facilitating competition in the electricity and gas industries requires non-discriminatory open access to the transmission and distribution network, for all producers and suppliers. Identify an incumbent network owner (firm 1) and a competitive supplier (firm 2). Given that mce_1 describes the incumbent's marginal cost of energy, then its price for final downstream supply is:

$$P_1 = mce_1 + ica + \pi,$$

where *ica* is the incremental cost of access provision, and π shows the profit mark-up to ensure financial viability Therefore the price of access is:

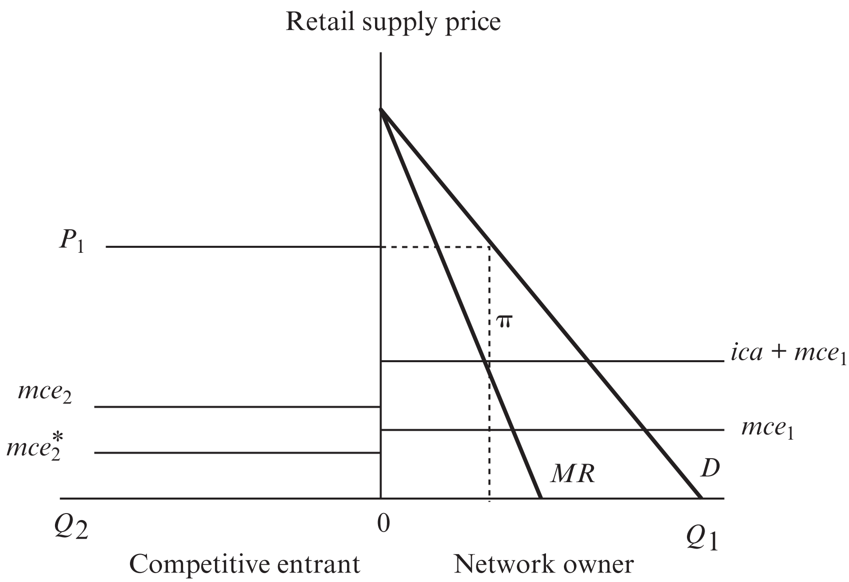


Figure 2.10 Unbundling products for access pricing

$$P_a = ica + \pi = ica + [P_1 - (mce_1 + ica)],$$

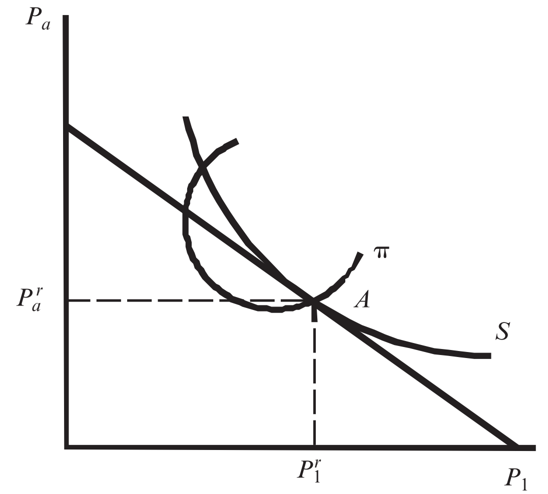
and this describes a version of the efficient component pricing rule (ECPR) (Vickers 1997). The competitive supplier will only enter the market if it has a lower marginal cost of energy, mce_2^* than the incumbent (Figure 2.10), but entry is inefficient at an entrant's energy source cost of mce_2. ECPR therefore discourages inefficient entry because the new supplier not only has to pay an access charge, it also has to pay the opportunity cost of access which includes the incumbent's lost profit. The model allows us to unbundle services, in this case distribution and supply. The ECPR model identifies efficient entry conditions, but assumes that regulatory issues are resolved elsewhere. Figure 2.10 shows efficient entry where the network owner is an unregulated monopolist.

Product differentiation can exist in the supply market, as in any other competitive market. Vickers extends the model to write:

$$P_a = ica + \sigma[P_1 - (mce_1 + ica)],$$

where σ is the displacement ratio, defined as the ratio of (a change in output sales for the incumbent with respect to the access price) to (a change in supply of access to new entrants with respect to the access price).

Three assumptions are made about the displacement ratio to ensure unity: homogeneous products; fixed coefficients technology (one unit of output requires one unit of access); and no bypass (the incumbent supplies all access via its distribution network). The first of these assumptions may be relaxed. Consequently when the demand for access by a new entrant increases by 1 unit, the incumbent will not see a 1 unit reduction in



Source: Derived from Vickers (1997).

Figure 2.11 Global price cap

demand for its product, because of customer inertia, brand loyalty, and the like, inducing $\sigma < 1$. Product differentiation will lower the access price relative to homogeneous products.

The regulatory issue of the network owner's profitability remains. Laffont and Tirole (1996) have suggested a global price cap in which the intermediate good (access) is treated as a final good and included in the computation of the price cap. This treats access and supply symmetrically in a Ramsey pricing framework. Laffont and Tirole contrast this with 'the general view that intermediate and final goods are to be treated asymmetrically' (pp. 244–5).

The efficiency gain of using the global price cap suggested by Laffont and Tirole can be neatly illustrated in Figure 2.11, which is derived from Vickers (1997). The final price and the access price are displayed on the horizontal and vertical axes. Separately regulated price caps are shown at point A as P_a^r and P_1^r. This pair of prices will generally lie on an iso-profit contour labelled π, and an indifference curve of consumer surplus labelled S. Consumer surplus improvements are represented by S contours closer to the origin, while profit gains to the firm are represented by π contours further from the origin. All of the area above the profit contour and below the consumer surplus contour represents price pairs which are more efficient than the pair at A. We can construct a global price cap: $wP_a + (1 - w)P_1 = \overline{P}$ through point A such that points between the locus and the π contour are more efficient than A without the consumer paying more in aggregate than at A. If the weights are proportional to the actual quantities consumed at A, the locus will be tangential to the S contour at A. Any chosen combination in the area between the locus and the profit contour will approximate to a more efficient entry-access allocation than the one implied by the separate price caps, and will yield a Ramsey pricing outcome. The incumbent will concentrate where it has a comparative advantage, reflecting Bertrand entry in ECPR.

A regulator may opt for maximum price limits to protect customers who will initially not benefit from competition. For access pricing this has the following effect, reflecting what Laffont and Tirole describe as the general asymmetric approach:

$$P_1 \leq \overline{P_1} \text{ and } P_a \leq \overline{P_a},$$

where the access price cap is determined by the distribution and transmission price controls. Firms would be expected to publish indicative charges well in advance of implementation, and efficiency requires that these are the same for each entrant to a particular supply market.

Access pricing may need floors and ceilings to prevent inefficient suppliers entering the market or to prohibit barriers to entry. Without use of a global price cap, Vickers worries about the distortion arising from partial regulation, a special case of this. If the access price is regulated, $P_1 - P_a$ will widen, increasing productive inefficiency, as less efficient rivals enter the market. To prevent predatory pricing, on the other hand, as a result of some competitive energy costs being allocated to the regulated business, suggests a constraint such as:

$$P - a \geq MC_1.$$

However, if a firm's distribution and supply business were separated into two companies, each with its own terms of license the possibility of cross-subsidy would no longer arise.

11 Conclusions

This chapter has attempted to make a broad survey of the theoretical core ideas in energy economics. The initial discussion used the idea of Pareto-efficient outcomes and social cost–benefit analysis to establish the benchmark competitive and efficient allocations of energy resources. An important ingredient is the choice of social discount rate which was first explained in terms of an optimal saving and growth model. The core ideas of efficient resource allocation were then applied to investments in new energy supply and capacity, and this was shown to be intimately related to the idea of marginal cost pricing. The measurement of marginal cost in multiple plants and multiple time period investment planning models of energy supply followed and was demonstrated in a spot pricing model with uncertainty. Having described the ideas of efficient resource allocation in an energy context, attention turned to the practical implementation in real-world energy markets. An architecture for efficient energy markets was suggested by Joskow (2006), and this was used as a context to investigate the role of entry by investor-owned firms into energy markets. Feasible competition was demonstrated with a finite number of entrants, but care is necessary in determining the optimal market design. One important aspect of this is the access pricing problem since much of energy supply is delivered through pipes and wires.

Notes

1. Incentive regulation of energy industries is treated in Chapter 21.
2. The fundamental theorems of welfare economics state that (i) every competitive equilibrium is a Pareto optimum, and (ii) for every Pareto optimum there is a competitive equilibrium.
3. The Pareto criterion states that an allocation of resources is optimal if no person can be made better off without making another worse off.
4. Since producer surplus is the difference between revenue and the area under the curve representing the horizontal summation of the marginal cost curves, it strictly excludes fixed cost and therefore is less than economic profit.
5. Strictly this should be the slope of the aggregate of the compensated demand curves.
6. There is no assumption that compensation is actually paid, otherwise the Pareto criterion itself would be satisfied.
7. In Figure 2.5, additional producer surplus from this capacity to change is zero, since the long-run marginal cost is constant.
8. Rees demonstrates that this is equivalent to comparing the NEC of consecutive programmes of identical investments in the different technologies where the investment programmes have a common lifetime factor.
9. Stoft (2002, pp. 48 and 136) discusses economic demand as the amount of power that would be consumed if the system were operating normally for all consumers. Shed load is included as part of demand.
10. The UK Pool market after privatisation, 1990–2000 is an example of this. Many US power markets remained in this situation subsequently (Joskow 2006; Wolak 2006).

References

Armstrong, M., S. Cowan and J. Vickers (1994), *Regulatory Reform: Economic Analysis and British Experience*, Cambridge, MA: MIT Press.

Baumol, W.J. and G. Sidak (1994), *Toward Competition in Local Telephony*, Cambridge, MA: MIT Press.

Berrie, T. (1983), *Power System Economics*, London: Peregrinus on behalf of Institution of Electrical Engineers.

Bohn, R., F. Schweppe and M. Caramanis (1983), 'Using spot pricing to co-ordinate deregulated utilities, customers and generators', in J. Plummer (ed.), *Electric Power Strategic Issues*, Arlington, VA: Public Utility Reports Inc., 265–82.

Boiteux, M. (1960), 'Peak load pricing', *Journal of Business*, **33**, 157–79.

Crew, M., C. Fernando and P. Kleindorfer (1995), 'The theory of peak load pricing: a survey', *Journal of Regulatory Economics*, **8**, 215–48.

Crew, M. and P. Kleindorfer (1979), *Public Utility Economics*, New York: St. Martin's Press.

Evans, J.E. and R. Green (2005), 'Why did British electricity prices fall after 1998?', Department of Economics Discussion Paper 05-13, University of Birmingham.

Green, R. (1996), 'Increasing competition in the British electricity spot market', *Journal of Industrial Economics*, **44**(2), 205–16.

Green, R. and D. Newbery (1992), 'Competition in the British electricity spot market', *Journal of Political Economy*, **100**(5), 929–53.

Joskow, Paul L. (2006), 'Introduction to electricity sector liberalization: lessons learned from cross-country studies', in F. Sioshansi and W. Pfaffenberger (eds), *Electricity Market Reform: An International Perspective*, Amsterdam: Elsevier, 1–32.

Klemperer, P. and M. Meyer (1986), 'Supply function equilibria in oligopoly under uncertainty', *Econometrica*, **57**(6), November, 1243–77.

Laffont, J.-J. and J. Tirole (1996), 'Creating competition through interconnection: theory and practice', *Journal of Regulatory Economics*, **10**(3), 227–56.

Littlechild, S.C. (1970), 'Marginal cost pricing with joint costs', *Economic Journal*, **80**(318), June, 323–35.

Mas-Colell, A., M.D. Whinston and J.R. Green (1995), *Microeconomic Theory*, Oxford: Oxford University Press.

MIT (2004), 'MIT report on Nuclear Power', Massachusetts Institute of Technology, Cambridge, MA.

Munasinghe, M. and J. Warford (1982), *Electricity Pricing*, Washington, DC: World Bank.

Newbery, D. (1985), 'Pricing policy', in R. Belgrave and M. Cornell (eds), *Energy Self Sufficiency for the UK?*, Aldershot: Gower, 77–113.

Powell, A. (1993), 'Trading forward in an imperfect market: the case of electricity in Britain', *Economic Journal*, **103**(417), 444–53.

Rasmusen, E. (1994), *Games and Information: An Introduction to Game Theory*, 2nd edn, Oxford: Blackwell.
Rees, R. (1973), *The Economics of Investment Analysis*, Civil Service College Occasional Papers, London: Her Majesty's Stationery Office.
Rees, R. (1984), *Public Enterprise Economics*, 2nd edn, London: Weidenfeld & Nicholson.
Stern, N. (2006), *Stern Review on the Economics of Climate Change*, London: H.M. Treasury.
Stoft, S. (2002), *Power System Economics: Designing Markets for Electricity*, New York: IEEE Press Wiley-Interscience.
Turvey, R. (1967), *Optimal Pricing and Investment in Electricity Supply*, London: Allen & Unwin.
Turvey, R. (1971), *Economic Analysis and Public Enterprises*, London: Allen & Unwin.
Turvey, R. and D. Anderson (1977), *Electricity Economics*, Washington, DC: World Bank.
Vickers, J. (1997), 'Regulation, competition and the structure of prices', *Oxford Review of Economic Policy*, **1**, 15–26.
Wenders, J.T. (1976), 'Peak load pricing in the electric utility industry', *Bell Journal of Economics*, **7**(1), Spring, 232–41.
Wolak, F.A. (2006), 'Why the United States has yet to benefit from electricity industry re-structuring (and what can be done to change this)', Western Power Trading Forum, New York Meeting, July 11, available at: http://www.stanford.edu/~wolak/ (accessed August 21 2008).

Energy

Demand / Supply Adequacy

Energy Vol. 15, No. 3/4, pp. 285-296, 1990
Printed in Great Britain

0360-5442/90 $3.00 + 0.00
Pergamon Press plc

2.6. DEMAND FORECASTING METHODOLOGIES: AN OVERVIEW FOR ELECTRIC UTILITIES+

AHMAD FARUQUI

Barakat, Howard & Chamberlin, Inc.
Oakland, California
U.S.A.

THOMAS KUCZMOWSKI

Energy Modeling Forum, Department of Engineering-Economic Systems, Stanford University
Stanford, California
U.S.A.

PETER LILIENTHAL

Energy Modeling Forum, Department of Engineering-Economic Systems, Stanford University
Stanford, California
U.S.A.

Abstract - Prior to 1973, most industries--including electric utilities--forecasted growth using rather straightforward time-trend approaches. While these approaches served the industry well during periods of steady and rapid growth, they failed to capture the underlying causal factors of growth. Thus, whey were unable to predict or explain the sudden changes in growth rates that have occurred since the 1973-74 oil embargo. Therefore, more sophisticated econometric and end-use models forecasting techniques were introduced to the utility industry. The growing awareness of the implications of a finite fossil fuel supply, as well as the increases in electricity rates, led forecasters to consider electricity prices and conservation issues explicitly in their models. In the industrial sector, changes in technology, structural changes in the economy, and fuel switching issues have led to an interest in process models.
The current trend in utility forecasting is toward strategic forecasting models. A key characteristic of strategic models is their ability to examine explicitly the factors and issues affecting future growth. This implies combining elements of the econometric approach with the technology detail found in the end-use/process models. Strategic models must be capable of doing more than merely forecasting future requirements. They must be able to provide planners with additional information on which policies to pursue to shape future loads.

1. ISSUES IN FORECASTING

Today's energy planning environment, particularly as it relates to industrial customers, is characterized by a number of important issues. First, estimating the future economic strength of the manufacturing industries is a key challenge to the industrial forecaster (1). Second, the purchasing patterns of industries' clients have changed. The advent of light weight, high performance materials is changing structurally the traditional intermediate materials sectors away

+An earlier version of this paper was presented at International Atomic Energy Agency course on Demand Forecasting at Argonne National Laboratory on September 21, 1988.

from firms producing basic metals to those producing composites, engineering ceramics and plastics, and specialty metals. Similarly, the advancement of electronics is rapidly changing the competitive advantage of traditional consumer product manufacturers in favor of those firms who have capitalized on new found consumer desires, and on firms which can effectively meet the expanded information needs of a dynamic economy. Finally, developments in new technologies are creating new markets. Thus, utilities must be able to assess the competitive position of industrial firms in their service area as well as that of the products those firms are producing. Only by examining both of these factors can a reliable, long-term forecast of electricity consumption be made.

The complexity of forecasting is further increased through the emergence of new technologies, particularly electricity-intensive technologies. As new technologies are introduced, the demand for inputs changes due to the requirements of the new technology and changes in relative competitive position of that industry as a whole. The key questions associated with new technologies are which ones will be adopted by industry and how will their adoption affect the input mix (2).

Since the early 1970s, economic, political, social, technological, and resource supply factors have combined to change the operating environment. As far as the electricity markets are concerned, many utilities are implementing demand-side management (DSM) programs. DSM consists of planning and implementing activities designed to influence customer use of electricity in ways that will produce desired changes in the utility's load shape, i.e., changes in the time pattern and magnitude of a utility's load. The approach is in contrast to traditional supply-side planning where utility capacity was built to satisfy a given level of demand. Many utilities are actively pursuing DSM programs (3). Figure 1 provides a framework for analyzing these issues. The diagram presents the scope of the INDEPTH modeling system which is described elsewhere in this chapter. Thus, an approach is required which can capture the influence of numerous, closely related issues: international competitiveness, product competition, consumer demand, technological change, and DSM programs. See (2) for additional perspectives.

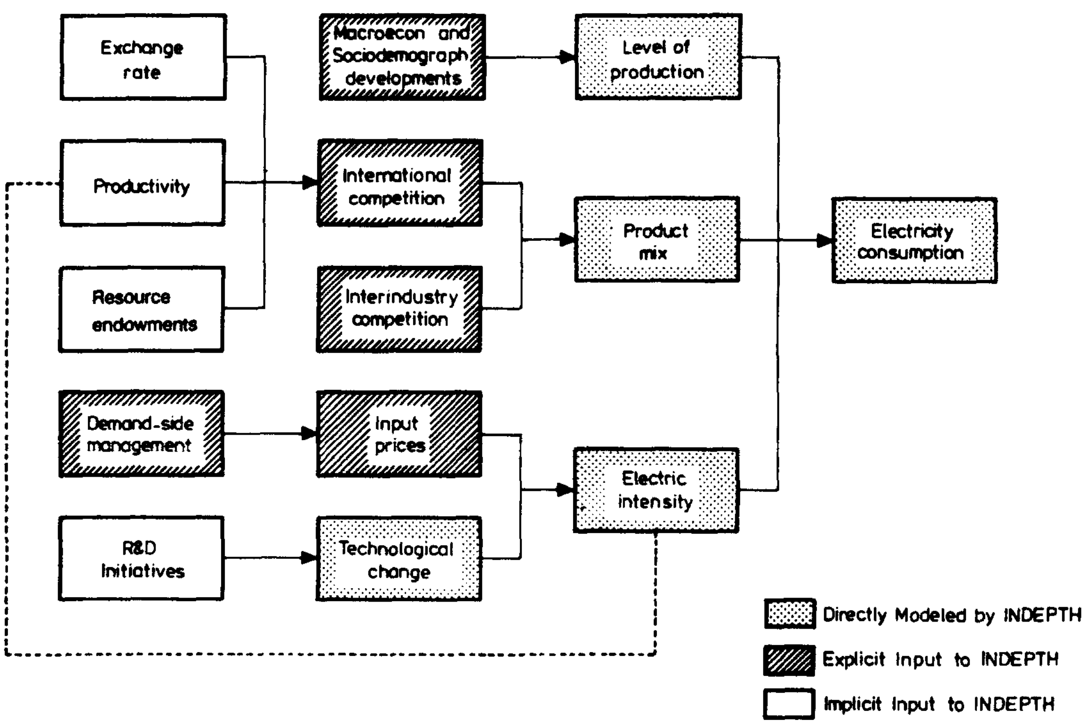


Figure 1. Structure of Industrial Electricity Demand

2. EVOLUTION OF FORECASTING METHODOLOGIES

Prior to 1973, most industries, including electric utilities, forecasted growth using rather straightforward time-trend approaches. While these approaches served the industry well during periods of steady and rapid growth, they failed to capture the underlying causal factors of growth. Thus, they were unable to predict or explain the sudden changes in growth rates that have occurred since the 1973-74 oil embargo. Therefore, more sophisticated econometric and end-use models forecasting techniques were introduced to the utility industry. The growing awareness of the implications of a finite fossil fuel supply, as well as the increases in electricity rates, led forecasters to consider electricity prices and conservation issues explicitly in their models. In the industrial sector, changes in technology, structural changes in the economy, and fuel switching issues have led to an interest in process models. The evolution of forecasting approaches over time is illustrated in Figure 2.

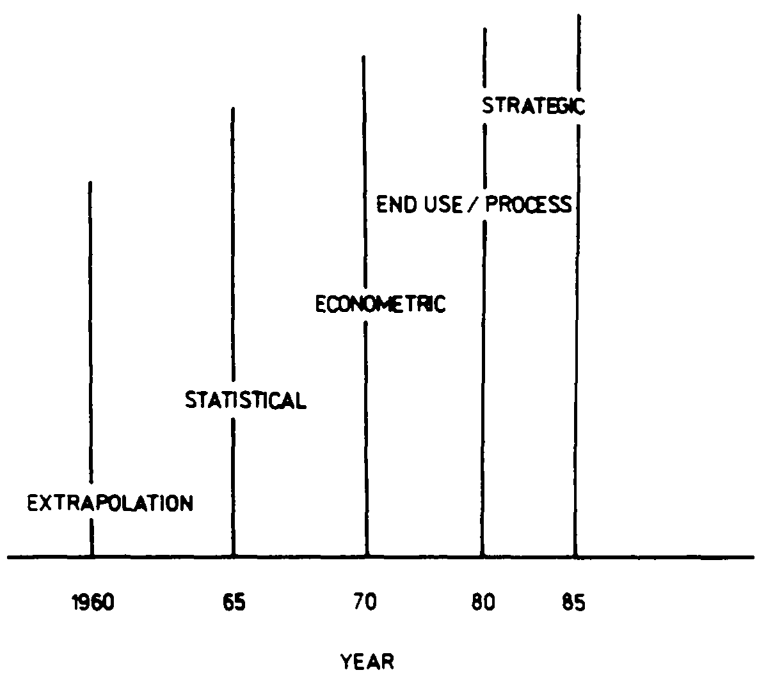


Figure 2. The Evolution of Forecasting Approaches

The current trend in utility forecasting is toward strategic forecasting models. A key characteristic of strategic models is their ability to examine explicitly the factors and issues affecting future growth. Strategic models recognize the impact that policy decisions can have on future loads. This requires detail--detail on the customers' operations, their current and potential demand for electricity, their competitiveness in the market place and their options with respect to production processes, switching alternatives, etc.

In the industrial sector, this implies combining elements of the econometric approach with the technology detail found in the end-use/process models. Strategic models must be capable of doing more than merely forecasting future requirements. They must be able to provide planners with additional information to help shape future demand. None of the modeling methods overcomes the uncertainty associated with forecasting.

3. CURRENT STATE OF PRACTICE

As noted, the methods used in forecasting have been changing rapidly since the early 1970s. Key findings from surveys of utility forecasters indicate:

1. Utilities spend significant resources on energy forecasting.
2. There is a growing desire to incorporate more detail on electricity use into the forecast.
3. Information on electricity consumption by industry type, customer equipment, and operations are generally available, although the quality of the information and the time frame over which it is available vary widely.

The overwhelming majority of utilities currently use an econometric approach to forecast industrial energy consumption. Survey results indicate that a combination of approaches is often used. In the sample of 14 utilities:

- Three use an aggregate econometric approach; that is, they produce forecasts for the industrial class as a whole.
- Ten use a detailed econometric approach; that is, the industrial class is disaggregated according to such factors as SIC, customer size, and growing industries and declining industries.
- One uses a customer survey as the sole basis for the forecast.
- One uses a customer survey to supplement other econometric forecasting methods.
- Two use end-use/process models for forecasting electric consumption of the major industries.

Because some utilities use more than one technique, the sum of the above responses exceeds the number of utilities interviewed.

The use of formal questionnaires to obtain information relevant to generating the industrial forecast is the exception rather than the rule. Only three of 14 respondents collect and incorporate customer information into the forecast through use of a formal questionnaire. The far more common practice is to collect customer-related information informally through customer representatives. Because industrial forecasts are typically generated at a more aggregate level of detail than the specific customer information, such information is incorporated into the forecast in an ad hoc manner. That is, the information is used judgmentally to adjust the statistically derived electricity consumption level up or down as appropriate. The practice of interviewing industrial customers, rather than relying on surveys, is quite a contrast to the residential sector where appliance surveys are conducted regularly by the industry.

The following information was collected regarding the overall familiarity of the utility industry with process models that estimate energy consumption by focusing on the underlying industrial processes and technologies. The familiarity of the respondents with the concept of process models is as follows:

- Twelve have gained familiarity with the models through seminars, conferences and/or the literature.
- Two used the process models in the preparation of their industrial forecast.
- Four expressed the opinion that process models are too data-intensive and cumbersome for use as forecasting tools at the service area level.
- Two have no familiarity with process models.

In addition to technology, industrial energy consumption is influenced by the level of economic activity. Economic drivers used by the respondents include:

- Final Reserve Board production index.
- Value of shipments.
- Service area employment.
- National/regional production levels for important industries in the service area (number of automobiles, tons of paper, etc.).

- Gross state product.

All but one of the respondents indicated they used outside economic forecasting services to provide them with at least the starting point for generating the drivers for their models. Four indicated that they had in-house economic services that regionalized or adjusted national forecasts to reflect service area conditions.

A common concern in the development of an industrial forecasting model is the availability of data, particularly with the increased tendency of the industry to incorporate more detail into the forecasting models. Of the 14 utilities interviewed, 12 classified their customers, and consequently, industrial electricity consumption, by at least the two-digit SIC level. Some of the customer classifications go back as far as the 1960s, others only to 1982. There is some doubt about the validity of the data from the earlier years due to misclassification, changes in customer product mix, and changes in the SIC categories.

In addition to consumption data, some models, such as process models, require information on industrial customer processes and equipment. One-third of the respondents indicated that their customer representative had good knowledge of their customer's electricity consuming equipment. Another third indicated that they obtained similar information through recently conducted surveys. The remaining one-third indicated that they had little knowledge of their customer's equipment and processes.

All but one respondent indicated a general level of satisfaction with their present forecasting procedures. This, however, does not imply that all of the forecasters' needs are currently being met. Specific needs, as expressed by the forecasters, include:

- More detail on the customer's operations to assess his business competitiveness.
- More detail on the pattern of consumption.
- More detail on consumption and end-uses to be used for marketing purposes.
- Better understanding of the impact of conservation on customer's consumption.
- Better understanding of the effect of technology changes on industrial consumption.
- Increased accuracy of economic variables driving the forecast.

Regarding DSM, it is interesting to note that for most respondents DSM in the industrial sector was limited exclusively to rate-related programs. Specific demand-side programs mentioned include:

- Time-of-use rates.
- Off-peak energy rate.
- Interruptible rate.
- Special metal melting rate.
- Economic development rate.
- Industrial customer audit program.
- Financial incentives for the purchase of energy efficient equipment.
- Educational activities on the efficient use of electricity.
- Marketing of conversions from fossil fuels to electricity.

Utilities devote considerable resources to the system-wide forecasting activity. The total level of effort devoted to forecasting residential, commercial, and industrial energy consumption, as well as the economic drivers, ranges from 2 to 40 person-years per year for the utilities interviewed. The average level of effort is 11 person-years per year.

4. SYSTEMS OF MODELS

A detailed review of the strengths and weaknesses of the alternative modeling approaches is described in (4). Based on this review, a hierarchical, menu-based approach appears well suited to meeting utility needs.

Such an approach is adopted in the INDEPTH modeling system, described elsewhere in this chapter. The concept is however perfectly general. The system should have some general properties such as the following. It should form an internally consistent, nested modeling system that would enable each utility to select an appropriate level of detail based on the availability of data and the specific characteristics of its service area. System of models consist of several levels of analysis. The broadest level will normally consist of a series of econometric models that produce electricity forecasts for the 20 two-digit major industry groups such as food products, lumber products, and pulp and paper products. These econometric models will provide complete coverage of the manufacturing sectors. The second level of models will focus on a few key industrial sectors through the use of process models. These models provide greater detail about the industry being modeled, but cannot, by themselves, be used to generate forecasts for the total manufacturing sector because only a few sectors are modeled. The third level of models is the equipment models. These models are designed to forecast change in market penetration of key end-use equipment, such as motors.

In the integrated modeling system, all the component models need to be supported by a common set of assumptions, external drivers, and a consistent logic. These common elements are necessary because the forecasts obtained from the second level--process models--are implicitly contained in the forecasts produced by the first level--econometric models. Similarly, the results of level three--the equipment models--are implicitly contained in the econometric and process models. Therefore, all levels of the system utilize the same macro-level economic forecasts to assure that the electricity forecasts are based on common perceptions of the underlying economic growth of the service area, of the structural effects on industry sectors, and of technological change. Therefore, a critical component of the modeling system is the data base and the methods to regionalize economic forecasts.

5. SELECTION OF MODEL SPECIFICATION

Selection of the particular model specification is based on satisfying several objectives. It is necessary to strike a balance between model specifications that are abstract, complex, theoretical, and difficult to implement with those that are overly simplistic, possess little statistical and theoretical justification, and are being used because of the expediency of implementation. The following objectives need to be considered:

- Validity--Is the model conceptually and theoretically sound? Are the relationships, specifications, and data accepted by the professional community? Do models of a similar nature possess a track record of accurate forecasting?
- Transferability--Is the model capable of being specified, estimated, and implemented in a way that makes it applicable to a variety of areas?
- Data availability--Are the data available for model estimation? Is the quality of the data sufficient to make key planning decisions?
- Output capabilities--Is the model capable of being formulated in such a way as to satisfy the need?
- Ease of understanding--Can the model structure and implementation be understood by critical decision makers who must have confidence in the model and its results to make decisions?
- Issue handling--Does the model allow examination of critical issues facing forecasters? Can the model address any or all of the following:
  - Structural change in product demands
  - Regulation
  - Technological change
  - Demand-side management issues such as rates, incentives, and education
  - Alternative input prices

A classification scheme useful for describing the trade-offs between the level of detail contained in models versus their ability to explain is shown in Figure 3.

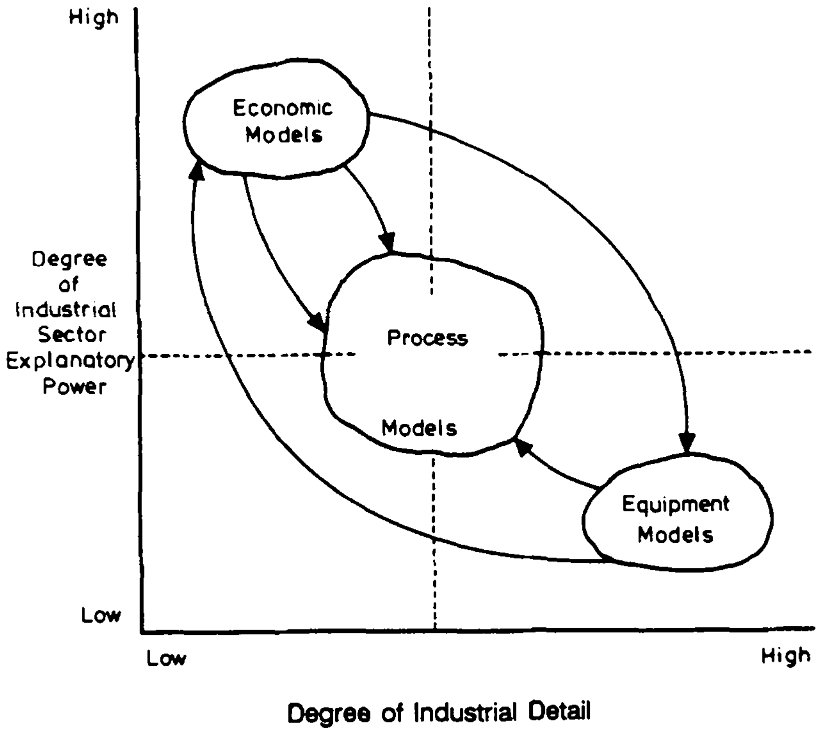


Figure 3. Positions of the Different Forecasting Models in Given Taxonomy

6. APPLICATION OF MODELS

Equipment Model Applications

The equipment models are used to examine the impact of technological change in key equipment common to most manufacturing establishments. Examples include electric motors, process heating equipment, and lighting fixtures. These models are run to analyze the potential impacts of DSM programs designed to promote the use of certain types of end-use equipment.

Equipment models when used for electric utilities will produce results in the form of electricity consumption by motors, electric lights, etc. The consumption of energy by this equipment will, however, already be contained within the forecasts made by the econometric and process models. Therefore, sector totals for each industry cannot be obtained by merely adding up the consumption levels forecasted by equipment models.

The implementation of the equipment models, therefore, requires two steps: (1) identification of equipment usage in the industrial sectors of other model levels, and (2) utilization of the equipment models in a simulation mode.

A bridge is built that describes the baseline level of equipment usage in each industrial sector in the econometric and process models. The bridge allows identification of the implicit baseline equipment consumption. By utilizing the equipment models in the simulation mode--with the baseline being the level of usage described in the bridge--the net change in equipment usage can be added to any other model result.

Econometric Model Applications

The econometric models are the most general and comprehensive of the models. Built at the two-digit SIC level, these models form the basic forecasting capability of most modeling systems. Through simple summation, the aggregate manufacturing consumption can be determined. All other models within the system modify and supplement the results obtained from the econometric model.

The macro-driven, service area economic forecasts, and decision-making assumptions used in the econometric models must be consistent with those used at the other levels. Only if the other levels are used in a simulation mode can different data or assumptions be used--and then, the results from the different data must be compared to baseline results from the data used in the econometric models.

Thus, the econometric models are critical tools for forecasting. Utility-specific applications will use the elasticity estimates derived from the system of cost-minimizing input demand models in single-equation models after regional calibration. The process and equipment models are used to analyze in greater details results which are implicitly contained in the econometric models.

Process Models Applications

The process models are for use by utilities interested in better understanding how a particular four-digit industry may evolve over time. Thus, the process choice models are used only if that industry segment is important in a local service area.

The process models yield a forecast of electricity consumption at a more detailed level than is produced by the econometric models. Therefore, the results of the process models cannot be added to the results of econometric models without careful consideration of possible double counting. The process models can be integrated with the econometric models as follows. Two sets of econometric models are specified. One for the set of entire two-digit SICs and another based on two-digit SICs except for the four-digit sector being modeled by the process models. For example, if a process model is built for SIC 3311--basic steel--then a special econometric model containing all activity in SIC 33 except 3311 is also built.

Econometric Versus Process Models

Application of econometric and process models is outlined on Figure 4. The driver model operating at the bottom level generates input (production mix) in monetary units. Therefore, calibration is required for the process model into the physical units.

Both econometric and process model have their specific data bases as listed in the figure. Behavioral lags representing an estimate of time required to achieve the equilibrium market share are often included in process models.

Model outputs can be described briefly as future demand trends from the econometric model and technology choice from the process model (as an optimization algorithm is applied).

Mixed Models Application

In the middle the mixed models are placed as described before. This type of modeling can use a combination of orthodox generic approaches described before for econometric and process models.

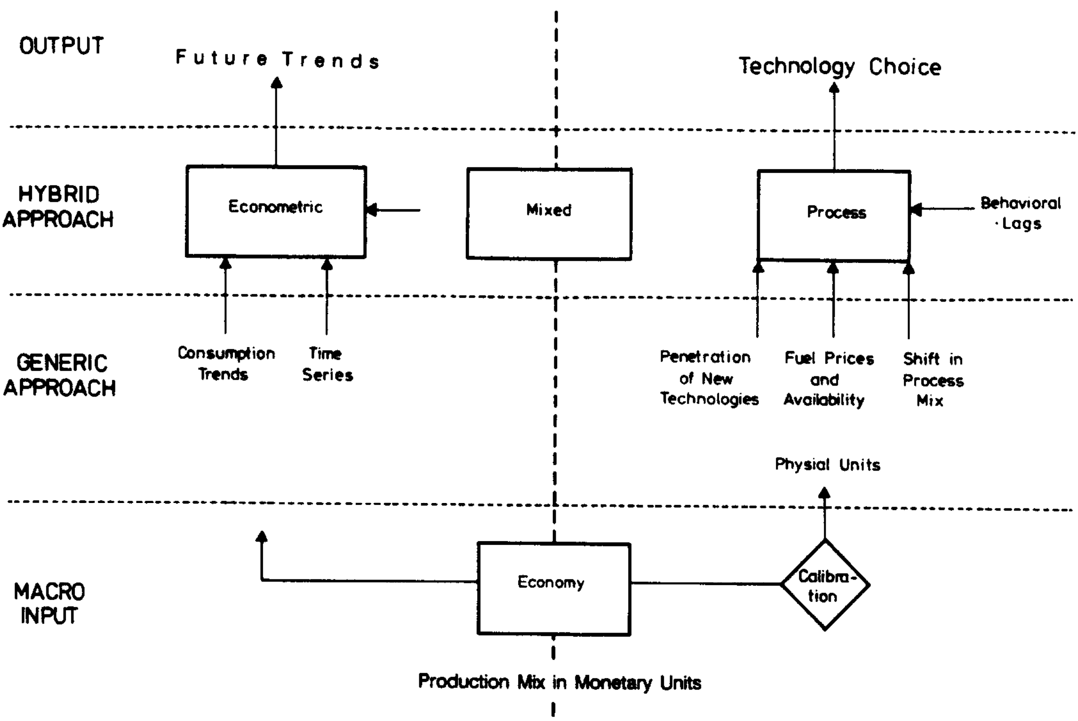


Figure 4. Econometric Versus Process Models

The mixed approach might be just one model, called hybrid in our taxonomy. It can also be a system of models, which can use separate econometric, process, and other types of models. INDEPTH is an example of such an integrated modeling system. In the most general case, it can be a set of different type of models employed to generate the demand.

Such a model set was used in an Energy Modeling Forum study to evaluate the impact of structural shift in the US economy on industrial energy demand. On the macroeconomic level, three economy-wide models were employed. To predict the major trends in the US economy, the PILOT model was used (see Chapter "Multisectoral Planning Models"). This is a macroeconomic model designed for long-term projections of both energy demand and economic output. The PILOT model is a large-scale dynamic input-output model of the US economy. The main goal of the model is to evaluate technological change, which is assumed to be the driving force of economic growth. The INFORUM model, an input-output macroeconomic model, was used for short-term projections only (i.e., through 1995).

Finally, the WHARTON model was used as a driver for industrial models. WHARTON generated sectoral outputs, which were used to forecast industrial energy demand. The PILOT model was calibrated to WHARTON's base-case scenario. It was impossible to use the PILOT model to generate energy demand by industrial sectors, as the aggregated energy services submodel is employed.

All the types of models represented in this study can be used to forecast the levels of demand for different energy carriers. However, the three models described above were designed generally to describe the macroeconomy. Therefore, to conduct the detailed research, the sectoral output levels obtained from these macroeconomic models were introduced as drivers to three different types of industrial sector models:

- Process (ISTUM)
- Hybrid (ORIM)
- Econometric (PURHAPS)

Observation on Models Behavior or Response

As a result of this comparative study, some observation on models behavior can be made. Preliminary observations were that the choice of method applied can bias the results. As a rule, the higher growth in demand level, especially for electricity, was obtained from the econometric model. The nature of the bias in the process-type models is to be overly optimistic on the rate of efficiency improvement in the input-output coefficients. The rationale of the econometric models is that our best guess of the future is provided by a study of past trends. Therefore econometric models can easily capture long-term regularities such as electrification and automation if they are contained in historical data.

During the process of detailed comparisons, it was found that differences in model outputs were not very large, after proper calibration. There are, however, some aspects of model types which might be of some interest for users.

The penetration of new technologies can be an extremely important factor in the future. This issue can be examined only by engineering process models.

Econometric models tested in the EMF 8 study show surprisingly low sensitivity to changes in the input prices, especially oil prices. This might be explained as a result of the relatively long time periods used by the models.

All models have built-in tendencies that favor or discourage use of certain inputs. All models show a trend toward electricity with virtually no substitution away from electricity toward fossil fuels. Process models display a tendency toward coal, possibly due to lower nonfuel costs of coal technologies.

In general, econometric models are highly dependent on time series data and are best suited for short-term projection. As a consequence, econometric models may not be appropriate in forecasting the effect of new technologies that depart from the historical pattern.

An excellent illustration of the difference between econometric and process models is afforded by the forecasted role of cogeneration by the models participating in the EMF 8 study. In Figure 5 both of the process models show declining intensity of purchased electricity in manufacturing while the other models show it to be increasing. This discrepancy was primarily due to self-generation.

The long term trend for cogeneration and self-generation had been declining for almost a century and the recent surge in activity had not yet been reflected in the data available to the modelers at the time of the study. Consequently, the forecasts of the econometric models reflected very little cogeneration activity in the future.

Because the process models base their forecasts more on future conditions than past trends, the favorable economics of cogeneration resulted in a very substantial penetration of this technology. This was the greatest single difference in the results of the two modeling methodologies.

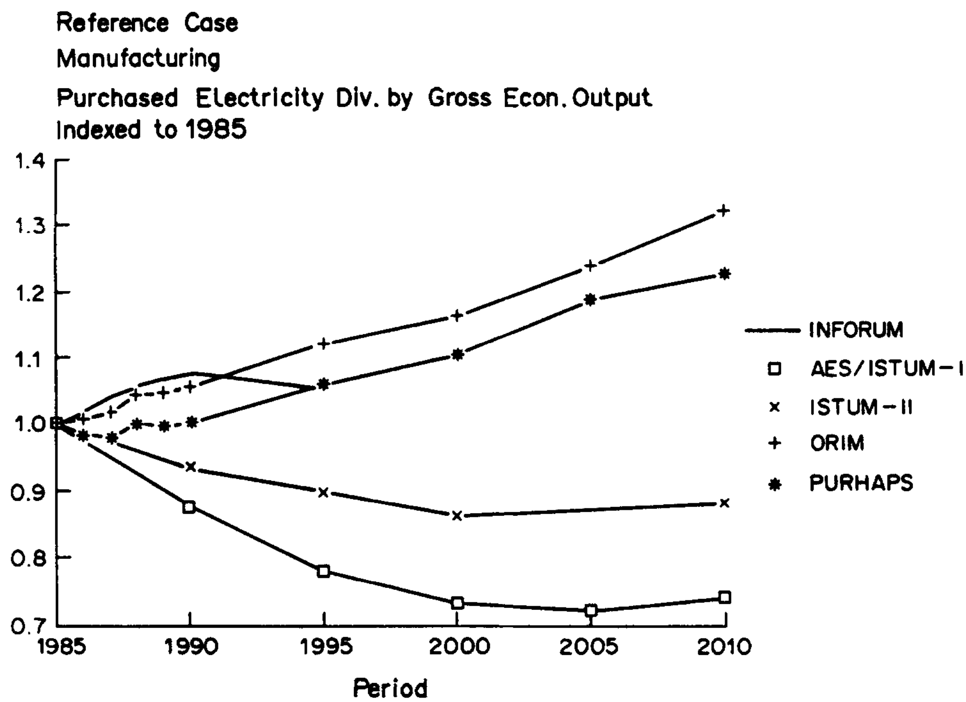


Figure 5

The inherent optimism concerning market penetration rates exhibited by process models stems from the difficulty of modeling institutional obstables to new technologies. In the case of cogeneration, the contractual and financial complexity of many of these projects has hindered projects that might otherwise have been economically attractive. Furthermore, the aggregate nature of national models obscured regional variations that limited cogeneration activity to those utility districts that needed capacity. These issues are typically handled by a logistic market penetration function but there is little theoretical basis for estimating lag parameters other than those associated with capital vintaging.

As was stressed before, the promising direction is to take advantage of the chief features of both methods. The models should not rely entirely on historical trends. Even recent history cannot point out future trends. Future trends are driven by many other factors such as penetration of new technologies, input prices and availability, shifts in processes, and product mix, etc. Econometric models are also plagued by the fact that supply constraints have been a major trend factor affecting consumption in the past in many countries. Thus, demand relationships cannot be easily identified through normal econometric techniques.

VIII. CONCLUSIONS

No one type of model can be recommended as the best one for a particular forecasting task. The choice is with the analyst, who can make it based on the experience and availability of techniques and resources. Often a system of models approach will be preferable over that based on a single model.

REFERENCES

Boyd, G. and T. Kuczmowski. (1989). "Impact of Structural Shift in the Economy on Energy Demand and Conservation" EMF Working Paper 8.14, J. Weyant (ed), Industrial Energy Demand, North Holland.

Battelle-Columbus Division. (1984-1987). Demand-Side Management, Vols I-V, EPRI EA/EM 2597, Palo Alto, CA.

Faruqui, A. and J. Broehl. (1987) (editors). The Changing Structure of American Industry and Energy Use Patterns, Columbus, OH: Battelle Press.

Lilienthal, P. (1986). "Overview of Model Structures." EMF Working Paper 8.13.

Resource Dynamics Corporation (1986). Electrotechnology Reference Guide, EPRI EM-4527, April.

Battelle-Columbus Division (1985). Industrial End-Use Planning Methodology (INDEPTH): Recommended Design, EPRI EA-4019, Palo Alto, CA: EPRI, Section 3.

Weyant, J.P. (1988) (editor). Industrial Energy Demand, North Holland.

Energy Vol. 15, No. 3/4, pp. 213–236, 1990
Printed in Great Britain

0360-5442/90 $3.00 + 0.00
Pergamon Press plc

2.1. ECONOMETRIC TECHNIQUES: THEORY VERSUS PRACTICE

PAUL J. WERBOS

Energy Information Administration+
Department of Energy, Washington D.C.
U.S.A.

Abstract - This paper introduces the basic concepts used in econometric modeling, and describes five prescriptions to avoid common real-world pitfalls in that style of modeling. The paper begins by comparing econometric modeling with other forms of modeling used in energy modeling and engineering. It describes what an econometric model is, and how to build one. It then gives a detailed explanation of many facets of the five prescriptions: pay attention to uncertainty; don't expect a free lunch when devising specifications; pay attention to prior information; don't expect to draw conclusions without adequate data; and check the historical track record of your model. The issues of generalization and robustness over time receive special attention; they are important in practice, and subtle in theory. Finally, the paper discusses model development in practice, building upon experience with PURHAPS, a model I developed for the Energy Information Administration (EIA).

1. BACKGROUND

Economic theory, in the United States, usually begins with simplifying assumptions like free markets, perfect competition, no externalities, and perfect foresight. After years of study, the advanced student is told how to modify this theory to address real problems in the real world, which are often quite different from the theory in important ways. Some students never quite make the adjustment.

Econometrics is very similar. This paper will introduce the novice to the basic assumptions and methods of econometrics, and then discuss problems which come up in modifying the theory to fit the real world.

Broadly speaking, there is no sharp dividing line between econometric models, engineering process models, statistical models, simple time-series models, systems dynamics models, etc. All these types of models are systems of equations designed to forecast or simulate whatever we want to forecast or simulate. The real difference lies in how we obtain information or parameters to plug into the models.

Some classes of models tend to rely on a priori information or indirect information about what we are forecasting; models of this sort include "pure" process models, classical systems-

+This paper expresses the views of the author, not those of EIA or DOE, though it was reviewed at EIA prior to submission. As this paper goes to press, the author's address has changed to: Room 1151, NSF, Washington D.C. 20550.

dynamic models and expert systems. Other classes are based on empirical data about exactly the kind of variables we are trying to forecast; this includes econometric models, time-series models, statistical models, (identified) control-theory models, and artificial neural networks.

In the energy business, the a priori models tends to be very complex, because people include lots of lower-level detail to create a feeling of earthy realism; however, the parameters are usually based on judgment or guessing, and it is hard to be sure the model will track actual trends. The good empirical models tend to be simpler, because they are usually limited to variables which are observed on a time-series basis; however, they are strongly rooted in empirical reality, if done right.

The a priori models sometimes seem easier to understand, at first, because they mimic concrete, well-known engineering processes (at least in part); however, because they contain so much detail, it is not always easy to know what causes the bottom-line forecasts to come out as they do, and the role of human behavior is often neglected or oversimplified. Empirically-based models are the reverse: the overall behavior is easier to understand, but the detailed reasons behind the trends -- both historically and in the forecast -- may require further analysis. A good researcher will learn how to combine both prior information and empirical information into a model, as this paper will discuss.

The relations between different kinds of empirical model are subtler.

On some level, there is no real difference between a statistical model, an econometric model, and a model developed by using the identification techniques of control theory; all three rely on the same core of theory, which this paper will discuss. "Simple time-series models" and artificial neural networks depend on the same theory as well, but they try to automate the process of coming up with a functional form; in effect, they assume that the user does not really understand the structure of the system he is studying, so that a computer can do the job as well as a person. This is a good assumption in some cases (as in recognizing patterns among thousands of variables which no one fully assimilates into his or her intuition), and a poor assumption in others (as in the study of physical phenomena for which the dynamics are well-understood).

2. GOALS OF THIS PAPER

In the United States and many other nations, econometrics is a major academic discipline, based on the idea that a careful analysis of historical data can be a good starting point for analyzing or projecting the future. Like any major discipline, econometrics has a long history, full of false starts, new perspectives, and hundreds of applications, some good and some bad.

This paper will present those concepts and rules of thumb which we have found most important, in practice, in a government organization concerned about the quality of its forecasts. No one can expect to become a first-class econometrician after reading one article; there are simply too many tricks and traps to learn. However, we will try to explain the key concepts, and cite books which elaborate on their application. We also hope to pinpoint those misunderstandings which are common among experienced practitioners, and we apologize to them that there is not enough space here to explain all the details. Unfortunately, these misunderstandings have often led to the creation of models which totally misrepresent the dynamics of the variables which they are supposed to predict.

This chapter will begin by saying what an econometric model is and how -- mechanically -- to build one. Next, it will discuss five major prescriptions for the correct use of econometric tools. Then it will discuss the use of these tools in practice at the Energy Information Administration (EIA). It will conclude with a very quick overview of the PURHAPS model, one of the econometric models I have developed for EIA.

3. WHAT IS AN ECONOMETRIC MODEL?

Strictly speaking, an econometric model is no different from any other forecasting model -- it is made up of any set of equations or formulas which can be used to predict the future. For example, consider the following simple model to predict population:

$$POP(t+1) = c \cdot POP(t) \tag{1}$$

This says that population in year t+1 is equal to a constant "c" multiplied by the population in year t. If you obtain your estimate of the constant c by asking your boss what c should be, or by studying textbooks on theology or ideology, then we would call this a judgmental model. If you obtain your estimate of c from small-scale case studies of controlled populations, we might call this an engineering model. If you have an historical time-series of data on population, for the state or nation whose population you are forecasting, and if you estimate c from that time-series in a rigorous way, then we would call this an econometric model. In principle, then, there is no such thing as an econometric model; there are only econometric methods for estimating parameters such as "c" in general models. A pure econometric model is simply a general model, in which all of the parameters have been estimated by econometric methods, based on empirical data.

Econometric methods were initially developed for use in economic forecasting. However, there is nothing in our discussion which will restrict their use to economics. Econometric methods have often been applied directly to forecasting social and political systems (Werbos, 1974; Werbos, 1977; Werbos and Titus, 1978). Human minds and computers which truly imitate human minds must also have a built-in capability to learn cause-and-effect relations by somehow analyzing a time-series of sensory experience; we have shown how econometric methods may be embodied directly into the wiring of such systems (Werbos, 1987a; Werbos, 1986a).

In general, people who use historical data or trends or track records to help them make decisions are making inferences about cause and effect. Like it or not, they are engaged in a form of statistical inference. Even if they say they are merely testing an hypothesis, or a relation, and not formulating a model, the fact is that they are estimating a model; the potential for error and uncertainty is merely less visible and harder to correct when they deny this fact. (Of course, some managers would prefer to hide such uncertainties from their superiors. If a superior really cannot understand econometrics, there is an art to using econometrics properly and then translating the results back into English, using graphics and discussions of percentage growth rates and historical analogues.)

4. HOW CAN AN ECONOMETRIC MODEL BE BUILT?

The first stage in building a model is to review the available data and concepts, as we will discuss further in the section on "Practice".

Next one must choose a computer package to work in, to implement the econometric methods. EIA generally prefers to use the SAS package (SAS, 1985a; SAS, 1985b) on its large computer, because of its superior flexibility and data-handling capabilities; however, Troll (1981) has also been used, because of the sophisticated econometric tools it contains (some developed under contract to us). On microcomputers, SAS is also available, but is relatively expensive at present; Lotus is widely used, and new packages from Wharton Econometric Forecasting Associates (of Philadelphia) and elsewhere may be used more in the future. We use SAS to estimate parameters such as "c" in equation 1, and to evaluate the overall degree of fit of equations such as 1 and alternatives to 1; then, when all the equations are estimated and selected, we usually program the forecasting itself in FORTRAN. Actually, SAS and Troll have the capability to simulate the model -- to generate forecasts - as well; we have used that capability only rarely, because our models have usually been too big to fit into those systems.

The next step is simply to use the package chosen. To estimate equation 1, for example, you would first locate a time-series of data on the variable POP, and load it into a SAS dataset using the SAS command DATA (SAS, 1985a). Then you would use a SAS command such as GLM (SAS, 1985b) to estimate "c" in equation 1, and to evaluate the error which this equation would have led to in forecasting the past. You could also use SAS to estimate alternative equations, and their errors, and you could select between equations based on their error. At that point, you have the equation, and you need only code it into a forecasting program.

The most common way to estimate a complex model, in econometrics, is to use "regression" or "least squares." Using regression, we estimate each equation of the model separately, one after another. For each equation, the regression command finds those values for the parameters which lead to the smallest possible error over the historical period you have data for. "Error" is defined as the sum over all observations of the square of (actual minus predicted). Regression also reports what the error is for the equation as estimated.

In actuality, most computer packages have two main regression commands available -- a linear regression command and a nonlinear regression command. (See Wonnacott and Wonnacott, 1977, Chapters 13 and 15, for more explanation.) To avoid complications, most economists use linear models such as the following two-equation model:

$$Y(t) = a \cdot Y(t\text{-}1) + b1 \cdot X_1(t) + \dots + bn \cdot X_n(t) + c \tag{2a}$$

$$Z(t) = c1 \cdot Y(t) + c2 \cdot Y(t\text{-}1) \tag{2b}$$

In equation 2a, Y(t) is the "dependent variable" -- the variable being predicted <u>in that equation</u>. Y(t-1) and $X_1(t)$ through $X_n(t)$ are the independent variables of that equation.

The term "c" is the "constant term" or "intercept;" note that equation 2b has no intercept. The parameters of the model are the constants a, b1 through bn, c, c1 and c2. The "endogenous variables" -- Y and Z -- are the variables being predicted <u>somewhere</u> in the model. Y(t-1) is a "lagged endogenous variable" (because Y is endogenous and because t-1 represents a "lagged" value, a previous year's value.) X_1 through X_n are "exogenous" because they are not endogenous.

This example is linear, because the dependent variable in every equation is predicted as a linear combination ("weighted sum") of the independent variables, plus an optional constant term. To estimate each equation in SAS, you need only use the linear regression command (GLM or something similar) once for each equation. You can be sure of quick results, and you do not have to give an initial guess for the values of the parameters. Each time, you only have to tell SAS the name of the dependent variable and the names of the independent variables. You also have to tell SAS whether you want a constant term in the equation, and whether you want SAS to print out all the diagnostic statistics anyone has ever thought of.

At first glance, equation 2a may appear somewhat abstract and unrealistic. Economic relations in the real world are often more complex. For example, even in a simple model of fuel oil demand (QOIL) as a function of residual oil prices one would not want to use the simple equation:

$$QOIL = a \cdot PRESID + b \cdot PDIST + c \cdot DISTSHARE \tag{3}$$

If you used this equation, by regressing QOIL on PRESID, PDIST, and DISTSHARE, you would expect to find that "a" and "b" are estimated as negative numbers, expressing the idea that higher prices lead to lower demand. However, with "a" and "b" negative, there will always exist a price so large that demand becomes negative, which is an absurd forecast. Likewise, the effect of changes in PDIST should depend on how large the distillate share is; PDIST and DISTSHARE have an "interaction effect." For these reasons, a better specification would be:

$$\mathrm{LOG(QOIL)} = a + b \cdot \mathrm{LOG(POIL)} \tag{4a}$$

$$\mathrm{POIL} = \mathrm{PDIST} \cdot \mathrm{DISTSHARE} + (1 - \mathrm{DISTSHARE}) \cdot \mathrm{PRESID} \tag{4b}$$

Equation 4a says that QOIL is a function of the weighted average price of fuel oil, POIL. It says that a given percentage change on POIL leads to a proportionate percentage change in QOIL; the factor of proportionality is just "b", the price elasticity of demand. (To see this, differentiate 4a or see Wonnacott and Wonnacott, 1977, Section 15-3). Equation 4a can be estimated easily in SAS by first using a DATA step to calculate:

$$\mathrm{LOGQOIL} = \mathrm{LOG(QOIL)}$$

$$\mathrm{LOGPOIL} = \mathrm{LOG(PDIST} \cdot \mathrm{DISTSHARE} + (1 - \mathrm{DISTSHARE}) \cdot \mathrm{PRESID},$$

and then calling the regression command and asking it to regress LOGQOIL on LOGPOIL. Equation 4b contains no parameters at all to estimate; it is called an "accounting identity" (as opposed to the "behavioral equation" 4a). Equation 4a is linear in the parameters a and b, but not in the original variables QOIL and POIL. Most econometric models are linear in the parameters but not in the original variables. Most of them also use tricks like the above to express economic relationships.

If equation 4a had actually been nonlinear in its parameters, then nonlinear regression could have been used. Nonlinear regression requires a lot more care and patience, depending on what computer package you use, but there is usually a way to make it work. Likewise, there are alternatives to regression which would require you to estimate the entire model as a system, together; to use these alternatives, you would have to type both equations into a single model file or command block.

Aside form their linearity, the models in equations 2 and equations 4 have two other simplifying features. First, they are "recursive". In economics, this means that they are really just simple formulas; you can calculate a forecast by plugging in values for the exogenous variables and lagged endogenous variables, and using the equations one after another like a formula or a recipe. Most econometric models are actually simultaneous, as in the following example:

$$\mathrm{LOG(SUPPLY)} = a + b \cdot \mathrm{LOG(GNP)} + c \cdot \mathrm{LOG(PRICE)} \tag{5a}$$

$$\mathrm{LOG(DEMAND)} = d + e \cdot \mathrm{LOG(GNP)} + f \cdot \mathrm{LOG(PRICE)} \tag{5b}$$

$$\mathrm{SUPPLY} = \mathrm{DEMAND},$$

where GNP is exogenous and where the model is used to forecast a PRICE that makes SUPPLY and DEMAND balance. to make a forecast, you cannot jus plug in GNP and PRICE on the right-hand side; you cannot, because you don't yet know that PRICE is. You have to solve this system of equations, as a set of three simultaneous equations in three unknowns. In fact, if you insert these three equations into Troll (1981), Troll will take care of this problem and give you a set of forecast which solve the equations.

Notice that it would be very dangerous to estimate a system like this by ordinary regression. If SUPPLY did equal SEMAND in all historical years, then you would get exactly the same set of parameters (d,e,f) when you use regression on 5b as you did (a,b,c) when estimating 5a; you would not really have two different equations. Even if the equations were very slightly different, you could not rely on what you get when you subtract one from the other (as required in solving them). In these kinds of situations, it is important to estimate the model as a system (Wonnacott and Wonnacott, 1977, chapter 22).

These situations arise in energy modeling, but the problem is usually not significant, mostly because we deal with dispersed system involving lagged responses. The systems estimation methods often lead to worse results, because e of their complexity, because the use of instrumental variables introduces random noise, and because of problems with "robustness" (discussed below). On the other hand, the simultaneity problem can be serious with fuels like LPG and other minor forms of oil, whose markets are very limited and respond quickly to price; our goal, in those cases, is to look for something like a "reduced form" model for each such fuel (e.g. in equations 5 first solve, to get log(Price)=g+h·LOG(GNP), and then estimate g and h).

Even the model in equations 5 still has one further simplification: all of the variables are assumed to be available for all historical years in you data base. (SAS will overlook a few missing values here and there, however.) It is possible to build econometric models which do not have this property, because they include "time-varying parameters" or "hidden variables;" however, this is not common at present, and the tools to estimate such models are hard to come by. One can work around this problem, to some degree; for example, if the population growth rate, "c" in equation 1, varies over time as a function of women's education (WED), then one might postulate that c=a+b·WED, and rewrite equation 1 as:

$$\mathrm{POP}(t+1) = a \cdot \mathrm{POP}(t) + b \cdot \mathrm{WED}(t) \cdot \mathrm{POP}(t) \tag{6}$$

Finally, for completeness, it should be emphasized that variables in an econometric model are not always simple time-series. Many authors will perform regressions on a data base of different observations at the same time, such as data from different states, and then use the results to predict the future. This is called forecasting based on cross-sectional analysis, and the results are usually unreliable at best, both in the short-term and in the long-term. For example, one of the first econometric equations ever studied was the classical consumption function:

$$C(t) = a + b \cdot Y(t) \tag{7}$$

where C is national consumption and Y is national income. In cross-sectional analysis, "a" was significantly larger than zero, and there seemed to be a large saturation effect in consumer spending. But in time-series, "a" was quite close to zero. For purposes of forecasting changes over time, the time-series version is the right one to use. In general, variations across space tend to be different from variations across time, and we have seen this lead to problems over and over again. (For example, see the discussion of "locational bias" in Werbos, 1983, Chapter 4.)

An ideal model should be able to account for variations over time and space both; however, without data from different times, it would be foolish to assume that one has an ideal model. Still, one can collect "pooled" data, which vary over time and space both, as we have often done (Werbos and Titus, 1978; Werbos, 1983). To use such data in packages like SAS can be slightly tricky, when you estimate a model containing lagged variables. In arranging our data (Werbos, 1983), we found it necessary to include a dummy year, 1973, to precede the years for which we had pooled data (1974-1981), and we inserted the SAS missing value code for all 1973 data. Observations 1 through 8 represented 1973 through 1981 in the first state, while 9 through 16 represented the second state, and so on. (Without this, the SAS "LAG" function would not have given us valid time lags.)

5. FIVE FUNDAMENTAL PRESCRIPTIONS

This section will provide a kind of back-door introduction to the theory underlying econometrics, by trying to explain five prescriptions for avoiding gross errors which are common even among professionalists.

- o Pay attention to uncertainty
- o Don't expect a free lunch when choosing specifications
- o Pay conscious attention to prior information
- o Don't expect to draw conclusions without adequate data
- o Check the historical track record of your model (This may be the most important)

Pay Attention to Uncertainty

None of the models above -- from equation 1 to equation 7-- say anything about uncertainty. They are all forecasting models, recipes for making base case projections. Even though there are many different schools of thought in statistics and econometrics, they all agree that uncertainty needs to be addressed explicitly as a central part of the analysis.

Broadly speaking, there are two major schools of thought here:

- o The purist school, which has done an admirable job of simplifying and unifying our understanding of statistical methods, and devising new and better and more elegant methods.
- o The utilitarian school, which has made live complicated and tricky all over again, by focusing on the intractable problems which occur in real-world forecasting. (This is quite different from the quick and dirty school, which pays more attention to deadlines than to quality problems either in theory or in the real world.)

Both schools have a great deal to contribute, but we incline towards the utilitarian school.

From the purist's point of view, regression simply cannot estimate equation 1 as it stands, as if it were a meaningful model of population growth. If you regress POP(t+1) on POP(t) with no constant term, then the model you are really estimating is:

$$\mathrm{POP}(t+1) = c \cdot \mathrm{POP}(t) + e(t) \tag{8}$$

where e(t) represents a random disturbance, governed (generated) by a normal probability distribution. We sometimes call e(t) "error," but statisticians like to think of it as something out there, in the real world, rather than an "error in the sense of "mistake.". Often we call e(t) "white noise," to make this view explicit. Equation 8 is a "stochastic model," because the assumptions about the random disturbance have been made explicit as an integral part of the model.

When we look at the noise term explicitly, we can see immediately that there is something implausible about the model in equation 8. Equation 8 assumes that the noise comes from the same probability distribution in all years, implying that we should expect the same general size range for the noise in all years. If population grows by a factor of 10 in the period under study, this could be a very poor assumption about the noise; as a practical matter, this assumption would lead to an estimate of "c" dominated by the experience of the last few years, disregarding the earlier data. It is more plausible to expect that the noise will represent a certain percentage of the population, and that its size range will grow in proportion to the population, as in the model:

$$\mathrm{POP}(t+1) = c \cdot \mathrm{POP}(t) + e(t) \cdot \mathrm{POP}(t) \tag{9}$$

In this equation, the overall noise terms -- e(t)·POP(t) -- grows in size in proportion to population; in other words, e(t) -- which now represents noise as a fraction of the population -- still comes from a fixed probability distribution (imposing a fixed size range).

Equation 9 cannot be estimated directly in regression. However, now that we have a complete stochastic mode, it is legitimate to divide both sides by POP(t), as we could with any algebraic equation; this yields the equivalent model:

$$\mathrm{POP}(t+1)/\mathrm{POP}(t) \;=\; c + e(t) \tag{10}$$

This can be estimated by regressing POP(t+1)/POP(t) on no independent variables plus a constant terms; in practice, this is just a matter of estimating the average value (mean) of POP(t+1)/POP(t). It is more conventional, however, to use a similar but slightly more plausible alternative to equation 9:

$$\mathrm{LOG}(\mathrm{POP}(t+1)) \;=\; c' + \mathrm{LOG}(\mathrm{POP}(t)) + e(t) \tag{11}$$

which is equivalent to:

$$\mathrm{LOG}(\mathrm{POP}(t+1))/\mathrm{POP}(t)) = c' + e(t) \tag{12}$$

More generally, equations like equation 8 -- which assume a constant size range for error when a constant size range is not plausible or does not fit the data -- are said to have a problem with "heteroscedasticity." This is a common problem, and algebraic transformations (like the above) are commonly used to overcome it. Sometimes, however, algebraic transformations are not a workable solution. For example, when the dependent variable is LOG(QOIL/QGAS), as in the standard "logit" specification for fuel choice, there is a heteroscedasticity problem which can only be resolved by resorting to weighted regression, which explicitly treats the size range of e(t) as a function of other variables; the theory is given in Pindyck and Rubinfeld (1976), and applied in the PURHAPS model (Werbos, 1983, p.12,64). (This correction would have been desirable, but far less necessary, if we had worked with a simple time-series showing no order-of-magnitude variations in fuel shares.)

Besides heteroscedasticity, there are other possible problems with the theory that e(t) is random and normal across time. For example, e(t) may be correlated with its previous value, e(t-1). When the standard Durbin-Watson test (available in SAS and other packages) gives a score much different from 2.0, it is conventional to use a different regression command -- regression with an autocorrelation correction -- to estimate the model under the assumption that $e(t)=r\cdot e(t+1)+a(t)$, where a(t) is random; if r -- the "autocorrelation parameter" -- is not significantly different from zero, one can go back to using conventional regression.

Recently, many statisticians have begun to recommend a more careful study of the model residuals, e(t), to see if they fit more complex "Box-Jenkins" models (Box and Jenkins, 1970). In theory, certain classes of Box-Jenkins models can represent the idea that forecast errors result from a combination of noise in the real world and noise in measuring what is happening in the real world. These kinds of models can reduce forecasting errors, but tests done for real-world multiple-equation models (Werbos, 1974; Werbos and Titus, 1978) suggest that it would be better to focus on the long-range track record of a model, as we will describe below. (Engineers have another way of estimating such stochastic models, but their formulation, unlike the statisticians' formulation, contains excess parameters and can almost never be uniquely estimated.)

All of these recommendations are based on the following fundamental theorem, an application of Bayes' Law, which underlies all inference from empirical data (in statistics or in other fields):

$$\Pr(\mathrm{Model}\,|\,\mathrm{Data}) \;=\; \Pr(\mathrm{Data}\,|\,\mathrm{Model}\cdot\Pr(\mathrm{Model})/\Pr(\mathrm{Data}) \tag{13}$$

This states that the probability of a model being true, after we have observed a certain history of data, is the product of three terms. One of these terms -- Pr(Data) -- is the same for all models, and has no effect on our relative choice between models. Another -- Pr(Data/Model) -- refers to the probability that we would have observed what we did in the data if the model were true. For stochastic models, like equations 8 through 12, this term can be calculated directly by calculating what e(t) would be in all the years of data (assuming a given estimate of the parameters as part of the "Model), and then using the normal probability distribution to calculate the associated probabilities. This term is called the likelihood function; it is a function of the parameter estimate, the model, and the data. The remaining term -- Pr(Model) -- represents the probability that a model would be true, a priori, before any statistical data are examined.

Most purists agree that it would be unscientific to account for Pr(Model) explicitly in statistical estimation. They argue that all possible models (and all possible sets of parameter estimates) should be treated as equally probable a priori. They argue that modelers should always estimate these models by finding parameter values which maximize the likelihood function. Most existing statistical packages do in fact maximize likelihood exactly (as in regression) or approximately (as in iterative methods which imitate the full information maximum likelihood command for estimating systems of equations).

Bayesian statisticians have argued that economists have important information, prior to statistical analysis, about the relative probability of different models and parameter values. We would agree, but would argue that the economists' information is very complex; it would be better to use the computer to produce a complete, graphic description of the likelihood functions -- the information found in the data -- and then count on the human being to account for his prior information after the statistical analysis is complete. This puts a heavy burden on the person doing the statistics, since it is not enough to just print our estimates of one final equation; it is essential to consider the range of uncertainty for all the parameter estimates, and to consider different ways of looking at the data.

Utilitarians (like us) go further, and argue that simple statistical models are never "true" in any absolute sense. They argue that your choice of estimation method should depend on the application of the estimates or forecasts. The overemphasis on definite, base case forecasts is a product of naive decision-makers, who have yet to understand well-known procedures for coping more honestly with uncertainty (Brown et al., 1974). Indeed, one may argue (Werbos, 1979) that probabilities, rather than expected outcomes, should be the main focus of long-range planning anyway; however, the efficient implementation of this principle involves many complexities (Werbos, 1987a; Werbos, 1986a). The utilitarian Raiffa has found that elite Americans tend to understate ranges of uncertainty by a factor of 3 or so, perhaps because they do not account for the limitations of the assumptions they use. This suggests a need for great care in using mathematical models built on expert judgement rather than empirical fact. Raiffa's followers, such as Rex Brown, have developed many techniques to train, improve and organize probability assessment by human judgement; nevertheless, the problem of bias remains difficult and fundamental. It is important that modelers help decision-makers think more clearly about alternative scenarios, rather than aggravate these biases. Even though it is very difficult to estimate probabilities objectively -- when technological and political forces are primary sources of uncertainty -- it should be possible to convey the nature of uncertainty in a useful way, and explain alternative viewpoints.

Utilitarians also tend to look for estimation methods which are likely to give more accurate forecasts even when it is hopeless to formulate a model which is "true" in an absolute sense; such methods are called "robust estimation methods," The problem of heteroscedasticity leads to a simple (though unconventional) example of robust estimation. Consider the simple model:

$$\text{ENERGY-USE}(s,t) = c \cdot \text{PRODUCTION}(s,t) + e(s,t) \tag{14}$$

where energy use is projected by state (s) and by the year (t). A purist would tell us to replace e(s,t) by e(s,t)·PRODUCTION(s,t), and they divide by PRODUCTION(s,t) to get a regression equation. If we do this, we are guaranteed that the percentage error in predicting energy use will average out to zero (i.e., positive and negative errors will balance out). If we had kept equation 14 as is, we would be guaranteed that the actual error in predicting energy use will average to zero; in other words, greater attention would be paid to bigger states. If the goal is to predict total energy use, then the latter is preferable; it would lead to more uncertainty in our estimate of c, in theory, but it would also guarantee that we are estimating that version of c which is right for our application. If we admit that equation 14 is only a simplification, then we have to accept that the version of c which minimizes one error measure will be different from the version which minimizes another. In technical terms, there is a tradeoff here between statistical efficiency (i.e., random uncertainty in our estimate of c) and statistical consistency (estimating the right c). Tradeoffs of this sort are quite common, and often require some sort of ad hoc compromise.

Don't Expect a Free Lunch When Choosing Specifications

Choosing the equations of a model is a difficult process, whether the model is econometric, judgmental, or engineering-based. The process is essentially the same for all three, except that econometricians normally restrict themselves to using variables for which they have data. Econometricians often start from a general theoretical model and translate it into its implications for observable variables; there is no need to represent the entire mechanism by which variable A affects variable B if the ultimate impact is represented correctly. Also, when doing econometrics, you usually consider several alternatives, and use empirical results to decide which version to select in the end. In fact, you typically try out new alternatives after you have studied the results and looked for explanations of what is going on.

There are some analysts who offer you a hope of forecasting without resort to this difficult process. They often suggest that "simple time-series analysis" or "simple econometrics" can be an alternative to the labor and uncertainty which comes with explicit models. In actuality, this is an illusion (though the explicit models of econometrics are simpler than most engineering models). For example, "simple Box-Jenkins analysis" (Box and Jenkins, 1970) offers more complicated models of noise than regression assumes; it essentially offers yet another complex correction to explicit models (Werbos, 1974). The vendors of "simple" analysis typically apply statistical methods to a simple forecasting model, such as:

$$Y(t+1) = a + b \cdot Y(t), \tag{15}$$

where Y is the variable you are trying to forecast. Admittedly, this model is sometimes worthwhile. Admittedly, simple models in general tend to be more robust than complex models, ceteris paribus. Some salesmen have suggested that this approach can be applied to electricity demand, to save utility planners from the pain of using models which require forecasts of local industrial growth, which are fraught with uncertainty. However, this economic uncertainty is real, and unavoidable; a forecaster can hide the uncertainty from his clients (which does them a disservice), but the economic uncertainties are there and will affect electricity demand. If economic growth is known to be central to electricity demand, then it should be reflected in the model. In general, the choice of a model should be based on a careful analysis of what is known about the variable being modeled, and what is shown in the data; there is no magical way to escape this process.

Forecasting problems in private industry are sometimes so complex that analysts cannot devise an adequate specification, even when data are plentiful. In such situations, a full-scale "neuron network" system may be useful. The best neuron network systems (Werbos, 1987a; Werbos, 1988; Werbos, 1989) are essentially equivalent to a massive automated search through all possible specifications -- linear and nonlinear -- to find that specification which minimizes some combination of forecast error and model complexity. The prior knowledge of the analyst is not used at all (except in the construction of the data base.) At EIA, we have yet to encounter such situations.

In one recent situation (IFS, 1986), we had access to a massive data base on fuel-switching in which we didn't know what to expect. Elaborate cross-tabulations in SAS were very useful in helping us form hypotheses, and then formulate and estimate econometric models.

All of these examples emphasize that a modeler should take the time to devise specifications carefully. Some students are willing to do this, but expect to be given exact rules on what kinds of specifications to use. Once again, they are looking for a kind of fee lunch, and they can find a few misleading papers in journals which give them the rules they are looking for. In actuality, the choice of specification should be based on a translation of your prior knowledge (Pr(Model) in equation 13) into mathematical equations; there is no set rule for what the equations must look like, but there are guidelines for how to do the translation. A good econometrician should have some familiarity with the guidelines for translation (Brown et al., 1974; Forrester, 1961) which have been developed for models in general. Econometricians have developed further guidelines, but they are too numerous to cover here; still, please do consider what happens to your model when the independent variables take on extreme values, and do consider whether the forecasts would change the way you want them to in response to a small change in the inputs (as a function of other inputs). Also consider whether the specification really could represent alternative points of view (e.g. large and small price elasticities) through different parameter estimates.

This notion of translation between human knowledge and mathematics is so vital that it merits several examples.

First of all, translation from English into mathematics may be compared with translation from Chinese into English. In chinese, one can make statements like "man see horse." In English, this could mean that "a man saw a horse", or that "every man sees a horse sometime in his life", or that "those three men are looking at a horse", or that "this man will see a lot of horses", etc. In order to translate from chinese into English, one has to decide what tense to give the verb "see", what number or article to put before the word "man", etc. A good translator will make these decisions based on a careful understanding of the context in which the statement appears. Even then, several interpretations may still make sense; in that case, the translator may go back to the author of the Chinese statement, and ask which alternative would be used. Note that the translator can state the alternatives to someone who only speaks Chinese; the Chinese language permits ambiguity, but does not require it.

An irresponsible translator would not try to understand the Chinese original; instead, he would follow a mechanical rule, such as assuming the present tense in every sentence which does not explicitly refer to the future or the past. Irresponsible translators can easily produce paragraphs in English which look downright silly (as in the instruction manuals which come with certain imported products). In translating from English into mathematics, one can produce silly mathematics just as easily, if one is not careful about the role of time in the equations.

Second, consider a question which the Energy Modeling Forum brought up in 1985; "How much fuel-switching has there been between oil and gas in response to prices in manufacturing?" Two modelers came up with completely different answers to this English-language question, based on the same set of data (Annual Survey of Manufactures) at the State level from 1974 to 1981. One modeler (David Reister of Oak Ridge) translated the English-language question as follows: "In any given year, was the market share of natural gas in manufacturing as a whole much greater in those States where the gas price was a smaller fraction of the oil price?" The other (myself) translated it as: "In any given industry, was the change in market share from one year to the next much greater in those States and years where the change in the price ratio was also great?" These are two different questions, and it is not surprising that they yield different answers. At EIA, we have tested both kinds of specifications, and were not surprised that the latter type led to much smaller forecasting errors.

Translation back from the statistics into intuitive terms is just as important. For example, a few years ago we reviewed a major paper on new car Miles per gallon (mpg), which developed a model which predicted that a doubling in gasoline prices would double mpg in the future. Working through their vehicle attribute equation, we discovered that this forecast depended on the assumption that new cars would be eight feet tall or eight feet long in order to get higher mpg.

Pay Conscious Attention to Prior Information

Often forecasts try to provide a "most likely" view of the future, based on all of the information available. Historical information is only part of that information. Expert judgement, private sector plans, and engineering information are also part of that information. In an ideal world, we should not have to choose between econometric forecasts versus engineering forecasts, and so on; we should develop forecasts which have the highest probability of coming true, conditional upon all three kinds of information. In an econometric model, this can be approximated by choosing specifications and altering parameters, where necessary, to reflect such information.

This kind of adjustment is a tricky process. There is a risk of confusing the final user, who may not be able to tell what comes from historical trends and what comes from adjustment. Also, when adjustments are made on the basis of judgments, political biases and wishful thinking easily enter in, and cause further confusion. Adjustments based on population wisdom which in turn depends on past history may represent a "double-counting" of history or far worse. Therefore, there is much value in having some forecasters -- such as academics -- produce pure econometric models and leave the discussion of other factors to their verbal discussion sections.

The Energy Modeling Forum (Werbos, 1987b) has given us some examples of how this kind of adjustment can be done and explained to the reader. For example, if an historical trend (reported, say, as c·year in an equation to predict energy intensity) can be clearly explained as the result of using a single technology throughout the historical period, and if we are quite sure that a radically different technology will come on and dominate the forecast period, then an econometric model should be adjusted to reflect our best knowledge of the new technology. Conversely, if new technologies are expected in the future, but are numerous and hard to predict exactly, and if there were also new technologies coming on line in the historical period, one is better off trusting the econometric model.

Don't Expect to Draw Conclusions Without Adequate Data

Econometric models, when estimated, make a statement about cause and effect relations. For example, in equation 6 above, if "b" were estimated as a negative number, this would say that increases in women;s education can reduce population growth by some amount. If b were estimated as a positive number, it would say the reverse. In either case, the estimated value may be a fluke, a coincidence due to a relative shortage of data. To see if this is likely, we need to examine the "standard error" of b, which is possibly the most important statistic printed out by standard regression programs. By and large (Wonnacott and Wonnacott, 1977), the true value of b will be equal to the estimated value plus or minus the standard error, in seventy percent of the cases; it will lie within two standard deviations in ninety-five percent of the cases. Old-fashioned statisticians would say that b "is not significantly different from zero" when the value $b=0$ was between these confidence limits; however, it is better simply to report what the standard errors are, to make it clear to the reader how big b still might be (given the limitations of the data).

Large standard errors may result from any of the following:

- o Lack of data. (Having four times as many independent observations cuts the standard errors in half. However, with pooled data, the observations -- e.g., from neighboring states -- may not be entirely independent, and the true standard errors may be larger than the reported ones.)

- o Lack of an adequate specification. (Cutting historical error in half cuts standard errors in half.)

- o Correlations between the independent variables, or "multicollinearity." (This represents a qualitative lack of data -- a lack of data on situations where the independent values have different values.)

Many social scientists do not appear to realize that standard errors really do account for the effects of multicollinearity. When two independent variables do correlate very strongly with each other, there is no magic procedure to solve the problem; an honest statistical analysis will simply report that there is not enough data available to decide which variable has an impact on the dependent variable. When there is strong multicollinearity (as hinted at by large standard errors), but no really strong correlation between two variables, one should suspect three-way patterns of correlation; to locate these kinds of patterns, one can perform an eigenvector analysis of the correlation matrix, and look for the eigenvector whose eigenvalue is closest to zero. (Belsley, Kuh and Welsch have discussed these kinds of diagnostics.) When there are many variables involved, and when forecasts will be made for situations where the independent variables continue to correlate with each other, methods like "ridge regression" may be better than ordinary regression when multicollinearity is suspected (Dempster et al., 1977). When a model <u>must</u> be estimated, but the data are inadequate, one must generally fall back on prior information (and flag the resulting uncertainty).

<u>Check the Historical Track Record Of Your Model</u>

There are many economists who test out alternative specifications, and publish whatever gets the highest "R–squared" score as printed out by SAS. This often leads to disaster, because R-squared scores are not comparable between equations which represent the dependent variable in different ways; for example, an equation which predicts energy per unit of <u>output</u> will often be more accurate than an equation which simply predicts energy, but will often have a lower R-squared score (because the dependent variable has less variance). The situation is even worse with complex statistical methods as used in fields like cost function estimation; there, the "adjusted R squares" are often aggregate constructs whose relation to forecast error may be quite tenuous.

As a first step, one can try to compare mean square error (MSE) across models, because it is reported by SAS and is more often comparable between equations. As a second step, one can simply use the <u>alternative</u> equations to predict the <u>same</u> basic variable (e.g., energy consumption), and calculate the average error; this can be done in SAS by using the "OUTPUT" option to output the regression predictions to a file, and by using the numerous SAS utilities to calculate the implied predictions of energy and their errors.

In practice, there is no substitute for trying to understand what is in the data, as directly as possible. Predictions and actual values should be plotted against time, where possible, and the differences explained. This provides a basis for going back and changing the model (or better understanding its weaknesses). Plots like these are important both in estimating a model and in explaining the model to others. Tukey of Princeton has written a book on <u>Exploratory Data Analysis</u>, describing additional techniques for better understanding the residuals graphically.

In the past, some econometricians have routinely used "dummy variables" (1 in some years and 0 in others) or other procedures to throw out "outliers," observations which are hard to explain using their forecasting model. (Some statisticians have also recommended maximizing the

1.5 power of error instead of the square error, which has much the same effect.) More recent authors, like Belsley, Kuh and Welsch, have stressed to need to study the outliers (and other "influential observations") rather than simply throw them out, because they may be crucial to what your model is trying to forecast and may be important as a guide to a better model. For example, the oil shortages of 1974 and 1979 were "outliers." but a model which ignores them is a poor guide to reality. Again, the hard-to-explain observations should be obvious in a plot of predictions and actuals, but more sophisticated tools exist for identifying them.

In practice, we have also found that there is no substitute for performing a "dynamic simulation" test of a model, if you are considering the use of a model which contains a lagged endogenous variable. This is quite different from evaluating the "predicted values" which come out of a standard regression command. For example, if you were estimating equation 1 over the period from 1967 to 1985, the "predicted population" for 1980 would be calculated as "c" multiplied by the actual population in 1979, in a regression package. In dynamic simulation, the prediction for 1980 is calculated as "c" multiplied by the prediction for 1979, which in turn is calculated as "c" times the prediction for 1978, and so on. The regression test would be appropriate, in theory, if predictions one year ahead were all that you care about. The dynamic simulation test would be better if you planned to forecast further out into the future, or if the real concern for policy is the eventual result several years into the future.

A purist would argue very strongly that the regression test is adequate, if one has faith in the truth of one's model. He would argue that those lacking in faith should look for better models. A utilitarian would argue that all models are oversimplifications, and that faith without tests is no way to do modeling. Experience has shown that cumulative error tends to be quite important (or even overwhelming) in models containing lagged endogenous variables, regardless of their theoretical virtues. More to the point, it has shown that such errors can be avoided either by specifications without lagged endogenous variables (if such can be found, with otherwise comparable MSE scores) or by a new form of robust estimation.

An example of the former comes from our PURHAPS model: by filtering the effect of energy prices over time, we can represent the notion of capital-embodied price responses just as effectively as do recent academic models based on lagged dependent variables (previous period energy intensity); however, our older version may be more robust. On the other hand, there are many models (especially models which assess changes induced by policy) which have much higher error levels and much crazier parameter estimates when lagged endogenous variables are not included.

To estimate models with lagged endogenous variables, several authors -- including Larry Klein of Wharton, and myself (Werbos, 1974) independently -- suggested several years ago that models could be estimated by directly picking parameters so as to minimize errors in dynamic simulation. This could be implemented in practice by doing the dynamic simulations on a PC package like Lotus, and adjusting the parameters by hand to minimize the error in dynamic simulation. This form of robust estimation may be somewhat extreme; however, we have found (Werbos and Titus, 1978; and Chapter 4 of Werbos, 1983) that it is possible to compromise between this approach and regression, and still allow for a noise term in the model (allowing for uncertainty). Dynamic robust estimation methods of this sort have cut errors in half in a number of applications (where lagged endogenous variables were important), and have even done better in short-term forecasting (Werbos, 1974; Werbos and Titus, 1978). The compromise method looks similar to "exponential smoothing" methods which are essentially equivalent to the Box-Jenkins methods discussed above; however, they weight the square error in different ways, and this difference in weighing will probably be the key feature even of more advanced methods along the same lines.

When calculating error for a system of equations, one may simply use a weighted sum of the error for different variables (including "filtered" variables). A utilitarian would argue strongly for doing this, and for weighing each dependent variable's error according to the importance of that variable in a larger context; he would not use the classic form of full information likelihood, which is very volatile and subject to singularity problems.

6. MODEL DEVELOPMENT IN PRACTICE

Econometric methods are used for many purposes in government and industry. Many people use them for causal analysis, their original purpose as described above.

In my division of the Energy Information Administration (EIA), we use these methods to build up models of energy demand by sector in the United States economy. These models are then used as part of a larger modeling system (IFFS) which currently projects energy supply, demand, and prices by year from 1985 to 1995. The base case projections are published in the Annual Energy Outlook, along with a few sensitivity cases. IFFS is also used in a variety of special studies which come up every year; for example, it may be used to predict energy demand and supply with and without an oil import tax. EIA, like Wharton Econometric Forecasting Associates and many other forecasting organizations, maintains both an annual and quarterly forecasting model. The problems of publishing annual forecasts are similar in all such organizations, where quality is a concern, and will be discussed here in general terms.

EIA has never used econometric models exclusively to generate forecasts. The goal is always to make forecasts which represent our best guess about the future, conditional upon various assumptions about the GNP or world oil prices. The GNP forecast is taken from an economic forecaster, such as Data resources, Inc. The world oil price is projected by another division of EIA, and alternative scenarios are developed to reflect the uncertainties in this projection. Our goal is to report the best guess we can, accounting for all sources of information, including historical data, engineering ada, the trade press, etc. In some sectors, we begin with econometric models, and in others we begin with engineering process models; in either case, we try to inform or calibrate the model, accounting for information from other sources.

In developing a new model, one rarely starts from scratch. There is usually an existing model that was used in the previous year to forecast the same general concept. Generally speaking, we go through six stages (when we have the resources to do things correctly):

- preliminary evaluation of the existing model(s)
- detailed literature review
- assessment of sources of uncertainty and how to reduce them
- data acquisition, econometric analysis, and the like
- coding and testing
- model maintenance

Preliminary Evaluation

The preliminary evaluation usually starts from the annual review of model forecasts which occurs as part of the annual forecasting cycle. Most reviewers have certain expectations about what the forecast should look like, based on a variety of sources. If a forecast is much different from these expectations, it is scrutinized further. (Unfortunately, it is more difficult to identify poor forecasts adjusted by brute force to match common expectations; however, when there are competing viewpoints represented in the process, based on deeply held professional or political orientations, any forecast may be different from what someone expects. It is common to be told within the same week, by different people, that a given elasticity is "absurdly high" and "absurdly low" both.) "Industry" is widely cited in this process, but in a haphazard way and in different directions.

Next, the analyst tries to explain why the forecast differs from expectations. Often, this explanation involves a plot of historical data, showing how the assumptions of the model perform compared with more popular assumptions. Simple plots and growth-rate calculations are vital to these explanations, which take up a major part of the analyst's time; translations back and forth between English, statistics, and simple plots or tables are central to the process.

A critical advantage of econometric methods at this point is that they do have an historical track record and a corresponding data base. Also, they may be simpler to understand and explain than the process models; they do not contain large files of assumptions which can carry hidden biases. Process models, on the other hand, can simply be adjusted to match the prior expectations. In the past, process models were more popular than they are now, because they could show large impacts from proposed regulations such as the Powerplant and Industrial Fuel Use Act. Models such as the Project Independence Evaluation System have come under great criticism for their optimism about oil imports, even though the authors worked hard to warn the reader in the document that the conditions for zero imports by 1985 (as requested by the policy-makers) would be very difficult to meet.

If the explanation for a forecast does not have enough information behind it, or is not communicated properly, adjustments are made, and the issue will be studied in more detail after the current forecasts are produced. (Econometric models can be adjusted and rerun at least as easily as process models, but the adjustments tend to be more visible.) Likewise, questionable or uncertain forecasts are usually revisited even if they are not initially modified. At times, a detailed review from the Quality Assurance Division may spark a preliminary evaluation. If the preliminary review of a model then suggests a major problem, the problem goes to the top of the priority list.

A great difficulty in this process is to sort out the difference between common expectations, political or other prejudices, and objective reality. When there are many layers of analysis between the raw data and the final publication, there is often a danger that the likelihood function will be multiplied <u>many times</u> by the same prior probability distribution, therefore biasing the publication a bit too heavily towards the priors; there is no cure for this problem other than improved communication.

In general, this process tends to be very instructive, but the details are rarely published; they tend to be viewed as too technical.

<u>Detailed Literature Review</u>

When there are major problems with an existing mode, the model may or may not be rewritten; that depends on the nature of the problem. The first step in reviewing the problem is a thorough review of alternative models and forecasts, and the information behind them. Uncertainty assessment is the next stage, but the literature review itself tries to find complete information for that next stage; thus, an informal uncertainty analysis is always being done.

The literature review requires an evaluation of the statistical methods used in the existing models. It requires an effort to understand what is going on, substantively, in the sector being modeled. It requires a great deal of thought about how <u>different</u> phenomena, in the real world, would influence different kinds of statistical analysis or model in different ways. (Among the important phenomena are things which bias different data collection efforts.) It requires skepticism and a search for evidence when judging the statements of other modelers and of "substantive experts" both. It requires a search for possible biases, as with models which make strong assumptions about the costs of new technologies. It requires a search for widely opposing points of view. Above all, it requires an effort to understand why different forecasts of the same variables from different sources come out differently.

After the literature review is largely complete, it is usually important to write a brief memo on the issues which have emerged, and to get feedback from other analysts.

Uncertainty Assessment

The most critical stage in modeling, most often neglected, is the analysis of first-order uncertainty. First-order uncertainties are those things which explain most of the differences between the existing models (or potential models). It is sad how often government agencies have spent millions of dollars on detailed, highly precise models, without checking to be sure that the gross, first-order sources of error have been thoroughly understood.

A key part of the first-order analysis is an approach to explaining the basic, first-order trends in past history. This includes a plan to resolve or reduce the uncertainties by statistical analysis or by the most reliable, empirically-based methods available. After the first-order uncertainty has been resolved, one can move on to the second-order uncertainty, and so on; at all times, no complexity needs to be added (on the conceptual level, anyway) unless it really helps reduce the basic uncertainties.

With behavioral effects, such as price elasticities, energy demand, or fuel-switching sensitivities, the historical data provide a good way to estimate model parameters; econometric techniques are appropriate. With engineering variables, such as synfuels costs, we have supported statistical cost-calibration analysis by Ed Merrow at the RAND Corporation; early studies, taking engineering estimates at face value, had been too optimistic. The key question about technology is how much changes in its rate of development are likely to change trends or price responses in the future; often, these changes introduce uncertainties both on the upside and the downside, uncertainties which it is difficult to resolve realistically. Changes in markets are often at least as important as changes in technologies as such.

In some cases, as with industry, the technological change has been so complex and so continuous (on average) that we use an econometric approach almost completely (except with petrochemicals). With transportation, a hybrid approach is used because of the unique role of new car miles per gallon (mpg). In the residential and commercial sectors, a key problem (not yet resolved) is to reconcile the conflicting studies which claim that conservation has mostly been due to lower thermometers (which won't continue when prices stabilize) or due to structural improvements in new buildings (which will continue as housing stocks roll over).

Uncertainty assessment cannot be done unless you know which forecast you are trying to evaluate. At EIA, our main concern is to get the national base-case forecasts right, and to get an accurate response to moderate price changes and economic growth. When new policy issues arise, a new assessment is needed for the new issue; at times, this may force the development of a new model. For example, a few years ago, when issues about contracts were considered fundamental to natural gas regulation issues, a new model was developed which specializes in that area; now, however, the empirical issues of base case accuracy are beginning to get more priority again.

According to industry representatives at the Energy Modeling Forum, private corporations often have a special need for State-level forecasts or the like. Unexpected national trends probably invalidate the State-level forecasts as often as any other problem does; likewise, behavioral assumptions at the State level may be a problem. For this reason, insights developed in national analysis may be useful to private industry. However, the forecasts themselves have rarely been subject to uncertainty analysis at the State level, even when State-level data are available; therefore, they are not to be taken at face value. EIA does publish regional forecasts, in special service reports; however, these forecasts are produced mainly because they are important to the accuracy of the national forecasts. The regional forecasts are checked, but are not evaluated in depth.

Analysis

The work at this stage should follow directly from the analysis of uncertainty. Econometric techniques are used at this stage more and more at EIA, in part because we have a comparative advantage in using data and in part because of cost considerations. (A thorough and conclusive analysis of even one technology is not cheap.)

Looking at models as boxes, we often ask what kind of real knowledge or information is really "in" the box; our econometric models contain primary information about what can be learned from history, based on the analysis stage. Because of the adjustment process, they contain a bit of secondary information that we borrow from other sources about markets and technologies. Good process models contain primary information about technologies, developed at the analysis stage; however, they are adjusted to match historical reality, and may be secondary users of econometric information. Some hybrid models or global models tend to be secondary users of both kinds of information.

Coding and Testing

This is straightforward conceptually; I wish it were so easy to do. At EIA, we code our models in FORTRAN, because even our econometric models are too complex for basic systems like Troll to handle at present. The statistical analysis is mostly done in SAS, because SAS is flexible and easy to use with complicated data bases; the DATA and PROC MATRIX components of SAS are especially powerful.

Model Maintenance

The uncertainty analysis and review described above continue for all models, even after they are established, on a regular cycle.

General Observations

After this exercise is completed, it often seems that the insights gained in modeling are more important to policy than are the forecasts themselves, if the job was done right. Good Bayesians know that the future is highly uncertain, and that any policy which is based on a definite expectation of a "baseline future" is likely to be a poor policy. People at EIA and elsewhere vary greatly in their willingness to describe and explain the full range of uncertainties, as opposed to defending the base case or presenting a small number of conservative sensitivity cases in line with popular expectations.

The steps needed to get good forecasts are also needed to refine one's understanding of cause and effect, as I pointed out in the discussion of maximum likelihood theory above. When cause-and-effect analysis is done outside of the forecasting context, it is often based on quick, casual statistical models or analysis by eyeball; it often derives incorrect conclusions because it fails to account for issues we learn about in forecasting. Unexpected futures are also important to cause-and-effect analysis, as they are to forecasting. It is a continuous challenge for modelers to find ways to exchange such insights and make them fully available to decision-makers.

Of EIA's model based reports, the Short-Term Energy Outlook is often cited as the most respected. These quarterly forecasts are based on simpler models, and are more completely econometric than the annual forecasts are; the model assumptions and historical track record have been communicated effectively in recent years, and there has been feedback from high levels of the government. The communication is more difficult for the annual model, for at least three reasons:

- o the annual model contains major process-model components (mainly on the supply side)
- o the published year-ahead forecasts are not a good basis for testing track records, because they are calibrated to the quarterly forecasts.
- o the annual model makes use of more detailed data sources, some of which have been discontinued, changed in scope, or collected at irregular intervals; while the models account for this, it is not possible to put all the relevant information on one neat plot that goes up to 1985

A key problem in comparing published long-term forecasts over time is to control for GNP assumptions and world oil price assumptions. For example, optimistic assumptions about economic growth, taken from other government components, can push up energy demand forecasts; energy use in industry tends to be very volatile with respect to the rate of growth, and few government forecasts ever show major recessions. EIA has published plots showing the track record of our model in predicting world oil prices as a function of OPEC capacity utilization; the track record has been surprisingly good, but earlier forecasts of OPEC capacity utilization led to major errors, due in part to economic assumptions.

7. EXAMPLE

The Purchased Heat and Power System (PURHAPS) is one of the models I have developed at EIA, following the approach described above. The core of PURHAPS is a system to predict purchased fuels and electricity in manufacturing in the United States, excluding refineries. This represents only half the energy covered in our industrial sector, but there is not enough space here to discuss all the rest. This part of PURHAPS contains good examples of the issues discussed above. For other -- simpler -- examples of these issues, see also our Transportation Energy Demand (TED) model (Werbos, 1986b) (which should be adjusted to reflect more recent analysis of alternative fuels (Werbos, 1987c).

EIA management decided it needed a new model of industrial energy use early in 1982, after several analysts in several divisions agreed that the existing model was unsatisfactory. The existing model, a version of the Oak Ridge Industrial Model (ORIM), was driven by "econometric equations of total energy use which had not been derived from empirical data, and which could not be estimated very easily; the fuel-switching was driven by a "process representation" which did not reflect differences between the different 2-digit industries, differences which seemed critical in our reading of the literature. ISTUM2 did have an adequate process representation, and was seriously considered for use; however, the cost of calibrating all the technology data, and doing the historical calibration we would require for our purposes, was enough to rule it out.

The decision was made to start out, at least, with an econometric model, and add engineering detail only when and as there was reason to do so. I proposed that we first get the basics right, and then consider more elaborate models of primary metals, chemicals, and paper after this point is reached. In later years, the possibility of more elaborate industry models was considered but never seemed to promise much reduction of uncertainty at acceptable cost; however, the system is still considering a suggestion that we buy data from SRI International which might allow a better treatment of petrochemical feedstocks, for which the usual data appear contradictory and problematic. (Some analysts have recommended that we model this sector equating ethylene with petrochemicals, but the complexity and flux of this industry suggests that this would not buy us anything better than the simple, aggregate approach we now use.)

Our initial analysis goal was two-fold:

- o To get total energy use right, at the national level, we wanted a model to track energy use in the individual industries (for which good historical data were available back to 1958) as well as possible.

- o To get fuel-switching right, we wanted a model to represent State-level tradeoffs within specific industries. Other EIA demand models do not carry State-level detail, but it was felt that fuel price tradeoffs vary a great deal from State to State in industry; also, the availability of State-level data from the Annual Survey of Manufacturers was critical.

Forecasting Total Energy By Industry

Based on our initial literature review, it was felt that Dale Jorgenson's work with Barbara Fraumeni was the best starting place for the national energy equations. (Other work by Ernie Berndt at MIT appeared very promising, but too difficult to update in time for the next Annual Energy Outlook.) Dale Jorgenson was given a small contract to update his database on the price and quantity of capital, labor, energy, and materials used in industry through to 1979, and to reestimate the equations he had published.

As part of our normal review process, we tested the resulting equations with Jorgenson's own data to see how well they would have performed in the past. The results were a great disappointment to myself and to Jorgenson; we have published a curve or two of the errors, but have otherwise tried to move on to something better. Some people suggested that the Jorgenson effort had been ill-conceived; however, the purpose of scientific research is to discover new information, as we had, and an updated database was needed in any case.

The original Jorgenson equations had been estimated as a complete system, using a system of constraints and estimation methods related to 3-stage least squares. We tried to reestimate Jorgenson's energy model, using only the energy equation and multiple regression; the errors were less, and some people wanted to go ahead and use the results, but the price elasticities did not make sense over time and the problem turned out to be unavoidable with the "translog" model being used.

We then estimated a similar equation, using the same data base and still controlling for the effects of capital prices, etc., but assuming a constant elasticity of energy prices on energy demand per unit of production. This worked better, and made much more sense in historical plots. It also showed much smaller price elasticities than the theoretical economists had been publishing in recent years (based largely on the translog or on data from 1958-74).

In the following year, we studied the residuals more carefully, and found some signs of bias. We explained this bias as a result of "capital-embodied price responses," which represent the effect of slow capital turnover on energy conservation; this explanation led to a much better fit, including average errors of about 1% in all the major industries in the postembargo period. The equation has no lagged endogenous variable to allow the possibility of cumulative error in forecasting several years out. (The new Transportation Energy Demand (TED) model has slightly larger errors -- 1.8% -- in forecasting total personal travel; however, because these errors are visible in a single plot of actual travel versus predictions from 1954 to 1983, the personal travel model is probably viewed as a bit more reliable.)

Forecasting Fuel Shares

Our analysis of fuel shares also produced a few surprises.

Coal use changed so slowly over time that we needed a nonlinear model (requiring nonlinear regression) to track it at all. Coal use turned out to be complementary to natural gas, to a very slight degree; at first, we thought that this was an artifact of our equations, but we found out that the historical fits got noticeably worse when we "corrected" the specification to cut out the complementarity.

With electricity use, we initially used a lagged endogenous variable to represent long-term price effects. However, we discovered that the parameter estimates were very unstable when we changed the equations slightly, because e of "robustness" problems; we then settled on something a little simpler.

With oil-gas switching and with electricity-fossil shifts, we predicted <u>changes</u> in shares as a function of <u>changes</u> in prices from one year to the next. Initially, we tried a simpler version with oil-gas switching, predicting shares as a function of prices, but a comparison of the two approaches verified our suspicion that the simpler version would be biased. States <u>initially</u> varied in fuel shares in 1974 because of differences in industrial mix which depend on other factors; an equation based on <u>changes</u> assumes that different States may have different fuel shares even when prices are equal, and it does lead to much smaller errors. (The MSE has been published for both versions, and is several times larger for the simpler version.)

With oil-gas switching and electricity shares both, we were worried that trends might have changed over time, or that long-term price effects might be larger than short-term price effects (i.e., that the simple specifications might have robustness problems). To test this, we tried specifications projecting changes in shares from 1974 and 1978 to year t as a function of price changes over the same period; little change was found. To test the trends, we used "dummy" variables (each set to 1 for a particular year and 0 in all others) in place of the constant term in these equations; we found little noticeable effect with gas, but we did decide to use the average trend over the last four years for electricity shares.

<u>Sources of Further Information</u>

Even the important details of this story would fill up too much of this book. For a concise description of the model, please consult Volume I of the PURHAPS documentation (Werbos, 1984). Volume III (Werbos, 1983) gives the story and the rationale behind the <u>initial version</u> of the model in much more detail (one chapter per equation). The nonmanufacturing portion of the model has changed since then. More recent information (including updates and adjustments) is given in a number of papers, mostly focusing on the substantive meaning of our forecasts (Werbos, 1987b; AEO, 1985; Werbos, 1987d).

<u>Structure of the Manufacturing Model</u>

The basic structure of PURHAPS is shown in Figure 1. For each of 17 manufacturing industries, 7 econometric equations were estimated, corresponding to the 7 in the chart. Each of the State-level equations is applied 51 different times in ever forecast year, once for each State; the results differ across States simply because the fuel prices, industrial production, and previous-year conditions differ across States. The State forecasts for total energy use are not simply "shared out" from the national total; in each industry, the national equation and the 51 State equations form a system of 52 simultaneous equations, each constituting a well-thought-out model, which are "solved" as a system.

The original Jorgenson model was a "KLEM" model, where "K" represents capital, "L" labor, "E" energy, and "M" intermediate materials. PURHAPS may be called a "KIM/LEO" model, because capital, materials, and time (I) are accounted for a national level, while labor, energy and output are also used at the state level.

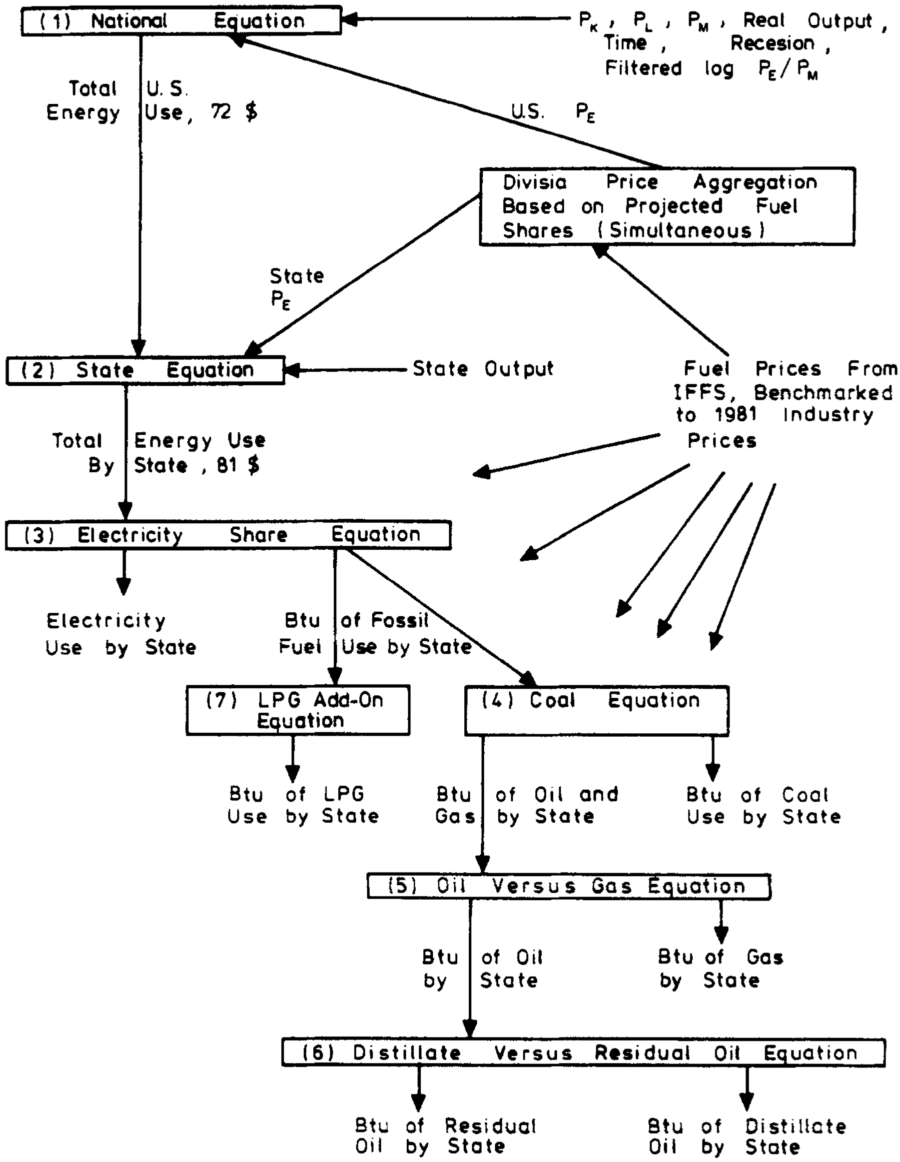


Figure 1. Structure of the Manufacturing Model

REFERENCES

Annual Energy Outlook (1985). DOE/EIA-0383(84), NEIC.

Box, G. E. P. and G. M. Jenkins (1970). Time-Series Analysis: Forecasting and Control, Holden-Day: San Francisco, Calif.

Brown, R. et al. (1974). Decision Analysis: An Overview, Holt, Rinehart and Winston: New York.

Dempster, A. et al. (1977). "A Simulation Study of Alternatives to Ordinary Least Squares", Journal of the American Statistical Association.

Forrester, J. (1961). Industrial Dynamics, MIT Press: Cambridge, Mass.

Industrial Fuel Switching (1986). DOE/EIA-0504, NEIC: Washington D.C.

Pindyck, R. S. and D. L. Rubinfeld (1976). Econometric Models and Economic Forecasting, McGraw-Hill: New York.

SAS Users Guide: Basics (1985a). SAS Institute: Cary, N.C.

SAS Users Guide: Statistics (1985b). SAS Institute: Cary, N.C.

Troll Users Guide (1981). MIT Center for Computational Economics and Statistics: Cambridge, Mass.

Werbos, P. (1974). "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. Dissertation, Harvard University, Cambridge, Mass.

Werbos, P. (1977). "Advanced Forecasting Methods For Global Crisis Warning and Models of Intelligence", General Systems Yearbook.

Werbos, P. (1979). "Changes in Global Policy Analysis Procedures Suggested by New Methods of Optimization", Policy Analysis and Information Systems.

Werbos, P. (1983). A Statistical Analysis of What Drives Industrial Demand. DOE/EIA-0420/3. National Energy Information Center (NEIC, 202-586-8800): EIA, Washington D.C. (Should be available at no charge, or contact the author at 202-586-1197.)

Werbos, P. (1984). Documentation of the PURHAPS Industrial Demand Model, Volume I. DOE/EIA-0420/1, NEIC: Washington D.C.

Werbos, P. (1986a). "Generalized Information Requirements of Intelligent Decision-Making System" in SUGI 11 Proceedings. SAS Institute: Cary, N.C. (A revised version, available from the author, is easier to read and contains more discussion of psychology.)

Werbos, P. (1986b). Documentation of the Transportation Energy Demand (TED) Model, DOE/EIA-MO13, NEIC: Washington D.C.

Werbos, P. (1987a). "Building and Understanding Adaptive Systems: A Statistical/Numerical Approach to Factory Automation and Brain Research", EIII Transactions on Systems, Man and Cybernetics.

Werbos, P. (1987b). "EMF 8: A Modeler's Perspective" in The Changing World Energy Economy, International Association of Energy Economists: Washington D.C.

Werbos, P. (1987c). Oil Dependency and the Potential for Fuel Cell Vehicles, SAE 871091, Society for Automotive Engineers: 400 Commonwealth Drive, Warrendale, Pa. 15096. (On request from author.)

Werbos, P. (1987d). "Industrial Structural Shift: Causes and Consequences for Electricity Demand", in A. Faruqui and J. Broehl, (eds.) The Changing Structure of American Industry and Energy Use Patterns, Battelle Press: Columbus, Ohio.

Werbos, P. (1988). "Backpropagation: Past and Future", Proceedings of the Second International Conference on Neural Networks. IEEE: New York.

Werbos, P. (1989). "Maximizing Long-Term Gas Industry Profits in Two Minutes In Lotus Using Neural Network Methods", IEEE Transactions SMC.

Werbos, P. and J. Titus (1978). "An Empirical Test of New Forecasting Methods Derived From a Theory of Intelligence: The Prediction of Conflict in Latin America", IEEE Transactions on Systems, Man and Cybernetics.

Wonnacott, T. H. and R. J. Wonnacott (1977). Introductory Statistics for Business and Economics, Wiley: New York.

ELSEVIER

Electrical Power and Energy Systems 26 (2004) 413–421

ELECTRICAL POWER & ENERGY SYSTEMS

www.elsevier.com/locate/ijepes

Power generation planning: a survey from monopoly to competition

Argyris G. Kagiannas*, Dimitris Th. Askounis, John Psarras

Decision Support Systems Lab (EPU-NTUA), Department of Electrical and Computer Engineering, National Technical University of Athens, 9, Iroon Polytechniou str., 15773, Athens, Greece

Received 21 February 2002; revised 14 October 2003; accepted 7 November 2003

Abstract

During the last two decades electric power generation industry in many countries and regions around the world has undergone a significant transformation from being a centrally coordinated monopoly to a deregulated liberalized market. In the majority of those countries, competition has been introduced through the adoption of a competitive wholesale electricity spot market. Short-term efficiency of power generators under competitive environment has attracted considerable effort from researchers, while long-term investment performance has received less attention. In this context, the paper aims to serve as a comprehensive review basis for generation planning methods applied in a competitive electric power generation market. The traditional modeling techniques developed for generation expansion planning under monopoly are initially presented in an effort to assess the evolution of generation planning according to the evolution of the structure of the electric power market.

Keywords: Power generation planning; Monopoly; Competition

1. Introduction

Throughout the world, the electricity industry is currently undergoing significant restructuring towards deregulation and competition. One impetus for this change is the belief that electricity generation no longer possesses the sub-additive cost properties of a natural monopoly due to technologically driven decreases in efficient plant sizes. As Banks [1] mentions, *this kind of thinking led to the question: if smaller installations are economically justified, then why should they be owned by a monopolist?* Diminishment of scale of economies in generation has made competition possible among power producers. This restructuring has broken the utility industry into generation firms who compete among each other to sell power, which is transmitted by a monopoly high-voltage transmission system to independent distribution firms and local customers. For generation firms, it is now very important to be able to analyse and to model the behavior of the market in order to make decisions with the highest level of information. This is true for existing firms, potential new investors or any entity interested in the electricity market [2].

Either under a more organized market scheme in which a market operator centralizes generation and demand bids, clears the market and leads the settlement process, or under a more decentralized one where physical bilateral contracts prevail, electric firms must assume much more risk and responsibility on their own decisions. These changes have drastically altered the nature of utility planning. Most generating capacity additions are now provided by non-utility generators [3]. A number of studies concerned with the competitive performance of electricity spot markets exist in the literature [4,5]. However, less attention has been paid to the long-run efficiency of restructuring—specifically, the area of investment in generation [5–7].

Optimal long-term generation expansion planning (GEP) is traditionally perceived as the determination of the minimum-cost capacity addition plan that meets forecasted demand within a pre-specified reliability criterion over a planning horizon (typically 20 years). Capacity expansion models have a long tradition in both the power sector and the operations research literature. They were one of the first applications of linear programming in the 1950s when the industry was operating under the regulated monopoly regime [8].

* Corresponding author. Tel.: +30-210-7723583; fax: +30-210-7723550.

E-mail address: akagian@epu.ntua.gr (A.G. Kagiannas).

0142-0615/$ - see front matter
doi:10.1016/j.ijepes.2003.11.003

Just as former monopolies must adapt to survive in newly deregulated competitive markets, so must existing planning tools if they are to remain useful in the new market environment. Dyner and Larsen [9] discuss the changes occurred in the electricity planning process, including methods from optimization to simulation and forecasting, as deregulation takes place and illustrate which new methods should be included in the electricity planning in the future. This paper focuses on generation planning, and aims to specify the main advancements occurred under a deregulated competitive electricity market, starting from the traditional concept in a central electricity market.

The rest of the paper is structured along four sections, as follows:

- Section 2 is a review of the GEP in a centralized monopolistic electricity market and the models mainly developed and used to deal with this problem.
- Section 3 emphasizes the need for new techniques for GEP under the new era of wholesale power competition
- Section 4 is a review of the literature regarding GEP under competition.
- Section 5 is the conclusions derived by the present study.

2. Generation planning in monopoly

The traditional aim of an electric power utility has focused on providing an adequate supply of electric energy at minimum cost. Various models for generation planning were developed to fulfill this function through optimization algorithms and probabilistic production costing (PPC) simulation. Recently the environment factor has been added to the problem of generation planning as a major new dimension [10].

The purpose of generation planning models is to determine the generation units to be constructed, the time to be constructed and the amount of power to be produced while the total cost (fixed and production cost) to a utility is minimised [11]. Anderson [12] reviews different types of mathematical programming models that have been used for generation planning, while Pokharel and Ponnambalam [13] discusses some of the essential issues of electricity generation expansion planning and develops a methodology to analyse electricity planning when the variables are deterministic and stochastic in nature.

Mathematically speaking, solving an optimal GEP problem is equivalent to find a set of optimal decision vectors which minimize an objective function under several constraints. The mathematical formulation of the traditional GEP problem has been presented by Park et al. [14], according to the one used in the WASP model [15]:

$$\min_{U_1,\ldots,U_T} \sum_{t=1}^{T} [f_t^1(U_t) + f_t^2(X_t) - f_t^3(U_t)] \tag{1}$$

$$\text{s.t. } X_t = X_{t-1} + U_t \tag{2}$$

$$\mathrm{LOLP}(X_t) < \varepsilon \ \ (t = 1, \ldots, T) \tag{3}$$

$$R_l \le R(X_t) \le R_u \ \ (t = 1, \ldots, T) \tag{4}$$

$$\underline{M_t^j} \le \sum_{i \in \Omega_j} x_t^i \le \overline{M_t^j} \ \ (t = 1, \ldots T,\ j = 1, \ldots, J) \tag{5}$$

$$0 \le U_t \le \bar{U}_t \ \ (t = 1, \ldots T) \tag{6}$$

where:

T	the number of years of the planning horizon
J	the number of fuel types
Ω_j	the index set for jth fuel-type plants
X_t	is the cumulative capacity vector (in MW) by plant type in year t (state vector)
x_t^i	the cumulative capacity of the ith plant type in year t
U_t	the capacity addition vector by plant type in year t (decision vector)
$\overline{U_t}$	the maximum construction capacity vector by plant type in year t
u_t^i	the capacity addition (in MW) of the ith plant type in year t
$LOLP(X_t)$	the loss of load probability with X_t in year t
$R(X_t)$	the reserve margin with X_t in year t
ε	is a reliability criterion expressed in LOLP
$\underline{R_u}$, R_l	the upper and lower bounds of the reserve margin
$\overline{M_t^j}$, $\underline{M_t^j}$	the upper and lower bounds of the jth fuel type's capacity in year t,
$f_t^1(U_t)$	the discounted investment costs associated with capacity addition U_t in year t
$f_t^2(X_t)$	the discounted fuel and O and M costs associated with capacity X_t in year t
$f_t^3(U_t)$	the discounted salvage value associated with capacity addition U_t in year t

To consider investments with longer lifetimes than a planning horizon, the linear depreciation option is used as in the WASP model [15]. The types of constrains that are mentioned above are as follows:

- Eq. (2) implies a state equation
- Eqs. (3) and (4) are related with reliability constraints (LOLP criterion and reserve margin bounds)
- Eq. (5) reflects the capacity mixes by fuel type
- Eq. (6) gives the yearly construction capabilities by plant type

Although the state vector X_t, and the decision vector U_t in the year t have dimensions of MW, we can easily convert these into vectors that have information on the number of each plant.

As obvious, a long-term GEP problem is a highly constrained non-linear discrete dynamic optimization problem that can only be solved by complete enumeration

[15–19]. Therefore, every possible combination of candidate options over a planning horizon must be examined to get the optimal plan, which leads to a computational explosion [14]. The high non-linearities of a GEP problem basically originated from the probabilistic production costing simulation and a set of physical/engineering non-linear constraints [15,16,20].

To solve this complicated problem, a number of methods were applied successfully during the past decades. Among them, dynamic programming (DP) is one of the most widely used algorithms in GEP [15–22]. Although DP-based approaches have some advantages over other algorithms, the so-called 'curse of dimensionality' [3,14] has interrupted the direct application of the conventional DP to practical GEP problems [15,16,18,19]. For this reason the WASP [15] and EGEAS [16] use a heuristic tunnel-based technique in the DP routine where users pre-specify state configurations and successively modify tunnels to arrive at a local optimum. Caramanis et al. [23] used dynamic programming (DP) to solve this investment problem in stages. At each stage, several polans or stages are identified and then investigated using PPC simulations, from which the best ones are saved to the next stage. Karaki et al. [10] present a GEP model based on tunnel dynamic programming (TDP) and PPC integrating a model for air effluents of generating units. The DP is based on the work of Caramanis et al. [23] with modification on the tunnel sizing procedure. The PPC simulation is based on combining a probabilistic generation model with the load duration curve of the system to deduce a risk model from which the expected energy not supplied is calculated along with the expected energies produced by the units. The model was applied in an expansion case study of the power generation system of Electricite du Liban.

Bloom [17] used generalized Bender's decomposition (GBD) algorithm to sub-divide the master GEP problem into a set of sub-problems, which are solved in an iterative way until the optimum cost is found. The master problem is solved using linear programming, and the sub-problems are solved using PPC simulation techniques. The modeling of environmental constraints is done, at the PPC level, as an energy limit on selected units. Simulating the operation of energy limited units is based on the economic equivalence of peak-shaving mode and the slice-in mode of operation [24]. Bloom and Gallant [25] present the 'facet algorithm' based on formulating the PPC calculation as a linear program, which can incorporate dispatch constraints as linear inequalities thus allowing the quick evaluation of binding constraints and the activation of new ones as needed. Masse and Gilbrat [8] applied a linear programming approach that necessitates the linear approximation of an objective function and constraints. Park et al. [26] applied the Pontryagin maximum principle whose solution lies in a continuous space. Also, there is some research on the development of static optimal mix by Levin and Zahavi [27] and Ramos et al. [28] based on non-linear programming frameworks.

Ammons and McGinnis [29] develop a comprehensive generation expansion planning model to determine the generation units to be constructed and the production level of each new and existing generation unit. Belgari and Laughton [30] and Sawey and Zinn [31] develop models for large-scale generation planning for the combination of generation and transmission operations. The above papers focus on relatively easy-to-quantify factors such as fixed and production cost [11]. Customer outage cost and utility outage cost are not considered in most of the generation planning literature. A few studies in the literature have considered utility and customer outage cost. Hobbs [3] incorporates scheduled utility and scheduled customer outage cost in the objective function of a model (i.e. the amount of the scheduled outage is a decision variable). Relative to scheduled outage, forced outage is unexpected and the resulting damage may be enormous. Wang and Min [32] consider the forced outage cost as a component of the total cost to a utility. However, in their model, only utility outage cost is considered.

The issues of risk and uncertainty that are usually associated with GEP were mainly addressed by stochastic optimization, decision analysis and tradeoff analysis [33, 34]. Gardner [35] presented a multi-stage stochastic electric utility planning model to compare the flexibility benefits of different electric utility resources when dealing with demand uncertainty. Felder [36] evaluates two financial approaches (risk adjusted discount rates and option theory) that incorporate risk in electricity resource planning.

A global optimization technique using a Genetic Algorithm (GA) has been successfully applied to various areas of power system such as economic dispatch [37,38], unit commitment [39–41], and others [42–44]. GA-based approaches for optimal GEP have several advantages. Naturally, they not only can treat the discrete variables but they overcome the dimensionality problem. In addition, they have the capability to search for the global optimum as well as quasi-optimums within a reasonable computation time. However, there exist some structural problems in the conventional GA such as the premature convergence and duplications among strings as evolution is progressing [45]. Park et al. [14] proposed a hybrid GA approach incorporating the tunnel-based DP scheme of WASP. The basic idea of this approach is based on the GA's capability to find the global optimum or quasi-optimums and the tunnel-based DP's high performance to find a local optimum. Fukuyama and Chiang [46] and Park et al. [47] applied a genetic algorithm to solve a test GEP problem, and showed promising results.

The multicriteria decision making models allows the evaluation of options against a wide range of criteria, grouped in a hierarchical structure. Mills et al. [48] present a tool that facilitates a multiobjective-driven electricity planning process, based on the Integrated Resource

Planning approach. Voropai and Ivanova [49] suggest a general approach to the multicriteria decision analysis on GEP, which unites different multicriteria analysis techniques in terms of different preference relations in the analysis of decisions and their combination. Beccali et al. [50] present a methodological tool in order to help the decision maker to synthesize a large set of variables, comparing the ELECTRE method to a fuzzy set approach. Martins et al. [51], present a multiple objective linear programming (MOLP) model for power generation expansion planning which accounts for demand-side alternatives in an equal footing with supply-side options. Mavrotas et al. [52] presented an approach based on mixed 0–1 MOLP model and applied to the Greek electricity generation sector for identifying the number of output of each type of power units needed to satisfy the expected electricity demand in the future. The core of the model is a branch and bound algorithm, which has been properly modified for the multiple-objective case and is capable of generating the whole set of efficient solutions.

Among the artificial intelligence techniques, expert or knowledge based systems have been the most successful, in particular as far as power system planning is concerned. In 1989, Zhang et al. [53] presented a bibliographical survey of expert systems in electric power systems. Madan and Bollinger [54] continued this work by presenting the application of artificial intelligence (mainly expert systems) to power systems. Balu et al. [55] and Adapa [56] concentrated on the application of expert systems in power system planning. While artificial intelligence is widely used in power planning, including transmission planning [57], distribution planning [58], voltage/reactive operation planning [59] and load forecasting [60], its application in GEP is still rare in the literature, to the best of our knowledge. David and Zhao developed a heuristic expert systems-DP (1989) [18] and applied the fuzzy set theory (1991) [19] to reduce the number of search states.

3. From monopoly to competition

Until the 1970s, the resource planner's task was just to determine the best size, timing and type of large central station generation plants to meet future electric loads [3]. The objective of the capacity expansion models is to select the mix of plants that minimizes the total cost of satisfying a time (and sometimes randomly) varying demand over a typical horizon of say twenty years [6]. During 1990s DSM options have been included in order to form the integrated resource planning (IRP) models [61–63]. Modeling demand side management programmes has been widely covered in the literature [64–67]. In addition, the planning procedure has become even more complex, through the disposal of several investment options [3]. A number of methodologies and models have been presented in the literature during the last two decades that deal with the above problem using several approaches of optimization techniques, as presented in Section 2.

However, the way that generation expansion planning has been approached and solved has been totally redirected through the introduction of competition and deregulation of electricity markets. The problem of power GEP has been reformulated from being cost-minimisation to profit-maximisation. The privatised approach evaluates a resource alternative's benefits according to its own revenue stream. This private revenue stream will depend on a number of factors (i.e. how privatisation is carried out and the type of market or contract). The greater uncertainty in load growth, fuel markets and government regulations have made the problem even more complex. Not only are future demands for electricity within a given region uncertain, but competition with other power generators also means that many utilities cannot take their market share for granted.

Historically, resource planners have focused on what resources should be chosen rather than the dual solution what prices should be charged. Prices and their effects on loads can be included in resource integration models by adding two types of constraints: a revenue requirements equation and a demand curve [68,69]. The first constraint relates the utility's price to its resource costs. Under the assumption of cost-based regulation of private utilities, wherein utilities are allowed to recover all costs prudently incurred, such a constraint would specify that the revenue received by the utility should cover its costs. The second constraint relates loads to prices. The more intense competition becomes, the more elastic that curve will be. These two constraints are generally non-linear. Revenue equals price times quantity demanded, which are now both decision variables. As a result, the models become more difficult to solve.

An important change in a traditional production cost model is the introduction of elasticity of the demand. In classic production cost models the demand was inelastic and had to be met (subject to a penalty for unserved load). Now, the equilibrium quantity is obtained by maximizing the total surplus, defined as the sum of consumer's and producer's surplus. As a result of the competitive environment that now utilities face, the needs of planners for optimization models have changed, in order to take into account mainly the uncertainty of the market factors evolution and the increasing competition.

The implications of competition in power generation expansion planning models have been widely addressed in the literature. Hobbs [3] stresses four major implications of increasing competition, including greater uncertainty regarding market sizes and prices, better understanding of the cost structure, response to price changes and behavior according to the actions of other suppliers of energy. The new planning approach is presented by Rosekrans et al. [70]. Ramos et al. [2] discuss the characteristics of traditional production cost models that remain relevant in the new regulatory framework. The detailed representation of the electric system operation and the quantitative character of the decision variables (unit output levels) of the traditional production cost models lend themselves to a Cournot analysis, which is one of the standard oligopoly equilibrium concepts. Murphy and Smeers [6], present the steps necessary to transform a capacity expansion model designed for a regulated monopoly company into one applicable to a perfectly competitive market. The first step is to introduce a demand model that accounts for the dependence of demand on prices. Assuming that the associated inverted demand function is the gradient of a utility function ($p(q) = \theta U(q)/\theta q$) the minimal cost capacity expansion problem can be readily extended into a nonlinear model where the objective is to minimize the net present value of producers and consumers surplus over a certain horizon.

4. Generation planning in competition

Outcomes in the electrical system will no longer depend upon traditional total cost minimization scheme, but rather on the interaction of individual, profit maximizing firms [41]. Each firm must look to maximize its production surplus (market revenues minus operation costs) in an uncertain context where its perception of risk, the behavior of the competitors, the ownership structures, the technology mix, as well as a multitude of other external, technical, economical and managerial factors heavily condition the market. The system behavior will therefore be characterised by the economic market equilibrium as a result of the interaction of all these factors. Market equilibrium defines a point of convergence of the market, provided that each participant behaves looking forward to maximizing its own profits [2].

Several areas of knowledge converge in modeling market equilibria, such as microeconomic theory [71] (Cournot and Bertrand models among others [72,73]), game theory (non-cooperative games [74]), mixed complementary problems (MCP) [75,76] and mathematical programming with equilibrium constraints (MPEC) [77]. Several papers have addressed the computation of the market equilibrium in the electric sector [78]. They differ in how each generating firm f anticipates that rivals will react to its decisions concerning either prices p or quantities q. In the following models we assume that all players get the market clearing price ('pay-your-bid' auctions operate differently). A comprehensive review of the types of strategic interactions that have been or could be included in power markets models has been presented by Day et al. [78], including:

- Pure competition (no market power)/Bertrand
- Generalised Bertrand Strategy ('Game in Prices')
- Cournot Strategy ('Game in Quantities')—The Cournot model is detailed in analyses by Newbery [79] and Borenstein and Bushnell [80].
- Collusion
- Stackelberg
- General conjectural variations
- Conjectured supply function (CSF)
- Supply function equilibria (SFE)

The general SFE approach was introduced by Klemperer and Meyer [81] and applied by Green and Newbery [4] to the electricity industry reforms in England and Wales (E and W). They use a simplified supply function equilibrium approach. Rudkevich et al. [82] extend this technique to the use of a stepwise supply function. Borenstein et al. [83] defined a general classification of the different markets and competitive equilibria in the electric industry. Green [84] used a linear version of this model and applied it to prospective divestitures of generation assets mandated by the regulator of the electricity industry in E and W. Bladick, Grant and Kahn [85], offered a generalization of the Green's model and extended the application to subsequent changes in the horizontal structure of the electricity market in E and W.

Borenstein and Bushnell [80] used a simulation model, which heuristically evaluates the California market under competition. The electric energy market is modeled using the Cournot equilibrium framework, where the companies are considered strategic or competitive fringe depending on their institutional characteristics. Each market equilibrium is calculated using an iterative algorithm that sets each strategic firm's production at its optimal level while holding constant the output of the other strategic firms. This process is repeated until no strategic company has the incentive to modify its production level given the production levels of the other strategic firms. Bushnell [86] extends this simulation model to include interperiod elements. He represents the equilibrium conditions analytically and his model achieves market equilibrium taking into account hydro scheduling decisions, which regard planning resources for multiple periods.

Hogan [87] models the profit maximization problem of each strategic firm as a nonlinear optimization problem that takes into account network constraints. The profit maximization of fringe companies is represented by incorporating their first order optimality conditions into the optimization problem of the strategic companies.

Ramos et al. [2] models the competitive behavior of the electric generation energy market by incorporating a set of constraints, naming the equilibrium constraints, into a traditional production cost model. These constraints reproduce the first-order optimality conditions of the strategic companies. Their approach represents the objective of profit maximisation, subject to oligopoly competition, while also keeping a high level of operational detail and without resorting to any kind of iterative procedure. All the system agents are represented in their model: the market operator, the electric generation firms, and the demand bidders.

Strategic investments is a first relevant notion for analyzing investments in restructured electricity systems.

Strategic investments are those that modify rival's actions. They are best interpreted in a two-stage decision context where investment decisions are made first while operations (generation, trading and sales) are decided in the second stage. Second stage uncertainties, when they are present, influence first stage decisions.

In one of the first models of this type, Spence [88] considers the case where an incumbent builds capacity in the first stage while the potential entrant invests in the second stage. Both operate in the market in the second stage. Spence assumes a single technology characterized by variable capacity and operations costs. The potential entrant incurs a fixed cost to enter the market, which the incumbent has already paid. There is no uncertainty in this model. The entrant optimizes its decision assuming that the incumbent will utilize all its capacity in the second stage, given the entrant effectively enters the market. The potential entrant decides to enter the market only if it can make a positive profit after paying for the fixed cost. The incumbent selects its capacity and operating levels in order to maximize its profit subject to the condition that it wants to bar the entrant from the market. Dixit [89] retains some elements of Spence's model and modifies others. As in Spence both the incumbent and the entrant resort to single technology characterized by variable capacity and operating costs and the fixed cost of entering the market. The incumbent is still the sole investor in the first stage but, in contrast with Spence, it can also add capacity in the second stage. The decision paradigm is also different from Spence's as the second stage market is of the Cournot type. The incumbent makes its first stage capacity decision to optimize its profit, taking the market equilibrium achieved in the second stage into account. There is no uncertainty in this model. From a computational point of view this nesting of an optimization and equilibrium problem results in a mathematical program subject to equilibrium constraints.

In contrast to both Spence and Dixit, Gabsewicz and Poddar [90], assume that the two firms may simultaneously enter the market. They do so by choosing their capacities in the first stage, and they cannot revise this choice in the second stage. The only second stage decisions are operational and as in Dixit's model, the second stage market equilibrium is Cournot. Gabsewicz and Poddar also modify the description of the technology by dropping the fixed cost to enter the market. Uncertainty is a key element of the Gabsewicz and Poddar model. They assume that the demand function is revealed in the second stage only and that the achieved equilibrium is contingent on this demand information. This implies that investments must be decided before knowing the intensity of the demand. Firms invest so as to reach to a Nash equilibrium in the first stage knowing the outcome of their decision in the second stage. This nesting of two equilibrium problems (a perfect subgame equilibrium) leads to an equilibrium problem subject to equilibrium constraints.

Von der Fehr and Harbord [5] present a two-stage game, where in the first stage n firms simultaneously enter the industry and choose the amount of capacity to install, then demand is realized and firms simultaneously submit offer prices, or bids at which they are willing to supply power. This model has similarities with those of Kreps and Scheinkman [91] and Davidson and Deneckre [92]. However, the difference between them locates in the way that market prices are determined.

These models provide a realistic framework for looking at investments in the restructured power sector. Murphy and Smeers [6] moved a few steps from economic concepts towards computable models of capacity expansion in restructured electricity systems by presenting three models. The first supposes a perfect competitive equilibrium. The second model (open loop Cournot game) extends the Cournot model, sometimes used for modeling operations in restructured electricity systems to include investments in new generation capacity. The third model (closed loop Cournot game) separates the investments and sales decision. It describes a situation where investments are decided in a first stage and sales occur in a second stage, both taking place in an oligopolistic market.

In reality markets may contain only a few players who will be well aware that their actions affect each other's decisions. Models of resource decision making and pricing that account for possible strategic behavior might yield more realistic assessment of prices within power markets. Both non-cooperative and cooperative game-theoritic models can play this role.

Non-cooperative Nash models have been used to analyse the effect of transmission costs upon competition among deregulated utilities in New York [93], to explore possible reactions of cogenerators to utility policies regarding purchased power [94] and to model bidding strategies in the UK electricity market [74].

Cooperative game models, in contrast, can provide estimates of the stakes each party has in coordinated planning and operations. The analysis by Gately [95] of the benefits of coordination among India's regional power boards is the classic study of this type. He used a mathematical programming-based generation expansion model to calculate the costs for each board if they operated autonomously, and then their costs if they cooperated. Those results allowed him to calculate the core of the cooperative game, defined as possible allocations of the benefits that would leave each party off than it could be by itself or in subcoalitions with any of the other parties.

5. Concluding remarks

The case of the California power crisis, during the summer months of 2000 and the winter months of 2000–2001, showed that the most important factor was the shortage of power supply relative to demand [96]. Of course, the design flaws in the Cal PX combined with a number of exogenous factors has undoubtedly lead the crisis. However, the specific event has emerged the procedure of long-term power generation expansion planning as one of the major concerns of power generators, considering the fact that potential regulatory flaws could not ensure that optimum short-term behaviour could lead the market actors to optimum long-term efficiency.

As Rosekrans et al. [70] stress, under competitive electricity markets, simulation and optimization models will be used in a different context. Rather than specifying the decisions for central planners to build an 'optimum' system, these models can be used to simulate the decisions of individual plant owners and developers, such as which plants are unprofitable and will retire and which resource options are likely to be profitable and thus enter the market. Such a market simulation should be of interest to private decision-makers in investigating the profitability of various courses of action, both by themselves and by their competitors. While the regulated monopoly model can easily be extended to deal with a perfectly competitive market, further extension to an imperfect competition context is much more complex, where power market is oligopolistic or in other words the number of suppliers is sufficiently small that each can influence prices [6].

Market power is an actively researched area in the literature on restructured electricity systems. Several models exist that look at the operations of a market with oligopolistic players when capacities are given. In contrast with this wealth of literature on market operations, very little is available when it comes to investment. To the best of our knowledge, market models dealing with both investments and operations in an oligopolistic electricity market are rare at this time. These models are stylized analytic models that do not lead to computable tools. What emerges from this study is in which level sophisticated techniques could really support in practice power generators who need to have in their availability a useful tool for GEP, even if the optimum plan is not developed.

Acknowledgements

The authors would like to acknowledge Prof. B.F. Hobbs, Prof. Y. Smeers and Dr P. Murto for the valuable provision of information and guidelines.

References

[1] Banks FE. Economics of electricity deregulation and privatization: an introductory survey. Energy 1996;21(4):249–61.

[2] Ramos A, Ventosa M, Rivier M. Modeling competition in electric energy markets by equilibrium constraints. Utilities Policy 1998;7(4): 233–42.

[3] Hobbs BF. Optimisation methods for electric utility resource planning. Eur J Oper Res 1995;83(1):1–20.

[4] Green RJ, Newbery DM. Competition in the British electricity spot market. J Political Economy 1992;100(5):929–53.

[5] von der Fehr NH, Harbord D. Capacity investment and competition in decentralized electricity markets. Memorandum from Department of Economics, University of Oslo, No. 1997. p. 27.

[6] Murphy FH, Smeers Y. Capacity expansion in imperfectly competitive restructured electricity markets. Harvard Electricity Policy Group, Center for Business and Government; May 2001.

[7] Grobman JH, Carey JM. Price caps and investment: long-run effects in the electric generation industry. Energy Policy 2001;29(7):545–52.

[8] Massé P, Gibrat R. Application of linear programming to investments in the electric power industry. Mgmt Sci 1957;3(1):149–66.

[9] Dyner I, Larsen ER. From planning to strategy in the electricity industry. Energy Policy 2001;29(13):1145–54.

[10] Karaki SH, Chaaban FB, Al-Nakhl N, Tarhini KA. Power generation expansion planning with environmental consideration for Lebanon. Int J Electr Power Energy Syst 2001; in press, uncorrected proof, available online 6 November 2001.

[11] Chung-Hsiao W, Jo MK. Generation planning with quantified outage costs. Electric Power Syst Res 2000;54(1):37–46.

[12] Anderson D. Models for determining least-cost investments in electricity supply. Bell J Econ Manag Sci 1972;3:267–85.

[13] Pokharel S, Ponnambalam K. Investment planning for electricity generation expansion. Int J Energy Res 1997;21(2):185–94.

[14] Park YM, Park JB, Won JR. A hybrid genetic algorithm/dynamic programming approach to optimal long-term generation expansion planning. Int J Electr Power Energy Syst 1998;20(4):295–303.

[15] Jenkins ST, Joy DS. Wien automatic system planning package (WASP)—an electric utility optimal generation expansion planning computer code. Oak Ridge, Tennesee: Oak Ridge National Laboratory; 1974. ORNL 4945.

[16] Electric Power Research Institute (EPRI), Electric generation expansion analysis system (EGEAS). Palo Alto, CA: EPRI EL-2561, EPRI; 1982.

[17] Bloom JA. Long-range generation planning using decomposition and probabilistic simulation. IEEE Trans Power Apparatus Syst 1982; 101(4):797–802.

[18] David AK, Zhao R. Integrating expert systems with dynamic programming in generation expansion planning. IEEE Trans Power Syst 1989;4(3):1095–101.

[19] David AK, Zhao R. An expert system with fuzzy sets for optimal planning. IEEE Trans Power Syst 1991;6(1):59–65.

[20] Lin M, Breipohl A, Lee F. Comparison of probabilistic production cost simulation methods. IEEE Trans Power Syst 1989;4(4):1326–34.

[21] Dapkus WD, Bowe TR. Planning for new electric generation technologies-a stochastic dynamic programming approach. IEEE Trans Power Apparatus Syst 1984;103(6):1447–53.

[22] Yang HT, Chen SL. Incorporating a multi-criteria decision procedure into the combined dynamic programming/production simulation algorithm for generation expansion planning. IEEE Trans Power Syst 1989;4(1):165–75.

[23] Caramanis M, Schweppe F, Tabors R. Electric generation expansion systems. 1(August): 6-30-42 EPRI EL-2561; 1981.

[24] Manhire B, Jenkins RT. A new technique for simulating the operation of multiple assigned-energy generating units suitable for use in generation system expansion planning models. IEEE Trans Power Apparatus Syst 1982;101(10):3861–9.

[25] Bloom JA, Gallant L. Modeling dispatch constraints in production cost simulation based on the equivalent load method. IEEE Trans Power Syst 1994;9(2):598–611.

[26] Park YM, Lee KY, Youn LTO. New analytical approach for long-term generation expansion planning based maximum principle and Gausian distribution function. IEEE Trans Power Apparatus Syst 1985;104(2): 390–7.

[27] Levin N, Zahavi J. Optimal mix algorithm with existing units. IEEE Trans Power Apparatus Syst 1984;103(5):954–62.

[28] Ramos A, Peres-Arriaga IJ, Bogas J. A nonlinear programming approach to optimal static generation expansion planning. IEEE Trans Power Syst 1989;4(3):1140–6.
[29] Ammons JC, McGinnis LF. A generation expansion planning model for electric utilities. Engng Econ 1985;30:205–25.
[30] Belgari F, Laughton MA. The combined costs method for optimal economic planning of an electrical power system. IEEE Trans Power Apparatus Syst 1975;46–96:1935–42.
[31] Saway RM, Zinn CD. Mathematical model for long range expansion planning of generation and transmission in electric utility systems. IEEE Trans Power Apparatus Syst 1977;PAS-96:657–65.
[32] Wang CH, Min KJ. An integrated resource planning model for utilities with quantified outage costs. Int J Electr Power Energy Syst 1998; 20(8):517–24.
[33] Gorenstin BG, Campondonico NM, Costa JP, Pereira MVF. Power system expansion planning under uncertainty. IEEE Trans Power Syst 1993;8(1):29–36.
[34] Merrill HM, Wood AJ. Risk and uncertainty in power system planning. Int J Electr Power Energy Syst 1991;13(2):81–90.
[35] Gardner DT. Flexibility in electric power planning: coping with demand uncertainty. Energy 1996;21(12):1207–18.
[36] Felder FA. Integrating financial theory and methods in electricity resource planning. Energy Policy 1996;24(2):149–54.
[37] Walters DC, Sheble GB. Genetic algorithm solution of economic dispatch with valve point loading. IEEE Trans Power Syst 1993;8(3): 1325–32.
[38] Chen PH, Chang HC. Large-scale dispatch by genetic algorithm. IEEE Trans Power Syst 1995;10(4):1919–26.
[39] Dasgupta D, McGregor DR. Thermal unit commitment using genetic algorithms. IEE Proc Part C, Generation, Transm Distribution 1994; 141(5):459–65.
[40] Shelbe GB, Maifeld TT, Brittig K, Fahd G. Unit commitment by genetic algorithm with penalty methods and a comparison of Lagrangian search and genetic algorithm—economic dispatch algorithm. Int J Electr Power Energy Syst 1996;18(6):339–46.
[41] Orero SO, Irving MR. Large scale unit commitment using a hybrid genetic algorithm. Int J Electric Power Energy Syst 1997;19(1): 45–55.
[42] Iba K. Reactive power optimization by genetic algorithm. IEEE Trans Power Syst 1994;9(2):685–92.
[43] Sundhararajan S, Pahwa A. Optimal selection of capacitors for radial distribution systems using a genetic algorithm. IEEE Trans Power Syst 1994;9(3):1499–507.
[44] Wong KP, Wong YW. Thermal generator scheduling using hybrid genetic/simulated-annealing approach. IEE Proc Part C, Generation, Transm Distribution 1995;142(4):372–80.
[45] Potts JC, Giddens TD, Yadav SB. The development and evaluation of an improved genetic algorithm based on migration and artificial selection. IEEE Trans Syst Man Cybernetics 1994;24(1):73–86.
[46] Fukuyama Y, Chiang H. A parallel genetic algorithm for generation expansion planning. IEEE Trans Power Syst 1996;11(2):955–61.
[47] Park YM, Park JB, Won JR. A genetic algorithms approach for generation expansion planning optimization. A Proceedings Volume from the IFAC Symposium: Control of Power Plants and Power Systems, Pergamon, Netherlands; 1996. p. 257–62.
[48] Mills D, Vlacic L, Lowe I. Improving electricity planning—use of a multicriteria decision making model. Int Trans Oper Res 1996;3(3/4): 293–304.
[49] Voropai NI, Ivanova EYu. Multi-criteria decision analysis techniques in electric power system expansion planning. Int J Electrl Power Energy Syst 2002;24(1):71–8.
[50] Beccali M, Cellura M, Ardente D. Decision making in energy planning: the ELECTRE multicriteria analysis approach compared to a fuzzy-sets methodology. Energy Conversion Mgmt 1998;39(16–18):1869–81.
[51] Martins AG, Coelho D, Antunes CH, Climaco J. A multiple objective linear programming approach to power generation planning with demand-side management (DSM). Int Trans Oper Res 1996;3(3/4): 305–17.
[52] Mavrotas G, Diakoulaki D, Papayannakis L. An energy planning approach based on mixed 0–1 multiple objective linear programming. Int Trans Oper Res 1999;6(2):231–44.
[53] Zhang ZZ, Hope GS, Malik OP. Expert systems in electric power systems—a bibliographical survey. IEEE Trans Power Syst 1989; 4(4):1355–61.
[54] Madan S, Bollinger KE. Applications of artificial intelligence in power systems. Electric Power Syst Res 1997;41(2):117–31.
[55] Balu NJ, Adapa R, Cauley G, Lauby M, Maratuku-Jam DJ. Review of expert systems in bulk power system planning and operation. Proc IEEE ISSN: 0018-9219 1992;80(5):727–31.
[56] Adapa R. Expert system applications in power system planning and operations. IEEE Power Engng Rev 1994;14(2):15–18.
[57] Shin JR, Park YM. Optimal long-term transmission planning by expert system approach. Proceedings IEEE Region 10 Conference on Computer, Communication, Control and Power Engineering ISBN: 0-7803-1233-3, Part 2 (of 5) Beijing, China; 1993. p. 713–7.
[58] Chandrashekara AS, Ananthapadmanabha T, Kulkarni AD. A neuro-expert system from planning and load forecasting of distribution systems. Int J Electr Power Energy Syst 1999;21(5):309–14.
[59] Kono Y, Doi A, Omichi Y. Development of a knowledge-based system for voltage/reactive power operation planning. Electr Engng Jpn 1991;111(6):25–35.
[60] Hsu YY, Ho KL. Fuzzy expert systems. An application to short-term load forecasting. IEE Proc Generation Transm Distribution ISSN: 0143-7046 1992;139(6):471–7.
[61] Baken JI, Lucas N. Integrated resource planning and environmental pricing in a competitive and deregulated electricity market. Energy Policy 1996;24(3):239–44.
[62] Swisher JN, Januzzi GM, Redlinger RY. Tools and methods for integrated resource planning—improving energy efficiency and protecting the environment. UNEP Collaborating Centre on Energy and Environment, Riso National Laboratory; 1997.
[63] Malik AS, Cory BJ. Integrated resource planning with consideration of dynamic costs of thermal units. Electric Power Syst Res 1999; 51(2):123–30.
[64] Vine EL. International DSM and DSM program evaluation: an indeep assessment. Energy 1996;21(10):983–96.
[65] Reddy BS, Parikh JK. Economic and environmental impacts of demand side management programmes. Energy Policy 1997;25(3):349–56.
[66] Papastamatiou P, Psarras J. Market modelling for assessment of demand side programs using the marginal cost. Int J Energy Res 2000; 24(10):887–900.
[67] Malik AS. Modelling and economic analysis of DSM programs in generation planning. Int J Electr Power Energy Syst 2001;23(5): 413–9.
[68] Rutz WL, Becker M, Wicks FE. Treatment of elastic demand in generation planning. IEEE Trans Power Apparatus Syst 1985;PAS-104/11:3092–7.
[69] Hobbs BF, Rouse RB, Hoog DT. Measuring the economic value of demand side resources in integrated resource planning models. IEEE Trans Power Syst 1993;8/3:979–87.
[70] Rosekrans S, Kirshner D, Marnay C. Issues in electricity planning with computer models: illustrations with Elfin and WASP. Utilities Policy 1998;7(4):210–9.
[71] Varian HR. Microeconomic analysis. New York: W.W. Norton and Company; 1992.
[72] Sherali HD, Soyster AL, Murphy FH. Stackelberg-Nash-Cournot equilibria: characterizations and computations. Oper Res 1983;31(2): 253–75.
[73] Kolstad CD, Mathiesen L. Computing Cournot-Nash equilibria. Oper Res 1991;39(5):739–48.
[74] Lucas N, Taylor O. Characterising generator behaviour: bidding strategies in the pool, a game theory analysis. Utilities Policy 1993; 129–35.

[75] Cottle WW, Pang JS, Stone RE. The linear complementarity problem. Boston: Academic Press; 1992.
[76] Ferris MC, Pang JS. Engineering and economic applications of complementary problems. Discussion Papers in Economics 1995–1997, Department of Economics, University of Colorado; 1995.
[77] Luo Z, Pang J, Ralph D. Mathematical programs with equilibrium constraints. Cambridge University Press; 1996.
[78] Day CJ, Hobbs BF, Pang JS. Oligopolistic competition in power networks: a conjectured supply function approach. University of California Energy Institute, POWER, PWP-090; 2002.
[79] Newbery DM. Power markets and market power. Energy J 1995; 16(3):39–66.
[80] Borenstein S, Bushnell J. An empirical analysis of the potential for market power in California's electricity industry. Power Conference on Electricity Restructuring. University of California Energy Institute; 1997.
[81] Klemperer PD, Meyer MA. Supply function equilibria. Econometrica 1989;57:1243–77.
[82] Rudkevich A, Duckworth M, Rosen R. Modeling electricity pricing in a deregulated generation industry: the potential for oligopoly pricing in a Poolco. Energy J 1998;19(3):19–48.
[83] Borenstein S, Bushnell J, Kahn E, Stoft S. Market power in California electricity markets. Utilities Policy 1995;5(3/4):219–36.
[84] Green R. Increasing competition in the british electricity spot market. J Ind Econ 1996;44(2):205–16.
[85] Baldick R, Grant R, Kahn E. Linear supply function equilibrium: generalizations, application, and limitations. University of California Energy Institute, POWER, PWP-078; 2000.
[86] Bushnell J. Water and power: hydroelectric resources in the era of competition in the western US. Power Conference on Electricity Restructuring. University of California Energy Institute; 1998.
[87] Hogan WW. A market power model with strategic interaction in electricity networks. Power Conference on Electricity Restructuring, University of California Energy Institute; 1997.
[88] Spence AM. Entry, capacity, investment and oligopolistic pricing. Bell J 1977;8:534–44.
[89] Dixit A. The role of investment in entry deterrence. Economic J 1980; 90:95–106.
[90] Gabsewicz JJ, Poddar S. Demand fluctuations and capacity utilization under duopoly. Economic Theory 1997;10(1):131–47.
[91] Kreps D, Scheinkman J. Cournot precommitment and Bertrand competition yield Cournot outcomes. Bell J Economics 1983;14: 326–37.
[92] Davidson C, Deneckere R. Long-run competition in capacity, short-run competition in price, and Cournot model. Rand J Economics 1986; 17:404–15.
[93] Hobbs BF. Network models of Bertrand and limit pricing equilibria in spatial markets. Oper Res 1986;34(3):410–25.
[94] Haurie A, Loulou R, Savard S. A two player game model of power generation in New England. IEEE Trans Automatic Control 1992; 37(9):1451–92.
[95] Gately D. Sharing the gains from regional cooperation: a game theoretic application to planning investments in electric power. Int Economic Rev 1974;15:195–208.
[96] Sioshansi FP. California's dysfunctional electricity market: policy lessons on market restructuring. Energy Policy 2001;29(9):735–42.

Renewable and Sustainable Energy Reviews 14 (2010) 1353–1366

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Multi-objective planning of distributed energy resources: A review of the state-of-the-art

Arturo Alarcon-Rodriguez *, Graham Ault, Stuart Galloway

Institute of Energy and Environment, University of Strathclyde, Royal College Building, 204 George Street, G1 1XW, Glasgow, UK

ARTICLE INFO

Article history:
Received 8 December 2009
Accepted 12 January 2010

Keywords:
Distributed energy resources
Distributed generation
Power distribution planning
Optimization methods
Multi-objective planning

ABSTRACT

The use of Distributed Energy Resources (DER) has been proposed as one of the possible solutions to today's energy and environmental challenges. The optimal integration of DER in distribution networks is essential to guarantee the best of resource, i.e. maximize their benefits, such as reduction of carbon emissions, reduction of network energy losses and to minimize the negative impacts, which can affect the network quality, cause network sterilization and increase investment and operation costs. Hence, DER planning is a multi-objective problem in which many objectives of interest, sometimes conflicting, need to be optimized simultaneously, and where a compromise for different perspectives (DER developer, Distribution System Operator, regulator) needs to be found. Appropriate multi-objective planning methods that consider technical, environmental and economic impacts of DER integration, and that are able to support a suitable model of stochastic DER and active networks, can provide a deep insight into the case specific and general advantages and drawbacks of DER. Consequently, the interest in multi-objective DER planning has increased in recent years, and a number of novel methods have been proposed in this area. This paper provides a timely review of the state-of-the-art in multi-objective DER planning, and discusses in detail the challenges, trends and latest developments in this field.

Contents

* Corresponding author.
E-mail address: aalarcon@eee.strath.ac.uk (A. Alarcon-Rodriguez).

1364-0321/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.rser.2010.01.006

1. Introduction

1.1. Background

In recent years, climate change has prompted international awareness about the impacts that electricity generation and the use of energy have on the environment. In this context, local generation of heat and electricity and the local use of renewable energy resources are considered as some of the most promising options to provide a more secure, clean and more efficient energy supply [1].

Distributed Generation (DG) is defined as "an electric power source connected directly to the distribution network or on the customer site of the meter" [2]. The most common DG technologies include Combined Heat and Power (CHP) generators, micro-turbine gas generators, solar photovoltaic generators (PV), wind generators and micro-hydro schemes [3]. At present, DG is considered within the broader concept of Distributed Energy Resources (DER), which also includes responsive loads and energy storage [4].

1.2. The need for optimal DER integration

Several benefits can be obtained when DER is correctly integrated. For example, DER located close to load centers and with a production coincident with demand reduces power flow in lines. This reduction in power flows results in an improvement of voltage profile, and in a decrease in the line losses [3]. Moreover, if DER produces power at peak times, network investments can be deferred [5]. Similarly, the reliability of the network can be increased by DER with constant production and connected to meshed grids, or by DER that are allowed to operate in islanded mode while connected to radial networks. In contrast, DER with a variable output, such as wind generators, or DER connected to radial networks and not allowed to work in islanded mode do not increase the reliability of the network [3].

Many of these technical effects translate to economic benefits for the Distribution System Operator (DSO) (e.g. reduction of line losses, investment deferral, increased reliability), or for the customer (e.g. increased reliability). The economic benefits for the DER owner arise from energy sales. Hence, for a DER developer maximizing the amount of energy traded, while keeping the system within technical operation limits, is paramount. From a societal perspective the use of renewable energy resources offsets fossil-fuel-based energy and provides a cleaner energy supply.

Distribution networks were designed deterministically for unidirectional power flows, from higher voltages to lower voltages, rather than to accommodate large penetrations of DER. As a result, wrongly located DER, DER whose production is not coincident with demand or DER whose capacity largely exceeds the capacity of the network, has negative effects, such as reverse power flows, increments in line losses and voltage rise [3]. DER located close to fault points contribute to the fault currents and might require the replacement of switchgear equipment [3]. Other impacts of DER include the degradation of voltage quality, by injecting power-electronic harmonics, and an increase in network instability, because of the low inertia of DER [6].

The distribution network must be kept within operational and design limits at all times to provide good-quality energy and avoid damage to the equipment. Hence, the technical impacts of DER can limit the installation of DER and restrict the associated economic and environmental benefits. In weak rural networks, where large amounts of renewable resources are expected to be located, voltage rise [7] and thermal capacity are the impacts limiting the integration of DER. In meshed urban networks, where large numbers of micro-CHP units could potentially be installed, thermal limits and fault levels are the most common constraining factors [3].

There are two management philosophies to keep the network within operational limits and to minimize the steady state impacts of DER: "fit-and-forget" and Active Network Management (ANM). Under a traditional "fit-and-forget" connection philosophy, the grid is reinforced to keep the system within deterministic operational limits. That is to say, the operational problems are solved at the planning stage. Strbac et al. [8] identifies that the "fit-and-forget" approach would require extremely costly reinforcements in the network to accommodate large penetrations of DER. Hence, this management philosophy is limiting for the integration of DER [4]. Thus, "a fundamental shift from passive to active network management" was proposed in recent years [9]. Under this management philosophy the operational problems are solved with the active management of the network and the DER. ANM has been shown to considerably increase the amount of DER that can be connected to the network without the need for reinforcement [4].

Under either management approach, the optimal integration of DER in the distribution grid is fundamental to guarantee the best use of resources, i.e. maximize benefits and minimize costs. The sub-optimal integration of DER under a "fit-and-forget" management will result in a requirement for additional and unnecessary transmission and distribution grid reinforcements, network 'sterilization', increased line losses and/or unattainable development and environmental targets [5,10,11]. Likewise, the sub-optimal integration of DER under active DER management will result in excessive energy curtailment, which could convert an economically feasible project into an unfeasible one [12], and restrict the further exploitation of renewable energy resources.

This paper reviews techniques for the optimal integration of DER. In particular, the area of multi-objective DER planning is examined in detail. In Section 2 the problem formulation is presented, and the dilemma between optimization and modeling discussed, some single-objective techniques are introduced. Section 3 examines the key concepts of multi-objective optimization, and introduces the two main types of techniques used in this area. Section 4 presents the critical review of multi-objective DER planning. A summarizing discussion and conclusions are provided in Section 5.

2. Planning of distributed energy resources

2.1. Problem formulation

DER planning is the structured process of optimizing DER type, size and/or location in order to achieve a set of objectives and subject to a set of constraints. A general representation of this problem can be expressed formally as:

$$\begin{aligned} \min \mathbf{F}(\mathbf{x}) = \min([f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})]) \\ \text{s.t.} \\ \mathbf{x} \in \Omega \\ \mathbf{G}(\mathbf{x}) = 0 \\ \mathbf{H}(\mathbf{x}) \leq 0 \end{aligned}$$

where f_i is the i^{th} objective function; m, the number of objectives; $\mathbf{x}$, the decision vector of DER location, sizes and types; Ω, the decision domain that defines the possible locations, sizes and types of DER (search space); $\mathbf{G}(\mathbf{x})$, the equality constraints, usually defined by the power flow equations of the network; $\mathbf{H}(\mathbf{x})$, the inequality constraints, usually technical limits of the equipment (e.g. voltage constraints, thermal constraints, short circuit limits, etc.), operating limits of DER (e.g. maximum capacity) or performance targets (e.g. reliability, emissions, maximum allowed curtailment).

This problem has nonlinear equality constraints defined by the power flow equations; hence, it is a non-convex optimization problem. It also has some nonlinear optimization objectives, such as line loss minimization. The planning variables are the discrete locations, sizes and types of DER and the topology of the network. As a result, DER planning is a non-convex combinatorial problem, with several local optima, and one global optimal solution. Non-convex, nonlinear, combinatorial problems are usually difficult to solve using traditional mathematical methods since these methods are designed to find local optima solutions [13].

The complexity of this optimization task is dealt with using two approaches. The first is to apply simplifying assumptions to the formulation of the problem. For example, linearization of the objective functions and constraints, relaxation of the constraints, reduction of the dimensions of the search space, assumption of the discrete nature of DER units as continuous, simplification of the time-variability of load and DER into snapshot analyses [12]. In this way, it is possible to solve the optimization problem using traditional mathematical programming methods, for which powerful programming methods are available (e.g. Linear Programming).

The second approach is based on the use of heuristic optimization techniques, such as Evolutionary Algorithms (EA). These heuristic techniques are well suited to deal with non-convex combinatorial problems [14] and can handle discontinuous search spaces. Moreover, they allow optimization of intricate non-differentiable objective functions. Hence, they enable more detailed modeling of the time-variability of DER. Though, the drawback of these techniques is that they only find an approximation of the global optimal solution in a limited time. This optimization/modeling dilemma is discussed further next. A thorough account of the development and application of heuristics techniques within power systems problems is given in [15].

2.2. DER planning: an optimization/modeling dilemma

When a real optimization problem, such as DER planning, is over-simplified the optimal solutions found are in fact sub-optimal, or as phrased by Irving and Song [16]: "a real solution to a non-problem". For instance, a simplistic model of renewable generators (e.g. a snapshot power-flow analysis, or the analysis of a single-day profile for wind generation) optimized with a very accurate optimization method (e.g. linear programming) will produce solutions that although accurate, will almost certainly be sub-optimal for the real problem, given the limited scope of the model. Similarly, a realistic model of DER is worthless when optimized with an inaccurate optimization method, i.e. "a non-solution to a real problem" [16]. For example, an extremely complex model of the power system (e.g. a minute-by-minute simulation of the power system) optimized by an ad hoc method (e.g. analysis of a few configurations chosen by hand) will clearly provide a solution that is not optimal, even with the realistic model used. These examples illustrate the optimization/modeling dilemma faced in the solution of real optimization problems (Fig. 1). Hence, if useful planning tools are to be produced for DER planning, the formulation of the problem should be as close to the upper-right corner of the figure as possible. Often, this requires a good compromise between the accuracy of the optimization method, and the detail of the model of DER and the network.

There are some key aspects that need to be considered when modeling the DER planning problem. First, some benefits and impacts of DER depend not only on the location and size of the generation, but also on the complex relationship of generation and demand over time. The interaction of diverse time-variant energy sources and demand is stochastic in nature. Adequate evaluation

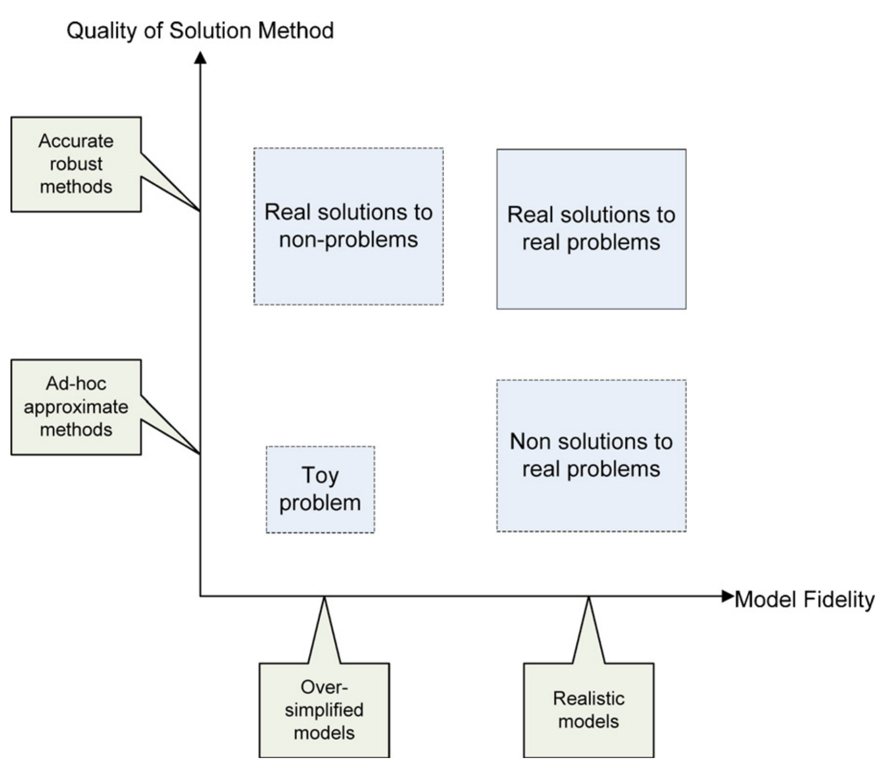


Fig. 1. Optimization/modeling dilemma. Adapted from [16].

methods, such as probabilistic load flow [17], stochastic simulation or Monte Carlo Simulation [18], can be used to evaluate the interaction of stochastic DER and demand.

Also, network access for DER has been traditionally allocated on a firm access basis. In this case a "worst-case scenario" analysis of maximum generation and minimum load is sufficient to evaluate some of the impacts of DER. However, some renewable generators, such as wind turbines, provide their maximum output for only very short periods of time [19], and as a result the use of a probabilistic analysis of DER integration provides a more objective evaluation of DER impacts and benefits [4].

Moreover, recent studies have shown that a non-firm integration of DER permits larger renewable energy production [20]. When non-firm access is considered, active management of the DER and the network is essential to minimize the network impacts and avoid expensive network reinforcements [4,21]. The modeling of the DER planning problem becomes more complex when the evaluation of controllable technologies is proposed.

A simplified deterministic approach limits the analysis of time-variant energy resources (e.g. wind and solar energy), or the consideration of controllable technologies (storage, ANM). Consequently, the optimal integration of DER must determine not only the optimal number, size and location of DER units, but also evaluate stochastic impacts of DER, and the possibility of actively controlling DER and the network.

2.3. Single-objective DER planning

In recent years diverse methods for optimizing the location, size and/or type of DER have been proposed, with particular emphasis on DG placement and sizing. Most DER planning methods focus on the optimization of a single objective. Some examples of single-objective methods are discussed next. One of the most common objectives found in literature is the minimization of line losses. Line loss minimization methods are based on analytical optimization techniques (e.g. [22]), mathematical programming techniques (e.g. [23]) and genetic algorithms (e.g. [24]). For simplification, few of these methods consider the stochastic nature of DER. Moreover, none of these techniques considers the impact of active networks in the analysis.

Other single-objective DER planning approaches focus on the minimization of total cost. Cost can be aggregated from different points-of-view. Hence, these techniques formulate the problem either from the perspective of a DER developer [25], from the perspective of a DSO that can invest in DER [26,27] or from the perspective of a DSO that cannot invest in DER and wants to minimize the cost of network reinforcements [28]. These methods are based on the use of traditional mathematical optimization techniques and genetic algorithms.

More recently, methods to quantify the network capacity, i.e. how much DER can be optimally connected without the need of reinforcements, have been proposed. These methods respond to the need to increase renewable DER installations at the minimum cost. For instance, Harrison and Wallace [11] present an Optimal Power Flow (OPF) approach to obtain the maximum DER capacity in predefined locations. This method was later upgraded to optimize both DER locations and size simultaneously, using a hybrid "GA-OPF" approach, where the Genetic Algorithm (GA) is used to solve the combinatorial problem, and the OPF solves the capacity allocation problem [29]. Keane and O'Malley explore a similar problem, and propose a linear programming technique to find the maximum capacity the can be installed using a firm connection [30], or to maximize the energy that can be harvested, minimizing network violations, in a non-firm integration [20]. Ochoa et al. [31] take the problem further, and propose a multi-period OPF to maximize the DER capacity that can be installed considering ANM. This method is based on non-linear programming. It allows the analysis of stochastic DER, and could be updated to maximize energy harvesting, instead of installed capacity.

A single-objective approach is often practical from a DER developer or a DSO point-of-view. DER developers can obtain information about the most promising locations for DER investments to maximize installed capacity, energy sales and revenue. Also, even if DSOs are not allowed to own and operate DG, as in most European counties, they can identify which locations, sizes and types are beneficial (or detrimental) for their system operation, and provide incentives for optimal network development [32].

2.4. Multi-objective DER planning

Diverse stakeholders participate in DER development and management. Hence, planning objectives can be formulated from different perspectives, e.g. the DER developer, the DSO, or civil society, ideally represented by the regulator [33]. Some of the DER planning objectives are naturally conflicting; consequently in some cases there is no single planning solution that will satisfy all stakeholders. For example, DER capacity maximization will produce an increase in line losses [34]; likewise, cost minimization of network investments conflicts both with capacity maximization and line loss minimization. Similarly, society's interest in low-carbon energy sources might conflict with the need for an affordable and reliable energy supply. A multi-objective approach helps to identify compromise solutions that benefit all stakeholders [34]. Moreover, multi-objective DER planning methods can provide valuable information about the correlations between the benefits and impacts of DER integration, and can inform the decision-making process [35]. From a high-level perspective, a multi-objective analysis of DER integration can help to inform incentives and policies to encourage DER developments in the places, sizes and types that ensure benefits and minimize the impacts of DER. Next, the key concepts of multi-objective optimization are discussed, and the main types of multi-objective optimization methods introduced.

3. Multi-objective optimization

3.1. Key concepts

When an optimization problem has a single objective the definition of "best solution" is one-dimensional and there is only a single best solution (or none, eventually). In contrast, a multi-objective problem with conflicting objectives has no single solution, but a set of optimal solutions. In this case, the multi-dimensional concept of "dominance" is used to determine if one solution is better than other solutions. A solution a is said to dominate a solution b if the following two conditions are true [36]:

- a is no worse than b in all objectives and
- a is better than b in at least one objective.

In this case b is said to be "dominated" by a, or alternatively, a is said to be "non-dominated" by b. A dominated solution is also said to be "sub-optimal". The solution to the multi-objective problem is the set of non-dominated solutions, known as the Pareto set. In terms of their objectives the Pareto set is referred as to the Pareto front, and these terms are sometimes used interchangeably. A solution belongs to the Pareto set if no improvement is possible in one objective without losing in any other objective. These concepts are illustrated in Fig. 2.

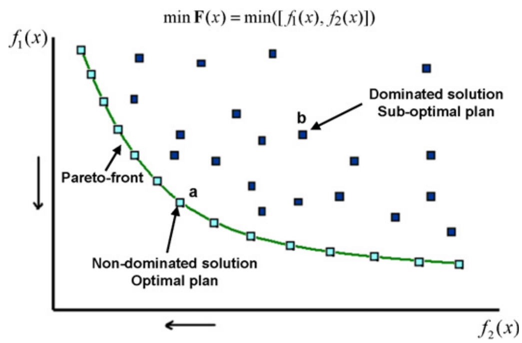


Fig. 2. A Pareto-front for a two-objective problem.

Finding a single solution for a multi-objective problem involves two stages: optimization and decision-making. Depending on the order in which these tasks are performed, there are two possible approaches to obtain a single solution for a multi-objective problem, as illustrated in Fig. 3. The first approach uses *a priori* preference information and single-objective optimization techniques. All objectives are aggregated into a single-objective function that is optimized (e.g. weighted-sum method), or alternatively, one 'master' objective is optimized and the rest of the objectives are considered as constraints. In these two cases the decision-making process precedes the optimization process (e.g. left-hand side of Fig. 3). These procedures can be very useful to find single solutions when detailed preference information is known *a priori* [36]. Deep knowledge of the problem is required to define an adequate aggregation method and weights, or master objectives and constraint levels, respectively.

When *a priori* information is not easily available, the process of obtaining as many solutions as possible in the Pareto set, i.e. *the multi-objective optimization process*, is critical. The information contained in the Pareto front elucidates compromise solutions between different stakeholders or trade-offs between incommensurable objectives. This knowledge facilitates a more informed decision-making process and provides deep insight into the problem. As a result, in the second approach the decision-making process takes place after the multi-objective optimization (e.g. right-hand side of Fig. 3). Initially, several solutions of the Pareto front are sought at once, and preference information is expressed afterwards (*a posteriori*). Some authors believe the second approach to be an "ideal" (or "true") multi-objective optimization approach for the following reasons:

- The method is more methodical, more practical and less subjective, compared with *a priori* approaches [36].
- It provides a wider range of alternatives to choose from, information that would have been lost is conserved; therefore, it permits more informed decisions [37].
- Since real problems are usually multi-objective, this approach permits a more realistic representation of practical problems [37].
- Through transparency the approach permits the generation of useful information about the problem being studied [38]; it is

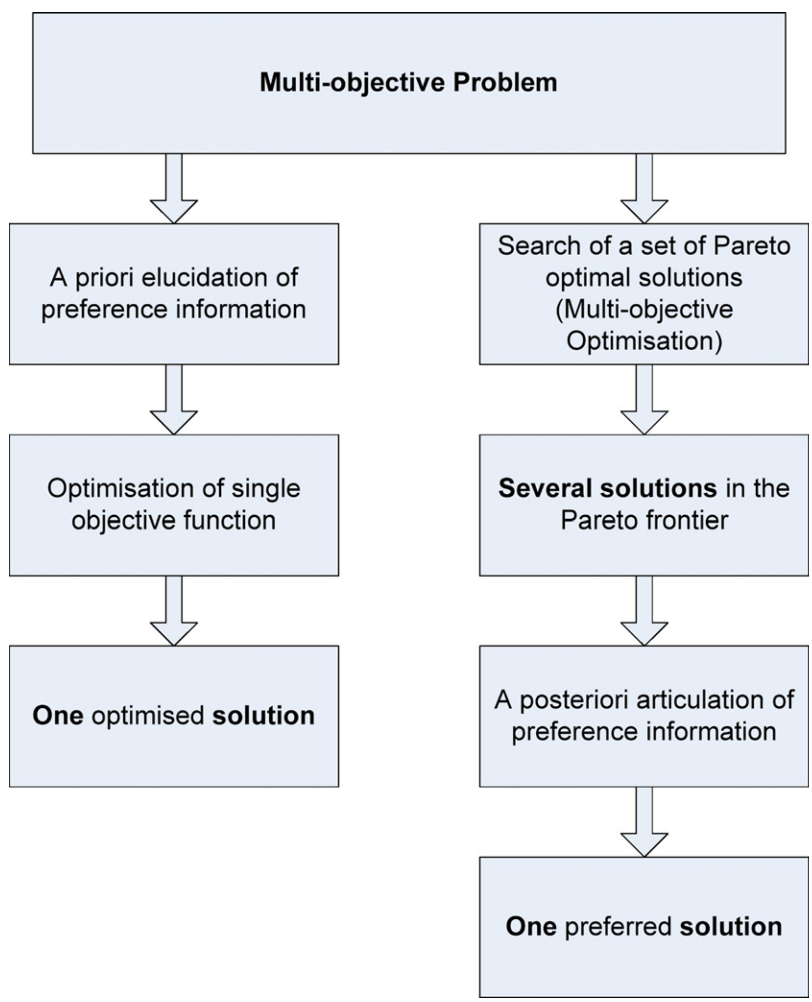


Fig. 3. Finding a single solution for a multi-objective problem.

possible to know the scope of every objective and to analyze the correlations between objectives.

Next, multi-objective optimization methods used to find the Pareto set are introduced. This introduction provides a theoretical background for the literature review of Section 4.

3.2. Multi-objective optimization methods

Normally, multi-objective problems have a large number of solutions defined by the Pareto front. Since finding all Pareto solutions is practically impossible, a subset of the Pareto set is usually looked for. Hence, multi-objective optimization is the process of finding as many solutions of the Pareto front as possible. Solving a multi-objective problem involves satisfying three areas [36,39]:

- *Accuracy*: To find a set of solutions as close to the real Pareto front as possible.
- *Diversity*: To find a set of solutions as diverse as possible.
- *Spread*: To find a set of solutions that "capture the whole spectrum" of the true Pareto front.

These requirements are exemplified in Fig. 4. The first case (Case 1) is able to obtain solutions that are accurate and capture the extent of the objectives; nonetheless, the set of solutions is not diverse. In the second case (Case 2), a diverse set of well-spread solutions is obtained, although these are not accurate. The solutions in the third case (Case 3) are accurate and diverse; however, the edges of the Pareto front are not explored. Finally, the fourth case (Case 4) illustrates the solution of an ideal algorithm.

Methods to find several Pareto set solutions are divided into two main groups [36]. The first group makes use of single-objective techniques and *a priori* information. Several solutions of the Pareto set are identified by changing the aggregation function or the master objective iteratively. The use of single-objective optimization methods for multi-objective optimization is known as the "classical approach" to multi-objective optimization. Two of the most common methods of this type are the weighted-sum method and the ε-constrained method [36]. These methods can be very useful when detailed preference information is known beforehand. However, these methods have their drawbacks, the weighted-sum method can prove to be very time consuming with a large number of objectives (i.e. a large Pareto set) and the solutions found will strongly depend on the shape of the Pareto frontier and the aggregation method [36]. Also, the weighted-sum method is unable to deal with non-convex Pareto fronts. Similarly, the ε-constrained has been classified as a "naïve" approach for multi-objective optimization [40], as it requires strong *a priori* knowledge of the problem [41], is time consuming (each single solution of the Pareto front requires several iterations) and it is not appropriate for a large number of objectives [41].

The second group of multi-objective optimization methods is based on Evolutionary Algorithms [36]. EA handle sets of possible solutions simultaneously, and as a result, permit identification of several solutions of the Pareto front at once. Hence, EA are recognized as a natural way of solving multi-objective problems efficiently. The first Multi-objective Evolutionary Algorithm (MOEA) was proposed in 1984. Since then, a large number of MOEA has been developed. Generally, these are classified as first-generation or second-generation MOEA [36]. The key characteristic

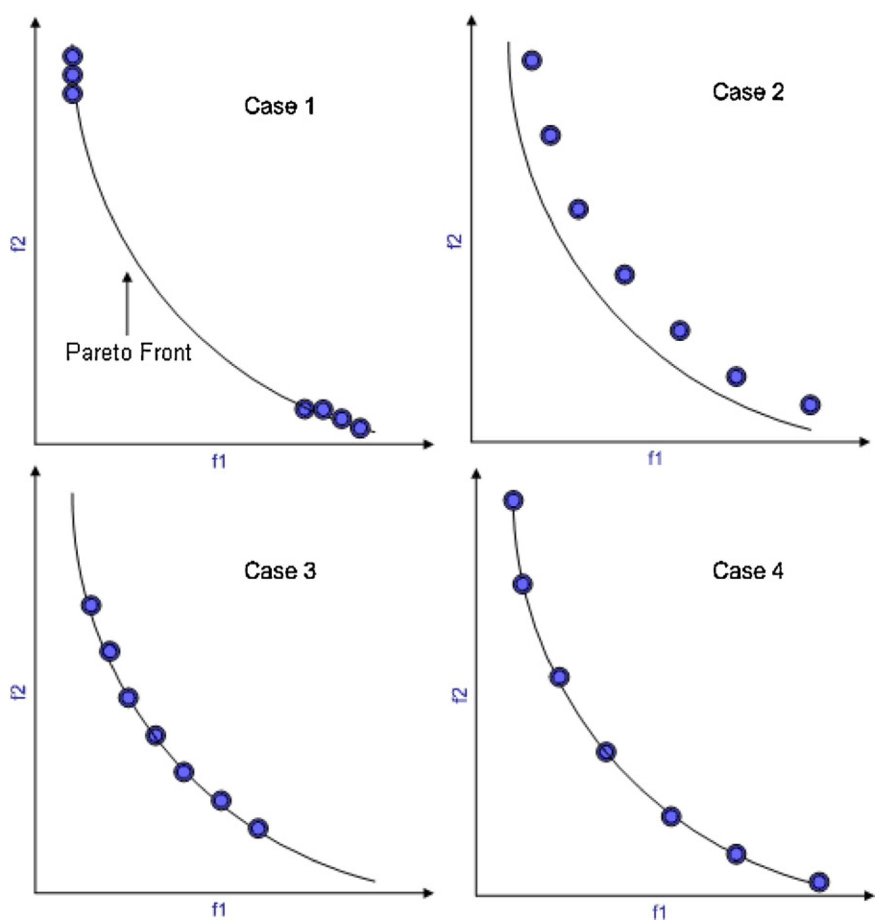


Fig. 4. Requirements of a multi-objective optimization problem.

of the second generation of MOEA is the use of elitism. Second-generation MOEA have been demonstrated to outperform their first-generation (non-elitist) counterparts [36]. A detailed account on the development of MOEA is presented in [12,36]. At present, two of the most recognized algorithms of the second generation are the Non Sorting Genetic Algorithm II (NSGA-II) [42] and the Strength Pareto Evolutionary Algorithm 2 (SPEA2) [43]. These algorithms include procedures to find an accurate, diverse and well-spread Pareto front. Hence, they guarantee to generate useful information for the subsequent decision-making process.

A growing number of authors have proposed multi-objective approaches for the DER planning problem, especially in the last six years. Initially, "classical" multi-objective optimization techniques were used. Then, the recognition that a "true" multi-objective approach provides a better way of solving the problem encouraged the use of specialized MOEA such as the ones aforementioned. In the next sections, these multi-objective DER planning approaches are reviewed. They have been grouped based on authors (or research groups) and this can be read as the 'schools' from which this thinking on DER planning optimization is emerging.

4. Multi-objective DER planning: a review

4.1. From the ε-constrained method to the NSGA-II method

Celli et al. [32] presented in the 2003 Power Tech Conference one of the first works to discuss the advantages of a multi-objective formulation for DG planning. This work proposes the use of a GA based ε-constrained method to find the best sizes and locations for DG to minimize several objectives. These objectives are: cost of reinforcements, cost of energy non-served, cost of power losses, cost of energy bought and a harmonic distortion index. In addition, technical constraints of the network are taken into account (voltage, line current and short circuit limits). The problem is analyzed from the point-of-view of a DSO that has no control over DG investments. Hence, Celli et al. mention that the information produced by the planning tool can be used to determine any incentives the utility could offer to DG developers.

This work was later extended and published in 2005 [35]. In this second publication, Celli et al. discuss load and DG variability. The objective function is evaluated by means of a probabilistic load-flow, previously developed by Celli et al. [44]. It can be inferred that this approach was also used in the publication reviewed in the previous paragraph. This probabilistic load flow assumes linear correlations among DG units, and between loads and DG units. Therefore, controllable DG units cannot be analyzed with this method. The probabilistic load flow used assumes that the probability distribution function (PDF) of all generators and loads is normally distributed. However, some DER cannot be accurately described by a normal PDF. For example, wind energy is usually described using Weibull or Rayleigh distributions [10]. Hence, in some cases the evaluation of planning attributes using this approach will be only approximated.

In 2005, Carpinelli et al. [45] extend the multi-objective approach in order to include uncertainties in DG energy production. Each one of the possible futures is formulated as a scenario. Subsequently, a "double trade-off method" is used. The double-trade-off method can be summarized in five steps:

1. Formulate the problem as a single-objective problem: use one objective of interest for the planner as the master objective, and the rest of the objectives as constraints.
2. For each objective chosen and for each scenario, apply the ε-constrained method [35] to find several Pareto solutions.
3. Evaluate the set of optimal solutions of each scenario in the remaining scenarios.
4. For each scenario, determine the set of non-dominated alternatives (conditional set).
5. Finally, find the global decision set: the alternatives that are not dominated in at least one future, that is, the union of the conditional sets.

The robustness of each of the alternatives in the global decision set is calculated and used to choose the best plans. The robustness of an alternative is defined as the proportion of scenarios where it belongs to the conditional set. That is, the alternatives with the highest robustness are those which belong to the Pareto front in most of the possible futures (scenarios).

The double-trade-off method is based on the trade-off analysis proposed by Burke et al. in 1988 [46] and it is a practical way to deal with uncertainties under a multi-objective perspective. The scenario technique is considered by Willis [47] as the only valid method to handle future uncertainties, especially in multi-objective problems. The work of Carpinelli et al. [45] analyses three minimization objectives: cost of energy losses, voltage profile and total harmonic distortion. The voltage profile objective is calculated as the mean deviation of voltage across the network. However, this might obscure localized benefits of DG, or alternatively, hide problems that are not solved by DG.

Subsequently in 2007, Carpinelli et al. [48] apply the double trade-off approach to the optimal sizing and siting of power-electronic interfaced (controllable DG). An inner optimization is used in every evaluation step of the Genetic Algorithm to determine the best operation mode of the power-electronic interface. This inner optimization has the objective of reducing harmonic distortion and improving voltage profile by managing reactive power. In this approach, the variability of DG is not addressed. Though, importantly, this paper illustrates the possibility of GA to accommodate inner optimization algorithms to handle controllable DER. It proposes the use of reactive power management to control voltage profiles. Nonetheless, because of the high R/X ratio of distribution lines, voltage magnitudes are also dependant on active power injections [49]. Hence, active power management of DG/DER could also be considered to manage voltage profiles [50].

Until 2008, largely all the multi-objective formulations proposed by the power systems research group of the University of Cagliari were based on the ε-constrained method. In Carpinelli et al. [45] the authors already recognized that *a priori* preferences could notably affect the final solutions. Moreover, in the 2008 PMAPS conference, Celli et al. [51] acknowledged that the use of true multi-objective approaches seems more effective than the ε-constrained method. So, in this latter work [51] the planning approach previously proposed in [35] is updated to a state-of-the-art Multi-objective Evolutionary Algorithm (NSGA-II [42]).

Celli et al. [51] also propose a problem formulation that can handle different types of generators simultaneously, and can incorporate optimization constraints using the concept of "constraint-dominance", proposed by Deb et al. [42]. Constraint-dominance extends the concept of dominance, discussed at the start of Section 3. A solution *a* is said to "constraint-dominate" a solution *b*, if any of the following conditions is true:

- Solution *a* is feasible and solution *b* is not.
- Solutions *a* and *b* are both infeasible, but solution *a* has a smaller overall constraint violation.
- Solutions a and *b* are feasible and solution *a* dominates solution *b*.

This concept provides a useful and parameter-less constraint handling technique that can be applied to other Multi-objective Evolutionary Algorithm.

Celli et al. [51] recognize that one of the drivers for DG is the environmental benefits that some DG technologies can provide; so, an environmental objective (minimization of total CO_2 emissions) is explicitly included. Hence, in this work, the authors provide a comprehensive formulation of the problem from the DSO perspective, which includes technical, economic and environmental objectives. DG and load time-variability are acknowledged, and DG production and load is evaluated using a probabilistic approach although, in the case study presented in the paper only simplistic daily load curves are used, ignoring seasonal variations of DG and load.

In the 2009 CIRED conference Celli et al. [33] argue, realistically, that DG investments are not decided by DSOs, in the current regulatory environment. As a result, they propose a multi-attribute analysis of random DG configurations. In this analysis no optimization is performed, but the dominance relationships between thousands of random solutions is evaluated to determine a sub-optimal Pareto front. Three objectives are chosen, one to represent the DSOs perspective, another one representing the DER developer point-of-view, and one representing civil society. Importantly, they incorporate diverse ANM schemes into the analysis. However, the probabilistic representation of wind production is simplified as a normal distribution, which will only provide approximate results, as already discussed at the start of this section. The multi-attribute analysis of random configurations provides a more realistic picture of possible DER developments. However, the optimization of DER configuration could inform DSOs and the regulator to encourage developments in an optimal manner, as initially discussed by Celli et al. [32].

In summary, through all the publications reviewed Celli, Carpinelli et al. highlight the advantages of using a multi-objective approach; they recognize that a multi-objective approach permits a better simulation of reality and that it can help in the decision-making process. They mention a key aspect: "a (planning) tool should leave the planner the faculty of choosing which aspects to consider in his search of the optimal solution" [35]. These publications brought the research community's attention towards the multi-objective nature of the DER planning problem. As a result, [35] is frequently cited in recent works in the area. Conversely, these approaches have some limitations. For example, the probabilistic approach used [44] cannot handle controllable DER units, and provides only an approximated representation of wind generation. Furthermore, while probabilistic information is available, the use of the probability of constraint violation as a planning objective/constraint is not investigated, even when new regulations favor the use of probabilistic treatment of constraints, for example the European Standard EN 50160 [52].

4.2. A multi-objective performance index

The work of Ochoa et al. [53] focuses on the technical impacts of DG. In 2005 the authors propose the use of a "multi-objective performance index" to evaluate various technical impacts of DG in unbalanced distribution networks: active power losses, maximum voltage drop and short circuit currents. This performance index is calculated as a weighted-sum of these technical impacts. In order to find the best locations for DG connections in distribution networks, Ochoa et al. [53] propose the use of a single-objective GA, and employ the weighted-sum index as the objective function. In this was the best locations that minimize DG impacts are determined. This paper recognizes that DSOs might not have control over DG investments, but that information about optimal DG locations could shape the nature of the contract between the DSO and the DER developer.

Subsequently, Ochoa et al. demonstrate the applicability of the multi-objective performance index to single DG/load scenarios [54] and to scenarios that include time-varying generation [55]. The analysis of two additional impacts is added: reserve capacity of conductors and reactive power losses. However, in both of these papers, the approach is limited to an evaluation of possible DG connections (exhaustive location of DG units in diverse nodes), rather than applying an optimization algorithm to find the best locations/sizes for DG. Even so, the approach is a powerful tool for DG impacts evaluation as it considers unbalanced networks, load and DG variability. Moreover, the authors acknowledge that other impacts (economic and environmental) could be included in the evaluation.

The multi-objective index evaluates several impacts. However, in the case of radial networks (as are most distribution networks in normal operation), it can be demonstrated that most of the impacts have a high positive correlation. For example, active and reactive line losses are concurrent. Similarly, line losses (active and reactive) and reserve capacity of conductors both depend on line flows. Likewise, line losses and maximum voltage levels have a positive correlation. As a result, the weighted-sum is measuring several times the same basic effect, i.e. the reduction of line flows. A "true" multi-objective formulation of the problem, explained in a previous section, and an analysis of objectives correlation (e.g. by means of Principal Component Analysis [56]) could identify these relationships and determine the minimum number of impacts that need to be analyzed [57].

The multi-objective index is a weighted-sum of the technical impacts; that is, a single value that represents not only the technical impacts of DG but also the point-of-view of the planner. The implications of using this weighted-sum are discussed in Ochoa's doctoral thesis [58], published in 2006. In this work, it is discussed that the major drawback of the weighted-sum approach is the difficulty of determining appropriate values for the weights when there is not enough information about the problem. So, Ochoa proposes a true multi-objective formulation of the problem. In this case, the first-generation Non Sorting Genetic Algorithm (NGSA) is used to locate a small number of fixed size wind turbines in order to maximize/minimize energy exports (for profit or energy independence, respectively) and minimize energy losses and short circuit limits. In this way, it is possible to investigate a compromise between DG benefits, and impacts. Ochoa mentions that while more objectives could be included in the formulation, care must be taken to guarantee that objectives are not concurrent.

The multi-objective index proposed by Ochoa et al. [54] was recently applied by Singh et al. [59] to investigate the effect of five different load models on the optimal placement of DG. This study used a snapshot analysis of the system, with a simplified representation of DG (constant power, at unity power factor). Results showed that different optimal locations and sizes were obtained with different load models. It concluded that the load model has "a decisive role" on the optimal placement and sizing of DG, demonstrating the importance of using an accurate model of the system being studied.

4.3. Multi-objective planning of stochastic DER and storage

Haesen et al. [60] discuss the drawbacks of single-objective formulations and recognize the advantages of a true multi-objective approach. Accordingly, in 2006 Haesen et al. [60] propose a multi-objective DER optimization based on the first-generation Strength Pareto Evolutionary Algorithm (SPEA). The objective function evaluation includes a simplistic simulation of daily DER production and load profiles; though the method permits the optimization of several types of DER simultaneously. This multi-objective DER planning approach is compared with the iterative use of a single-objective method, previously proposed by Haesen et al. [61]. The comparison shows that single weighted-sum solutions are better than the ones found in the Pareto front by SPEA, but that in contrast the whole Pareto front provides a wider range of possible solutions to choose from. Also, each weighted-sum solution is highly sensitive to the set of parameters chosen. Therefore, if a single solution is sought, inaccuracy in any weight will lead the search towards mistaken regions of the Pareto set and produce a sub-optimal plan. As a result, Haesen et al. suggest the use of both methods to gain insight into the planning problem. However, finding each single weighted-sum solution requires as many iterations as finding the whole Pareto front using SPEA.

Importantly, in this work, Haesen et al. recognize that in cases when attributes cannot be converted to cost accurately, when all costs cannot be aggregated into a single parameter or when a larger number of objectives are analyzed, the "true" multi-objective optimization becomes essential. This argument is exemplified by a case study that analyses four very distinct planning objectives: line loss minimization, minimization of the main grid energy flow (as a proportional measure of reliability), DER installation cost and the gas distribution grid investments. Finally, this paper proposes the use of bi-objective plots [36] to examine correlations or conflicts between objectives. This visualization technique becomes extremely useful when the number of objectives is greater than three.

In the next paper of Haesen et al. [62], the use of traditional mathematical optimization techniques for planning time-variant DER is studied. The DER planning problem is formulated as an iterative Mixed Integer Quadratic Programming problem. Traditional optimization techniques require mathematical formulations of the objective functions. These formulations can only include deterministic profiles. As a result, the authors conclude that traditional optimization techniques cannot model the stochastic aspects of DER and load effectively. In addition, the authors identify that some objectives (e.g. voltage sags, reliability) cannot be formulated as a mathematical function of DER type, placement and size.

As a result, Haesen et al. [62] highlight that GA can handle objectives that are too complex to be reasonably formulated in an analytical expression. So, the use of Monte Carlo Simulation (MCS) in the objective evaluation is suggested, instead of the daily profiles simulation used in [59]. The MCS method produces an accurate evaluation of the stochastic performance of DER (e.g. wind production) and load without the need for an analytical formulation. Moreover, it permits the evaluation of other objectives that are difficult to formulate analytically (e.g. reliability). In the approach proposed, MCS consists of the simulation of a number of different yearly profiles. The planning methodology for stochastic DER is summarized in a further paper [63] published in 2007. In this work, the authors recognize that an optimization approach should be as adapted to the problem as possible, a clear reference to the optimization/modeling dilemma already discussed.

The GA-MCS approach provides a practical way of evaluating topologies with stochastic DER, however two trade-offs can be identified. First, the optimization/modeling trade-off [15]: GA permit the evaluation of more realistic models, but the convergence towards global optima cannot be reached in limited time. In contrast, analytical expressions are able to find the global optima (with appropriate parameters), yet, they are limited to evaluate simplified models. The second trade-off relates to the accuracy of the MCS. The accuracy of MCS evaluations depends on the number of trials or years simulated [18]. So, accuracy improves but to the detriment of the speed of the GA, and vice versa. Though, the GA-MCS evaluation time needs to be put in perspective. First, planning is not an "online" task and optimization studies can be performed at the same time as other studies. Second, and more importantly, it is possible to get results and insights that otherwise would have never been obtained.

The SPEA planning approach is used by Haesen et al. [64] to analyze the incorporation of a single controllable energy storage unit into a distribution grid with stochastic DER. This work was presented in the CIRED 2007 conference. In this case, an inner optimization algorithm is used in the objective evaluation stage of the GA to optimize the operation of the storage unit. Simultaneously, the external multi-objective optimization is used to optimize the rating (power) and capacity (energy) of the storage unit. This inner optimization offers a practical method to optimize controllable energy storage *when DER units are already installed*. However, the approach is not able to optimize stochastic units that can be controlled (e.g. curtailment of wind generators, dispatch of CHP units), or the simultaneous optimization of stochastic and controllable units.

In the CIRED 2007 conference presentation [65], Haesen proposed the use of Principal Component Analysis (PCA) [66], a powerful method to reduce the dimensions of a multi-objective problem and analyze multiple objective correlations. Moreover, Haesen et al. [64] explore the use of probabilistic measures of DER impacts; it uses the 95-percentile of the probability distribution for the maximum voltage deviation as one of the planning objectives.

In summary, the method proposed by Haesen et al. permits the multi-objective optimization of diverse types of time-variant DER or controllable energy storage. It provides a flexible platform in which different planning objectives can be formulated. Though, it is not possible to infer from the works published how network constraints are treated in the multi-objective formulation. No constraint management under the multi-objective formulation is described.

4.4. A multi-objective planning framework for controllable and stochastic DER

In a work published in 2009, Alarcon-Rodriguez et al. [50] extend the framework proposed by Haesen et al., reviewed in the previous section. Alarcon-Rodriguez et al. [50] present a flexible planning framework for optimizing controllable and stochastic DER. The method has three key elements:

- An outer multi-objective optimization algorithm based on the second-generation SPEA2. Previous studies had shown that the SPEA algorithm is outperformed by both the NSGA-II [42] and SPEA2 [43].
- A stochastic simulation algorithm for the evaluation of stochastic DER.
- An inner optimization algorithm for the evaluation of controllable DER, formulated as a linear Optimal Power Flow.

In this work, the formulation of the outer optimization algorithm (SPEA2) permits the optimization of different types of stochastic and controllable DER simultaneously. The stochastic simulation algorithm permits the evaluation of stochastic DER, either using historical data of DER production, or using weather models to produce this data. The accuracy of the stochastic evaluation will depend on the number of events evaluated [18]. Hence, accurate evaluations will require longer computation times. The inner OPF can be adapted to evaluate energy storage or different ANM schemes (e.g. active power dispatch, reactive power dispatch, active voltage control). Moreover, Alarcon-Rodriguez et al. [50] use the probability of voltage violations in the system as one of the planning objectives. It has been suggested that a probabilistic analysis permits a more objective evaluation of DER impacts [4]. In addition, the minimization of carbon emissions is explicitly formulated as a planning objective. Previous work of Alarcon-Rodriguez et al. presented in 2006 [67] and published later in 2008 [68] proposed a multi-attribute analysis of DER, using a MCS evaluation, with a flexible treatment of constraints, and the explicit formulation of environmental objectives.

Other planning objectives considered in the case study in [50] are: the minimization of line losses, the minimization of extra energy dispatch, the minimization of energy curtailment, and the minimization of the DER penetration level. This last objective might seem counterintuitive, as single-objective techniques aim to maximize DER penetration; nonetheless, it is necessary in a multi-objective approach to determine the optimal attainment level of each of the planning objectives for each level of penetration of DER [12].

The planning framework proposed in [50] was extended by Haesen et al. [69] to compare network reinforcement and DER as alternative planning options. The effects of different tariff schemes in the objectives of the DSO and DER developers were examined. This work was presented in 2009. The concept of constraint-dominance, already discussed in a previous section, was incorporated in the SPEA2 fitness evaluation step. Constraint-dominance permits the formulation of any planning attribute as a planning constraint, extending the flexibility of the framework. In the case study presented by Haesen et al. [69], the use of probabilistic constraints is proposed, following recent European Regulations. The EN50160 Power Quality norm which has to be guaranteed by the DSO in many European countries requires the grid voltage at LV to remain within 10% of the nominal voltage for 95% of the time [52].

In summary, the planning framework proposed by Alarcon-Rodriguez, Haesen et al., which is extensively discussed in [12,70], provides a flexible platform, in which diverse impacts of DER integration can be analyzed, either as planning objectives or planning constraints. Probabilistic measures of DER impact can be evaluated, and any number of DER types can be incorporated in the analysis. The framework can be modified to analyze different ANM schemes.

A drawback of the proposed methodology is that it is inherently computationally expensive. Though, this can be said of any approach based on EA with inner MCS objective evaluations, as mentioned in a previous section. SPEA2 requires the evaluation of thousands of potential solutions (i.e. chromosomes) to obtain a good approximation of the Pareto set. At the same time, the stochastic simulation of each chromosome requires from hundreds to thousands of evaluations to get accurate estimations of the planning attributes. Moreover, OPF evaluations (for controllable DER units) tend to take more than simple power flows to analyze uncontrollable units. Though, this long evaluation time needs to be considered in perspective, as already discussed, the method enables to produce information that cannot be obtained with simplified approaches.

4.5. Other multi-objective approaches

The methods reviewed next do not belong to any of the "schools" previously introduced. Their key contributions and limitations are highlighted:

Pelet et al. [71] study the optimization of the design parameters of an integrated energy system (diesel and PV generators) for a remote community. Detailed analytical formulations are used for the diesel engines and PV operation, cost and emissions calculation. Two objectives are used: total cost and CO_2 emissions. The authors use a "true" multi-objective formulation, based on a second-generation MOEA. They argue that keeping the two objectives separated enables more informed design decisions, as it is possible to find and rank the best integrated solutions, which are both cost effective and less polluting. Moreover, the conflict between cost and environmental benefits is recognized with the conclusion that clean solutions are more expensive.

Harrison et al. [34] use the OPF approach presented in [11] to evaluate the incentives provided to DSOs and DER developers for loss reduction and reinforcement deferral. Two different objective functions are analyzed. Each one reflects the point-of-view of a DG developer and a DSO, respectively, both trying to maximize their net benefits. A multi-objective formulation based on the ε-constrained method is presented. Moreover, a multi-period OPF is proposed, which evaluates a load duration curve to provide a better estimation of losses. Harrison et al. show that DG developers and DSOs have conflicting objectives and that a multi-objective formulation can effectively replicate different perspectives of the DG planning problem. Moreover, this work demonstrates that incentives do have a major impact on stakeholders' optimal locations and sizes for DG. For example, DG developers are not directly exposed to the effect of losses, so they try to maximize capacity and profit. On the other hand, DSOs have a loss reduction incentive that outweighs the benefit of connecting DG. Subsequently, they would prefer smaller DG investments that provide a larger reduction in losses, to the detriment of a DG developer's profit. A trade-off analysis enables the identification of several possible compromise solutions. A similar analysis is made for reinforcement deferral incentives. A limitation of the proposed approach is that DG is considered as a firm supply of energy, operating constantly at rated power. This restricts the analysis of time-variant generators such as renewable DG and heat-led CHP. In addition, the ε-constrained method has some drawbacks, which were already discussed.

Mori and Yamada [72] present an approach based on SPEA2 to optimize *distribution network expansion planning*. This approach considers DG as an option for the planner, together with possible substations and lines. It aims at minimizing three objectives: power losses, cost of new equipment and voltage deviation. The cost objective only considers installation costs and it does not take into account operating costs of DG (fuel, O&M). So, the optimal solution could be more expensive in the long-term. In addition, the problem disregards the time-variability of DG. The whole planning exercise is made in terms of peak power. As a result, only a single type of DG can be handled by the formulation (i.e. constant power). Nonetheless, an important point of this work is that it demonstrates that SPEA2 provides better solutions than NSGA-II in the case study presented, although SPEA2 computational time is slightly higher than NSGA-II.

Haghifam et al. [73] also assume that DG is a constant power source. The authors propose an approach based on NSGA-II. The planning objectives include total cost (net present value of energy bought from the transmission system, DG installation and operation), technical and economic risks. The novelty of this work is that it proposes to minimize the maximum risk of constraint violation as one of the planning objectives. In this case, load behavior uncertainty is modeled using fuzzy numbers. The risk of voltage constraint violations is calculated as the fuzzy possibility of voltage constraint violation. The economic risk is treated similarly: the uncertainty of market price of energy is modeled using fuzzy numbers. Then, the fuzzy possibility of DG being a more expensive solution is calculated and minimized. Fuzzy numbers permit the representation of uncertain variables for which limited information is available. Therefore, a quasi-probabilistic formulation of the problem is possible. An analogy can be made between the fuzzy "possibility" of constraint violation and the more elaborated "probability" of constraint violation. However, the calculation of this latter requires more detailed information about the load behavior (e.g. load curve duration, load profile, load model).

Ahmadi et al. [74] also propose to use the NSGA-II algorithm to find the optimal combination of DG units in a network. Three planning objectives are optimized: to minimize total cost, to minimize line losses and to improve voltage profile. Though, the approach is simplistic. The case study mentions five types of DG, including PV generators, though, the time-variability of DG is not mentioned in the paper, nor modeled in the approach proposed. Since only snapshot analyses are used results can be expected to be unrealistic.

Zangeneh et al. [75] base the optimization on the NSGA-II method. The approach includes the minimization of economic objectives: the total cost of DER installation and operation, the cost of losses and the cost of extra purchased power. It also formulates an environmental objective: the maximization of avoided emissions. In addition, this work outlines a simple methodology for choosing a single solution from the Pareto set, i.e. the decision-making process, discussed in the next section. After the optimization, all objectives are aggregated into a single parameter. Hence, the multi-objective problem is formulated from the perspective of a DSO that can invest in DER as a planning option. It also includes a budget constraint, reflecting that the planner must make the best use of scarce resources. Although the authors decide to use numerically similar weights, it is clear that the weights could reflect the planner's preferences more strongly. Monte Carlo Simulation is proposed as a means to evaluate the uncertainty in some parameters, such as DER costs and energy prices, prior to the decision-making process. This use of MCS must not be confused with the use of MCS to evaluate the stochastic behavior of DER, mentioned in previous sections. A critical shortcoming of this work is that it does not model the variability of DER. It fails to acknowledge that some DER, such as PV and wind turbines, might not be available at peak demand times with any level of certainty.

4.6. Multi-criteria decision-making methods

The review of the previous sections shows that most of the approaches aim at generating a large number of non-dominated solutions, namely the multi-objective optimization process. The decision-making process of choosing correct *a priori* weights or selecting a single solution *a posteriori* is only explored briefly, or not mentioned at all. Several decision-making techniques exist in literature. When a number of attributes are analyzed, these techniques are referred as to Multi-criteria Decision Making (MCDM), or Multi-criteria Decision Analysis (MCDA). This is a vast research area. A recent review of the state-of-the-art of MCDA techniques was compiled by Figueria et al. [76]. Also, the application of MCDM to energy planning problems is studied by Hobbs and Meier [77]. Similarly, Loken [78] and Pohekar and Ramachandran [79] have reviewed the use of MCDM for energy planning.

There are some examples of application of MCDA and MCDM techniques to multi-objective DER planning. Tang and Tang [80] propose a weighted-sum for the optimization of DG location and size. It analyses four distinctive objectives. Although the DG modeling is still simplistic (a snapshot analysis), the authors addressed the problem of how to chose appropriate weights, depending on the planner preferences. They propose to use the Analytical Hierarchy Process (AHP) for this purpose, a recognized decision-making method [77]. Zangeneh and Jadid [81] in contrast, focus on how to obtain diverse solutions of the Pareto front using a single-objective minimization. They propose to use the Normal Boundary Intersection method to generate evenly distributed solutions in the Pareto set. They consider three objective functions, the total cost of DG (installation and operation), the cost of energy losses and the cost of energy not served. The modeling of the DG planning problem is still very simplified. Moreover, it is not clear from the case study proposed if this approach would be more effective than specialized MOEA (such as SPEA2 or NSGA-II) in finding a well spread, diverse and accurate Pareto front.

Barin et al. [82] explore the problem of choosing the best planning option from a "previous list of viable places for installation". Hence, they focus on the *a posteriori* multi-criteria decision-making problem, assuming that the multi-objective optimization problem has already been solved. They propose to use the Bellman–Zadeh algorithm. Following this algorithm, each objective of every solution is normalized to obtain a *normalized membership* value between 0 and 1 (0 been the worst performance, and 1 being the best performance). Weights are assigned to each membership value, according to its importance. Then, solutions are ranked according to the highest value of its worst normalized membership. This can be understood as a mini–max approach, in which the maximum distance to the goal (i.e. the worse performance in a weighted and normalized objective) is minimized. An interesting aspect is that the method facilitates the use of qualitative criteria for DG planning, such as security, physical space and vandalism. Qualitative criteria, such as social acceptability, are commonly hard to quantify numerically, but important to consider in some DG developments.

5. Discussion and conclusions

5.1. Discussion

From the review in Section 4, some trends can be identified in terms of the optimization methods utilized and the detail of the DER model. It can be observed that "classical" approaches, such as weighted-sum or the ε-constrained method are being gradually replaced by state-of-the-art MOEA, particularly second-generation SPEA2 and NSGA-II (Fig. 5). Classical approaches remain a useful option when detailed preference information is known *a priori*, and when the objective of the planning exercise is to find a single solution that represents a single point-of-view. The use of multi-attribute analysis has been proposed when the exogenous nature of DER investments is recognized. Multi-attribute analysis is not an optimization method, though it can be used to evaluate multiple impacts of unplanned (or random) DER developments.

The key characteristic of second-generation MOEA is the use of elitism [36], as already mentioned. Elitist MOEA have been

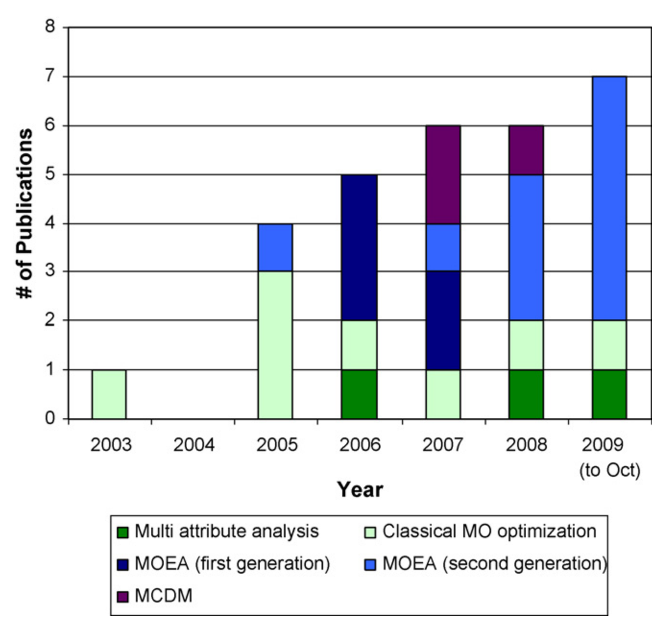


Fig. 5. Number of multi-objective DER planning publications per type of method used and per year.

demonstrated to out-perform non-elitist MOEA, hence, it is expected that in the coming years specialized MOEA of the second generation will be used widely in this area, while the use of first-generation MOEA should diminish. The concept of "constraint-dominance" permits the handling of constraints within the multi-objective formulation. Although initially proposed to be used with NSGA-II, this concept can be applied to any MOEA. At present it is used only by a small number of authors, though, it is expected that "constraint-dominance" will be more widely adopted in coming years.

MOEA permit the identification of a large number of Pareto front solutions, and provide information on the trade-offs and correlations between planning objectives. Moreover, EA permit the evaluation of complex models of DER. As a result, it is possible to observe that the detail of the DER models has evolved together with the use of MOEA. For example, most publications reviewed propose methods to evaluate the stochastic behavior of DER and load, either by means of probabilistic load flow, MCS, or stochastic simulation. Probabilistic load flow provides a fast evaluation of the stochastic behavior of the power system, though it limits the evaluation of controllable DER, and requires simplifying assumptions of the PDF of some DER, as already discussed. In contrast, the use of MCS and stochastic simulation permits a more accurate representation of some DER, such as wind generation, and can evaluate controllable DER. Though, a major drawback of the MOEA-MCS framework is the long computational evaluation time required. Hence, a key avenue for research is the parallel-computing implementation of MOEA-MCS DER planning approaches, and the use of clustering techniques to reduce the number of power flow evaluations.

Some authors still base the analysis of one unique DER solution on a single snapshot analysis of the power system. It must be emphasized that snapshot analyses are not appropriate to model most DER, which are variable in nature. Hence, these approaches are limited and only applicable to very specific DER types and circumstances (e.g. back-up DG units, capacity assessment of networks). As the use of renewable energy resources (most of which are variable in nature), heat-led CHP and Demand Side Management (DSM) become more widespread, simplistic models for DER planning will not be sufficient to generate useful knowledge.

Few techniques consider the controllability of DER. The few approaches that do illustrate that the use of MOEA permits the incorporation of "inner optimization" algorithms in the objective evaluation. The inner optimization algorithms, formulated commonly as an Optimal Power Flow algorithm, facilitate the simulation of controllable energy storage, controllable loads and controllable DER units. It is expected that the possibility of inner optimization algorithms will be exploited more widely, as the concept of active management of DER, Demand Side Management (DSM), and smart networks becomes widespread.

The papers reviewed highlight the benefits of a multi-objective formulation, and a wide range of technical and economical objectives are formulated. This demonstrates the flexibility provided by the multi-objective approach. Most of the authors recognize that one of the drivers for DER development is the environmental benefit(s) that can be obtained from an adequate integration of these technologies. As a result, new approaches incorporate explicitly environmental objectives. Moreover, the variety of case studies proposed highlight that multi-objective DER planning can be used to study different incentives schemes, analyze different impacts from a single stakeholder perspective or to determine compromise solutions that benefit different stakeholders. However, evidence of real applications in support of decision-making has still to be published.

5.2. Conclusions

In a future where a larger share of energy will be supplied from distributed sources, and where potentially a larger number of stakeholders will be involved, multi-objective planning tools will be needed to provide compromise solutions, and guide the optimal development of the system. Hence, multi-objective DER planning is a novel area that has gathered increased interest in recent years. This paper has presented a critical review of the state-of-the-art of multi-objective DER planning methods. Key aspects of the techniques that need to be considered in implementation have been highlighted and recent trends in the area have been discussed.

Some future avenues for the research in multi-objective DER planning can be identified. For example, multi-objective DER planning methods have yet to be applied to analyze the wide implementation of controllable loads and DSM. DSM and load controllability will gain prominence in a future where the impacts of energy use will be more carefully scrutinized and managed. Moreover, the use of electric vehicles (EVs) has been proposed as one solution to reduce carbon emissions from transport. EVs can be utilized as a sizeable and distributed form of electrical energy storage. The analysis of the impacts of EVs in the power system can be formulated within the multi-objective framework, where different perspectives of the problem can be represented.

In all cases, an adequate level of detail must be provided in the optimization models of active DER and active networks, in order to provide realistic solutions. Moreover, the focus must lie not only in the supply of electricity; as the interaction of electricity networks, gas networks, heat networks and DER will be essential in a future with a more decentralized and decarbonized energy supply. In addition, this research area will benefit greatly from publications that study real case studies in which the support to multifaceted decision-making scenarios is demonstrated.

Acknowledgments

This research was partially funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under the SUPERGEN 3–Highly Distributed Power Systems Consortium (www.supergen-hdps.org/) and by the Transition Pathways to a Low Carbon Economy Consortium (http://www.lowcarbonpathways.org.uk/) funded by EPSRC and Eon UK.

References

[1] Energy white paper: meeting the energy challenge. Department of Trade and Industry UK (DTI); 2007.
[2] Ackermann T, Andersson G, Soder L. Distributed generation: a definition. Electric Power Systems Research 2001;57(April (3)):195–204.
[3] Jenkins N, Allan R, Crossley P, Kirschen D, Strbac G. Embedded generation. London: Institution of Electrical Engineers; 2000.
[4] Pecas Lopes JA, Hatziagyriou N, Mutale J, Djapic P, Jenkins N. Integrating distributed generation into electric power systems: a review of drivers, challenges and opportunities. Electric Power Systems Research 2007;77(July (9)):1189–203.
[5] Mendez VH, Rivier J, De la Fuente JI, Gomez T, Arceluz J, Marin J, et al. Impact of distributed generation on distribution investment deferral. International Journal of Electrical Power and Energy Systems 2006;28(May (4)):244–52.
[6] Thong VV. Impact of Distributed Generation on Power System Operation and Control. PhD thesis. Katholieke Universiteti Leuven, Leuven, Belgium; May 2006.
[7] Pepermans G, Driesen J, Haeseldonckxc D, Belmans R, D'haeseleer W. Distributed generation: definition, benefits and issues. Energy Policy 2005;33(April (6)):787–98.
[8] Strbac G, Ramsay C, Pudjianto D. Integration of Distributed Generation into the UK Power System. Summary Report. DTI Centre for Distributed Generation and Sustainable Electrical Energy; March 2007
[9] Djapic P, Ramsay C, Pudjianto D, Strbac G, Mutale J, Jenkins N, et al. Taking an active approach. IEEE Power & Energy Magazine 2007;(July/August):68–77.

[10] Mendez VH, Rivier J, Gomez T. Assessment of energy distribution losses for increasing penetration of distributed generation. IEEE Transactions on Power Systems 2006;21(May (2)).
[11] Harrison GP, Wallace AR. OPF evaluation of distribution network capacity for the connection of distributed generation. IEE Proceedings Generation Transmission & Distribution 2005;152(January (1)):115–22.
[12] Alarcon-Rodriguez AD. A Multi-objective Planning Framework for Analysing the Inegration of Distributed Energy Resources. PhD Thesis. Institute of Energy and Environment, University of Strathclyde; April 2009.
[13] Song YH, Irving MR. Optimization techniques for electrical power systems—part 2 heuristic optimization methods. Power Engineering Journal 2001; 15(June (3)).
[14] Miranda V, Srinicasan D, Proenca DM. Evolutionary computation in power systems. International Journal of Power & Energy Systems 1998;20(February (2)):89–98.
[15] Lee KY, El-Sharkawi MA, editors. Modern heuristic optimization techniques with applications to power systems. IEEE Power Engineering Society (02TP160); 2002.
[16] Irving MR, Song YH. Optimization techniques for electrical power systems—part 1 mathematical optimization methods. Power Engineering Journal 2000; 14(October (5)).
[17] Chen P, Chen Z, Cak-Jensen B. Probabilistic load flow: a review. In: Proceedings of the 3rd international conference on electric utility deregulation and restructuring and power technologies (DRPT 2008); 2008.
[18] Billinton R, Allan RN. Reliability evaluation of power systems. Springer; 1996, ISBN 0306452596.
[19] Zhou Q, Bialek JK. Generation curtailment to manage voltage constraints in distribution networks. IET Generation Transmission & Distribution 2007;(3): 492–8.
[20] Keane A, O'Malley M. Optimal utilization of distribution networks for energy harvesting. IEEE Transactions on Power Systems 2007;22(February (1)).
[21] Currie RAF, Ault GW, McDonald JR. Methodology of economic connection capacity for renewable generator connections to distribution networks optimized by active power flow management. IEE Proceedings Generation Transmission & Distribution 2006;153(July (4)).
[22] Wang C, Hashem M. Analytical approaches for optimal placement of distributed generation sources in power systems. IEEE Transactions on Power Systems 2004;19(November (4)).
[23] Narayan S, Rau SM, Yih-heu Wan M. Optimum location of resources in distribution planning. IEEE Transactions on Power Systems 1994;9(November (4)).
[24] Kim JO, Nam SW, Park SK, Singh C. Dispersed generation planning using improved Hereford ranch algorithm. Electric Power Systems Research 1998;47(October (1)):47–55.
[25] Willis HL, Scott WG. Distributed power generation. Planning and evaluation. New York, USA: Marcel Dekker; 2000, ISBN: 0-8247-0336-7.
[26] Dugan RC, McDermott TE, Ball GJ. Planning for distributed generation. IEEE Industry Applications Magazine)2001;(March/April).
[27] El-Khattam W, Hegazy YG, Salama MMA. An integrated distributed generation optimization model for distribution system planning. IEEE Transactions on Power System 2005;20(May (2)).
[28] Celli G, Pilo F. Optimal distributed generation allocation in MV distribution networks. In: 22nd IEEE PES international conference on power industry computer applications PICA 2001; 2001.p. 81–6.
[29] Harrison GP, Piccolo A, Siano P, Wallace AR. Hybrid GA and OPF evaluation of network capacity for distributed generation connections. Electric Power Systems Research 2008;78(March (3)):392–8.
[30] Keane A, O'Malley M. Optimal allocation of embedded generation on distribution networks. IEEE Transactions on Power Systems 2005;20(August (3)).
[31] Ochoa LF, Keane A, Dent C, Harrison GP. Applying active network management schemes to an Irish distribution network for wind power maximization. In: Proceedings 20th international conference on electricity distribution CIRED 2009; 2009.
[32] Celli G, Ghiani E, Mocci S, Pilo F. A multi-objective formulation for the optimal sizing and siting of embedded generation in distribution networks. In: Power tech conference proceedings, 2003 IEEE, vol. 1; 2003.
[33] Celli G, Pilo F, Soma GG, Gallanti M, Cicoria R. Active distribution network cost/benefit analysis with multi-objective programming. In: Proceedings 20th international conference on electricity distribution CIRED 2009; 2009.
[34] Harrison GP, Piccolo A, Siano P, Wallace AR. Exploring the trade-offs between incentives for distributed generation developers and DSO's. IEEE Transactions on Power Systems 2007;22(May (2)).
[35] Celli G, Ghiani E, Mocci S, Pilo F. A multi-objective evolutionary algorithm for the sizing and siting of distributed generation. IEEE Transactions on Power System 2005;20(May (2)).
[36] Deb K. Multi-objective optimization using evolutionary algorithms. John Wiley and Sons; 2001, ISBN 047187339X.
[37] Savic D. Single-objective vs. multi-objective optimization for integrated decision support. In: Proceedings of the first biennial meeting of the international environmental modeling and software society, vol. 1; 2007. p. 7–12.
[38] Zitzler E. Two decades of evolutionary multi-objective optimization: a glance back and a look ahead (Presentation). In: IEEE symposium on computational intelligence in multi criteria decision making (MCDM); 2007.
[39] Konak A, Coit DW, Smith AE. Multi-objective optimization using genetic algorithms: a tutorial. Reliability Engineering and System Safety 2006;91: 992–1007.
[40] Coello CA. An updated survey of GA-based multi-objective optimization techniques. ACM computing Surveys 2002;32(June (2)).
[41] Leyland G. Multi-objective Optimization Applied To Industrial Energy Problems. Doctoral Dissertation. Ecole Poly-technique Fédérale de Lausanne; 2002.
[42] Deb K, Paratap A, Agarwal S, Meyarivan T. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 2002;6(April (2)).
[43] Zitzler E, Laumanns M, Thiele L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103. Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich; May 2001.
[44] Celli G, Mocci S, Pilo F, Cicoria R. Probabilistic optimization of MV distribution networks in presence of distributed generation. In: Proceedings of the 14th PSCC; 2002.
[45] Carpinelli G, Celli G, Mocci S, Pilo F, Russo A. Optimization of embedded generation sizing and siting by using a double trade-off method. IEE Proceedings Generation Transmission & Distribution 2005;152(July (4)).
[46] Burke WJ, Schweppe FC, Lovell BE. Trade off methods in system planning. IEEE Transactions on Power Systems 1988;3(August (3)).
[47] Willis HL. Power distribution planning reference book. New York, USA: Marcel Dekker; 2004, ISBN: 0-8247-4875-1.
[48] Carpinelli G, Celli G, Mocci S, Pilo F, Proto D, Russo A. Multi-objective programming for the optimal sizing and siting of power-electronic interfaced dispersed generators. In: Proceedings of the 2007 IEEE power tech conference; 2007.
[49] Ackermann T, Garner K, Gardiner A. Embedded wind generation in weak grids—economic optimization and power quality simulation. Renewable Energy 1999;18:205–21.
[50] Alarcon-Rodriguez AD, Haesen E, Ault GW, Driesen J, Belmans R. Multi-objective planning framework for stochastic and controllable distributed energy resources. IET Journal of Renewable Power Generation 2009;3(June (2)):227–38.
[51] Celli G, Mocci S, Pilo F, Soma G. A multi-objective approach for the optimal distributed generation allocation with environmental constraints. In: Proceedings of the 10th conference on probabilistic methods applied to power systems PMAPS; 2008.
[52] European Standard EN 50160. Voltage Characteristics of Electricity Supplied by Public Distribution Systems.
[53] Ochoa LF, Padilha-Feltrin A, Harrison GP. Evaluation of a multi-objective performance index for distribution systems with distributed generation. In: Proceedings 18th international conference on electricity distribution CIRED 2005; 2005.
[54] Ochoa LF, Padilha-Feltrin A, Harrison GP. Evaluating distributed generation impacts with a multi-objective index. IEEE Transactions on Power Delivery 2006;21(July (3)).
[55] Ochoa LF, Padilha-Feltrin A, Harrison GP. Evaluating distributed time-varying generation trough a multi-objective index. IEEE Transactions on Power Delivery 2008;21(April (2)).
[56] Smith LI. A Tutorial on Principal Components Analysis. http://reflect.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf; 2002.
[57] O'Rourke N, Hatcher L, Stepanski EJ. A step-by-step approach to using SAS for univariate & multivariate statistics. SAS Publishing; 2005, ISBN 1590474171, 9781590474174.
[58] Ochoa LF. Desempenho de Redes de Distribuição com Geradores Distribuídos (Performance of Distributions Networks with Distributed Generation), Doctoral Dissertation, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista "Julio de Mesquita Filhio"; November 2006.
[59] Singh D, Singh D, Verma KS. Multiobjective optimization for DG planning with load models. IEEE Transactions on Power Systems 2009;24(February (1)).
[60] Haesen E, Driesen J, Belmans R. A long-term multi-objective planning tool for distributed energy resources. In: IEEE PES power systems conference & exposition; 2006.p. 741–7.
[61] Haesen E, Espinoza M, Pluymers B, Goethals I, Van Thong V, Driesen J, et al. Optimal placement and sizing of distributed generator units using genetic optimization algorithms. Electrical Power Quality and Utilisation Journal 2005;11(1):97–104.
[62] Haesen E, Deconinck G, Driesen J, Belmans R. Planning of distributed energy resources: traditional optimization tools versus evolutionary algorithms. In: Influence of distributed and renewable generation on power system security—CRIS; 2006.p. 190–9.
[63] Haesen E, Driesen J, Belmans R. Robust planning methodology for integration of stochastic generators in distribution grids. IET Renewable Power Generation 2007;1(1). pp. 25–32 116.
[64] Haesen E, Driesen J, Belmans R. Long-Term Planning for Small-Scale Energy Storage Units (Paper) CIRED 2007. In: 19th international conference and exhibition on electricity distribution; 2007.
[65] Haesen E, Driesen J, Belmans R. Long-Term Planning for Small-Scale Energy Storage Units (Presentation) CIRED 2007. In: 19th international conference and exhibition on electricity distribution; 2007.
[66] Lindsay IS. A Tutorial on Principal Components Analysis. http://reflect.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
[67] Alarcon-Rodriguez AD, Ault GW, Currie RAF, McDonald JR. Planning the development of highly distributed power systems. In: 2nd international conference of distributed energy resources; 2006.

[68] Alarcon-Rodriguez AD, Ault GW, Currie RAF, McDonald JR. Planning highly distributed power systems: effective techniques and tools. International Journal of Distributed energy Resources 2008;4(January (1)).
[69] Haesen E, Alarcón-Rodríguez AD, Driesen J, Belmans R, Ault GW. Opportunities for active DER management in deferral of distribution system reinforcements. In: 2009 IEEE Power Systems Conference & Exposition; 2009.
[70] Haesen E. Multi-Objective Optimization of the Integration of Stochastic Distributed Energy Resources in Electricity Grids. PhD Thesis. KU Leuven; May 2009.
[71] Pelet X, Favrat D, Leyland G. Multi-objective optimization of integrated energy systems for remote communities considering economics and CO_2 emissions. International Journal of Thermal Sciences 2005;44:1180–9.
[72] Mori H, Yamada Y. An efficient multi-objective meta-heuristic method for distribution network expansion planning. In: Proceedings of the 2007 IEEE power tech conference; 2007.
[73] Haghifam MR, Falaghi H, Malik OP. Risk-based distributed generation placement. IET Renewable Power Generation 2008;2(2):252–60.
[74] Ahmadi M, Yousefi A, Souroudi A, Ehsan M. Multi objective distributed generation planning using NSGA-II. In: Proceedings 13th international power electronics and motion control conference (EPE-PEMC 2008); 2008.
[75] Zangeneh A, Jadid S, Rahimi-Kian A. Promotion strategy of clean technologies in distributed generation expansion planning. Renewable Energy 2009;34(December):2765–73.
[76] Figueira J, Greco S, Ehrgott M, editors. Multiple criteria decision analysis: state of the art surveys. Springer-Verlag; 2005.
[77] Hobbs BF, Meier P. Energy decisions and the environment: a guide to the use of multicriteria methods. Springer; 2000, ISBN 079237875X, 9780792378754.
[78] Loken E. Use of multicriteria decision analysis methods for energy planning problems. Renewable and Sustainable Energy Reviews 2007;11(September (7)):1584–95.
[79] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning. Renewable and Sustainable Energy Reviews 2004;8(August (4)):365–81.
[80] Tang X, Tang G. Multi-objective planning for distributed generation in distribution network. In: Third international conference on electric utility deregulation and restructuring and power technologies, DRPT 2008; 2008.
[81] Zangeneh A, Jadid S. Normal boundary intersection for generating pareto set in distributed generation planning. In: Proceedings of international power engineering conference; 2007. p. 1077–82.
[82] Barin A, Pozzatti LF, Canha LN, Machado RQ, Abaide AR, Farret FA, et al. Analysis of multi-objective methods applied to distributed generation systems allocation. In: First international conference on power engineering, energy and electrical drives—Powereng; 2007.p. 243–8.

Arturo Alarcon-Rodriguez received his Bachelors degree in electromechanical engineering from the Universidad Privada Boliviana, Cochabamba, Bolivia in 2001. He obtained a MSc degree in electrical power engineering with business, in 2005, and a PhD in Electrical Engineering, in 2009, from the University of Strathclyde, Glasgow, UK. Since 2008 he is a research assistant at the Institute of Energy and Environment at the University of Strathclyde. His research interests include distributed energy resources, power systems planning and multi-objective optimization methods.

Graham Ault received his Bachelors in Electrical and Mechanical Engineering (1993) and PhD in Electrical Power Systems (2000) from the University of Strathclyde, Glasgow, UK. Since 1996 has been researching power system planning and operations issues relating to distributed energy resources in distribution systems. He is currently a reader for energy and environment at the University of Strathclyde.

Stuart Galloway obtained his PhD in mathematics from the University of Edinburgh, UK, in 1998. He is a lead investigator on several major collaborative research programmes in the UK (HiDef, EPSRC/EON Transition Pathways). His current research interests include the application of optimization techniques to power engineering problems, the modelling of novel electrical power systems, market simulation and energy economics. Currently, he is a Senior Lecturer in the Institute for Energy and Environment at the University of Strathclyde.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Renewable and Sustainable Energy Reviews
8 (2004) 365–381
ELSEVIER
RENEWABLE & SUSTAINABLE ENERGY REVIEWS
www.elsevier.com/locate/rser

Application of multi-criteria decision making to sustainable energy planning—A review

S.D. Pohekar *, M. Ramachandran

Birla Institute of Technology and Science (BITS), Pilani 333 031, India

Received 1 December 2003; accepted 19 December 2003

Abstract

Multi-Criteria Decision Making (MCDM) techniques are gaining popularity in sustainable energy management. The techniques provide solutions to the problems involving conflicting and multiple objectives. Several methods based on weighted averages, priority setting, outranking, fuzzy principles and their combinations are employed for energy planning decisions. A review of more than 90 published papers is presented here to analyze the applicability of various methods discussed. A classification on application areas and the year of application is presented to highlight the trends. It is observed that Analytical Hierarchy Process is the most popular technique followed by outranking techniques PROMETHEE and ELECTRE. Validation of results with multiple methods, development of interactive decision support systems and application of fuzzy methods to tackle uncertainties in the data is observed in the published literature.

Keywords: Multi-objective optimization; Multi-criteria decision making; Decision support systems; Sustainable energy planning

Contents

* Corresponding author. Tel.: +91-01596-245783; fax: +91-01596-244183.
E-mail address: sdp@bits-pilani.ac.in (S.D. Pohekar).

1364-0321/$ - see front matter
doi:10.1016/j.rser.2003.12.007

1. Introduction

Energy planning using multi-criteria analysis has attracted the attention of decision makers for a long time. The methods can provide solutions to increasing complex energy management problems. Traditional single criteria decision making is normally aimed at maximization of benefits with minimization of costs. These methods provide better understanding of inherent features of decision problem, promote the role of participants in decision making processes, facilitate compromise and collective decisions and provide a good platform to understanding the perception of models' and analysts' in a realistic scenario. The methods help to improve quality of decisions by making them more explicit, rational and efficient. Negotiating, quantifying and communicating the priorities are also facilitated with the use of these methods.

During the 1970s, energy planning efforts were directed primarily towards energy models aimed at exploring the energy–economy relationships established in the energy sector. The main objectives followed were to accurately estimate future energy demand. A single criteria approach aimed at identifying the most efficient supply options at a low cost was popular [1,2]. In the 1980s, growing environmental awareness has slightly modified the above decision framework [3]. The need to incorporate environmental and social considerations in energy planning resulted in the increasing use of multicriteria approaches.

Multi-objective linear programming is another planning methodology used for illustrating the trade-off between environmental and economic parameters and for assisting in the selection of a compromise solution [4,5]. It was popular in energy planning with conventional fuels in the 1970s. However, after the oil shock of 1973, a thought was given for energy conservation and energy substitution. Renewable energy sources are being promoted for a wide variety of applications worldwide. They are free from any form of pollution and are capable of substituting for conventional fuels in most of the applications. However, the contribution of these sources is very low, despite considerable technological development and their increasing competitiveness with respect to conventional fuels. This compels the planners and decision makers to identify the barriers for penetration and suggest interventions to overcome them. It is therefore felt that, along with the necessary policy measures, the wide exploitation of sustainable energy should be based on a completely different conception of energy planning procedure. The role of different actors in decision making thus becomes important. Methods of group decision are therefore of primary interest for the implementation of decision sciences in real-life problems.

Multi-criteria decision making (MCDM) methods deal with the process of making decisions in the presence of multiple objectives. A decision-maker is required to choose among quantifiable or non-quantifiable and multiple criteria. The objectives are usually conflicting and therefore, the solution is highly dependent on the preferences of the decision-maker and must be a compromise. In most of the cases, different groups of decision-makers are involved in the process. Each group brings along different criteria and points of view, which must be resolved within a framework of understanding and mutual compromise. Applications of MCDM include areas such as integrated manufacturing systems [6], evaluations of technology investment [7], water and agriculture management [8,9] in addition to energy planning [10–12].

2. Overview of multi-criteria decision making (MCDM) methods

Multi-Criteria Decision Making is a well known branch of decision making. It is a branch of a general class of operations research models which deal with decision problems under the presence of a number of decision criteria. This major class of models is very often called MCDM. This class is further divided into multi-objective decision making (MODM) and multi-attribute decision making (MADM) [13]. There are several methods in each of the above categories. Priority based, outranking, distance based and mixed methods are also applied to various problems. Each method has its own characteristics and the methods can also be classified as deterministic, stochastic and fuzzy methods. There may be combinations of the above methods. Depending upon the number of decision makers, the methods can be classified as single or group decision making methods. Decision making under uncertainty and decision support systems are also prominent decision making techniques [14].

These methodologies share common characteristics of conflict among criteria, incomparable units, and difficulties in selection of alternatives. In multiple objective decision making, the alternatives are not predetermined but instead a set of objective functions is optimized subject to a set of constraints. The most satisfactory and efficient solution is sought. In this identified efficient solution it is not possible to improve the performance of any objective without degrading the performance of at least one other objective. In multiple attribute decision making, a

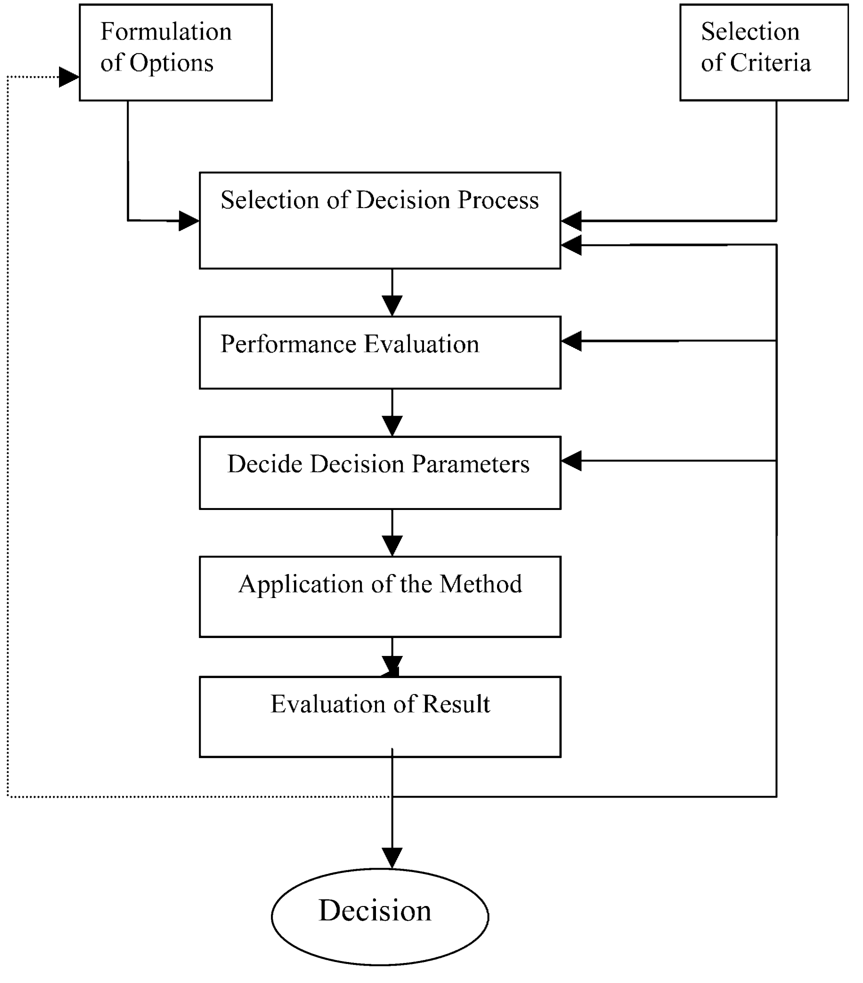


Fig. 1. Multicriteria decision process.

small number of alternatives are to be evaluated against a set of attributes which are often hard to quantify. The best alternative is usually selected by making comparisons between alternatives with respect to each attribute. The multicriteria decision process is as shown in Fig. 1. The different methods are described as follows.

2.1. Weighted sum method (WSM)

The WSM is the most commonly used approach, especially in single dimensional problems. If there are M alternatives and N criteria then the best alternative is the one that satisfies the following expression:

$$A^*_{WSM} = Max\sum_{i}^{j} a_{ij}w_j \quad \text{for} \quad i = 1, 2, 3, ...M \tag{1}$$

where $A^*{}_{WSM}$ is the WSM score of the best alternative, N is the number of decision criteria, a_{ij} is the actual value of the i^{th} alternative in terms of the j^{th} criterion, and w_j is the weight of importance of the j^{th} criterion. The total value of each alternative is equal to the sum of products. Difficulty with this method emerges when it is applied to multi-dimensional decision-making problems. In combining different dimensions, and consequently different units, the additive utility assumption is violated [15].

2.2. *Weighted product method (WPM)*

The WPM is very similar to WSM. The main difference is that instead of addition in the model there is multiplication. Each alternative is compared with the others by multiplying a number of ratios, one for each criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion. In general, in order to compare the alternatives A_K and A_L the following product is obtained:

$$R(A_K/A_L) = \sum_{j=1}^{N} (a_{kj}/a_{Lj})^{w_j} \tag{2}$$

where N is the number of criteria, a_{ij} is the actual value of the i^{th} alternative in terms of the j^{th} criterion, and w_j is the weight of importance of the j^{th} criterion. If $R\ (A_K/A_L)$ is greater than one, then alternative A_K is more desirable than alternative A_L (in the maximization case). The best alternative is the one that is better than or at least equal to all the other alternatives [16].

2.3. *Analytical hierarchy process (AHP)*

Analytical Hierarchy Process (AHP) is developed by Saaty [17,18]. The essence of the process is decomposition of a complex problem into a hierarchy with goal (objective) at the top of the hierarchy, criterions and sub-criterions at levels and sub-levels of the hierarchy, and decision alternatives at the bottom of the hierarchy. Elements at given hierarchy level are compared in pairs to assess their relative preference with respect to each of the elements at the next higher level. The verbal terms of the Saaty's fundamental scale of 1–9 is used to assess the intensity of preference between two elements. The value of 1 indicates equal importance, 3 moderately more, 5 strongly more, 7 very strongly and 9 indicates extremely more importance. The values of 2, 4, 6, and 8 are allotted to indicate compromise values of importance. Ratio scale and the use of verbal comparisons are used for weighting of quantifiable and non-quantifiable elements. The method computes and aggregates their eigenvectors until the composite final vector of weight coefficients for alternatives is obtained. The entries of final weight coefficients vector reflect the relative importance (value) of each alternative with respect to the goal stated at the top of hierarchy. A decision maker may use this vector due to his particular needs and interests. To elicit pair wise comparisons performed at a given level, a matrix A is created in turn by putting the result of pair wise comparison of element i with element j into the position a_{ji} as below.

$$A = \begin{bmatrix} a_{11} & a_{12} & . & a_{1n} \\ a_{21} & a_{21} & . & a_{2n} \\ a_{n1} & a_{n2} & . & a_{nn} \end{bmatrix} \tag{3}$$

After obtaining the weight vector, it is then multiplied with the weight coefficient of the element at a higher level (that was used as criterion for pair wise comparisons). The procedure is repeated upward for each level, until the top of the hierarchy is reached. The overall weight coefficient, with respect to goal for each decision alternative is then obtained. The alternative with the highest weight coefficient value should be taken as the best alternative. One of the major advantages of AHP is that it calculates the inconsistency index as a ratio of the decision maker's inconsistency and randomly generated index. This index is important for the decision maker to assure him that his judgments were consistent and that the final decision is made well. The inconsistency index should be lower than 0.10. Although a higher value of inconsistency index requires re-evaluation of pair wise comparisons, decisions obtained in certain cases could also be taken as the best alternative.

2.4. *Preference ranking organization method for enrichment evaluation (PROMETHEE)*

This method uses the outranking principle to rank the alternatives, combined with the ease of use and decreased complexity. It performs a pair-wise comparison of alternatives in order to rank them with respect to a number of criteria. Brans et al. [19] have offered six generalized criteria functions for reference namely, usual criterion, quasi criterion, criterion with linear preference, level criterion, criterion with linear preference and indifference area, and Gaussian criterion. The method uses preference function $P_j\ (a,\ b)$ which is a function of the difference d_j between two alternatives for any criterion j, i. e. $d_j = f(a,\ j) - f(b,\ j)$, where $f(a,\ j)$ and $f(b,\ j)$ are values of two alternatives a and b for criterion j. The indifference and preference thresholds q' and p' are also defined depending upon the type of criterion function. Two alternatives are indifferent for criterion j as long as d_j does not exceed the indifference threshold q'. If d_j becomes greater than p', there is a strict preference. Multi-criteria preference index, $\pi(a, b)$ a weighted average of the preference functions $P_j\ (a,\ b)$ for all the criteria is defined as

$$\pi(a,b) = \frac{\sum_{j=1}^{J} w_j P_j(a,b)}{\sum_{j=1}^{J} w_j} \tag{4}$$

$$\phi^{+}(a) = \sum_{A} \pi(a,b) \tag{5}$$

$$\phi^{-}(a) = \sum_{A} \pi(b,a) \tag{6}$$

$$\phi(a) = \phi^{+}(a) - \phi^{-}(a) \tag{7}$$

where w_j is the weight assigned to the criterion j; $\phi^{+}(a)$ is the outranking index of a in the alternative set A; $\phi^{-}(a)$ is the outranked index of a in the alternative set A; $\phi(a)$ is the net ranking of a in the alternative set A. The value having maximum $\phi(a)$ is considered as the best.

$$a \text{ outranks } b \text{ iff } \phi(a) > \phi(b), a \text{ is indifferent to } b \text{ iff } \phi(a) = \phi(b)$$

2.5. The elimination and choice translating reality (ELECTRE)

This method is capable of handling discrete criteria of both quantitative and qualitative in nature and provides complete ordering of the alternatives. The problem is to be so formulated that it chooses alternatives that are preferred over most of the criteria and that do not cause an unacceptable level of discontent for any of the criteria. The concordance, discordance indices and threshold values are used in this technique. Based on these indices, graphs for strong and weak relationships are developed. These graphs are used in an iterative procedure to obtain the ranking of alternatives [20]. This index is defined in the range (0–1), provides a judgment on degree of credibility of each outranking relation and represents a test to verify the performance of each alternative. The index of global concordance C_{ik} represents the amount of evidence to support the concordance among all criteria, under the hypothesis that A_i outranks A_k. It is defined as follows.

$$C_{ik} = \sum_{j=1}^{m} W_j c_j(A_i A_k) / \sum_{j=1}^{m} W_j \tag{8}$$

where W_j is the weight associated with j^{th} criteria. Finally, the ELECTRE method yields a whole system of binary outranking relations between the alternatives. Because the system is not necessarily complete, the ELECTRE method is sometimes unable to identify the preferred alternative. It only produces a core of leading alternatives. This method has a clearer view of alternatives by eliminating less favorable ones, especially convenient while encountering a few criteria with a large number of alternatives in a decision making problem [21].

2.6. The technique for order preference by similarity to ideal solutions (TOPSIS)

This method is developed by Huang and Yoon [22] as an alternative to ELECTRE. The basic concept of this method is that the selected alternative should have the shortest distance from the negative ideal solution in geometrical sense. The method assumes that each attribute has a monotonically increasing or decreasing utility. This makes it easy to locate the ideal and negative ideal solutions. Thus, the preference order of alternatives is yielded through comparing the Euclidean distances. A decision matrix of M alternatives and N criteria is formulated firstly. The normalized decision matrix and construction of the weighted decision matrix is carried out. This is followed by the ideal and negative-ideal solutions. For benefit criteria the decision maker wants to have maximum value among the alternatives and for cost criteria he wants minimum values amongst alternatives. This is followed by separation measure and calculating relative closeness to the ideal solution. The best alternative is one which has the shortest distance to the ideal solution and longest distance to negative ideal solution.

2.7. Compromise programming (CP)

Compromise Programming defines the best solution as the one in the set of efficient solutions whose point is the least distance from an ideal point [23].The aim is to obtain a solution that is as close as possible to ideal. The distance measure used in CP is the family of L_p-metrics and is given as

$$L_p(a) = \sum_{j=1}^{j} w_j^p \left| f_j^* - f(a) \right| / \left| M_j - m_j \right| \tag{9}$$

where L_p (a) is the L_p metric for alternative a, f (a) is the value of criterion j for alternative a, M_j is the maximum (ideal) value of criterion j in set A, m_j is the minimum (anti ideal) value of criterion j in set A, f_j^* is the ideal value of criterion j, w_j is the weight of the criterion j, p is the parameter reflecting the attitude of the decision maker with respect to compensation between deviations. For $p = 1$, all deviations from f_j* are taken into account in direct proportion to their magnitudes meaning that there is full (weighted) compensation between deviations

2.8. Multi-attribute utility theory (MAUT)

Multi-attribute Utility Theory takes into consideration the decision maker's preferences in the form of the utility function which is defined over a set of attributes The utility value can be determined by determination of single attribute utility functions followed by verification of preferential and utility independent conditions and derivation of multi-attribute utility functions. The utility functions can be either additively separable or multiplicatively separable with respect to single attribute utility. The multiplicative form of equation for then utility value is defined as follows.

$$1 + ku(x_1, x_2, \ldots x_n) = \prod_{j=1}^{n} (1 + kk_j u_j(x_j)) \tag{10}$$

Here j is the index of attribute, k is overall scaling constant (greater than or equal to -1), k_j is the scaling constant for attribute j, $u(.)$ is the overall utility function operator, $u_j(.)$ is the utility function operator for each attribute j [24].

3. Multi-criteria decision making applications in energy planning

The application areas of MCDM in energy planning presented in this section are renewable energy planning, energy resource allocation, building energy management, transportation energy management, planning for energy projects, electric utility planning and other miscellaneous areas. The comparison of MCDM methods applicable to energy planning are discussed in the literature. Hobbs and Meirer [25] compared the methods with respect to simplicity of applications and feasible expected outcomes, Huang and Poh [26] discussed the methods used in energy and environmental modeling under uncertainties, Lahdelma et al. [27] discussed these methods for environmental planning and management. The commonly applied MCDM methods out of the above are multi-objective optimization, AHP, PROMETHEE, ELECTRE, MAUT, fuzzy methods and decision support systems (DSS). More than one MCDM method is also applied in many application areas to validate the results [28–30].

A review of the published literature is presented here with a view to highlighting the applications areas and trends. A classification of published literature before 1990 and beyond 1990 is also presented to highlight suitability of the methods in changed global scenario. Six generalized application areas and a miscellaneous area presented here have common features of minimization of cost benefit ratios, high degrees of uncertainties in formulating the problems, incommensurable units and the need to handle socio-economic aspects in planning. Renewable energy planning and energy resource allocation refers to compilation of feasible energy plan and dissemination of various renewable energy options. The key factors applicable are investment planning, energy capacity expansion planning and evaluation of alternative energies. Building energy management refers to design, selection, installation and building energy management options in a multi-criteria environment. The application normally deals with quantitative issues. Transportation system applications include evaluation of alternative strategies for pollution control, elimination of old polluting vehicles, choosing between private and public transport etc. The key features of transportation applications are of a high concern for socio-economic reasons. Project planning refers to site selection, technology selection and decision support in renewable energy harnessing projects. The objectives are arriving at a Pareto optimal solution for technology selection, sizing, execution, investment planning. Optimal electrical dispatch scheduling, deciding power generation mix, optimum electricity supply planning are the applications of electric utility planning using MCDM. Miscellaneous applications include desalination plant selection, solid waste management.

It can be observed from the surveyed literature that AHP is the most popular method for prioritizing the alternatives, followed by PROMETHEE and ELECTRE. Multi-objective programming is also very widely used to formulate alternative plans. Fuzzy MCDM methods are also adopted for considering the uncertainties in energy planning. Decision support systems are becoming popular in energy planning and resource allocation with the advent of the latest computational aids.

3.1. Multi-objective optimization

This method is very widely used in energy resource allocation, energy planning and electric utility applications. Maximization of cost benefit ratio to arrive at optimum resource allocation in rural areas [31], national level energy planning [32] are amongst a few applications. The application areas have common features of higher investment costs, higher project durations, conflicting objectives and uncertainty. Energy security and social benefits are prominent objectives in energy planning with these methods. These techniques are also used for sustainability evaluation of power plants [33], deciding optimum mix of renewable energy technologies at various sectors [34–37]. Renewable energy planning with energy environment linkages [38], economic constraints, technology limitations etc. are the main features of applications surveyed . Applications to various national level issues [39–42] and household energy issues [43,44] are also among the prominent application areas. Multi-objective optimization also finds applications in building energy management [45]. The issues identified are building material design [46], building performance design [47,48], building arrangement design [49], and building shape design [50,51]. Regional energy supply optimization [52,53], desalination power plant selection [54,55], electricity distribution planning using fuzzy approaches [56,57] are also worth mentioning. Genetic algorithms are also applied to electric utility planning and building energy management problems [46]. An analysis of utilizing multi-objective optimization reveals that the methods are being used for a wide variety of applications after 1990. These may be due to the advent of sophisticated computational aids available and increased need for larger socio-economic considerations in energy planning.

3.2. Decision Support System (DSS)

These are sophisticated, interactive and computer aided techniques for aiding the decisions [58]. These can support complex problems that would be otherwise difficult to handle. Knowledge based DSS can support the decision makers in selecting criteria, alternatives and trade-offs, thus making the energy planning simple. The identified DSS use MCDM methods for arriving at interim results. The applications of DSS in energy planning developed are solid waste management [59], transportation energy management [60], electricity production alternatives [61], building energy management [62] and renewable energy project planning [63].

3.3. Multi-criteria decision making methods

The Multi Attribute Utility Theory is developed to help decision makers assign utility values to outcomes by evaluating these in terms of multiple attributes and combining individual assignments to obtain overall utility values. It is observed that MAUT is not very extensively used in energy planning. This may be due to requirements of interactive decision environment required in formulating utility functions, complexity of computing scaling constants using the algorithm [64]. Selecting portfolios for solar energy projects [65], energy policy making [66], environmental impact assessment [67] and electric power system expansion planning [68] are the applications identified in the literature. A few numbers of studies are observed using this method after 1990.

The outranking methods belonging to ELECTRE family are popularly used in energy planning. These methods are also used in renewable energy DSS after 1990 [62,69,70].Other common application areas include electric utility planning, building energy management and project planning. These methods are also applied to selection of thermal power plant location by eliminating certain sites [71], renewable energy plant selection [72,73], selecting pollution control technologies [74] and transportation energy planning [75,76]. Though various versions of ELECTRE are developed ELECTRE III is found to be widely used in energy planning applications.

Outranking methods belonging to PROMETHEE category are also extensively used in energy planning. These methods provide a scientific basis to arrive at multi-criteria preference index by calculating the strengths and weaknesses of alternative actions. This method is used in energy project planning and applications to geothermal site selection [77,78] and small hydro site selections [79]. Other application areas are impact analysis of energy alternatives [80,81], old vehicle elimination [75,76] and building product designs [82]. Different versions of PROMETHEE are in use and PROMETHEE II has been extensively used after 1990.

Analytical Hierarchy Process is very widely used in energy planning. This may be due to provisions of converting a complex problem into a simple hierarchy, flexibility, intuitive appeal, its ability to mix qualitative as well as quantitative criteria in the same decision framework [83] and use of computational aids leading to successful decisions in many domains [84]. Though a there are number of shortcomings [85], the method is popularly used in renewable energy planning [86–90], energy resource allocation [91], transportation energy planning [92], project planning [93] and electric utility planning [94–96]. The applications surveyed have the main objectives of priority setting and have features such as less number of criteria, interaction with decision makers etc. The correctness of AHP has been established by comparing it with other MCDM methods. The method is used with modifications during post 1990.

In addition to the above discussed methods, preference desegregation method is also used for energy analysis and policy making studies [97]. Fuzzy set programming is used for a variety of applications after 1990. A few of the application areas surveyed are solar system evaluation [98,99], power systems [100–103] and wind site selection [104].

Table 1
Classification of MCDM methods by application areas

Applications	Multi-objective	Methods					Total number
		MAUT	AHP	PROMETHEE	ELECTRE	Others	
Renewable energy planning	[2,33–38]	[65,66]	[87,90,95,96]	[70,75,79]	[69,70]	[62,97–99,104]	***22***
Energy resource allocation	[43,88,89,95]		[88–91]			[1–32,39]	***10***
Building energy management	[46,47,49–51]			[82]	[45,62]		***8***
Transportation energy systems			[76,92]	[76]	[75]	[60]	***5***
Project planning			[93]	[77,78]	[71–74]		***7***
Electric utility planning	[4,5,53,101]	[67,68]	[94–96]			[59,98,100,102]	***12***
Others	[56]				[74]		***2i***
Total number	***22***	***4***	***14***	***7***	***10***	***13***	

Numbers in square brackets refer to reference numbers.

Table 2
Classification of MCDM methods by year of publication

Year of Publication	Upto 1990	Beyond 1990	Total Number
Multi-Objective	[2,4,5,40,47,53,89]	[25,33–38,43,46,49–51,56,95,101]	***22***
MAUT	[65,66]	[67,68]	***4***
AHP	[75,76,94]	[87–93,95,96]	***14***
PROMETHEE	[75,76,79]	[70,77,78,82]	***7***
ELECTRE	[71,79]	[45,62,69,70,72–75]	***10***
Others	[31,32]	[25,39,59–60,62,97]–100, 102,104]	***13***
Total Number	***19***	***48***	

Numbers in square brackets refer to reference numbers.

It can be observed from the studies (Tables 1 and 2) that multi-objective optimization accounts for 29% of the identified studies, followed by AHP (20%), ELECTRE (15%), PROMETHEE (10%). Miscellaneous methods including DSS and fuzzy methods account for a share of 20% in energy decision making applications. The number of MCDM applications surveyed upto 1990 is 29% and beyond 1990 is 69% approximately. The methods are observed to be highly popular for renewable energy planning (34%), followed by electric utility planning (19%), energy resource allocation (15%), building energy management (13%) and project planning (12%).

4. Conclusion

A feview of the published literature on sustainable energy planning presented here indicates greater applicability of MCDM methods in changed socio-economic scenario. The methods have been very widely used to take care of multiple, conflicting criteria to arrive at better solutions Increasing popularity and applicability of these methods beyond 1990 indicate a paradigm shift in energy planning approaches. The methods are observed to be most popular in renewable energy planning followed by energy resource allocation. It is observed that Analytical Hierarchy Process is the most popular technique followed by outranking techniques PROMETHEE and ELECTRE. Validation of results with multiple methods, development of interactive decision support systems and application of fuzzy methods to tackle uncertainties in the data is observed in the published literature.

Acknowledgements

The authors gratefully acknowledge the contributions of Dr. K. S. Raju, Group Leader–Civil Engineering of BITS, Pilani for his constructive criticism and support.

References

[1] Samouilidis J, Mitropoulos C. Energy economy models—a survey. European Journal of Operations Research 1982;25:200–15.

[2] Meirer P, Mubayi V. Modeling energy-economic interactions in developing countries-a linear programming approach. European Journal of Operations Research 1983;13:41–59.

[3] Nijcamp P, Volwahsen A. New directions in integrated energy planning. Energy Policy 1990;18(8):764–73.
[4] Kavrakoglu I. Multi-objective strategies in power system planning. European Journal of Operations Research 1983;12:159–70.
[5] Schulz V, Stehfest H. Regional energy supply optimization with multiple objectives. European Journal of Operations Research 1984;17:302–12.
[6] Putrus P. Accounting for intangibles in integrated manufacturing-non-financial justification based on analytical hierarchy process. Information Strategy 1990;6:25–30.
[7] Boucher TO, McStravic EL. Multi-attribute evaluation within a present value framework and its relation to analytic hierarchy process. The Engineering Economist 1991;37:55–71.
[8] Ozelkan EC, Duckstein L. Analyzing water resources alternatives and handling criteria by multicriterion decision techniques. Journal of Environmental Management 1996;48:69–96.
[9] Raju KS, Pillai CRS. Multicriterion decision making in performance evaluation of irrigation projects. European Journal of Operational Research 1999;112(3):479–88.
[10] Afgan NH, Carvalho MG. Sustainable assessment method for energy systems. Boston: Kluwer Academic Publishers; 2000.
[11] Afgan NH, Gobaisi D, Carvalho MG, Cumo M. Sustainable energy management. Renewable and Sustainable Energy Reviews 1998;2:235–86.
[12] Afgan NH, Carvalho MG, Hovanov NV. Sustainability assessment of renewable energy systems. Energy Policy 2000;28:603–12.
[13] Climaco J, editor. Multicriteria analysis. New York: Springer-Verlag; 1997.
[14] Gal T, Hanne T, editors. Multicriteria decision making: Advances in MCDM models, algorithms, theory, and applications. New York: Kluwer Academic Publishers; 1999.
[15] Solnes J. Environmental quality indexing of large industrial development alternatives using AHP. Environmental Impact Assessment Review 2003;23(3):283–303.
[16] Chang YH, Yeh CH. Evaluating airline competitiveness using multi-attribute decision making. Omega 2001;29(5):405–15.
[17] Saaty TL. The analytic hierarchy process. New York: McGraw-Hill; 1980.
[18] Saaty TL. Decision making for leaders. Pittsburgh: RWS Publications; 1992.
[19] Brans JP, Vincke Ph, Mareschal B. How to select and how to rank projects: the PROMETHEE method. European Journal of Operations Research 1986;24:228–38.
[20] Roy B. Métodologie multicritère d'aide la décision. Collection Gestion, Paris: Economica; 1985.
[21] Goicoechea A, Hansen D, Duckstein L. Introduction to multi objective analysis with engineering and business application. Wiley: New York; 1982.
[22] Huang CL, Yoon K. Multi attribute decision making: methods and applications. New York: Springer-Verlag; 1981.
[23] Zeleny M. Multiple criteria decision making. New York: McGraw-Hill; 1982.
[24] Keeny RL, Raiffa H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley; 1976.
[25] Hobbs BF, Meirer PM. Multicrerion methods for resource planning: an experimental comparison. IEEE Transactions on Power Systems 1994;9(4):1811–7.
[26] Huang JP, Poh KL, Ang BW. Decision analysis in energy and environmental modeling. Energy 1995;20(9):843–55.
[27] Lahdelma R, Salminen P, Hokkanen J. Using multicriteria methods in environmental planning and management. Environmental Management 2000;26(6):595–605.
[28] Karni R, Feigin P, Breiner A. Multicriterion issues in energy policy making. European Journal of Operational Research 1992;56:30–40.
[29] Mirasgedis S, Diakoulaki D. Multicriteria analysis vs. externalities assessment for the comparative evaluation of electricity generation systems. European Journal of Operational Research.; 1997;102(2):364–79.
[30] Salminen P, Hokkanen J, Lahdelma R. Comparing multicriteria methods in the context of environmental problems. European Journal of Operational Research 1998;104:485–96.

[31] Christensen JM, Vidal R. Project evaluation for energy supply in rural areas of developing countries. European Journal of Operations Research 1990;49:230–46.
[32] Lootsma FA, Boonekamp PGM, Cooke RM, Van Oostvoorn F. Choice of a long term strategy for the national electricity supply via scenario analysis and multi-criteria analysis. European Journal of Operational Research 1990;48:189–203.
[33] Afgan NH, Carvalho MG. Multi-criteria assessment of new and renewable energy power plants. Energy 2002;27:739–55.
[34] Iniyan S, Sumanthy K. An optimal renewable energy model for various end-uses. Energy–The International Journal 2000;25:563–75.
[35] Suganthi L, Williams A. Renewable energy in India—a modeling study for 2020–2021. Energy Policy 2000;28:1095–109.
[36] Sinha CS, Kandpal TC. Optimal mix of technologies in rural area: the cooking sector. International Journal of Energy Research 1991;15:85–100.
[37] Cormico C, Dicorato M, Minoia A, Trovato M. A regional energy planning methodology including renewable energy sources environmental constraints. Renewable and Sustainable Energy Reviews 2003;7:99–130.
[38] Borges AR, Antunes CH. A fuzzy multiple objective decision support model for energy-economy planning. European Journal of Operational Research 2003;145(2):304–16.
[39] Chedid R, Mezher T, Jarrouche C. A fuzzy programming approach to energy resource allocation. International Journal of Energy Research 1999;23:303–17.
[40] Psarras J, Capros P, Samoulidis JE. Multi-criteria analysis using a large scale energy supply LP model. European Journal of Operations Research 1990;44:383–94.
[41] Ramanathan R. A multi-criteria methodology to the global negotiations on climate change. IEEE Transactions on Systems, Man and Cybernetics—Part C: Applications and Reviews 1998;28:541–8.
[42] Ramanathan R. Selection of appropriate greenhouse mitigation options. Global Environmental Change 1999;9:203–10.
[43] Ramanathan R, Ganesh LS. Multi-objective analysis of cooking-energy alternatives. Energy 1994;19(4):469–78.
[44] Ramanathan R, Ganesh LS. Energy alternatives for lighting in households: an evaluation using an integrated goal programming-AHP model. Energy 1995;20(1):63–72.
[45] Blondeau P, Allard F. Multicriteria analysis of ventilation in summer period. Building and Environment 2002;37(2):165–76.
[46] Wright JA, Loosemore HA, Fermani R. Optimization of building thermal design and control by multi-criterion genetic algorithm. Energy and Buildings 2002;34:959–72.
[47] D'Cruz NA, Radford AD. Multi-criteria model for building performance and design. Building and Environment 1987;22(3):167–79.
[48] Flourentzou F, Roulet CA. Elaboration of retrofit scenarios. Energy and Buildings 2002;34(2):185–92.
[49] Klemm K, Marks W, Klemm AJ. Multicriteria optimization of the building arrangement with application of numerical simulation. Building and Environment 2000;35(6):537–44.
[50] Marks WM. ulticriteria optimization of shape of energy-saving buildings. Building and Environment 1997;32(4):331–9.
[51] Jedrzejuk H, Marks WO. ptimization of shape and functional structure of buildings as well as heat source utilization-partial problems solution. Building and Environment 2002;379(11):1037–43.
[52] Schulz V, Stehfest H. Regional energy supply optimization with multiple objectives. European Journal of Operational Research 1983;17:302–12.
[53] Amagai H, Leung P. Multiple criteria analysis for Japan's electric power generation mix. Energy Systems and Policy 1989;13:219–36.
[54] Afgan NH, Darwish M, Carvalho MG. Sustainability assessment of desalination plants for water production. Desalination 1999;124:19–31.
[55] Akash BA, Al-Jayyousi O, Mohsen M. Multi-criteria analysis of non-conventional energy technologies for water desalination in Jordan. Desalination 1997;114:1–12.

[56] Matos MA. Fuzzy filtering method applied to power distribution planning. Fuzzy Sets and Systems 1999;102(1):53–8.
[57] Levitin G, Tov SMT, Elmakis D. Genetic algorithm for open-loop distribution system design. Electric Power Systems Research 1995;32(2):81–7.
[58] Turban E. Decision support and expert system. New York: Prentice Hall; 1995.
[59] Fiorucci P, Minciardi R. Solid waste management in urban areas: Development and application of a decision support system. Resources Conservation and Recycling 2003;37(4):301–28.
[60] Brand C, Mattarelli M, Moon D, Calvo RW. STEEDS: A strategic transport-energy-environment decision support. European Journal of Operational Research 2002;139(2):416–35.
[61] Gandibleux X. Interactive multicriteria procedure exploiting a knowledge-based module to select electricity production alternatives: The CASTART system. European Journal of Operational Research 1999;113(2):355–73.
[62] Roulet CA. ORME: A multi-criteria rating methodology for buildings. Building and Environment 2002;37:579–86.
[63] Georgopoulou E, Sarafidis F, Mirasgedis S, Zaimi S, Lalas DP. A multiple criteria decision-aid approach in defining national priorities for greenhouse gases emissions reduction in the energy sector. European Journal of Operational Research 2003;146(1):199–215.
[64] Dyer JS. Remarks on analytical hierarchy process. Management Science 1990;36:249–58.
[65] Golabi K, Kirkwood CW, Sicherman A. Selecting a portfolio of solar energy projects using multiattribute preference theory. Management Science 1981;22(2):174–89.
[66] Jones M, Hope C, Hughes R. Amulti-attribute value model for the study of UK energy policy. Journal of Operations Research Society 1990;41(10):919–29.
[67] McDaniels TL. A multiattribute index for evaluating environmental impact of electric utilities. Journal of Environmental Management 1996;46:57–66.
[68] Voropai NL, Ivanova EY. Multicriteria decision analysis technique in electric power system expansion planning. Electrical Power and Energy Systems 2002;24:71–8.
[69] Beccali M, Cellura M, Mistretta M. Decision-making in energy planning-application of the ELECTRE method at regional level for the diffusion of renewable energy technology. Renewable Energy 2003;28(13):2063–87.
[70] Georgopoulou E, Sarafidis Y, Diakoulaki D. Design and implementation of a group DSS for sustaining renewable energies exploitation. European Journal of Operational Research 1998;109:483–500.
[71] Barda OH, Dupluis J, Lencoini P. Multicriteria location of thermal power plants. European Journal of Operations Research 1990;45:332–46.
[72] Haralambopoulos DA, Polatidis H. Renewable energy projects: structuring a multi-criteria group decision-making framework. Renewable Energy 2003;28:961–73.
[73] Georgopoulou E, Lalas D, Papagiannakis L. A multicriteria decision aid approach for energy planning problems: the case of renewable energy option. European Journal of Operational Research 1997;3:38–54.
[74] Hokkanen J, Salminen P. Choosing a solid waste management system using multicriteria decision analysis. European Journal of Operational Research 1997;98:19–36.
[75] Tzeng GH, Shiau TA. Energy conservation strategies in urban transportation: application of multi-criteria decision making. Energy Systems and Policy 1987;11:1–19.
[76] Tzeng GH, Tsaur SH. Application of multicriteria decision making to old vehicle elimination in Taiwan. In Proceeding of International Conference on MCDM Taipei. Taiwan, 1990: 4:268–283.
[77] Goumas MG, Lygerou V, Papayannakis LE. Computational methods for planning and evaluating geothermal energy projects. Energy Policy 1999;27:147–54.
[78] Goumas MG, Lygerou V. An extension of the PROMETHEE method for decision-making in fuzzy environment: ranking of alternative energy exploitation projects. European Journal of Operational Research 2000;123:606–13.
[79] Mladineo N, Margeta J, Brans JP, Mareschal B. Multicriteria ranking of alternative locations for small scale hydro plants. European Journal of Operational Research 1987;31:215–22.
[80] Siskos J, Hubert PH. Multi-criteria analysis of the impacts of energy alternatives: A survey and a comparative approach. European Journal of Operational Research 1983;13:278–99.

[81] Tzeng GH, Shiau TA, Lin CY. Application of multi-criteria decision making to the evaluation of new energy system development in Taiwan. Energy 1992;17(10):983–92.

[82] Teno TFL, Marseschal B. An interval version of PROMETHEE for comparison of building products' design with ill defined data on environmental quality. European Journal of Operational Research 1998;109:522–9.

[83] Wedley WC. Combining qualitative and quantitative factors -an analytical hierarchy approach. Socio-economic Planning Sciences 1990;24:57–64.

[84] Golden BL, Wasil EA, Harker PT. Analytical hierarchy process. Berlin: Springer-Verlag; 1989.

[85] Dyer JS, Fishburn RE, Steuer RE, Wallenenius J, Zionts S. Multiple criteria decision making, multiattribute utility theory: the next ten years. Management Science 1992;38:645–54.

[86] Mohsen MS, Akash BA. Evaluation of domestic solar water heating system in Jordan using analytical hierarchy process. Energy Conversion and Management 1997;38(18):1815–22.

[87] Wang X, Feng Z. Sustainable development of rural energy and its appraisal system in China. Renewable and Sustainable Energy Reviews 2002;6:395–404.

[88] Ramanathan R, Ganesh LS. Energy resource allocation incorporating quantitative and qualitative criteria: an integrated model using goal programming and AHP. Socio Economic Planning Sciences 1995;29:197–218.

[89] Ramanathan R, Ganesh LS. A multi-objective programming approach to energy resource allocation problems. International Journal of Energy Research 1990;17:105–19.

[90] Elkarni F, Mustafa I. Increasing the utilization of solar energy technologies (SET) in Jordan: analytic hierarchy process. Energy Policy 1998;21:978–84.

[91] Hobbs BF, Horn GTF. Building public confidence in energy planning: a multi-method MCDM approach to demand side planning at BC gas. Energy Policy 1997;25(3):357–75.

[92] Yedla S, Shreshtha RM. Multicriteria approach for selection of alternative option for environmentally sustainable transport system in Delhi. Transportation Research part A 2003;37:717–29.

[93] Chattopadhyay D, Ramanathan R. A new approach to evaluate generation capacity bids. IEEE Transactions on Power Systems 1998;13:1232–7.

[94] Rahman S, Frair LC. A hierarchical approach to electric utility planning. International Journal of Energy Research 1984;8:185–96.

[95] Zong W, Zhihong W. Mitigation assessment results and priorities for China's energy sector. Applied Energy 1997;56(3/4):237–51.

[96] Akash BA, Mamlook R, Mohsen MS. Multi-criteria selection of electric power plants using analytical hierarchy process. Electric Power Systems Research 1999;52:29–35.

[97] Diakoulaki D, Zopounidis DC, Doumpos M. The use of a preference desegregation method in energy analysis and policy making. Energy 1999;24(2):157–66.

[98] Mamlook R, Akash BA, Mohsen MS. A neuro-fuzzy program approach for evaluating electric power generation systems. Energy 2001;26:619–32.

[99] Mamlook R, Akash BA, Nijmeh S. Fuzzy set programming to perform evaluation of solar system in Jordan. Energy Conversion and Management 2001;42:1717–26.

[100] Srinivasan D, Liew AC, Chang CS. Applications of fuzzy systems in power systems. Electric Power Systems Research 1995;35(1):39–43.

[101] Brar YS, Dhillon JS, Kothari DP. Multiobjective load dispatch by fuzzy logic based searching weightage pattern. Electric Power Systems Research 2002;63(2):149–60.

[102] Mills D, Vlacic L, Lowe I. Improving electricity planning—use of a multicriteria decision making model. International Transactions in Operational Research 1996;3(3/4):293–304.

[103] Beccali M, Cellura M, Ardente DD. Decision making in energy planning: the ELECTRE multicriteria analysis approach compared to a fuzzy-sets methodology. Energy Conversion and Management 1998;39(16-18):1869–81.

[104] Skikos GD, Machias AV. Fuzzy multi criteria decision making for evaluation of wind sites. Wind Engineering 1992;6(4):213–28.

Available online at www.sciencedirect.com
ScienceDirect
Renewable and Sustainable Energy Reviews
11 (2007) 1584–1595

RENEWABLE
& SUSTAINABLE
ENERGY REVIEWS

www.elsevier.com/locate/rser

Use of multicriteria decision analysis methods for energy planning problems

Espen Løken*

Department of Electrical Power Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway

Received 28 October 2005; accepted 23 November 2005

Abstract

Most decision making requires the consideration of several conflicting objectives. The term multiple criteria decision analysis (MCDA) describes various methods developed for aiding decision makers in reaching better decisions. Energy planning problems are complex problems with multiple decision makers and multiple criteria. Therefore, these problems are quite suited to the use of MCDA. A multitude of MCDA methods exists. These methods can be divided in three main groups; value measurement models, goal, aspiration and reference level models, and outranking models. Methods from all of these groups have been applied to energy planning problems, particularly in the evaluation of alternative electricity supply strategies. Each of the methods has its advantages and drawbacks. However, we cannot conclude that one method generally is better suited than the others for energy planning problems. A good alternative might be to apply more than one method, either in combination to make use of the strengths of both methods, or in parallel to get a broader decision basis for the decision maker. Until now, studies of MCDA in energy planning have most often considered energy networks with only one energy carrier. More advanced energy systems with multiple energy carriers have been neglected, even though this field ought to be suitable for use of MCDA due to its high complexity, many decision makers and many conflicting criteria.

Keywords: Energy planning; Multiple criteria decision analysis (MCDA)

*Tel.: +47 73 59 44 44; fax: +47 73 59 42 79.
E-mail address: espen.loken@elkraft.ntnu.no.

1364-0321/$ - see front matter
doi:10.1016/j.rser.2005.11.005

Contents

1. Introduction

When making decisions, decision makers (DMs) always try to choose the optimal solution. Unfortunately, a true optimal solution only exists if you are considering a single criterion. In most real decision situations, basing a decision solely on one criterion is insufficient. Probably several conflicting and often non-commensurable objectives should be considered. Because of this, it is impossible to find a genuine optimal solution, a solution that is optimal for all DMs under each of the criteria considered [1].

Multiple criteria decision making (MCDM) is a generic term for all methods that exist for helping people making decisions according to their preferences, in cases where there is more than one conflicting criterion [2]. Using MCDM can be said to be a way of dealing with complex problems by breaking the problems into smaller pieces. After weighing some considerations and making judgments about smaller components, the pieces are reassembled to present an overall picture to the DMs [3].

Another term which is often used is multiple criteria decision analysis (or aid) (MCDA). The reason for using 'decision analysis' or 'decision aid' instead of 'decision making' is to emphasize that the methods should aid DMs in making better decisions. The methods themselves cannot make the actual decisions. The aim of MCDA methods is to help DMs organize and synthesize the information they have collected, so that they feel comfortable with and confident in their decisions. By using MCDA methods, DMs should feel that all important criteria have been properly accounted for. This should help to reduce the post-decision regret [4]. Ideally, the MCDA methods will help the DMs to understand and identify the fundamental criteria in the decision problem, and avoid making important decisions out of habit.

Energy planning is a field that is quite suitable for MCDA methods because it is subject to many sources of uncertainty, long time frames and capital-intensive investments [5], along with featuring multiple DMs and many conflicting criteria. The complexity in the planning of local energy systems is discussed in more detail in Ref. [6]. Before the 1970s, little effort was made in the formal planning of energy systems. The oil crisis in the 1970s resulted in more emphasis being placed on identifying efficient supply options. However, most studies were based only on cost minimization [7]. In the 1980s, the public started to become more aware of environmental issues. Consequently, it was necessary to start incorporating environmental considerations in energy planning [8]. This led to a more comprehensive use of MCDA methods. Subsequently, it has become common to include other criteria in the studies, such as reliability, land use, aesthetics and human health concerns [9].

The purpose of this review article is to provide an overview of some of the most important MCDA methods that have been proposed over the years. I will present examples of how different methods have been applied for energy planning purposes. The examples have been chosen to give a broad overview of all the methods that have been used for energy planning. The main advantages of the different methods, as well as the difficulties that they may be subject to, will also be evaluated. In the end, I will argue that MCDA can be a very useful tool for the planning of local energy systems with multiple energy carriers and multiple energy resources, even though no MCDA studies have examined this type of problem until now.

2. Multicriteria decision analysis methods

Over the years, hundreds of MCDA methods have been proposed [10]. The methods differ in many areas—theoretical background, type of questions asked and type of results given [11]. Some methods have been created particularly for one specific problem, and are not useful for other problems. Other methods are more universal, and many of them have attained popularity in various areas. The main idea for all the methods is to create a more formalized and better-informed decision making process.

I will start this section by providing guidelines on the selection of the most appropriate method for a given problem. Thereafter, I will present some of the most well known MCDA methods.

2.1. Choosing an MCDA method

When choosing an MCDA method, there are many criteria to consider. The most important is to find a method that measures what it is supposed to measure (validity). Different methods are likely to give different results, so a method that reflects the user's 'true values' in the best possible way should be chosen. In addition, the method must provide the DMs with all the information they need, and the method must be compatible with the accessible data (appropriateness). The method must also be easy to use and easy to understand [10]. If the DMs do not understand what is happening inside the methodology, they perceive the methodology like a black box. The result may be that the DMs do not trust in the recommendations from the method. In that case, it is meaningless to spend time applying this method.

2.2. Classifying MCDA methods

There are many possible ways to classify the existing MCDA methods. In this review, I have chosen the same classification as Belton and Stewart used in their book [4]. According to Ref. [4], there are three broad categories (or schools of thought):

- Value measurement models.
- Goal, aspiration and reference level models.
- Outranking models (the French school).

In the next sections, I will describe the main characteristics of the three categories, and I will present some of the most important methods that belong to each group. For more detailed descriptions of the methods, I recommend Ref. [4], or specific literature for each method written by the developers of the various methods.

2.2.1. Value measurement models

When using value measurement methods, a numerical score (or value) V is assigned to each alternative. These scores produce a preference order for the alternatives such that a is preferred to $b\,(a \succ b)$ if and only if $V(a) > V(b)$. When using this approach, the various criteria are given weights w that represent their partial contribution to the overall score, based on how important this criterion is for the DM(s). Ideally, the weights should indicate how much the DM is willing to accept in the tradeoff between two criteria [4,12,13].

The most commonly used approach is an additive value function (multiattribute value theory (MAVT)):

$$V(a) = \sum_{i=1}^{m} w_i\, v_i(a), \tag{1}$$

where $v_i(a)$ is a partial value function reflecting alternative a's performance on criterion i. The partial value function must be normalized to some convenient scale (e.g. 0–100). Using Eq. (1), a total value score $V(a)$ is found for each alternative a. The alternative with the highest value score is preferred. MAVT is a pretty simple and user-friendly approach where the DM—in cooperation with the analyst—only needs to specify value functions and define weights for the criteria to get very useful help with his decision [4].

The multiattribute utility theory (MAUT) first proposed in detail by Keeney and Raiffa [14] can be said to be an extension of MAVT. MAUT is a more rigorous methodology for how to incorporate risk preferences and uncertainty into multicriteria decision support methods. When using this approach, multiattribute utility functions $U(a)$—where the risk preferences are directly reflected in the values—must be established instead of value functions [4,14].

The analytical hierarchy process (AHP) developed by Saaty [15] has many similarities to the multiattribute value function approach. Belton and Stewart [4] described AHP "as an alternative means of eliciting a value function". However, they pointed out that the two methods rest on different assumptions on value measurements, and that AHP is developed independently of other decision theories. Of these reasons, many of the proponents of AHP claim that AHP is not a value function method [4]. However, both MAUT and AHP present their results as cardinal rankings, which mean that each alternative is given a numerical desirability score. Consequently, the results from the two methods are directly comparable.

The major characteristic of the AHP method is the use of pair-wise comparisons, which are used both to compare the alternatives with respect to the various criteria and to

Table 1
Fundamental scale

1	Equally preferred
3	Weak preference
5	Strong preference
7	Very strong or demonstrated preference
9	Extreme importance
2, 4, 6, 8	Intermediate values

estimate criteria weights [4,13]. In the pair-wise comparisons, a special ratio scale (Table 1) constructed by Saaty [15,16] is used:

The results from all the comparisons are put into matrices. From these matrixes, an overall ranking of the alternatives can be aggregated. The alternative with the highest overall ranking is preferred to the others [13]. The mathematical procedure that is used to calculate the overall rankings is quite complex (more details can be found for instance in Ref. [15]), and the procedure is, therefore, normally performed with specially designed computer programs.

2.2.2. Goal, aspiration and reference level models

Alternatives to value measurement methods are goal programming (GP), the aspiration level and the reference level methods. Often GP is used as a common abbreviation for all these approaches, and this simplification is used also in this article. When using GP approaches, we try to determine the alternatives that in some sense are the closest to achieve a determined goal or aspiration level [4]. Often the GP approach is used as a first phase of a multicriteria process where there are many alternatives. In that case, GP is used to filter out the most unsuitable alternatives in an efficient way.

Mathematically, we can say that the idea in the GP methods is to solve the inequalities $z_i + \delta_i \geqslant g_i$, where z_i is the attribute values, δ_i is the non-negative deviational variables and g_i is the goals (a desirable level of performance) for each criterion i. The aim is to find a feasible solution that minimizes the vector of deviational variables. If it is possible to find a solution where $\delta_i = 0$ for all i, this will be the recommended solution. In most cases, this is not the case, and another solution must be found. The simplest method for this purpose is to minimize the weighted sum of deviations $\sum_{i=1}^{m} w_i\,\delta_i$ [4], where w_i is the importance weight and δ_i is the deviation of criterion i.

A more advanced possibility is to use the so-called Tchebycheff norm, where the aim is to minimize the maximum weighted deviation, i.e. to minimize $\max_i\{w_i\,\delta_i\}$. It means that the focus is always placed on the relatively worst performance area [4].

GP methods are well-suited for the use of interactivity. There are many possible methods. I will only give a brief explanation of some of them. A well-used interactive method is the method of displaced ideals, as proposed by Zeleny [17]. The concept in this method is to minimize

$$\left[\sum_{i=1}^{m}[w_i\,\delta_i]^p\right]^{\frac{1}{p}}, \qquad (2)$$

for different values of p. p is a constant that decides the penalty for greater deviations compared to smaller deviations. After the DM has been presented for solutions for various values of p, he is supposed to eliminate clearly undesirable solutions. This is called displacement of ideals. After the displacement, the procedure will be repeated until the difference between the ideal solution[1] and compromise solution are acceptably small [4,11].

In the STEM approach (also called the step method) proposed by Benayoun [18], the ideal solution is used as a goal for each criterion, and deviations are found by the Tchebycheff norm explained above. The weights for the criteria are not specified by the DM, but are calculated by the relative range of values available on each criterion. Consequently, the weights are only giving a normalization of the objective function to some convenient scale, i.e. 0–100. When a possible solution is found, the DM is asked which of the calculated values he finds satisfactory and which he finds unsatisfactory. In the next loop, the unsatisfactory values will be improved, while the satisfactory values are "sacrificed". This is repeated until the DM is happy with the proposed solution [4].

The basic idea in the technique for order preference by similarity to ideal solutions (TOPSIS) method is to compare the alternative solutions with the ideal and anti-ideal solutions. The best solution is the solution with the highest so-called "relative closeness to the ideal solution," which is a proportion between the Euclidean distances to the ideal and anti-ideal solutions [8,19].

2.2.3. Outranking models

In outranking models, the alternatives are compared pair-wise to check which of them is preferred regarding each criterion. When aggregating the preference information for all the relevant criteria, the model determines to what extent one of the alternatives can be said to outrank another. We can say that an alternative *a* outranks an alternative *b* if there is enough evidence to conclude that *a* is at least as good as *b* when taking all criteria into account [4]. The methods based on this way of thinking are often called the French school. The two main families of methods in the French school are ELECTRE and PROMETHEE. Below, I will give a brief explanation of these methods.

The family of ELECTRE methods was developed as an alternative to the utility function and value function methods. Details of the ELECTRE methods can, e.g. be found in Ref. [20]. The most common ELECTRE method in energy planning problems is ELECTRE III, so I will concentrate on that one is this review. The main idea in ELECTRE III is to choose alternatives that are preferred for most of the criteria. However, alternatives which are very unfavorable for any of the criteria should not be chosen, even if this alternative is favorable for most of the other criteria. The method makes use of the so-called indifference thresholds and strict preference thresholds. These thresholds are used to calculate concordance and discordance indices. From these indices, we can calculate graphs for strong and weak relationships, and these graphs are used to rank the alternatives through an iterative process. The method is sometimes not able to find the best alternative. However, it is often useful to apply the ELECTRE III method in the beginning of the decision process to produce a shortlist of the best alternatives. These alternatives can then go through further analysis by using another, more detailed method [4,21].

[1] In the world of multicriteria, an ideal solution is a theoretical solution where all the criteria have been respectively maximized or minimized.

An alternative outranking approach is the PROMETHEE method, developed by Brans and his co-workers [22]. In this method, a pair-wise comparison of alternatives is performed to make up a preference function for each criterion. Based on the preference function, a preference index for *a* over *b* is determined. This index is a measure of support for the hypothesis that *a* is preferred to *b*. It is defined as a weighted average of preferences on the individual criteria. The preference index is used to make a valued outranking relation which determines a ranking of the alternatives [4,8].

3. MCDA in energy planning

As mentioned in the introduction, energy planning is a field very suitable for MCDA methods. Over the last years, many applications of MCDA methods for energy planning problems have been published. In this section, I will give some examples that describe use of various MCDA methods for energy planning problems.

3.1. Value measurement models

Value measurement models have been used in various application areas in energy planning problems, especially for choosing/ranking energy strategies or technologies. Some of the applications have been evaluating alternative electricity supply strategies, using either analytical hierarchical process (AHP) [17,19], an AHP-similar method [23] or MAUT [24,25]. MAUT has also been used for an energy supply optimization process [26]. Hobbs et al. have done some interesting studies where they have compared various methods for collecting weights in MAVT analyses for evaluating demand-side management (DSM)[2] programs [10], and in the choice of an energy resource portfolio [11]. In Ref. [11], the MAVT approaches were also compared to a GP approach.

Buehring et al. [24] emphasized that the MAUT process in itself has many benefits for the DMs. They claimed that the process of assessing utility functions will help the DMs to identify the most important issues, generate and evaluate alternatives, resolve judgment and preference conflicts among the DMs and identify improvements to the impact. Siskos and Hubert [27] were more concerned about the drawbacks of the MAUT approach in their description of various MCDA methods. They claimed that MAUT presents many complications in the decision process, especially concerning the assessment of probabilities and attaching utilities to the criteria. To establish utility functions is a difficult and cumbersome task, because most DMs do not have a good perception of their own risk preferences [28]. However, MAUT is one of few MCDA methods designed especially for handling risk and uncertainties.

Advantages and shortcomings of the AHP method were discussed by Ramanathan and Ganesh [29]. They claimed that the main reasons for the AHP method's popularity are its simplicity, flexibility, intuitive appeal and its ability to handle both quantitative and qualitative criteria in the same framework. However, the method also has some drawbacks. According to Ref. [29], the main disadvantage is that AHP is very time-consuming when the number of alternatives and/or criteria is large, as is often the case in

[2]DSM activities are designed to encourage the customers to reduce their energy consumption and/or change their energy usage pattern. Such activities can to some extent be introduced as an alternative to increase the energy production.

energy problems. Another, often criticized problem, for instance Refs. [30–34], with the AHP method is the conversion from verbal to numerical judgments given by the fundamental scale (Table 1). It seems like the conversion table tends to overestimate preference differences [33]. There is also a lot of other criticism raised against the AHP method which are covered in more detail, e.g. in Ref. [32].

3.2. *Goal, aspiration and reference level models*

Another approach that has been used for energy planning studies is goal programming. The most commonly used GP method in energy planning problems seems to be the method of displaced ideals. The method has, e.g. been used for energy supply optimization [35], comparing different electricity generations systems from an environmental point of view [36] and for choosing an energy resource portfolio [11]. In these last two studies, the method of displaced ideals was compared to a monetization method[3] [36] and to a number of value-based methods [11], respectively.

Other GP methods that have been used for energy planning are the STEP method, which was used for energy resource allocation [37], and the TOPIS method, which was used for evaluation of alternative electricity supply strategies [19]. Ramanathan and Ganesh [29] has used the weighted sum of deviations to solve an energy resource allocation problem.

A reason to use GP techniques is that GP is less subjective than value theory and utility theory. In addition, GP offers a very straightforward procedure that DMs find easy to understand [29]. A third advantage is that many of the GP methods are suitable for being implemented directly into LP solvers [35]. It means that MCDA can be included into already existing one-criterion optimization models in a simple way. However, there is also a lot of criticism raised of GP, especially regarding the assignment of weights, the determination of goals and the normalization of the variables [29]. Another main disadvantage with the GP approach is that each criterion needs to be associated with an attribute defined on a measurable scale, which means that the methods are generally not able to handle non-quantitative criteria [4,29]. Therefore, GP must be combined with other techniques if qualitative criteria are going to be included in a study.

Pokharel and Chandrashekar [8] presented some advantages of the STEP method. According to them, the STEP method is the only method that allows direct comparison among the alternate solutions. This is supposed to help DMs to be aware of "the impact of a preference for an objective function on the solution". In addition, they found the STEP method easy to understand and to implement. However, they also found some drawbacks of the method. The main drawback is that the method requires that the DMs are able, precisely, to define their goals at each iteration. Furthermore, the method will—as other GP methods—in some cases present dominated solutions as being optimal.

3.3. *Outranking models*

Outranking models seems to be popular for energy planning problems. Outranking was used in many studies for evaluation of alternative electricity supply strategies (demand side management was also included in some of them). The most popular outranking

[3]In a monetization method, all criteria are translated into monetary values so that they can easily be compared.

methods in these evaluations is PROMETHEE II [7,17] and ELECTRE III [27,38–40]. PROMETHEE II has also been used for evaluating alternative strategies concerning geothermal energy usage [41].

Some of the main advantages of the outranking methods are that they provide a deep insight in the problem structure, they model the DM's preferences in a realistic way by recognizing hesitations in the DM's mind, and they are able to treat uncertainties in various ways [7,41]. In addition, it is claimed that the representation of the results from the outranking methods is simpler and easier to understand than the results from other MCDA approaches, such as MAVT [40].

A main difference between PROMETHEE II and ELECTRE III is the calculation procedure that is used. PROMETHEE II has a transparent calculation procedure, which is easy for DMs to understand [7], while the DMs often find the calculations from ELECTRE III too complex. Consequently, the ELECTRE method ends up as a 'black box' which feels unsatisfactory for the DMs [40,41].

The outranking methods are normally not used for the actual selection of alternatives, but they are very suitable for the initial screening process (to categorize alternatives into acceptable or unacceptable) [13]. After the screening process, another method must be used to get a full ranking or actual recommendations among the alternatives.

3.4. Combination of methods

Some researchers have tried to combine use of different MCDA methods. The AHP method has been especially popular to combine with other methods. Tzeng et al. [17] combined the use of AHP and PROMETHEE II, while Yang and Chen [19] combined AHP and TOPSIS in their evaluations of energy strategies. Ramanthan and Ganesh [29] integrated AHP and the GP method called the weighted sum of deviations for an energy resource allocation problem in India.

A proper combination of two (or more) methods might be very favorable. Such integration will help to make use of the strengths of both the methods. Moreover, even though both methods have some limitations, their limitations might be complementary. Ramanathan and Ganesh [29] argue that GP and AHP are well-suited to combine for a resource allocation problem. It is likely that suitable combinations of MCDA methods can be found also for other types of problems.

4. Conclusions and suggestions for further work

This literature review has shown that energy planning is a field that is quite suitable for the use of MCDA. I have shown that there exists a multitude of MCDA methods, and that many of these methods have been applied to energy planning purposes. Choosing among all the MCDA methods that exist can be said to be a multicriteria problem. Each of the methods has its own advantages and drawbacks, and it is not possible to claim that any one of the methods is generally more suitable than the others are. Different DMs will always disagree about which methods are most appropriate and valid.

The choice of method mostly depends on the preferences of the DM and the analyst. It is important to consider the suitability, validity and user-friendliness of the methods. It is also important to realize that use of different methods will most probably give different recommendations. This should not lead to the conclusion that there is anything wrong with any of the methods. It just means that the different methods work in different ways.

Hobbs and Horn [10] emphasized that choice of method can significantly affect judgment decisions. They claimed that change of method often makes more difference than change of the person that is applying the method. Hobbs and Horn [10] and Hobbs and Meier [11], therefore, concluded that ideally more than one multicriteria method should be used in a decision making process. This will give the DMs a broader decision basis. Additionally, DMs should be allowed to reflect upon and change their values after they get the first results from the methods. Accordingly, Hobbs and his colleagues proposed an interview process and a discussion among the DMs after the first collection of weights. During the interview and discussion, inconsistencies among the methods should be discovered. According to Buehring et al. [24], individuals will be more likely to discuss their judgments after they have been through a formalized decision making process. The extra effort required by the use of more than one method and the implementation of an interview process is not large compared to the potential benefits, which include enhanced confidence in the decision and a more reliable process [11].

In this review article, I have given many examples of how different MCDA methods have been utilized for energy planning. All the studies I have presented consider different aspects of energy networks with only one energy carrier (which was electricity in most of the studies). The majority of the studies are at a high planning level, such as a regional or even national level.

What seem to be missing in the research until now, are multicriteria studies on local energy systems with multiple energy carriers. Such combined energy systems with infrastructure such as networks for electricity, district heating and natural gas are common all over the world. In the past, these infrastructures were normally planned and commissioned by independent companies. It is believed that synergetic effects might be lost when such infrastructures are planned independently. Consequently, planning tools that can evaluate and analyze alternative energy carriers in mutual combination will give some benefits.

There is no doubt that if properly applied, MCDA can be a valuable tool also for planning of combined energy systems. Such systems may include several energy resources (hydro, oil, gas, garbage, etc.) and several energy carriers (electricity, district heating, natural gas, hydrogen, etc.) combined in a complex network with various conversion, storage and transportation technologies [6]. Often, there is more than one DM in such systems, and each of them will probably have many conflicting objectives which they would want to include in the study. In sum, these aspects make for a very complex problem. Because of this complexity, it is difficult for the DMs to get the full overview of their problem without using some decision-aid systems. Due to the conflicting objectives, some kind of MCDA should be well-suited. The problem, however, will be to choose which of the multitude of MCDA methods are most suitable for this type of problem.

Acknowledgements

I would like to acknowledge Nancy Bazilchuk for all her help and constructive feedback during the work with this review article. I would also like to acknowledge my main supervisor Professor Arne T. Holen.

References

[1] Ehrgott M, Gandibleux X. Introduction. In: Ehrgott M, Gandibleux X, editors. Multiple criteria optimization: state of the art annotated bibliographic surveys. Boston: Kluwer Academic Publishers; 2002.

[2] Bogetoft P, Pruzan PM. Planning with multiple criteria: investigation, communication and choice. København: Handelshøjskolens forlag; 1997.

[3] Dodgson J, Spackman M, Pearman A, Phillips L. DTLR multi-criteria analysis manual. UK Department for Transport, Local Government and the Regions;2001.

[4] Belton V, Stewart TJ. Multiple criteria decision analysis: an integrated approach. Boston: Kluwer Academic Publications; 2002.

[5] Huang JP, Poh KL, Ang BW. Decision analysis in energy and environmental modeling. Energy 1995;20(9):843–55.

[6] Catrinu M, Bakken BH, Holen A. Modelling local energy systems from a multicriteria perspective. Presented at the 17th international conference on efficiency, optimization, simulation and environmental impact of energy and process system. Guanajuato, Mexico;2004.

[7] Georgopoulou E, Sarafidis Y, Diakoulaki D. Design and implementation of a group DSS for sustaining renewable energies exploitation. Eur J Oper Res 1998;109(2):483–500.

[8] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning—a review. Renew Sustain Energ Rev 2004;8(4):365–81.

[9] Hobbs BF. Optimization methods for electric utility resource planning. Eur J Oper Res 1995; 83(1):1–20.

[10] Hobbs BF, Horn GTF. Building public confidence in energy planning: a multimethod MCDM approach to demand-side planning at BC gas. Energ Policy 1997;25(3):357–75.

[11] Hobbs BF, Meier PM. Multicriteria methods for resource planning: an experimental comparison. IEEE Trans Power Syst 1994;9(4):1811–7.

[12] Stewart TJ. A critical survey on the status of multiple criteria decision making theory and practice. Omega 1992;20(5-6):569–86.

[13] Greening LA, Bernow S. Design of coordinated energy and environmental policies: use of multi-criteria decision-making. Energ Policy 2004;32(6):721–35.

[14] Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley; 1976.

[15] Saaty TL. The analytic hierarchy process: planning, priority setting, resource allocation. New York: McGraw-Hill; 1980.

[16] Saaty TL. Decision making with dependence and feedback: the analytic network process; the organization and prioritization of complexity. Pittsburgh, PA: RWS Publications; 1996.

[17] Tzeng G-H, Shiau T-a, Lin C-Y. Application of multicriteria decision making to the evaluation of new energy system development in Taiwan. Energy 1992;17(10):983–92.

[18] Benayoun R, de Montgolfier J, Tergny J, Larichev O. Linear programming with multiple objective functions: step method (STEM). Math Prog 1971;1:366–75.

[19] Yang HT, Chen SL. Incorporating a multi-criteria decision procedure into the combined dynamic programming/production simulation algorithm for generation expansion planning. IEEE Trans Power Syst 1989;4(1):165–75.

[20] Roy B. Multicriteria methodology for decision aiding. Nonconvex optimization and its applications, vol. 12. Boston, MA, Dordrecht, Netherlands;1996.

[21] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning—a review. Renew Sustain Energ Rev 2004;8(4):365–81.

[22] Brans JP, Vincke P, Mareschal B. How to select and how to rank projects: the PROMETHEE method. Eur J Oper Res 1986;24(2):228–38.

[23] Lootsma FA, Boonekamp PGM, Cooke RM, Van Oostvoorn F. Choice of a long-term strategy for the national electricity supply via scenario analysis and multi-criteria analysis. Eur J Oper Res 1990;48(2):189–203.

[24] Buehring WA, Foell WK, Keeney RL. Examining energy/environment policy using decision analysis. Energ Syst Policy 1978;2(3):341–67.

[25] Pan J, Teklu Y, Rahman S, de Castro A. An interval-based MADM approach to the identification of candidate alternatives in strategic resource planning. IEEE Trans Power Syst 2000;15(4):1441–6.

[26] Schulz V, Stehfest H. Regional energy supply optimization with multiple objectives. Eur J Oper Res 1984;17(3):302–12.
[27] Siskos J, Hubert P. Multi-criteria analysis of the impacts of energy alternatives: a survey and a new comparative approach. Eur J Oper Res 1983;13(3):278–99.
[28] Botterud A. MCDM in the E-transport model. Working paper. NTNU;2004.
[29] Ramanathan R, Ganesh LS. Energy resource allocation incorporating qualitative and quantitative criteria: an integrated model using goal programming and AHP. Socio-Econ Plan Sci 1995;29(3):197–218.
[30] Finan JS, Hurley WJ. Transitive calibration of the AHP verbal scale. Eur J Oper Res 1999;112(2):367–72.
[31] Løken E, Botterud A. Planning of mixed local energy distribution systems: a comparison of two multi-criteria decision methods. Presented at the 28th annual IAEE international conference. Taipei, Taiwan;2005.
[32] Salo AA, Hämäläinen RP. On the measurement of preferences in the analytic hierarchy process. J Multi-Criteria Decision Anal 1997;6(6):309–19.
[33] Huizingh EKRE, Vrolijk HCJ. A comparison of verbal and numerical judgments in the analytic hierarchy process. Organ Behav Human Decision Process 1997;70(3):237–47.
[34] Pöyhönen M, Hämäläinen RP, Salo AA. An experiment on the numerical modelling of verbal ratio statements. J Multi-Criteria Decision Anal 1997;6(1):1–10.
[35] Oliveira C, Antunes CH. A multiple objective model to deal with economy–energy–environment interactions. Eur J Oper Res 2004;153(2):370–85.
[36] Mirasgedis S, Diakoulaki D. Multicriteria analysis vs. externalities assessment for the comparative evaluation of electricity generation systems. Eur J Oper Res 1997;102(2):364–79.
[37] Pokharel S, Chandrashekar M. A multiobjective approach to rural energy policy analysis. Energy 1998;23(4):325–36.
[38] Beccali M, Cellura M, Ardente D. Decision making in energy planning: the ELECTRE multicriteria analysis approach compared to a FUZZY-SETS methodology. Energy Convers Manage 1998;39(16–18):1869–81.
[39] Beccali M, Cellura M, Mistretta M. Decision-making in energy planning. Application of the Electre method at regional level for the diffusion of renewable energy technology. Renew Energ 2003;28(13):2063–87.
[40] Georgopoulou E, Lalas D, Papagiannakis L. A multicriteria decision aid approach for energy planning problems: the case of renewable energy option. Eur J Oper Res 1997;103(1):38–54.
[41] Haralambopoulos DA, Polatidis H. Renewable energy projects: structuring a multi-criteria group decision-making framework. Renew Energ 2003;28(6):961–73.

Espen Løken is a 27-years-old Ph.D. student at the Norwegian University of Science and Technology (NTNU). He finished as Master of Science in Energy and Environment Engineering in June 2003 and started working as a Ph.D. student in August 2003 at the Department of Electrical Power Engineering at NTNU. He is a part of the research project called "Sustainable Energy Distribution Systems" (SEDS). The preliminary title of his doctoral thesis is "Multi-Criteria Decision Methods for Planning and Operation of Energy Distribution Systems".

Applied Energy 87 (2010) 1134–1140

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

A multicriteria approach to evaluate district heating system options

S. Ghafghazi [a], T. Sowlati [a,*], S. Sokhansanj [b,c], S. Melin [b,d]

[a] Department of Wood Science, University of British Columbia, 2931–2424 Main Mall, Vancouver, BC, Canada V6T1Z4
[b] Department of Chemical and Biological Engineering, University of British Columbia, 2360 East Mall, Vancouver, BC, Canada V6T1Z3
[c] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[d] Delta Research Corporation, Delta, BC, Canada

A R T I C L E   I N F O

Article history:
Received 1 May 2009
Received in revised form 12 June 2009
Accepted 15 June 2009
Available online 12 July 2009

Keywords:
District heating systems
Multicriteria decision making
PROMETHEE
Group decision making
Renewable energy
Biomass

A B S T R A C T

District energy systems, in which renewable energy sources may be utilized, are centralized systems to provide energy to residential and commercial buildings. The aim of this paper is to evaluate and rank energy sources available for a case of district heating system in Vancouver, Canada, based on multiple criteria and the view points of different stakeholders, and to show how communication would affect the ranking of alternatives. The available energy sources are natural gas, biomass (wood pellets), sewer heat, and geothermal heat. The evaluation criteria include GHG emissions, particulate matter emissions, maturity of technology, traffic load, and local source. In order to rank the energy options the PROMETHEE method is used. In this paper, two different scenarios were developed to indicate how the communication between the stakeholders would affect their preferences about criteria weights and would change the ranking of alternatives. The result of this study shows that without communication the best energy source for the considered district energy system is different for different stakeholders. While, addressing concerns through efficient communication would result in a general consensus. In this case, wood pellet is the best energy alternative for all the stakeholders.

1. Introduction

1.1. Background

The total primary and secondary energy consumption in Canada was 7643.2 petajoules in 2006 [1]. Long travel distances, cold climate, energy-intensive industrial base, relatively low energy prices, and a high standard of living were the main reasons for high consumption of energy [2]. Space and water heating account for 60–80% of the energy consumed in the residential, commercial, institutional, and public administration sectors [3]. Energy consumption in these sectors represents 36% of the total energy use in Canada [1]. Energy strategies regarding space and hot water heating in Canadian provinces have remarkable impact on the total consumed energy and consequently on the resulted environmental impacts.

District energy systems, which provide energy for space and hot water heating to buildings, have several advantages compared to the decentralized ones. These advantages include: (1) increased energy and performance efficiencies through implementing advanced equipment and maintaining them professionally, (2) reduced life cycle costs, and (3) augmented control over environmental impacts [4]. Although the history of district energy systems in Canada goes back to the 19th century, this sector is still growing in Canada as more municipalities and communities turn to district energy each year in order to conserve energy, mitigate climate change, and secure supply of energy [5].

District energy systems can have access to a wider range of energy sources compared to decentralized systems. Renewable energy sources such as wood biomass, sewer heat and geothermal heat can be more economically and efficiently exploited in larger systems in district energy systems. The type of energy to be used in a district energy system broadly identifies the characteristics of the system, such as the heating technology, system efficiency, capital investment, operating costs, system emissions, etc. The suitability of a district energy system, which depends on the type of energy source, characteristics of the system, and district's objectives and requirements, should be assessed carefully. Typically, different alternatives are available for district energy systems that should be evaluated based on economic, technical, environmental and social factors. These factors may be quantitative or qualitative. Moreover, the importance of these factors may be different for various stakeholder groups involved in the decision making, as they may have different and sometimes conflicting interests and objectives. The need to incorporate different factors and the viewpoints of various actors in the analysis has promoted the use of multicriteria approaches in energy planning. These approaches provide a

* Corresponding author. Tel.: +1 604 822 6109.
E-mail address: taraneh.sowlati@ubc.ca (T. Sowlati).

0306-2619/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2009.06.021

better understanding of the decision problem, help reaching a compromised decision, and facilitate the negotiation and communication among different stakeholders [6].

Various multicriteria decision making (MCDM) methods have been applied in energy planning. Review of more than 90 published papers in the energy planning field by Pohekar and Ramachandran [6] showed that beyond 1990, and especially in the area of renewable energy planning, Analytic Hierarchy Process (AHP), PROMETHEE and ELECTRE were the most commonly used MCDM methods.

This study focuses on the utilization of different energy sources, including renewable ones, in a district heating system in Vancouver, Canada. Four energy options of natural gas, biomass (wood pellets), sewer heat, and geothermal heat are ranked based on six important criteria using the PROMETHEE method. Two different scenarios are considered to show how the ranking of alternatives for stakeholder groups would change when communication among them is facilitated. It is assumed that communication between the stakeholders would change their preferences about the criteria which eventually would affect the ranking of alternatives. The objective of this study is to assess the effect of communication between legitimate stakeholders in the decision making process when multiple decision criteria exist, using the considered case as an example.

1.2. Literature review

Traditionally, the selection of energy alternatives was based only on cost minimization [7]. However, the need to incorporate the environmental and social impacts of energy options and viewpoints of different actors in the analysis promoted the use of multicriteria decision making (MCDM) methods [8]. A wide range of MCDM methods have been applied in the energy planning area. Selection of an appropriate MCDM method is an MCDM problem itself [9]. In an MCDM approach, first it is necessary to define the problem clearly, then identify realistic alternatives. It is important to define the actors involved in the decision making, select the evaluation criteria, and evaluate each alternative according to the set of criteria. Next, an MCDM method is selected to aggregate the performance of each alternative. Most MCDM methods require weighting of the selection criteria. In real-life problems, a weighting method that is reliable and easy to apply should be used. The application of the MCDM method provides a ranking of alternatives [7].

AHP, PROMETHEE and ELECTRE are the most common MCDM methods used in energy planning. AHP has been used successfully for alternative energy source selection [10–12], and energy resource allocation [13,14].

The ELECTRE method has been applied for renewable energy planning [7], energy planning [15], choosing the most suitable heating system for buildings [16], and community energy modernization and development planning [17].

Pohekar and Ramachandran [18] used the PROMETHEE method to rank alternative cooking energy sources in India. Through sensitivity analysis, they concluded that dissemination of a parabolic solar cooker in India was not merely upon its economic appeal, but several technical, social, and behavioral factors should also be addressed simultaneously. Haralambopoulos and Polatidis [19] used the PROMETHEE II method to rank exploitation scenarios for geothermal resources in the island of Chios, Greece. The emphasis has been on finding the agreement and conflicting points among different stakeholders. The PROMETHEE method is also used in ranking sustainable electricity generation technologies and energy exploitation scenarios of low temperature geothermal fields in Greece [20,21]. A group decision support system in which PROMETHEE is used for ranking the renewable energy sources is designed and presented by Georgopoulou et al. [8].

2. Methods

2.1. The PROMETHEE method

In this paper, the PROMETHEE II method is used to rank the energy sources available for the considered case based on stakeholders' preferences. The PROMETHEE method introduced by Brans and Vincke [22] belongs to the group of outranking methods.

In order to better explain the PROMETHEE method, suppose a multicriteria problem as:

$$\{f_1(a), f_2(a), \ldots, f_h(a), \ldots, f_k(a) | a \in K\} \tag{1}$$

where K is a (finite) set of possible alternatives, and $f_h(a)$, $h = 1, 2, \ldots, k$, is the value of alternative a for criterion h. Ideally, a decision maker is interested in finding an optimal alternative $\hat{a}$ which dominates all other alternatives (has the highest value for all criteria compared to other alternatives) so $f_h(\hat{a}) \geqslant f_h(a), \forall a \in K, \forall h$. In general, such an optimal solution does not exist, and indeed the dominance relationship between the alternatives defined as: a dominates b iff $f_h(a) \geqslant f_h(b), \forall h \in \{1, 2, \ldots, k\}$ is poor between all the two-by-two alternatives. Outranking methods such as PROMETHEE try to enrich the dominance relationship between the alternatives.

Considering two alternatives a and b, the preference structure can be defined as:

$$\begin{cases} aPb & \text{iff } f_h(a) > f_h(b) \\ aIb & \text{iff } f_h(a) = f_h(b) \end{cases} \tag{2}$$

aPb means that alternative a is preferred over alternative b, if alternative a is performing better than alternative b with regard to criterion h, and *aIb* means that alternatives a and b are indifferent with regard to criterion h.

The PROMETHEE method gives a numerical value between 0 and 1 to the preference relationship in Eq. (2) by introducing the preference function $P(a, b)$ such that:

$$P(a,b) = \begin{cases} 0 & \text{if } f_h(a) \leqslant f_h(b) \\ p[f_h(a), f_h(b)] & \text{if } f_h(a) > f_h(b) \end{cases} \tag{3}$$

where $0 < p[f_h(a), f_h(b)] \leqslant 1$. For practical applications, it is then reasonable to assume that:

$$p[f_h(a), f_h(b)] = p[f_h(a) - f_h(b)] \tag{4}$$

Let $D_h(a, b)$ be the difference between alternative a and alternative b for criterion h as shown in Eq. (5):

$$D_h(a,b) = f_h(a) - f_h(b) \tag{5}$$

Brans and Vincke [22] recognized six types of preference functions that are most common in the real case situations. In this paper, the usual preference function is used, which is:

$$p[f_h(a), f_h(b)] = \begin{cases} 0 & \text{if } D_h(a,b) \leqslant 0 \\ 1 & \text{if } D_h(a,b) > 0 \end{cases} \tag{6}$$

As an example, suppose that the cost of energy for option a is \$1000,000 less than that for option b, then preference of alternative a over alternative b is 1 and preference of alternative b over a is 0.

Then, the PROMETHEE method uses the weighted preference index $\pi(a, b)$ to give an integrated overall preference of alternative a over b, shown in Eq. (7):

$$\pi(a,b) = \frac{\sum_{h=1}^{k} w_h P_h(a,b)}{\sum_{h=1}^{k} w_h} \tag{7}$$

where w_h is the relative importance of criterion h, which is defined by the decision makers. To build the outranking relation among the alternatives, PROMETHEE introduces three outranking measures for each alternative as follows:

- Outgoing flow $\phi^+(a) = \sum_{x \in K} \pi(a, x)$. The larger $\phi^+(a)$, the more alternative a outranks the other alternatives in the set K,
- Incoming flow $\phi^-(a) = \sum_{x \in K} \pi(x, a)$. The smaller $\phi^-(a)$, the less alternative a has been outranked by other alternatives in the set K.,
- Net flow $\phi(a) = \phi^+(a) - \phi^-(a)$.

PROMETHEE II considers the net flow for each alternative $a \in K$ to find the total preorder (complete ranking) such that:

- a outranks $b(aPb)$ iff $\phi(a) > \phi(b)$,
- a is indifferent to $b(aIb)$ iff $\phi(a) = \phi(b)$.

In summary, to rank alternatives using the PROMETHEE II method, the analyst needs to identify the alternatives/criteria matrix, which is called the decision matrix, the relative importance of criteria over each other, and the preference functions for each criterion.

In this study, Decision Lab. 2000 software [23] is used which offers the PROMETHEE I and II methods and provides graphical interface for these methods.

2.2. The expected value method to determine criteria weights

The selected criteria usually do not have equal importance and different actors may perceive their importance differently. Different methods may be used to extract the decision makers' preferences. The direct method of assigning weights to criteria is the simplest one. Georgopoulou et al. [8] used an indirect method based on a hierarchical ranking of criteria. In this paper, the Expected Value method [24] is used to extract the criteria weights. This method estimates the weights based on the decision makers' preferred ranking of the criteria. If there are k criteria in the analysis which are ranked in ascending order of importance based on the decision maker's preference, then the expected values in Eq. (8) are assigned as criteria weights.

$$
\begin{aligned}
E(w_1) &= \frac{1}{k^2} \\
E(w_2) &= \frac{1}{k^2} + \frac{1}{k(k-1)} \\
&\vdots \\
E(w_{k-1}) &= \frac{1}{k^2} + \frac{1}{k(k-1)} + \cdots + \frac{1}{k \cdot 2} \\
E(w_k) &= \frac{1}{k^2} + \frac{1}{k(k-1)} + \cdots + \frac{1}{k \cdot 2} + \frac{1}{k \cdot 1}
\end{aligned} \tag{8}
$$

where $E(w_i)$ is the expected value of the ith criterion and is used as the weight for that criterion, and k is the number of criteria.

3. Case study

A district heating system to provide hot water to 350,000 m^2 floor area of a newly developed community in the city of Vancouver, British Columbia (BC) was evaluated by the City.[1] The supplied hot water would be used for space heating and providing hot water to the buildings within the community. The share of floor area usage for this community is approximately 40% residential, 40% office area and the rest would be hospital, commercial retail, and hotel/casino. The annual heat demand of the connected buildings to the district heating system was estimated to be 28,000 MWh, with the peak demand to be 10 MW [25]. It was considered to install a base-load system to provide about 68% of the annual heat demand and to use a low capital cost system with secure technology and fuel supply such as a natural gas boiler alongside the base-load system for peaking and backup. The considered split of energy supply between the two systems was 2.5 MW for the base-load system and a 10 MW natural gas boiler be used for peaking and backup [26]. Having a separate base-load system would give the opportunity of exploiting alternative energy sources to meet the majority of the community's energy demand throughout the year. Moreover, using alternative renewable energies for the base-load system in district energy centers help the province meet its greenhouse gas (GHG) reduction plan. Specifically, it conforms to the actions (4) "Evaluate opportunities for renewable energy in public facilities" and (13) "Encourage the use of district energy systems" of British Columbia's GHG Action Plan [27].

BC's municipalities have access to a wide range of energy sources. Natural gas is one of the most popular sources of energy all around BC. Its utilization is both cheap and easy because of the well developed network and infrastructure available in the province. There are also other locally available energy sources to the communities in BC such as sewer heat, geothermal, and biomass. Sewer heat recovery and geothermal heat exchange systems are new, yet proven technologies. Wood biomass, as a renewable energy source, is getting increasing attention in BC. Relatively cheap fuel price, low capital intensive equipment, GHG neutrality, and abundant resources in the province are the advantages of wood biomass which makes it a favorable energy source for energy projects in BC. Clean electricity produced by hydro power plants in the province also provides relatively cheap energy to communities.

Waste incineration and waste energy from industrial processes can be also utilized in district heating systems [28] provided that these sources are practically available to the district heating center, which was not the case for the considered district heating center. Alternative energy production systems such as Combined Heat and Power (CHP) systems would also enable district heating centers to produce heat as well as electricity more efficiently. Feasibility and performance of CHP systems with heat storage possibility for district heating systems are confirmed [29,30]. None the less, the focus of this study was only on heat generation and CHP option was beyond the objectives set for the considered district heating center.

Considering the available energy sources to energy centers in Vancouver, BC, four energy options of sewer heat, geothermal heat, biomass (wood pellets), and natural gas were evaluated by the City of Vancouver to provide the base-load energy requirement of the district heating system [31]. During the feasibility study stage, several studies including heat demand, base-load and backup system capacities, cost and emission analyses were performed by consultants for the City. Several actors influenced the decision directly and indirectly. The final decision, which was based partly on studies' results and partly by the influence of different actors, was to choose the sewer heat option.

In this study, the mentioned case with the same energy source options is considered. Based on analyzing the study reports and comprehending the actual decision making process and actors involved, we specify six important criteria and three stakeholder groups (actors). This paper presents the ranking of energy source options for the district heating system considering all important factors using PROMETHEE. Two different scenarios are examined and indicate how consensus can be reached as the result of proper communication between different actors.

[1] The data and information used in this paper regarding the considered district energy center were provided to the authors by the City of Vancouver [25,26,28,31,32]. Since these references are internal reports and not available to the public, the information from these reports which was used in this study is presented clearly in the text (Section 3) and Table 1.

3.1. Stakeholders

Usually in the decision making process where multiple decision makers with diverse backgrounds and viewpoints exist, if not impossible, it is difficult to reach at a single, globally agreed upon decision. Therefore, the decision makers and stakeholders should be in communication with each other from the early stages of the project in order to discuss their objectives, values and/or concerns. MCDM methods can provide valuable basis for decision makers' discussions throughout this phase.

In deciding about projects with public benefits or those that might be of public concern, such as energy decisions, it is important to involve or open a discourse with the public. Some important purposes of public involvement are: to inform the public, to reflect public values in the decisions, to consider the impacts that might be overlooked, and to provide 'due process' [13]. Moreover, failure to involve the public in the decision process from the early stages, sometimes results in strong oppositions at the final stages of the decision making process from the public pressure groups such as community associations, NGOs, and the media.

In this study, based on the real case, three groups are identified who affect the decision on the base-load energy source for the district energy system including: (1) developer, (2) environmental group, and (3) community representative group. Developer is responsible for the design and construction of the district energy center. Technical information about the considered energy sources, generated by the developer, would be reviewed by the environmental group for obtaining any required permissions from the City. Also, there has to be no objection from the community representative groups to issue permission on a selected energy source by the City. Therefore, the decision authority situation in this case implies that the decision on the choice of the base-load energy source is affected by the environmental group and, to a greater extent, the community representative group.

3.2. Alternative/criteria matrix

In order to compare energy systems against each other, various criteria could be considered. Considered criteria depend to a large degree on the situation and nature of the case, provided that the performance of energy systems varies with regard to the considered criterion. Sometimes these criteria are referred to as "sustainability indicators" meaning that these criteria identify the degree of sustainability of the energy systems [32]. Usually these criteria are classified into economic, environmental, technological, and social subgroups. They may be stated based on quantitative values or a given qualitative measure. Normally, those with well established quantitative measures such as investment cost, payback period, efficiency rate, and system emissions are stated based on quantitative units. Stakeholders' judgmental values such as continuity and predictability of an energy technology, contribution to regional development, and contribution to employment opportunities creation can be shown on a scaled measure; for example, scale of 1–10, 1 being the worst and 10 being the best performance of an alternative [33].

In this research, four energy system alternatives are considered which include biomass (wood pellets) combustion, sewer heat recovery, geothermal exchange, and natural gas boiler systems. For evaluation of energy systems, six criteria are considered as follows:

- *Costs (economic factor, quantitative value).* Considered costs are the present value (2005 base year) of the plants (including a 2.5 MW base-load system and a 10 MW peaking and backup natural gas system) at 10% discount rate. The cost of the plant includes land, building, major equipment, electrical and mechanical installations, soft costs (engineering, construction management and supervision), 7% provincial sales tax, contingency cost at 10%, maintenance cost, and operating cost (fuel and/or electricity cost and staffs' salary) over the 25-year service life of the system. Costs and benefits that are common and equal for different energy alternatives such as grid development, energy sales, and sale taxes are not considered [26]. For the biomass energy source, wood pellet price is considered to be 6.5 (CAD/GJ) with an inflation rate of 2 (%/yr) [34]. For the sewer heat option, the extra costs of the sewage system include: redundant self cleaning screens, booster pumps, backwash pits, transfer pumps, electrical infrastructure, incremental building, stainless steel wetted interconnecting piping between treatment system and heat pump evaporator. For geothermal option, the cost of geothermal well includes wells, buried piping between wells and plant, pumps, and electrical infrastructure as required [26,31].
- *Total GHG emission of the system (global environmental impact, quantitative factor).* This is the CO_2 equivalent emission of the 2.5 MW base-load system and 10 MW peaking and backup natural gas system. Biomass is considered to be GHG neutral. Reported number for biomass option also includes the GHG emissions associated with the road transportation of biomass (wood pellets) from the nearest producing facility (275 Kms) to the district heating system. For the electricity used in sewer heat and geothermal options heat pumps, a GHG factor of 205 (tonnes GHG/GWh) is considered for 30% of the electricity consumed [26]; that is, 70% of the generated electricity is assumed to come from GHG free hydro generators while 30% of it is generated in power plants with GHG emission equal to 205 (tones GHG/GWh).
- *Particulate Matter (PM) emission of the system (local environmental impact, quantitative factor).* This includes particulate matter less than or equal to 2.5 μm in diameter. Number reported is the total PM2.5 emission produced by the facility without emission control system [35].
- *Maturity of the technology (qualitative factor).* This is a technical factor which shows the risks associated with the installation, handling and future break downs of the system [33]. This criterion is considered as a qualitative measure based on a five-point scale (1 = very low, 5 = very high).
- *Local source (qualitative factor).* Whether or not the energy source is available within the community was considered as a binary criterion (0 = locally not available, 1 = locally available). The community representative group saw geothermal heat and sewer heat options as local energy sources that would have secure supply over the service life of the facility and if not implemented, would be wasted.
- *Traffic load (qualitative factor).* One of the community's concerns, raised particularly about the biomass option, was trucking in the biomass to the facility and taking out the remained ash. The community representative group was concerned whether trucking of this fuel would have major traffic burden for the community. Therefore, this factor is considered as a binary value which is 1 for the biomass option and 0 for other options.

Table 1 shows the alternatives/criteria matrix of the decision problem.

4. Scenarios

In this paper, two scenarios are examined to evaluate the impact of communication and transparency of information among the stakeholders on the final decision about the most suitable energy option of a district energy center. The first scenario represents

Table 1
Alternatives/criteria matrix.

Criteria	Units	Alternatives			
		Natural gas	Biomass	Sewer heat	Geothermal
Cost	10^3 CAD $	16,875	14,688	19,041	23,521
GHG emission	Tonne/yr	7875	2564	3635.2	4081.28
$PM_{2.5}$	Tonne/yr	0.14	2.40	0.04	0.04
Maturity of technology	Qualitative scale (1–5)	5	4	1	2
Local source	Binary value (0, 1)	0	0	1	1
Traffic load	Binary value (0, 1)	0	1	0	0

the actual decision process for selecting the energy option for the district heating system. In this scenario, criteria weights are extracted based on stakeholders' preferences and ranking of criteria using the Expected Value method. The second scenario assumes there is communication between stakeholders and information generated during the feasibility study is available to all stakeholders. This would affect stakeholders' preferences and their ranking of criteria, and consequently the final decision in a multicriteria analysis.

4.1. Scenario I

As the developer carried out the feasibility analysis of the district energy center, studying the economics and GHG emissions of different energy options available to the energy center, they could confirm the superior performance of the biomass option. The recommendation of the developer was to install a 2.5 MW biomass combustion system for the base-load and a 10 MW natural gas boiler for peaking and backup. In the beginning of the implementation phase when the developer required regulatory permissions for the energy center, the environmental and the community representative groups became more aware of the planned energy facility. At this stage, the idea of utilizing biomass in the energy center was rejected by both the environmental and the community representative groups. Community representative group's concerns were mainly regarding the negative effect of biomass utilization on the local air quality and traffic load due to biomass transportation to the facility. Environmental group's review of the energy options identified that the feasibility study carried out by the developer had not addressed such issues as particulate matter emissions from the biomass combustion system which was central for acquiring the air quality permission. Because of the time limit the developer had in delivering the project, the public process required for addressing the issues raised by other stakeholders could not be fulfilled. Therefore, the next economic option which did not have any objection from the other involved stakeholders, i.e. sewer heat recovery system, was chosen to provide the base-load heat for the energy center.

Based on the review of memorandums and comments received from the three stakeholder groups involved in the decision process, below ranking of decision criteria was inferred for each stakeholder:

1. Developer:

$$\begin{aligned} &Cost > Maturity\ of\ technology > GHG\ emissions \\ &\quad > PM\ emissions > Local\ source = Traffic\ load \end{aligned} \tag{9}$$

2. Environmental group:

$$\begin{aligned} PM\ emissions > GHG\ emissions > Cost& \\ = Maturity\ of\ technology = Local\ source& \\ = Traffic\ load& \end{aligned} \tag{10}$$

3. Community representative group:

$$\begin{aligned} PM\ emissions > Local\ source = Traffic\ load > Cost& \\ = Maturity\ of\ technology = GHG\ emissions& \end{aligned} \tag{11}$$

Table 2 summarizes the criteria weights considered for each stakeholder in scenario I using Eq. (8). In this scenario, the number of criteria (k) is 6 and in a situation when criteria have equal importance, the average of weights is considered. For example, the criteria weights for the two criteria "Local source" and "Traffic load" considered for the developer based on the ranking expressed in Eq. (8) would be the average of the criteria weights when *Local source > Traffic load* and *Local source < Traffic load.*

4.2. Scenario II

In this scenario, in an ex-post attempt, it is tried to address and evaluate the major concerns of the environmental and community representative groups about the biomass burning facility, i.e. particulate matter emissions from the biomass burning facility and traffic load due to biomass transportation.

The issue of traffic burden of biomass fuel supply over the service life of the energy center which was of great importance for the community representative group was evaluated based on a fuel supply/ash disposal model [34]. The result of this analysis indicated that the supply of wood pellets can be secured through long term contract with a local wood pellet supplier. Moreover, the amount of wood pellets required for the biomass plant with a boiler of 2.5 MW would be about 5800 tonnes/year. This volume of wood pellets could be scheduled for one truck of 40 tonnes capacity every 3 days. Also, ash disposal could also be scheduled such

Table 2
Criteria weights considered in Scenario I.

Stakeholders	Criteria					
	Cost	Maturity of technology	GHG emission	$PM_{2.5}$ emission	Local source	Traffic load
Developer[a]	0.41	0.24	0.16	0.10	0.045	0.045
Environmental group[b]	0.0875	0.0875	0.26	0.41	0.0875	0.0875
Community representative group[c]	0.06	0.06	0.06	0.41	0.21	0.21

[a] The criteria weights are obtained from Eq. (8) with k equals to 6 and ranking stated in Eq. (9).
[b] The criteria weights are obtained from Eq. (8) with k equals to 6 and ranking stated in Eq. (10).
[c] The criteria weights are obtained from Eq. (8) with k equals to 6 and ranking stated in Eq. (11).

that the same truck is used so that no additional truck would be required to commute to the facility [34]. This result was communicated to the authorities and the insignificant effect of biomass transportation to the energy center on the local traffic was confirmed.

In another study, air quality assessment of the surrounding area of the biomass facility was carried out by the City [35]. It was shown that with commitment to the air quality standards and with the means of installing appropriate emission control systems, the emissions of the district energy facility can be within the emission guideline of Vancouver (BC). The considered emission control system including a cyclone system coupled with an Electro Static Precipitator (ESP) would constitute about 5% of the total cost of the biomass plant estimated in Table 1; while collection efficiency of particulate matter emissions of this configuration is 93–98% [36]. This study estimated that through implementing the indicated emission control system, the contribution of the biomass burning facility to the ambient air level of particulate matter emission would be, at the worst case, less than 2% and well below the most stringent emission level [35].

The insignificant effect of the biomass burning activity on the ambient air quality and traffic load, when communicated and approved by the stakeholder groups, was reflected in Scenario II by omitting the two respective criteria from the decision matrix. Taking out the PM emissions and traffic load criteria from the criteria list, the criteria ranking by each stakeholder would change as follows:

1. Developer:

$$\begin{aligned} &\mathit{Cost} > \mathit{Maturity\ of\ technology} > \mathit{GHG\ emissions} \\ &\quad > \mathit{Local\ source} \end{aligned} \tag{12}$$

2. Environmental group:

$$\begin{aligned} &\mathit{GHG\ emissions} > \mathit{Local\ source} = \mathit{Cost} \\ &\qquad = \mathit{Maturity\ of\ technology} \end{aligned} \tag{13}$$

3. Community representative:

$$\begin{aligned} &\mathit{Local\ source} = \mathit{Cost} = \mathit{Maturity\ of\ technology} \\ &\qquad = \mathit{GHG\ emissions} \end{aligned} \tag{14}$$

The criteria weights shown in Table 3 are obtained from Eq. (8) with k equals to 4 and equally important criteria are averaged as was done in the first scenario.

5. Results and discussion

Table 4 shows the ranking of the alternatives obtained by PROMETHEE II method for the three stakeholder groups in Scenario I and Scenario II. The outcome of the PROMETHEE II method for Scenario I explicitly shows that stakeholders' interpretations about the best option are very diverse when communication is not facilitated and major concerns of the biomass burning facility is not addressed. The fact that biomass is the worst and second worst option for the community representative and environmental groups, respectively, stems from the concern of these groups about the effects of biomass burning facility on the local activities. Also, it can be seen that sewer heat recovery option is the best option for the environmental and community representative groups. Therefore, it can be expected that sewer heat recovery option be chosen since, as mentioned, decision authority of these groups is higher than that of the developer. This result was similar to what was observed in reality. The ranking of alternatives in Scenario II has changed, as the main points of conflict between the stakeholders are addressed because of the proper communication among them. This affected the stakeholders' preferences and ranking of criteria and resulted in a general agreement.

Therefore, to reach a consensus, all the legitimate stakeholders should be identified accurately and their perceptions, viewpoints, and concerns should be obtained and addressed in the feasibility studies of potential district energy systems. The accurate data and information produced during the feasibility study of the system should then be conveyed to stakeholders properly. In order to include all important factors in the analysis, a proper decision making tool that can incorporate quantitative and qualitative factors should be applied.

6. Conclusions

In this paper, four energy options of natural gas, biomass (wood pellets), sewer heat recovery, and geothermal exchange to provide

Table 3
Criteria weights considered in Scenario II.

Stakeholders	Criteria			
	Cost	Maturity of technology	GHG emission	Local source
Developer[a]	0.52	0.27	0.15	0.06
Environmental group[b]	0.16	0.16	0.52	0.16
Community representative group[c]	0.25	0.25	0.25	0.25

[a] The criteria weights are obtained from Eq. (8) with k equals to 4 and ranking stated in Eq. (12).
[b] The criteria weights are obtained from Eq. (8) with k equals to 4 and ranking stated in Eq. (13).
[c] The criteria weights are obtained from Eq. (8) with k equals to 4 and ranking stated in Eq. (14).

Table 4
Ranking of alternatives for each stakeholder based on PROMETHEE II.

	Ranking			
Stakeholders	1	2	3	4
Scenario I				
Developer	Biomass (Φ = 0.47)	Natural gas (Φ = 0.21)	Sewer heat (Φ = −0.13)	Geothermal (Φ = −0.54)
Environmental group	Sewer heat (Φ = 0.35)	Geothermal (Φ = 0.14)	Biomass (Φ = −0.17)	Natural gas (Φ = −0.54)
Community group	Biomass (Φ = 0.37)	Natural gas (Φ = 0.30)	Sewer heat (Φ = −0.27)	Geothermal (Φ = −0.40)
Scenario II				
Developer	Biomass (Φ = 0.7)	Sewer heat (Φ = 0.22)	Natural gas (Φ = −0.22)	Geothermal (Φ = −0.71)
Environmental group	Biomass (Φ = 0.63)	Sewer heat (Φ = 0.12)	Geothermal (Φ = −0.31)	Natural gas (Φ = −0.45)
Community group	Biomass (Φ = 0.68)	Sewer heat (Φ = −0.08)	Natural gas (Φ = −0.22)	Geothermal (Φ = −0.38)

the base-load heat demand of a district energy system in Vancouver (BC) were compared. The PROMETHEE II method was used to rank the alternatives against six criteria of cost, GHG emissions, PM emissions, maturity of technology, traffic load, and local source. Two scenarios were investigated to indicate how the consensus between the stakeholder groups involved in the district energy project can be reached through good communication during the feasibility study and decision making process. The first scenario represented the real case of decision making where there was no communication between stakeholders. This fact was reflected in the analysis by the decision makers' preferences and ranking of criteria by them. Based on the ranking, criteria weights were assigned using the Expected Value method. The PROMETHEE results conformed to the decision made in the real case. Despite advantages of utilizing biomass in the considered district energy system, such as low capital cost, advanced and low risk burning technology, and GHG neutrality, it was not chosen as the best option since the concerns of local groups were not addressed properly during the feasibility study phase. The sewer heat option was selected due to the higher decision authority of the opposing groups and the fact that the information about biomass option was not communicated to other stakeholders well. In the second scenario, it was assumed that communication was facilitated among the stakeholders and concerns of the stakeholders were addressed during the decision process. Therefore, the traffic load and PM emission criteria were omitted from the analysis. Based on the decision makers' ranking of criteria and applying the Expected Value method, the importance of criteria was derived. The PROMETHEE results showed a general agreement among stakeholders. The top ranked alternative in this scenario for all stakeholders were the same indicating that transparency and communication would help reaching consensus.

References

[1] Statistics Canada. Report on energy supply and demand in Canada – 2006 Revision; 2009 <http://www.statcan.gc.ca/pub/57-003-x/57-003-x2006001-eng.pdf>.

[2] Environment Canada. Energy consumption; 2008. <http://www.ec.gc.ca/soer-ree/English/Indicator_series/new_issues.cfm?issue_id=11&tech_id=1#graph4>.

[3] Cuddihy J, Kennedy C, Byer P. Energy use in Canada: environmental impacts and opportunities in relationship to infrastructure systems. Can J Civ Eng 2005;32(1):1–15.

[4] Harvey LDD. A handbook on low energy buildings and district energy systems: fundamentals, techniques and examples. London (Sterling, VA): Earthscan; 2006.

[5] Gilmour B, Warren J. The new district energy: building blocks for sustainable community development. Online handbook; 2008, <http://cdea.ca/resources/highlights-guidelines-and-manuals/UES%20Handbook%20Final%2021-01-08.pdf/view>.

[6] Pohekar SD, Ramachandran M. Application of multi-criteria decision making to sustainable energy planning—a review. Renew Sustain Energy Rev 2004;8(4):365–81.

[7] Georgopoulou E, Lalas D, Papagiannakis L. A multicriteria decision aid approach for energy planning problems: the case of renewable energy option. EJOR 1997;103(1):38–54.

[8] Georgopoulou E, Sarafidis Y, Diakoulaki D. Design and implementation of a group DSS for sustaining renewable energies exploitation. EJOR 1998;109:483–500.

[9] Al-Shemmeri T, Al-Kloub B, Pearman A. Model choice in multicriteria decision aid. EJOR 1997;97(3):550–60.

[10] Xiaohua W, Zhenmin F. Sustainable development of rural energy and its appraising system in China. Renew Sustain Energy Rev 2002;6(4):395–404.

[11] Nigim K, Munier N, Green J. Pre-feasibility MCDM tools to aid communities in prioritizing local viable renewable energy sources. Renew Energy 2004;29(11):1775–91.

[12] Jaber JO, Jaber QM, Sawalha SA, Mohsen MS. Evaluation of conventional and renewable energy sources for space heating in the household sector. Renew Sustain Energy Rev 2008;12(1):278–89.

[13] Hobbs B, Horn G. Building public confidence in energy planning: a multimethod MCDM approach to demand-side planning at BC gas. Energy Policy 1997;25(3):357–75.

[14] Ramanathan R, Ganesh LS. Energy resource allocation incorporating qualitative and quantitative criteria: an integrated model using goal programming and AHP. Soc Econ Plann Sci 1995;29(3):197–218.

[15] Beccali M, Cellura M, Ardente D. Decision making in energy planning: the ELECTRE multicriteria analysis approach compared to a fuzzy-sets methodology. Energy Convers Manage 1998;39(16–18):1869–81.

[16] Mroz TM, Thiel T. Evaluation of a heating system for buildings using multiple criteria decision analysis. Arch Civ Eng 2005;51(2):281–98.

[17] Mroz TM. Planning of community heating systems modernization and development. Appl Therm Eng. 2008;28(14–15):1844–52.

[18] Pohekar SD, Ramachandran M. Multi-criteria evaluation of cooking energy alternatives for promoting parabolic solar cooker in India. Renew Energy 2004;29(9):1449–60.

[19] Haralambopoulos DA, Polatidis H. Renewable energy projects: structuring a multi-criteria group decision-making framework. Renew Energy 2003;28(6):961–73.

[20] Doukas H, Patlitzianas KD, Psarras J. Supporting sustainable electricity technologies in Greece using MCDM. Resour Policy 2006;31(2):129–36.

[21] Goumas M, Lygerou V. An extension of the PROMETHEE method for decision making in fuzzy environment: ranking of alternative energy exploitation projects. EJOR 2000;123(3):606–13.

[22] Brans JP, Vincke P. A preference ranking organisation method: (the PROMETHEE method for multiple criteria decision-making). Manage Sci 1985;31(6):647–56.

[23] Brans JP, Mareschal B. How to decide with PROMETHEE; 2008 <http://www.visualdecision.com>.

[24] Nijkamp P, Rietveld P, Voogd H. Multicriteria evaluation in physical planning. North-Holland: Amsterdam; 1990.

[25] FVB Energy Incorporation. Demand forecast for a neighbourhood energy utility. City of Vancouver, Vancouver, BC; 2006.

[26] Compass Resource Management Ltd. Business analysis for a neighbourhood energy utility. False Creek Precinct, Vancouver. Compass Resource Management Ltd., Vancouver, BC; 2006.

[27] BC's Ministry of Energy, Mines and Petroleum Resources & BC's Ministry of Environment, Lands and Parks. British Columbia greenhouse gas action plan: meeting the challenge of climate change; 1995.

[28] Sahlin J, Knutsson D, Ekvall T. Effects of planned expansion of waste incineration in the Swedish district heating systems. Resour Conser Recycl 2004;41(4):279–92.

[29] Rolfsman B. Combined heat-and-power plants and district heating in a deregulated electricity market. Appl Energy 2004;78(1):37–52.

[30] Streckiene G, Martinaitis V, Andersen AN, Katz J. Feasibility of CHP-plants with thermal stores in the German spot market. Appl Energy 2009;86(11):2308–16.

[31] FVB Energy Incorporation. Potential heat sources for a neighbourhood energy utility. False Creek Precinct, City of Vancouver, Vancouver, BC; 2006.

[32] Afgan NH, Carvalho MG, Hovanov NV. Energy system assessment with sustainability indicators. Energy Policy 2000;28(9):603–12.

[33] Cavallaro F, Ciraolo L. A multicriteria approach to evaluate wind energy plants on an Italian island. Energy Policy 2005;33(2):235–44.

[34] Trent Berry. Biomass permit analysis. City of Vancouver, Vancouver, BC; 2006.

[35] Humphries B, Koscher C, Walker J. Air quality assessment for the False Creek neighbourhood energy utility. Vancouver (BC): Levelton Consultants Ltd.; 2006.

[36] Yoo KH, Lee JS, Oh MD. Charging and collection of submicron particles in two-stage parallel-plate electrostatic precipitators. AS&T 1997;27(3):308–23.

Multi-Agent Simulation of Generation Expansion in Electricity Markets

Audun Botterud, *Member IEEE*, Matthew R. Mahalik, Thomas D. Veselka, Heon-Su Ryu, *Member IEEE*, and Ki-Won Sohn

***Abstract* — We present a new multi-agent model of generation expansion in electricity markets. The model simulates generation investment decisions of decentralized generating companies (GenCos) interacting in a complex, multidimensional environment. A probabilistic dispatch algorithm calculates prices and profits for new candidate units in different future states of the system. Uncertainties in future load, hydropower conditions, and competitors' actions are represented in a scenario tree, and decision analysis is used to identify the optimal expansion decision for each individual GenCo. We test the model using real data for the Korea power system under different assumptions about market design, market concentration, and GenCo's assumed expectations about their competitors' investment decisions.**

***Index Terms*—Electricity Markets, Generation Expansion, Agent-Based Modeling, Probabilistic Dispatch, Decision Analysis.**

I. INTRODUCTION

Traditional generation expansion planning in electrical power systems is usually based on centralized least-cost planning, subject to reliability constraints. However, the centralized least-cost planning approach does not reflect how investment decisions are made in today's electricity markets, where several generating companies (GenCos) are competing with each other, both in short-run operations and long-run investments. Some would argue that a well-functioning electricity market would converge toward the optimal expansion plan from a system's perspective in the long run. A competitive market should provide correct investment incentives through price signals in short- and long-term markets. Others, however, would contend that the independent and decentralized decision-making process in restructured electricity markets leads to suboptimal expansion plans. Several important factors, such as market power, limited information about competitors current and future actions, low demand-side participation, inadequate market design, and increased financial risk, cause the expansion decisions to deviate from the optimal plan.

The submitted manuscript was created by UChicago Argonne, LLC, under Contract No. DE-AC02-06CH11357 with the U.S. Department of Energy.

A. Botterud, M.R. Mahalik, and T.D. Veselka are with the Decision and Information Sciences Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, USA (e-mails: abotterud@anl.gov, mahalik@anl.gov, tdveselka@anl.gov).

H-S. Ryu and K-W. Sohn are with the Power Planning Department and the Market Operations Department, Korea Power Exchange, 167, Samseong-Dong, Gangnam-Gu, Seoul 135-791, Korea (e-mails: maxima@kpx.or.kr, kwson@kpx.or.kr).

It is still too early to judge all the long-term consequences of power industry restructuring from historical data, because of the large time horizon involved in capacity expansion. However, there is clearly a need to develop new modeling approaches to improve our understanding of the long-term price and investment dynamics in restructured electricity markets.

From a modeling point of view, the centralized least-cost expansion planning perspective is convenient, since one objective function can be used to optimize the entire system. The generation planning problem can then be solved using standard optimization methods, such as dynamic programming. Several models have been developed for traditional least-cost generation planning, e.g. the WASP model [1]. Modeling of generation investments in restructured electricity markets is a fairly new area of research. It is a challenge to model the strategic business interactions between competing GenCos, and at the same time include sufficient detail in the technical representation of the power system. In the literature we find some examples of generation planning models for restructured electricity markets based on game theory [2]. System dynamics [3], real options theory [4], and agent-based modeling [5] have also been applied to analyze GenCos' investment decisions.

In this paper we, present a novel model for analyzing generation expansion decisions in electricity markets. We use agent-based modeling to simulate the decentralized decision-making processes underlying GenCos' investment decisions. In the model, GenCos are represented as independent and decentralized agents interacting with each other in a complex, multidimensional environment. A convolution algorithm is used to simulate the market operation of current and future generation system configurations, taking into account thermal generators' forced outage rates and scheduled maintenance needs. A peak-shaving algorithm is used to represent hydropower dispatch. Uncertainties in future load growth, hydropower availability, and competitors' expected future investment decisions are represented with scenario trees. Finally, decision analysis is used to model each individual GenCo's investment decision. The model can simulate generation expansion decisions over a multiyear time period.

The paper has the following structure. First, we describe the algorithm of the new multi-agent generation expansion model. We then present results from testing of the model using realistic data for the power system in South Korea, where generation expansion decisions are simulated under a number of different assumptions about market structure and design. Conclusions and directions for future work are provided in the end.

II. Model Description

Argonne National Laboratory has spent several years developing an agent-based model for electricity markets. So far, the main focus of the Electricity Market Complex Adaptive Systems (EMCAS) model has been on short-term hourly simulations (see [6] and [7] for a description of EMCAS, with an example of an application in [8]). The development of the expansion model presented in this paper facilitates analysis of long-term investment aspects within the same multi-agent modeling framework.

A. *Overview of the Expansion Model*

The overall structure of the simulated decision making process is illustrated in Fig. 1. The model runs for a number of decision years. Within each decision year, each GenCo makes a forecast of future market conditions, in which it assesses potential investments in new generation capacity, taking into account the impact on the profitability of its own existing portfolio of plants. The actual system developments may deviate from the GenCos' expectations. Hence, as in the real markets, optimality is not guaranteed, neither from a GenCo nor from a system perspective. Currently, the GenCos consider investments only in thermal generation during the simulation. However, investments in other technologies, such as hydro- and wind-power, may be specified as external inputs. Plant retirements, regardless of the technology type, can also be specified as external inputs.

After GenCos have formulated expansion plans in a decision year, the plans are made publicly available. Based on an assumed technology-specific construction period, the new units come online in the system at a future date. For each decision year, the GenCos learn about the actions of their competitors through their announcements of new investment projects. The latest information about the current system, capacity retirements, and announced capacity additions are always taken into account by the GenCos in the assessment of new investment alternatives. However, information about expansion plans is not shared among the GenCos within the decision year. Hence, competitors' expansion decisions may be very different from what the individual GenCos originally forecasted.

A decision year simulation is performed to evaluate prices, GenCo profits, and generation system reliability within the decision year, based on the current system configuration. At the end of the decision year, expansion decisions of all GenCos are aggregated and the system is updated with the latest information about completed projects, retirements, and new announcements. Load growth rates are exogenous inputs to the model. There are two types of load growth rates: the first is the actual load growth rate, which is simulated for each decision year. This rate is unknown to the GenCos, until the after a decision year has been simulated. The second rate is used by the GenCos in their forecasts and investment decision making and can consist of several scenarios, as explained below. It may deviate from the actual simulated load growth.

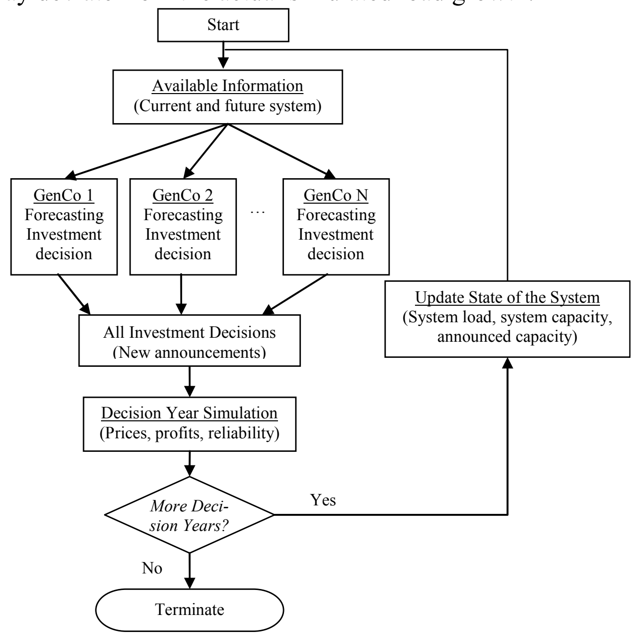


Fig. 1. Overview of simulated decision-making process in multi-agent expansion model.

Each GenCo uses the same general decision model. However, several of the parameters that go into the model, such as a GenCo's decision preferences, the probabilities of load and hydropower scenarios, and available investment alternatives may vary among the companies. At the same time, the GenCos will learn about the decisions of their competitors during the simulation. This will also contribute to differentiate the investment strategies applied by the various companies in the system. Another key component to investment decisions is that GenCos may have distinctly different portfolios of existing supply assets. One GenCo may estimate that it is profitable to build a certain new technology because it will have little or no impact on the profitability of its existing supply portfolio, while another GenCo may estimate that the same technology would not be profitable because it would have a large detrimental impact on its existing assets.

B. *Uncertainty in Load Growth and Hydropower Generation*

Load growth is an important driver for future prices and the need for capacity expansion in the system. There is usually considerable uncertainty regarding future load levels in the system. This uncertainty is represented in the model through scenarios describing the annual percentage change in the system load for each year in the forecast period. The hourly loads specified for the initial year are scaled for each forecast year depending on the load growth scenario.

In a system with considerable hydropower, the uncertain inflow of water into the system is also an important factor that must be considered. This uncertainty can be modeled by specifying a number of hydropower availability scenarios for all hydropower plants in the system. In the dispatch algorithm, the hydro generation is modeled with a peak-shaving logic, where the amount of peak-shaving within each week depends on the hydropower scenario. Other renewable and non-dispatchable resources (e.g. wind, biomass, waste) are represented with an hourly time series for generation that is subtracted from the forecasted loads.

C. Competitor Expectations

In a decision year, the GenCos know all the existing capacity in the system and what has been announced by their competitors in previous years. However, when forecasting prices and profits over the lifetime of a new unit, the GenCos also need to anticipate what investments their competitors are likely to make further into the future, i.e. beyond what has already been announced. To model future investments from other GenCos, we assume that each GenCo has an aggregate view of how much new capacity the rest of the market will add to the system over time. The representation of others' anticipated investments consists of the total installed capacity and the technology mix of the new competitor plants. Both of these characteristics are, of course, highly uncertain at the time a GenCo makes its investment decision. We, therefore, model the anticipated installed capacity and technology mix from others as scenarios. The first competition layer represents the anticipated total amount of new installed capacity that competitors will build over time. The second competition layer represents the technological composition of this new competitor capacity. The result is a scenario-tree structure used to represent uncertainties in load growth, hydropower conditions, and competitors' expansions (Fig. 2).

The new capacity built by others is linked to a GenCo-specific system reserve margin target that represents a GenCo's expectation about future system reserve margins. A GenCo assumes that the total investments from the competitors will cover a certain percentage of the required capacity needed to maintain the system reserve target. The competitor capacity type can be one of several specified candidate technologies. Hence, each GenCo can derive a complete competitor expansion plan based on the parameters described above for all scenarios in Fig. 2.

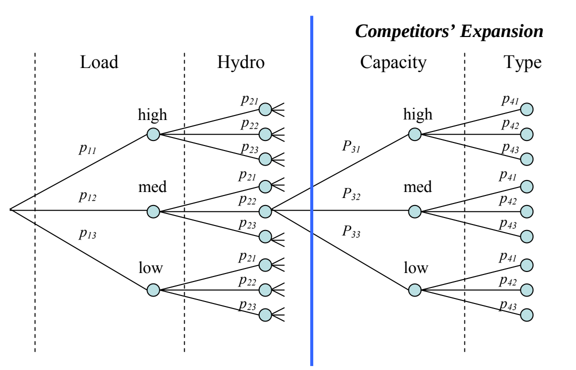


Fig. 2. Scenario tree for uncertainties in load growth, hydro conditions, and competitors' expectations.

To save computation time in the dispatch simulation we assume that all GenCos use the same scenario definitions for load growth and hydropower conditions. In the GenCos' representation of competitors' decisions, the capacity levels are defined individually for each GenCo, as explained above, whereas the definition of capacity types is the same for all GenCos. However, the probabilities are specified individually for each GenCo over all four layers in the scenario tree. The scenario probabilities are currently exogenous inputs to the model and kept constant during the simulation. However, in future versions the idea is that the GenCos can learn and update these probabilities during the simulation.

Prices and profits must be calculated for all GenCos' units over all leaves in the scenario tree. Computational efficiency is, therefore, of major importance in the dispatch algorithm, which is outlined below.

D. Probabilistic Dispatch: Prices, Profits, Reliability

A probabilistic dispatch algorithm based on the traditional Baleriaux-Booth method [9] is used to model forced outages in thermal units and their impact on prices and reliability for a given system configuration. An equivalent load, L_e, represents the load that a unit will serve accounting for outages of units that are lower in the merit order dispatch. L_e can be defined as:

$$L_e = L_s + L_r \tag{1}$$

where

L_s	original system load	[MW]
L_r	forced (random) component of unit outages	[MW]

The cumulative probability distribution of the equivalent load is found by convoluting each thermal unit's forced outages into the original system load. This is done in merit order, based on the units' marginal production cost. A single load level is evaluated at a time. Hence, the cumulative distribution for the initial load is a vertical line (Fig. 3). As units are convoluted into this curve, the resulting equivalent load curve is transformed into one that has an upper elongated tail. The resulting cumulative probability distribution function for L_e is calculated recursively, based on (2). The probability of a thermal unit being the marginal producer in the system is also determined. We assume that all thermal units bid their marginal production cost. Therefore, the price probability is given by (3). The probability of having energy not served (ENS) and, therefore, price being equal to a regulatory price cap, P_{CAP}, is given by (4). An illustration of the convolution process and the price distribution calculation for a given load level, L_s, in a simple system with two units of equal size is shown in Fig. 3.

$$F_n(L_e) = p_n F_{n-1}(L_e) + q_n F_{n-1}(L_e - C_n) \tag{2}$$
$$f(MC_n) = F_{n-1}(TC_{n-1}) - F_n(TC_n) \tag{3}$$
$$f(P_{CAP}) = F_N(TC_N) \tag{4}$$

where

$F_n(L_e)$ cumulative probability distribution for L_e, $F_0(L_e \leq L_s) = 1$, $F_0(L_e > L_s) = 0$

$f(MC_n)$ probability price equals marg. cost unit n, MC_n

$f(P_{CAP})$	probability price equals price cap, P_{CAP}	
C_n	capacity of generating unit n, n = 1..N	[MW]
TC_n	sum capacity of generating units 1.. n	[MW]
p_n	probability unit n in operation	
q_n	probability unit n on forced outage (p_n+q_n=1)	

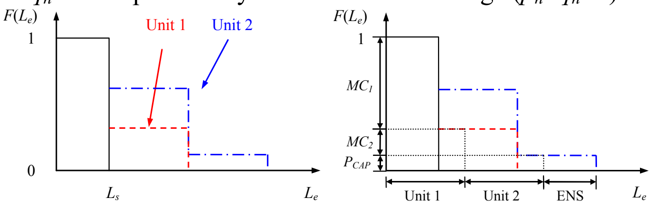


Fig. 3. Calculation of cumulative distribution for equivalent load (left), and price distribution (right) for a given load level, L_s. ENS = Energy Not Served.

Probabilistic convolution is done for each month. Planned maintenance of the thermal units is taken into account. A monthly maintenance scheduling routine is used, which minimizes the maximum monthly loss of load probability in each year. Hydropower and non-dispatchable generation is subtracted from the original hourly loads within the month, using a peak-shaving algorithm for hydro power. Price distributions are calculated for a sample of the resulting thermal loads, and the results are aggregated into a monthly price distribution. Note that it is necessary to perform the recursive convolution only for the maximum thermal load in the month over all load/hydro scenarios. The resulting convolution curves are stored in tables with small discrete load steps. The price distribution for lower load levels can easily be derived from the probabilities stored in the convolution table for the maximum thermal load. Monthly price distributions are calculated for each load/hydropower scenario throughout the planning period, taking into account the monthly maintenance plan. However, the underlying convolution tables need to be updated only for each new forecast period (i.e. when the portfolio of thermal plants in the system changes, due to either retirements or new announced capacity). Furthermore, all GenCos use the same convolution tables when evaluating profitability of new units. This greatly improves the computational efficiency of the model.

The aggregated monthly price distributions are used to calculate the profitability of new candidate units. In addition, the impact of a new unit on the profitability of a GenCo's existing thermal and hydropower units due to potential reduction in prices is also estimated. The resulting cost and revenues for a new candidate plant, discounted over all months in the payback period and calculated for all scenarios in the scenario tree (Fig. 2), are used as input to the GenCo's investment decision.

The unannounced capacity in the GenCos' expectations about competitors' future investments is not included in the convolution procedure described above, as this information is GenCo specific. However, an approximation is made to take into account how this capacity influences prices and candidate unit profit in the different competitor expectation scenarios for each GenCo. A GenCo's own unannounced new capacity is handled in a similar manner within each decision year.

E. Decision Analysis

Decision analysis is used to identify the preferred investment decision for each individual GenCo. Multi-attribute utility theory (MAUT) is used to calculate the expected utility from all possible investment decisions, including not investing at all. The optimal decision according to MAUT is to choose the alternative with the highest expected utility. The underlying assumption is that a decision maker's preferences can be quantified in terms of a multi-attribute utility function. The utility function takes into account the decision maker's risk preferences and the trade-offs between different objectives. The theoretical background for MAUT is thoroughly described by Keeney and Raiffa in [10].

We use the additive form of the multi-attribute utility function, i.e., the total utility for an alternative equals the weighted sum of the single attribute utilities, as shown in (5). An exponential form is used for the single-attribute utility functions, as shown in (6). The corresponding risk parameters indicate risk preferences for the individual attributes. If β is zero, the decision maker is risk-neutral. A negative β means risk aversion, whereas a positive β means a risk-seeking attitude. The upper and lower limits of each attribute refer to the maximum and minimum values considering all candidate technologies.

The trade-off weights and the risk parameters are specified as input for each GenCo and can be used to represent different preferences among the market participants.

$$u(\boldsymbol{x}) = \sum_{i=1}^{m} k_i \cdot u_i(x_i) \tag{5}$$

$$u_i(x_i) = 1/(1-e^{\beta_i}) \cdot \left\{1 - e^{\beta_i (x_i - \underline{x}_i)/(\overline{x}_i - \underline{x}_i)}\right\} \tag{6}$$

where

$u(\boldsymbol{x})$	total utility for attribute set $\boldsymbol{x} = x_1, x_2, ..., x_m$
$u_i(x_i)$	utility for single attribute, i = 1,2, ..., m
k_i	trade-off weight, attribute i
β_i	risk parameter, attribute i
$\overline{x}_i$	upper limit, attribute i
$\underline{x}_i$	lower limit, attribute i

Currently, three attributes can be taken into account in the model: 1) Profit over unit payback period, i.e. (discounted revenue) – (discounted cost); 2) Profit ratio over unit payback period, i.e. (discounted profit)/(discounted cost); and 3) Market share, measured in terms of capacity at a certain time in the future. These attributes are calculated for all the leaves in the scenario tree (Fig. 2). The expected utility for an alternative is then calculated over all leaf scenarios based on the probabilities in the tree.

In each decision year, a GenCo must decide how many units to build of each candidate unit technology type. The number of possible alternatives can therefore become very high. To reduce the discrete search space, we limit the GenCo to choose only one plant at a time. In the algorithm, the GenCo therefore calculates the expected utility for one unit of all its candidate technologies. The unit with the highest expected utility is chosen. The process is repeated with plants already selected added to the GenCo's fleet of existing units. The iterative selection process continues within the same decision year until the GenCo's choice is to not build more plants, or until an imposed constraint on the GenCo's annual capacity expansion is reached.

F. Flowchart of Expansion Code

A flowchart describing the main parts of the multi-agent expansion code is given in Fig. 4. Note that in the decision year loop, steps 3–6 are done only once in each decision year, and the results are used by all GenCos. In contrast, the calculation of competitor expectations (step 2) and candidate unit evaluation and decision analysis (step 7) are done individually for each GenCo.

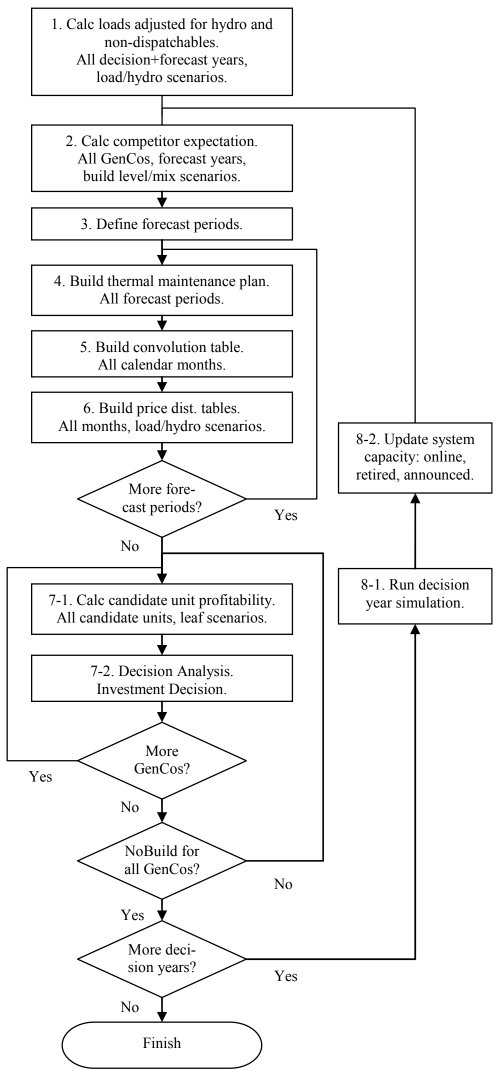


Fig. 4. Flowchart of multi-agent expansion algorithm.

III. Case Study: Korea Power System

We have tested the new expansion model in collaboration with Korea Power Exchange (KPX), using real data for the Korea power system. A selection of results is presented below. Note that the only purpose of the case study was to test the new EMCAS expansion model. None of the results are used for actual planning purposes by KPX.

A. Assumptions for Korea Power System

The technical specifications for the power system and the load forecast assumptions are based on the 3rd Basic Plan for Korea Long-Term Power Supply and Demand [11]. A 15 year simulation period is used, starting from 2006. Table 1 shows the expected long-term load growth for the Korea power system within this period. The peak load is expected to be gradually saturated in the far future. In the expansion model, we use the growth rates in Table 1 in the decision year simulations, whereas the GenCos' forecasted growth rates, which are used as input to their price and profit projections, are set to 2.5% until 2014 and 2.0% afterwards.

TABLE 1
Long-Term Load Forecast for Korea Power System

Year	Peak(MW)	Growth Rate
'05	54,631	-
'06	56,681	3.8%
'08	61,132	3.5%
'10	64,605	2.6%
'12	67,120	1.8%
'14	68,832	1.1%
'16	70,049	0.8%
'18	71,025	0.7%
'20	71,809	0.6%

The installed capacity in the Korea power system in 2005 was about 62.7GW. An additional 20.8GW is under planning and construction and will be built by 2020 (Table 2). Nuclear and coal capacity each account for about 30% of total capacity. About 20 % of capacity consists of Natural Gas Combined Cycle (NGCC) plants. There is also a small amount of hydropower generation and other renewable generation in the system. The capacity in Table 2 comprises a total of 127 units, which are all represented individually in the input data.

TABLE 2
Existing Capacity and Units Under Construction [MW, %]

Technology	Existing Cap. (as of 2005)	Under Construction (until 2020)	Retiring Cap.	Share (%)
Nuclear	17,716	6,800	-	31.7
Coal	17,965	6,540	1,525	29.7
NGCC	16,449	1,500	1,537	21.2
Oil	4,662	200	2,643	2.9
Hydro	3,829	2,400	-	8.0
COGEN	1,382	1,983	-	4.3
Renewable	210	1,433	-	2.1
Other.	52	9	-	0.1
Sum	62,265	20,865	5,705	100
TOTAL SUM	77,425			

There are many existing GenCos in the Korea power market. As shown in Table 3, the share of the nuclear company, KHNP, is about 32%, and the share of the five major coal companies is about 53% of the installed capacity. The capacity shares of the other existing companies are small. For simplicity, we use an aggregate representation for the small companies. In addition to the existing companies, we also include two new GenCos (new entrants) in the expansion simulations.

We used five candidate units (Table 4), whose bid prices are based on the production cost. It is assumed that the nuclear company can build only Nuclear 1400. The five coal companies can build both Coal (870, 1000) and NGCC (500, 700). The NGCC companies and the new entrants can build NGCC (500, 700) only. The technical data of candidate units is shown in Table 4. Nuclear units have the highest capital and lowest operating cost, and vice versa for the NGCC unit. The expected forced outage rates (EFOR) are around 5% for all candidate units. A 7.5 % discount rate was used for all candidate units and GenCos.

TABLE 3
EXISTING COMPANIES AND THEIR CAPACITY SHARE (AS OF 2020)

Entrants	Name	Resource	#of units	Share(%)
Existing GenCos	KHNP	Nuclear & Hydro	26	32.4
	NADO	Coal & NGCC & PS	15	12.4
	JUBU	Coal & NGCC & PS	14	8.4
	SEBU	Coal & NGCC & PS	18	10.0
	NABU	Coal & NGCC & PS	19	10.7
	DOSE	Coal & NGCC & PS	20	11.5
	PSCP	NGCC	4	2.3
	GSEP	NGCC	2	1.3
	GSPW	NGCC	2	1.2
	MYUC	NGCC	1	0.7
	KPWR	NGCC	2	1.3
	SUJA	Hydro	1	1.3
	COGEN	NGCC & Oil	1	4.3
	Renewable	Wind, LFG, etc.	1	2.1
	Others	Others	1	0.1
New Entrants	DARM	NGCC	0	0
	SKES	NGCC	0	0

TABLE 4
CANDIDATE UNITS CHARACTERISTICS. EXCHANGE RATE: 1 KWON ≈1$.

Technology	Fuel Cost (kWon/Gcal)	Payback Period(year)	Construction Period (Year)
NGCC 500	35.4	20	3
NGCC 700	35.4	20	3
Coal 870	9.5	25	7
Coal 1000	9.5	25	7
Nuclear 1400	1.4	30	10

We used a simplified scenario tree structure in the simulations presented here, with only one load growth scenario and one hydropower scenario (based on actual hydropower data for 2005). For the competitor expectations, we used one build level scenario, and three build type scenarios (NGCC 500, NGCC 700, and Coal 1000 with equal probability for each type). Other simulation parameters are summarized in Table 5 (base case). The GenCo's Own Build Limit is a constraint on how much each GenCo can build within each decision year, as a percentage of the total capacity required in the system to meet the expected reserve margin.

TABLE 5
SIMULATION PARAMETERS

Parameter	GenCo		
	NGCC	Coal	Nuclear
Reserve Margin Parameter	30%	30%	30%
Genco's Own Build Limit	12%	12%	12%
Competitor Expansion	100%	95%	55%
Decision Analysis Attribute	Profit Ratio	Profit Ratio	Profit Ratio
Risk Preference	Neutral	Neutral	Neutral

B. Case Study Simulations and Results

We first simulated a base scenario, where the input parameters, as shown in Table 5, were calibrated to obtain results similar to a reference expansion plan for Korea from the WASP model [11]. A number of additional scenarios were simulated, where results were compared to the base case. Below we present results from sensitivity analyses of the energy market price cap, the GenCo's expectation about competitors' future expansion decisions, and the effect having no any new entrants investing in new capacity.

1) Base Case

The simulated generation capacity expansion in the base case is shown in Fig. 5. We can see that the GenCos invest in the NGCC 700, Coal 1000, and Nuclear 1400 technologies. The two new entrants (SKES, DARM) build most of the new NGCC capacity. JUBU is the GenCo with most coal expansion, whereas KHNP builds two nuclear plants, which come online toward the end of the simulation period. From the simulated prices and reserve margin (Fig. 6), we see that the price gradually decreases and stabilizes around 60 kWon/MWh, whereas the reserve margin grows towards a level of approximately 20%.

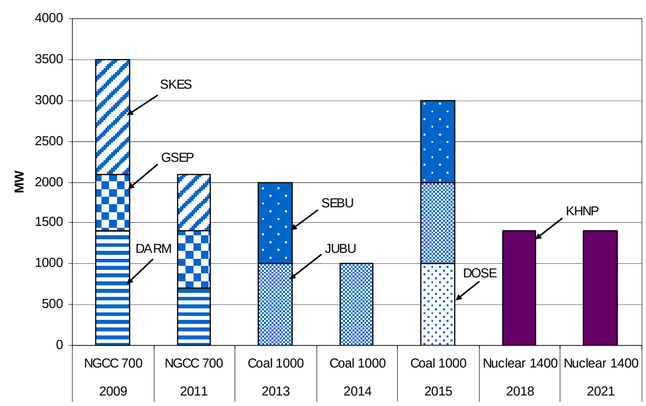


Fig. 5. Base case expansion by technology, GenCo and online year.

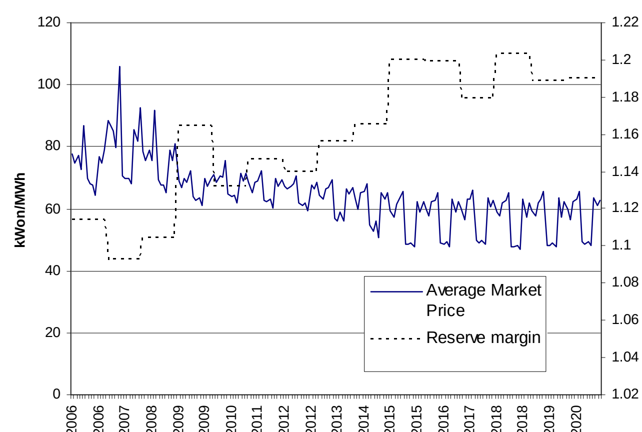


Fig. 6. Average monthly market price and annual reserve margin. Base case.

2) Sensitivity to Lower Energy Price Cap

The level of the energy price cap is very important for the incentive to invest in new generation capacity, since it determines the price and GenCos' income during periods with shortage of supply. Ideally, the price cap should be set to a high value equal to the value of lost load. However, regulators tend to set a lower price cap in electricity markets to avoid very high prices. In the base case we used a price cap of 999 kWon/MWh. We repeated the simulations with lower price caps to analyze the effect on expansion decisions, prices, and system reliability. Table 6 shows that a lower price cap reduces the investments in new generation capacity. This is because GenCos are less willing to invest in new generation capacity due to lower expected profitability. Investments in new NGCC plants seem to be most sensitive to the price cap, probably because this technology is dispatched less than coal and nuclear plants and is, therefore, more dependent on the profit during hours of scarcity. Furthermore, the investment decisions of new entrants are apparently more sensitive to the price cap than are those of the existing GenCos.

TABLE 6
EXPANSION BY GENCO FOR DIFFERENT PRICE CAPS

Genco	Base case NG7/ CO10/ NU14	Price cap = 750 NG7/ CO10/ NU14	Price cap = 500 NG7/ CO10/ NU14	Price cap = 300 NG7/ CO10/ NU14
New Entrants	6 / 0 / 0	4 / 0 / 0	4 / 0 / 0	0 / 0 / 0
Existing NGCC	2 / 0 / 0	2 / 0 / 0	1 / 0 / 0	0 / 0 / 0
Existing Coal	0 / 6 / 0	0 / 6 / 0	0 / 4 / 0	0 / 1 / 0
Existing Nuclear	0 / 0 / 2	0 / 0 / 2	0 / 0 / 2	0 / 0 / 1
Sum(MW)	14,400	13,000	10,300	2,400

Fig. 7 shows that simulated prices go up as a function of lower price cap, particularly with a price cap as low as 300 kWon/MWh. Hence, the simulations show that a regulatory policy of setting a low price cap, which aims to protect the end-users from high prices in the short-run may, in fact, lead to increasing prices in the long run because of a lower rate of investments. The simulated reserve margin also goes down, and in the 300 kWon/MWh scenario it actually drops to a level close to zero. The results illustrate the importance of designing a market with adequate incentives for investments in new generation capacity. An interesting extension of the analysis would be to consider the effect on investments from different capacity adequacy policies, such as capacity markets.

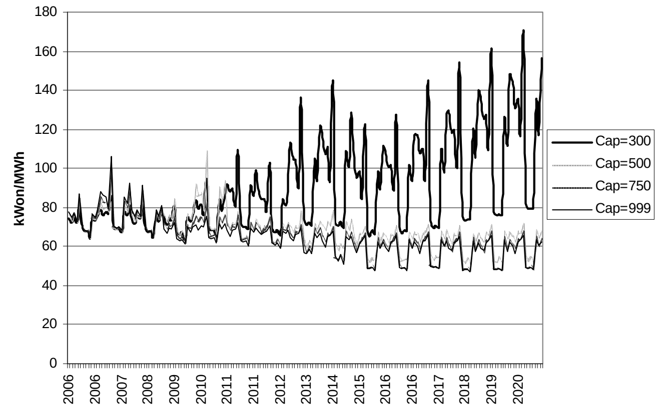


Fig. 7. Average monthly market price for different energy price caps.

3) Sensitivity to Competitor Expansion Expectations

The GenCos' expectations about competitors' expansion plans are important for their investment decisions, as outlined in Section II.C. To study the representation of competitor expectations in more detail, we changed the competitor unannounced expansion level parameters for the six coal GenCos. It was set to 95% for these GenCos in the base case, i.e. each GenCo expects that future unannounced expansion from all competitors will add up to 95% of what is required to meet the expected system reserve margin of 30% (Table 5).

When the competitor expansion parameter for coal GenCos is reduced to 90%, the level of investment for these companies increases compared to the base case (Fig. 8). This is because the coal GenCos now forecast lower rates of investment from their competitors, which in turn means that their projections of future prices and profits from ther own units increase. In contrast, when the competitor expansion expectation is increased to 100%, the coal GenCos build less capacity (Fig. 9). In fact, GenCo NADO and DOSE, whose existing capacities are higher than those of other coal GenCos, will build no new units. At the same time, other GenCos invest in more NGCC capacity than in the base case, which makes up for parts of the reduction in the new coal capacity.

The simulated prices and reserve margins are also affected by the changes in the competitor expectation parameter. Prices go down and the reserve margin goes up compared to the base case in the 90 % competitor expectation scenario, and vice versa for the 100% scenario.

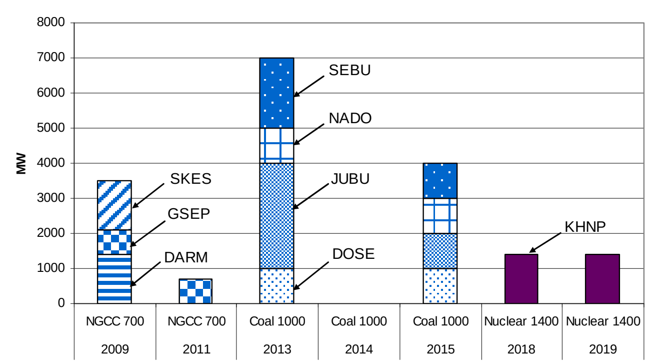


Fig. 8. Expansion by technology and GenCo with coal GencCs' competitor expansion parameter reduced to 90%.

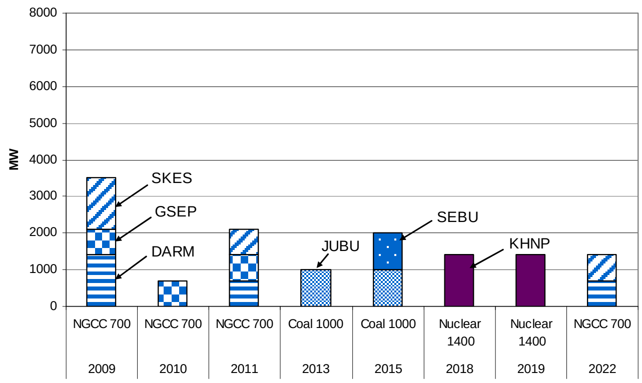


Fig. 9. Expansion by technology and GenCo with coal GenCos' competitor expansion parameter increased to 100%.

4) No New Entrants

Finally, we looked at the effect of removing the two new entrants (DARM and SKES) from the expansion simulation. It turned out that this had a profound effect on the results; the total level of new investments decreased dramatically. No new NGCC plants were built (i.e., a reduction from eight to zero NGCC 700 plants compared to the base case). At the same time, the number of new Coal 1000 plants dropped from six to four. KHNP still builds the two new nuclear plants, although they come online one and two years later than in the base case.

The reduction in new capacity leads to a major increase in prices, as shown in Fig. 10. The results from this scenario serve to illustrate the important role of new entrants in electricity markets. The new entrants can clearly lower the thresholds for investment and thereby contribute to keep prices at a competitive level.

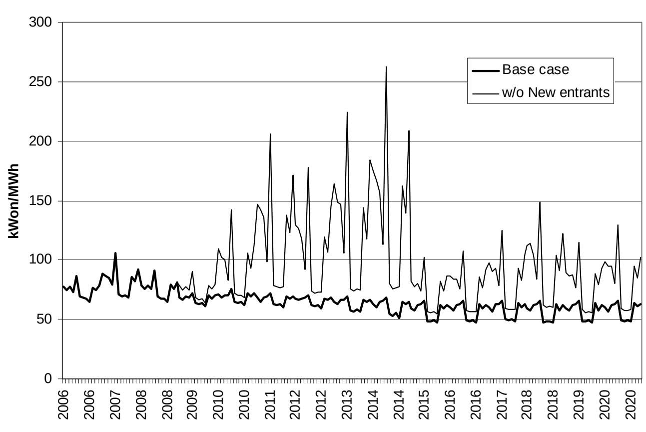


Fig. 10. Simulated average market price in scenario without new entrants.

IV. Conclusion

The multi-agent expansion model presented in this paper simulates the complex interaction between decentralized and profit-maximizing GenCos in restructured electricity markets. The presented results from test simulations of the Korea power system shows that the model can provide important insights into the long-term development of generation investments, prices, and reliability in real-world systems. Important issues regarding market design, GenCo decision preferences, and market concentration, can be analyzed. Such results can not be obtained with traditional generation expansion models.

We see a number of interesting extensions to the model, including: 1) Revision of the probabilistic dispatch logic to account for strategic bidding; 2) Simulate the effect of different capacity adequacy policies, such as installed capacity markets; 3) Model transmission constraints and location of new generating plants; and 4) Introduce more advanced learning and adaptation, so that GenCos adjust their forecast of the future depending on what they learn during the simulation.

V. References

[1] International Atomic Energy Agency (IAEA), "Wien Automatic System Planning (WASP) Package: A Computer Code for Power Generating System Expansion Planning. Users' Manual", IAEA, Vienna, 2001.

[2] P.-O. Pineau and P. Murto, "An Oligopolistic Investment Model of the Finnish Electricity Market", *Annals of Operations Research*, Vol. 121, pp.123–148, 2003.

[3] A. Ford, "Waiting for the Boom: A simulation study of power plant construction in California", *Energy Policy*, Vol. 29, pp. 847-869, 1999.

[4] A. Botterud and M. Korpås, "A stochastic dynamic model for optimal timing of investments in new generation capacity in restructured power systems", *Int. Journal of Electrical Energy and Power Systems*, Vol. 29, pp. 163-174, 2007.

[5] E. Gnansouno, J. Dong, S. Pierre, A. Quintero, "Market Oriented Planning of Power Generation Expansion using Agent-Based Model", In *Proc. of 2004 Power Systems Conference and Exposition*, Vol. 3, pp. 1306 – 1311, New York, NY, USA, 2004.

[6] G. Conzelmann, M. North, G. Boyd, V. Koritarov, C. Macal, P.R. Thimmapuram, T.D. Veselka. "Simulating Strategic Market Behavior Using an Agent-Based Modeling Approach. In *Proc. of the 6th European IAEE conference*, Zurich, Switzerland, 2004.

[7] V. Koritarov, "Real-World Market Representation with Agents: Modeling the Electricity Market as a Complex Adaptive System with an Agent-based Approach", *IEEE Power & Energy Magazine*, Vol. 2, No. 4, pp. 39-46, 2004.

[8] A. Botterud, V. Koritarov, P.R. Thimmapuram, "Multi-Agent Simulations of the Electricity Market in Central Europe", in *Proc. of the 26th USAEE/IAEE North American Conference*, Ann Arbor, MI, USA, 2006.

[9] R.R. Booth, "Power System Simulation Model Based on Probability Analysis", *IEEE Transactions on Power Apparatus and Systems*, Vol. 91, pp. 62-69, 1972.

[10] R.L. Keeney and H. Raiffa, *Decisions with Multiple Objectives. Preferences and Value Tradeoffs*, Cambridge University Press, 1993.

[11] Ministry of Commerce, Industry and Energy (MOCIE) Korea, "The 3rd Basic Plan of Long-term Electricity and Demand", to be published in Dec., 2006.

VI. Biographies

Audun Botterud (M'41576374) was born in Stange, Norway, on June 17, 1974. He received M.Sc. in Industrial Engineering (1997) and Ph.D. in Electrical Power Engineering (2003), both from the Norwegian University of Science and Technology. He was working at SINTEF Energy Research, Trondheim, Norway, from 1998 to 2000. In 2005 he joined Center for Energy, Environmental, and Economic Systems Analysis (CEEESA) at Argonne National Laboratory (ANL). His research interests include power system planning and economics, stochastic optimization, decision analysis, and agent-based modeling.

Matthew Mahalik is a software engineer who has been with CEEESA at ANL since 2000. Much of his work has focused on developing the GTMax and EMCAS models. Mr. Mahalik is proficient in the leading software languages, and is also proficient in the development of linear, non-linear, and integer programs that are frequently used for CEEESA models. Much of his recent focus was on agent-based modeling techniques and hydropower modeling in support of CEEESA's EMCAS software. Mr. Mahalik holds a degree in Computer Science from the University of St. Francis.

Thomas D. Veselka is an Energy Systems Engineer in CEEESA at ANL. He has provided technical support and managed projects related to electric utility systems and the environment for over 20 years. His work includes extensive hydropower systems analyses as related to power markets.

Heon-Su Ryu (M'40209135) was born in Korea, on June 03, 1969. He received B.S, M.S., and Ph.D. degrees in Electrical Engineering from Yonsei University, Korea, in 1992, 1996, and 2002, respectively. In 2002, he joined the Korea Power Exchange, and is now with the Power Planning Department. Dr. Ryu is a member of the IEEE Power Engineering Society. His interests are in the areas of generation and transmission planning and power system stability analysis.

Ki-Won Sohn was born in Korea, on April 24, 1975. He received B.S degree in Electrical Engineering from Dongguk University, Korea, in 1999. In 2001 he joined the Korea Power Exchange, and is now with Market Operation Department.

Modeling

externalities of Energy

ELSEVIER

Energy Policy 28 (2000) 1127–1136

ENERGY POLICY

www.elsevier.com/locate/enpol

Comparison of methodologies for externality assessment

Lotte Schleisner*

Systems Analysis Department, Risø National Laboratory, Box 49, DK-4000 Roskilde, Denmark

Received 13 July 1999

Abstract

The production of energy causes different kinds of damage to the environment depending on the specific type of technology used in producing a given energy supply. The common term that expresses the costs of these environmental damages is externalities. These costs are not included in the cost and price structure faced by the producer and the consumer. During the last few years, external costs related to power production technologies have been calculated making use of different methodologies. The external costs may turn out to be very different for the same fuel cycle depending on the methodology that has been used to assess the externalities. The article will focus on some of the most important reasons for differences in the numbers. To illustrate the importance of knowing the exact data and assumptions used, two studies using the same approach and with integrated computer models are compared. The models are based on the same concept with air dispersion modules and dose–response functions for the calculation of impacts. Although the models are comparable, the resulting external costs turn out to differ with a factor of five in the two studies for the same power plant due to different assumptions, different dose–response functions used and different impacts included in the studies. In the paper the most important differences to be aware of will be illustrated.

Keywords: Externalities; Dispersion; Mortality

1. Introduction

Choosing one energy option over another may influence many aspects of society and the environment, which should be accounted for if we want to procure the highest benefits for society. These impacts on society or environment, which have not previously been accounted for, are termed externalities. Externalities related to energy production are, in general, defined as the costs of the damages the energy production give rise to, that are not accounted for by the producers or consumers of energy. In other words, damages not reflected in the market price. Normally, when one thinks of externalities related to energy, the externalities are environmental. A frequently cited example is the loss of production in fisheries due to the spillage of pollutants in rivers, as a direct result of energy use. Public health, agriculture and ecosystems are other examples of aspects of society affected by the use of energy by others. The effects may be positive (external benefits) or negative (external costs), and their consideration may make some energy options more attractive than others in spite of their higher costs, or vice versa.

During the past several years, external costs related to power production technologies have been calculated using various methodologies. Some studies have used a "top–down" approach, while others are based on a "bottom–up" approach. Some studies are based on a life cycle assessment, including all impacts from the extraction of materials for manufacturing to disposal, while others assess only impacts related to the fuel cycle.

Differences in methodologies may also be noticed in the quantification and valuation procedure. Some studies rely on previous estimates, which are not site-specific; others rely on abatement costs. Still other studies use the damage function approach, where the impact from each burden related to the technology is identified, and the damage caused by the burden is quantified and monetised.

An important aspect to consider when estimating externalities based on earlier studies is that some studies include only regional and local impacts and do not take into account the global impacts related to greenhouse gases.

Considerable uncertainty arises when considering the global externalities regarding the time horizon for the greenhouse effect, choice of dose–response function and

* Tel.: + 45-4677-5115; fax: + 45-4677-5199.
E-mail address: lotte.schleisner@risoe.dk (L. Schleisner).

0301-4215/00/$ - see front matter
PII: S0301-4215(00)00084-7

monetisation values. Assumptions on famine and the monetisation of human life may be the totally dominant factor in estimating external costs.

The external costs may turn out to be very different for the same fuel cycle, depending on the methodology used to assess the externalities. As a consequence of this, it seems rather important to be aware of the methodology that is made use of, which impacts are included and the monetary values used in a given study before utilising the external costs from a specific study as a policy measure.

2. External costs from different studies

In Table 1, the external costs from seven studies have been compared to show the large differences in results from different studies, assessing the same fuel cycle. The studies have been chosen in order to cover old, well-known studies as well as new, unknown, but interesting ones. Some of the new studies are based on results from earlier ones, while others implement new ideas concerning the methodology. Most of the studies chosen are bottom-up studies. In the table, the results from the different studies have been translated to mECU/kWh year 1995.

The results from the US-EC study are very low. One reason for this is that the global warming effect is not included in the results. The results from the Swiss study are rather high compared with those from the other studies. Looking at the natural gas fuel cycle the results in the ExternE study are high compared to the other studies. The reason for this is that external costs related to CO_2 are included in this study, while CO_2 is not included in the New York study, and in the IEA study CO_2 is captured. Both the Swiss and Hohmeyer studies use a top-down approach, and both result in rather high external costs.

The comparison shows the importance of possessing knowledge of which kind of methodologies have been used, which impacts are included, etc., to explain why the numbers vary so much in different studies for the same fuel cycle. It is evident at the outset that the impacts, damages and externalities are very project specific. For example, emissions from an integrated gasification combined cycle coal plant are considerably lower than from a pulverised fuel plant. The specifications of the plant to be analysed will in this way affect the magnitude of the externalities. The specifications include installed pollution abatement technologies and their efficiencies as well as stack height and other source parameters that are used in atmospheric transport modelling. All of these parameters may be problematic when they are used to define future technologies.

3. Overview of two studies for detailed analysis

Two studies have been selected for detailed analysis. The two studies use basically the same methodology and include both computer models based on the same concept. Nevertheless, the comparison will show large differences in the external costs, pointing out the importance of knowing which data and assumptions the study is based upon.

The following overview gives a description of the two studies in regard to which methodology has been used, the impacts included, valuation methods, etc.

3.1. *ExternE national implementation*

The objective of the ExternE National Implementation project (CEC, 1995), (Schleisner and Nielsen, 1997a) has been to establish a comprehensive and comparable set of data on externalities of power generation for all EU

Table 1
External costs in mECU/kWh year 1995 for different fuel cycles for the studies chosen (1.2US$(1992) = 1 ECU (1995))[a]

	Coal /oil	Natural gas	Nuclear	Wind	Biomass
ExternE (Schleisner and Nielsen, 1997a)		NGCC: 7.1–80		Off-shore: 0.7–3.6 On land: 0.6–2.6	Biogas: 4.4–16.1
IEA (ETSU, 1994)	PC: − 0.6–5.4	NGCC: 0.6–2.3 IGCC: 1.6–3.9			
New York (Rowe *et al*., 1995)	PC: 4.5 FB: 0.9	NGCC: 0.2			Wood: 3.5
US-EC (Oak Ridge, 1992)	Coal: 0.4–1.0 Oil: 0.1–0.2	0.01–0.2	0.1–0.2		Wood: 1.6
India (Bhattacharyva, 1997)	Coal: 9.4				
Swiss (Ott, 1997)	Oil:99.6–158	NGCC: 68–101	4.8–11.5		
Hohmeyer (Hohmeyer, 1988)	Fossil fuels: 7.4–40	Fossil fuels: 7.4–40	7.8–78.3	On land:0.1	

[a]PC: pulverised coal, FB: fluidised-bed coal, NGCC: natural gas combined cycle, IGCC: integrated gasification combined cycle.

member states and Norway. The tasks include the application of the ExternE methodology to the most important fuel cycles for each country. The study is from 1997. A wide range of technologies has been analysed, covering more than 60 cases for 15 countries and 11 fuel cycles, including fossil fuels, nuclear and renewable technologies.

The methodology used for assessing externalities of the fuel cycles selected is a bottom-up methodology with a site-specific approach; i.e. it considers the effect of an additional fuel cycle, located in a specific place. The study estimates the damage costs related to different fuel cycles.

Quantification of impacts is achieved through the damage function approach, an approach that proceeds sequentially through a pathway, where emissions and other types of burdens, such as risk of accident, are quantified and followed through to impact assessment and valuation. The study employs a unified approach to ensure compatibility between results. This is being achieved through the use of the EcoSense model, which assesses the environmental impacts and resulting external costs from electricity generation systems. The computer model includes an environment database at both a local and regional level with data on population, crops, building materials and forests. The model system also incorporates two air transport models, enabling local and regional scale modelling to be made. Also a set of impact assessment modules, based on linear dose-response relationships, and a database of monetary values for different impacts are included in the model. There is no model for ozone included in the software, but ozone is estimated by assuming a simple relationship to NO_x.

Local, regional as well as global impacts are assessed. The monetisation values used for CO_2 have been estimated by employing two different models (Schleisner and Nielsen, 1997b, Appendix 1). Four different values have been used: 3.8 ECU/t CO_2, 18 and 46 ECU/t and 139 ECU/t CO_2. The estimate given in Table 1 is based on a CO_2 value of 18 ECU/t.

The underlying principle behind the economic valuation is to obtain the willingness to pay by the affected individuals in order to avoid a negative impact, or the willingness to accept the impact. A limited number of goods — crops, timber, building materials, etc., — are directly marketed. However, many of the more important goods of concern are not directly marketed. These include human health, ecological systems and non-timber benefits of forests. Alternative techniques have been developed for valuation of such goods, the main ones being hedonic pricing, travel cost methods and contingent valuation.

For the valuation of health risk, a value of 3.1 MECU has been used for the value of a statistical life (VSL). This value has been used for valuing fatal accidents and mortality impacts in climate change modelling. In the case of deaths arising from illness caused by air pollution, the years of life lost (YOLL) approach has been used. YOLL depends on a number of factors, such as how long it takes for the exposure to result in illness and the survival time for the individuals.

3.2. The New York electricity externality study

In this study (Rowe *et al.*, 1995), the EXMOD model is used, developed at the Tellus Institute in Boston. The model is built up in the same way as the European EcoSense model. The EXMOD model is an American model, which models air dispersion from locations in New York State to receptor cells throughout the northeastern US and eastern Canada. The study is from 1995.

It is a bottom–up study, also based upon "The damage function approach". Here damage costs are estimated for 23 new electric resource options within coal, oil, natural gas, nuclear, municipal solid waste, hydroelectric, biomass, wind, solar and demand side management. Default air emission rates, land use and other characteristics are specified for each facility in the model; however, these characteristics may be replaced. The air dispersion models in EXMOD are annual average and simple peak models used by US regulatory agencies. The two models are used to predict short-range dispersion changes (<50 km) and long-range changes (50–1500 km) covering local and regional ranges. Also, ozone models are included driven by changes in NO_x concentrations. So far the model does not compute CO_2 damages (i.e. EXMOD implicitly assumes 0\$/t CO_2). However, it is possible to include other values for CO_2.

Impact calculations are based on dose–response parameters in EXMOD with default high, central and low parameter values. Based on a review of the literature, EXMOD uses a central VSL estimate of 4.0 million \$ for individuals under 65 years, and a central estimate of 3.0 million \$ for those 65 years or older. The argument for the decrease with age of VSL is that the years of expected remaining life does decrease with age. Thus, life expectancy and health status tend to decrease with age, so that the quality of life is reduced. The model only includes VSL for the valuation of health damages. YOLL is not included.

The study uses control cost valuation to estimate the environmental cost associated with various air emissions. For other impacts the study uses the contingent valuation method.

3.3. Comparison of results from ExternE and the New York study

A comparison of the impacts and damage costs related to air emissions calculated in the two studies using the EXMOD model and the EcoSense model has been made for the same plant. The plant is a pulverised coal-fired plant with a capacity of 300 MW. The impacts from this plant have been calculated in EXMOD as well as in EcoSense. However, EXMOD only includes data for emission levels and population for a part of the USA, while EcoSense includes data only for Europe. Therefore, the same plant has been located in two different sites. Using EXMOD, the plant is located in the Capital District of New York State, which is a suburban site outside of Albany, while the same plant in EcoSense is located in Roskilde, Denmark. The external costs have been estimated using the methodologies and valuation methods described above. The costs estimated in Table 2 are central estimates.

Comparing the externalities for the same power plant estimated in the two studies using models based on the same concept, we see that the externalities are five times higher in the ExternE study than in the New York study. The difference in the external costs in the two studies reflects differences in impacts, differences in monetary values included in the two studies, differences in dose–response functions used, and finally differences in location of the plants.

The most apparent differences in the estimates are the extent of the greenhouse gas effect and the estimation of mortality. The greenhouse gas effect is not included in the New York study (by default monetised to zero), but in the ExternE study four different values of CO_2 have been estimated. In the above table a value of 18 ECU/t CO_2 has been used. Excluding the global warming effect the estimate in EcoSense is still three times higher than the estimate in EXMOD.

The external costs of mortality are four times as high in ExternE as in the New York study. EcoSense normally uses the YOLL approach; the figures in brackets are based on the VSL approach. In EcoSense mortality includes as well chronic as acute mortality, while EXMOD only covers acute mortality. Including as well chronic mortality as the global warming effect in EXMOD, the estimate in EcoSense becomes less than the estimate in EXMOD.

The emission of ozone causes mortality as well as morbidity cases for the population at large and also affects crops. The quantification and valuation of the emission of ozone has been included in the US EXMOD model, while in the case of the EU EcoSense model quantification and valuation of the emission of ozone has not been included. Instead, damages due to ozone are calculated based on the NO_x emissions related to the plant. The difference in crops is a result of ozone.

Other impacts are impacts like visibility loss, which is included in EXMOD, but not in the EcoSense model. Apart from global warming, human health is the dominant impact in both models.

Table 2
Central estimates of external costs for a coal-fired plant

Externalities	The New York study (mECU/kWh)	ExternE (mECU/kWh)
Human health	2.42	9.27
Mortality	1.71	7.97 (32.46)
Morbidity	0.70	1.30
Crops	0.002	0.134
Materials	0.10	0.22
Other impacts	0.32	0
Greenhouse gas effect	0	6.10
Total	2.84	15.72 (40.21)

4. Harmonising the estimates of effects on human health

In the following, the reasons for the differences in the estimates of the effect on human health using the two models will be analysed. The estimate of EcoSense as well as the estimate of EXMOD will be decomposed in order to harmonise the external costs from the two models.

In Fig. 1 the estimates of the effect on human health based on the two models have been decomposed in seven steps. The steps correspond to seven categories of differences:

- Inclusion of greenhouse gases.
- Inclusion of ozone.
- YOLL versus VSL.
- Inclusion of chronic mortality.
- Difference in monetised values.
- Difference in dose–response functions.
- Difference in morbidity impacts.

As illustrated in the figure the first four steps are related to the specific methodology and considerations used in the two studies, while the last three steps are related to the values used in the studies.

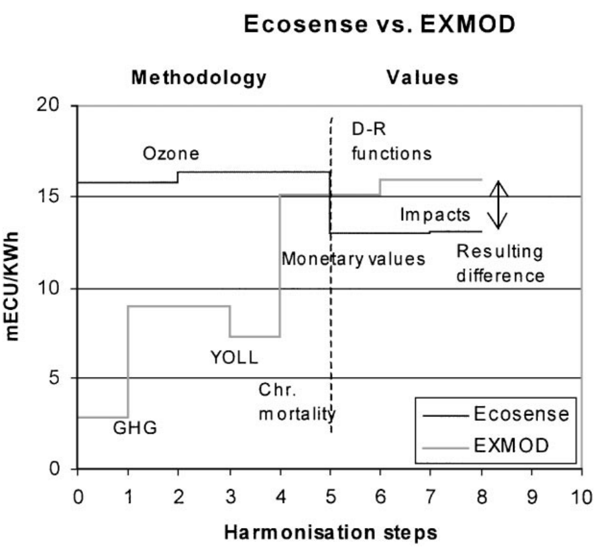


Fig. 1. Differences in estimates of the effect on human health.

Before decomposition EXMOD starts with a central value of 2.84 mECU/kWh, while EcoSense starts at a value of 15.72 mECU/kWh (Table 2).

The first step of harmonisation is the inclusion of greenhouse gases in the EXMOD estimate. Including the value of this impact from EcoSense makes the external costs in the EXMOD line rise considerably. The next step is the inclusion of ozone impacts in EcoSense; however, including the value of this impact from EXMOD results in only a small increase in the EcoSense line.

The third step of the harmonisation is estimating mortality using YOLL. Using YOLL instead of VSL in EXMOD lowers the external costs for EXMOD. The fourth step of harmonisation, which seems to be a very important methodological factor, is chronic mortality. Chronic mortality is not included in EXMOD; including the value of this impact from EcoSense increases the external costs in the EXMOD line considerably, and the external costs for EcoSense and EXMOD come rather close to each other.

The monetary values used in the two models differ in some cases. The fifth step of harmonisation is therefore to include the monetary values from EXMOD in EcoSense, lowering the EcoSense value, and the EXMOD values become higher than the EcoSense values. The final two steps toward harmonisation are to include the same dose-response functions and morbidity impacts in the two models, which is shown in the last part of the figure. However, these differences are small compared to the other differences.

Having adjusted for the above-mentioned parameters there is a difference of 3 mECU/kWh in the two estimates. Most of this difference may be attributed to the different locations of the plants, which affect population density and background level of emissions. This will be analysed in a later paragraph.

4.1. *Discussion of mortality and morbidity estimates*

A more detailed explanation for the differences in mortality and morbidity will be given below. The results of including greenhouse gases on human health have already been discussed above and will not be discussed in this section.

4.2. *Mortality*

In Table 3, the mortality impacts, monetary values and damage costs are shown as a central estimate for a pulverised coal-fired plant using the EXMOD model. For comparison, the monetary values used in EcoSense are used for the same impacts.

The last column shows the external costs for mortality calculated in EXMOD using the monetary values from EcoSense. Using these monetary values results in an increase in mortality damage costs of 17%, verifying that using the monetary value for VSL from EcoSense gives higher results.

The mortality impacts and damages have been calculated for the same plant using EcoSense (Table 4). EcoSense normally uses the YOLL approach with much smaller monetary values than VSL.

Comparing Tables 3 and 4, the external costs of mortality are more than 4 times higher in the ExternE study, although the YOLL approach is used. Using the VSL approach would result in 19 times as high external costs in ExternE as in the New York study.

Comparing the results the most obvious reason for the large difference in mortality impacts for the two models, beside the YOLL approach using other monetary values, is the inclusion of chronic mortality in EcoSense. (Chronic mortality is people dying from a long-cycle pain evoked by emissions.). In EXMOD mortality is only

Table 3
Mortality impacts and damages using EXMOD, central estimate

		EXMOD			EcoSense	Eco/EXMOD
		Impacts	Mon value (ECU)	Damage (mECU/kWh)	Mon value (ECU)	Damage (mECU/kWh)
Mortality over 65	NO_x	0.377	2.497 mio	0.5512	3.1 mio	0.6843
	PM_{10}	0.2139	2.497 mio	0.3127	3.1 mio	0.3882
	SO_2	0.0764	2.497 mio	0.1117	3.1 mio	0.1387
	Total	**0.6673**		**0.9757**		**1.2111**
Mortality under 65	NO_x	0.0336	3.330 mio	0.0655	3.1 mio	0.0610
	PM_{10}	0.01845	3.330 mio	0.0360	3.1 mio	0.0335
	SO_2	0.00453	3.330 mio	0.0088	3.1 mio	0.0082
	Total	**0.0566**		**0.1103**		**0.1027**
Mortality	Ozone	0.385	2.747 mio	0.6192	3.1 mio	0.6988
Mortality total		**1.1089**		**1.7052**		**2.0126**

Table 4
Mortality impacts using EcoSense, central estimate

		EcoSense		
		Impacts	Mon value (ECU)	Damage (mECU/kWh)
Chronic YOLL	PM_{10}	5.73	84330	0.48
	Nitrate	43.15		3.64
	Sulphate	43.97		3.71
	Total	**92.85**		**7.83**
Acute YOLL	SO_2	0.90	84330	0.14
Mortality total		**93.75**		**7.97**

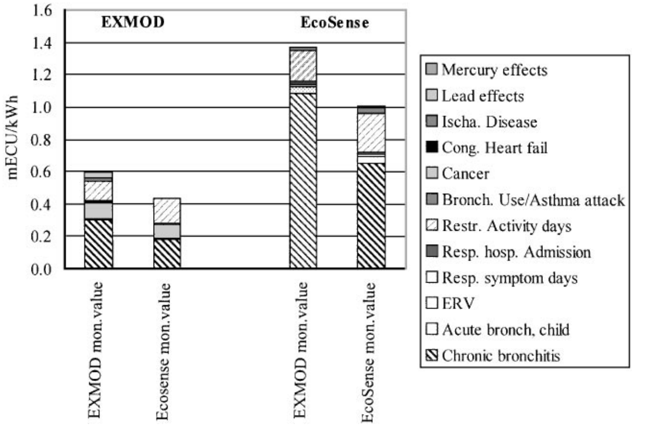


Fig. 2. Damage costs for morbidity calculated in EcoSense and EXMOD for the same power plant, central estimate.

related to acute mortality. Another important factor is that impacts due to ozone are included in EXMOD, but not in EcoSense (ozone has been included in a later version of EcoSense).

4.3. Morbidity

In order to compare the differences in externalities related to morbidity for the two computer models, the morbidity impacts, monetary values and damage costs have been analysed. The morbidity impacts caused by ozone have been excluded from the analysis, as these impacts are omitted in the EcoSense model.

In Fig. 2 the damage costs for the same plant are compared using the EXMOD and EcoSense models. The first two columns in the figure represent the external costs calculated in EXMOD, the first column with monetary values from EXMOD, the second with monetary values from EcoSense. The last two columns represent the external costs calculated in EcoSense, the first column with monetary values from EXMOD, the second with monetary values from EcoSense. The figure shows two important facts:

(1) The damage costs are higher when the EcoSense model is used than when the EXMOD model is used
(2) The monetisation values in EXMOD are higher than in EcoSense.

re(1) The figure shows more than a doubling of the damage costs using EcoSense rather than EXMOD. The most important different is for chronic bronchitis. However, chronic bronchitis is the dominant impact in both models, but much larger in EcoSense than in EXMOD. Also, restricted activity days are important, having a higher effect in EcoSense than in EXMOD. Restricted symptom days account for 16% of the damage costs using the EXMOD model, while they are negligible using the EcoSense model. Other impacts have lesser significance in both models.

re(2) Analysing the results from the EcoSense model as well as from EXMOD the externalities are around 30% higher using the EXMOD monetisation values than using the EcoSense monetisation, applying that the monetary values used in EXMOD in general are higher than the values used in EcoSense. The most dominating monetary value is chronic bronchitis, which results in a 66% higher damage using EXMOD values than using EcoSense values. However, when the same monetary values for the two models are used, much higher morbidity costs are encountered with the EcoSense model.

5. Why are the external costs different for the same plant?

In the previous chapter the reasons for the differences in the estimates of the effect on human health using two comparable models have been analysed. Seven categories of differences have been explained. After the above discussion these seven categories may be gathered in four parameters of importance, as shown in Fig. 3:

- Difference in impacts.
- Different dose–response functions.
- Different monetary values.
- Difference in delta concentration and population for the US and Europe.

The four parameters are depending on each other. However, in the following the importance of the parameters has been explained individually.

5.1. Difference in impacts

The impacts included in two studies may differ, which may affect the total external costs estimated in the studies considerably. On a superior level mortality is included in both studies. However, EXMOD includes only acute mortality, while EcoSense includes acute as well as chronic mortality. This results in 19 times larger externality costs in EcoSense as in EXMOD concerning mortality.

Also in morbidity impacts there are differences between the two models. Table 5 shows the morbidity impacts estimated in EcoSense and in EXMOD. The table illustrates that more impacts are included in EcoSense than in EXMOD. Looking closer at the impacts it may be possible to make some assumptions in order to analyse the results from the two models.

Restricted activity days and chronic bronchitis for adults are directly comparable in the two models and the amount of impacts is close to each other. Bronchodilator usage does not exist in EXMOD, but is for comparison regarded as an asthma attack in EXMOD, being somewhat higher in EXMOD.

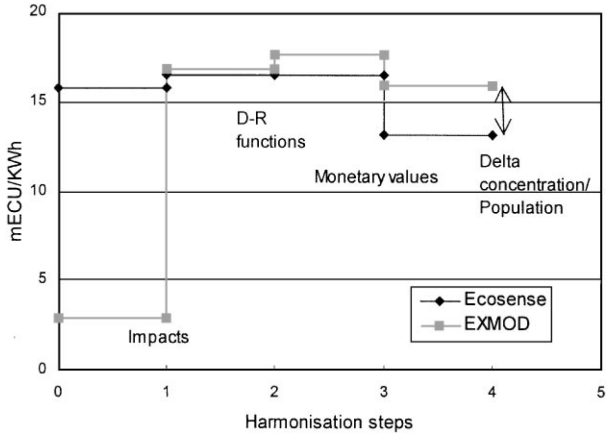


Fig. 3. Important reasons for differences in external costs.

For comparison of respiratory symptom days, in EcoSense asthmatic cough for adults and children must be included in respiratory symptoms days. Still the amount of respiratory symptoms days is much larger in EXMOD. Again NO_x is the dominating source. Acute bronchitis for children (EXMOD) does not exist in EcoSense, however, analysing the dose-response functions chronic bronchitis for children in EcoSense is similar to acute bronchitis in EXMOD. The amount of impacts is more than four times higher in EcoSense than in EXMOD, which apparently is a result of differences in emission concentrations and population. The difference in cases of impacts of chronic bronchitis is illustrated in Fig. 1. Emergency room visits are comparable in the two models; however, the number of visits is much smaller in EcoSense. For respiratory hospital admission also cerebrovascular hospital admissions and hospital visits for children with croup are included in EcoSense in order to make the impacts comparable.

Lead health effects as well as radiation are impacts only included in EXMOD, while congestive heart failure (> 65) and ischaemic heart disease (> 65) is represented only in EcoSense.

5.2. Different dose–response functions

It is not only necessary to check if the same impacts are included in two studies or the impacts are comparable, but also to analyse the dose–response functions used to

Table 5
Morbidity impacts in EcoSense and EXMOD, central estimate

EcoSense	Impacts	EXMOD	Impacts
Congestive heart fail (> 65)	0.54		
Ischa. Heart disease (> 65)	0.51		
Restricted activity days	3215	Restricted activity days	3440
Chronic bronchitis, adults	6.19	Case of chronic bronchitis	2.95
Chronic bronchitis, children	88	Children, acute bronchitis	20
Bronchodilator use, adults	735	Asthma attack	1330
Bronchodilator use, children	147		
Asthmatic cough, adults	756		
Asthmatic cough, children	254		
Low resp. symptom, adults	273		
Low resp. symptom, children	195	Respiratory symptoms days	19460
Chronic cough, children	112		
Respiratory hosp. Admission	0.47	Respiratory hosp. admission	2.5
Cerebrov. Hosp. Admission	1.14		
hosp. Visits child. Croup	6.6		
ERV for COPD	1.6	Emergency room visit	17
ERV for asthma	1.5		
Cancer	0.001	Survivable cancer	0.0005
		Lead health effects	1157
Mercury health effects	0	Mercury health effects	602
		Radiation	0.025

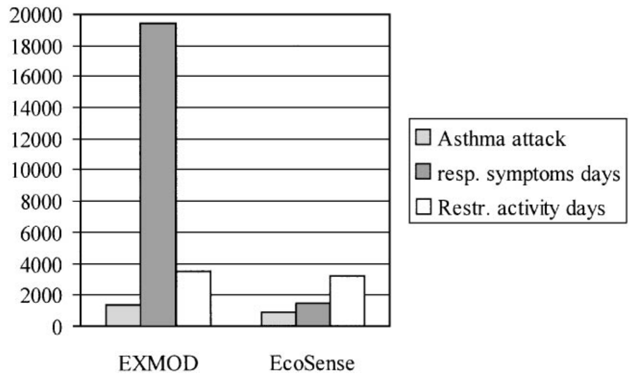


Fig. 4. Cases of impacts calculated in EXMOD and EcoSense.

estimate the damages. Looking at the above-mentioned assumptions it is obvious that some of the compared dose-response functions differ, resulting in differences in the amount of impacts.

Some of the morbidity impacts calculated for the same plant in EcoSense as well as in EXMOD have been compared in Fig. 4. The figure shows the very large difference in the cases of respiratory symptoms days in the two models. This is also visible in Fig. 2, where respiratory symptom days are important in EXMOD, but not visible in EcoSense. The reason for the large difference in impacts in the two models is the dose–response functions used to define the impacts. In EXMOD the dose-response function for respiratory symptom days is related to the total population, while in EcoSense the function is related to the asthmatic population, being 3.5% of the total population. Taking the large number of cases into consideration, the damage costs related to respiratory symptom days are small due to a low monetary value, illustrated as *D–R* functions in Fig. 1.

5.3. *Different monetary values*

Comparing the morbidity results in Fig. 2 on a superior level using the EXMOD model the total damage costs caused by morbidity are 28% larger using the EXMOD monetary values instead of using the values from EcoSense. This shows the importance of considerations concerning the monetary values used.

The morbidity damages estimated in EXMOD have been decomposed in Fig. 5 using monetary values from EXMOD as well as EcoSense. Comparing the results from the two models the morbidity externalities are higher using EcoSense. However, many of the impacts are monetised higher in EXMOD. Chronic bronchitis, emergency room visits, respiratory symptom days and respiratory hospital admissions are all monetised higher in EXMOD than in EcoSense, while restricted activity days and asthma attacks are valued highest using EcoSense.

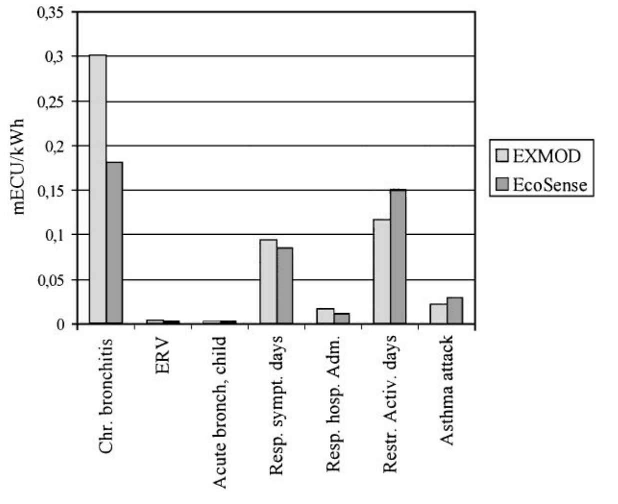


Fig. 5. Morbidity damages calculated in EXMOD using monetary values of EXMOD and EcoSense.

5.4. *Difference in delta concentration and population for US and Europe*

As illustrated in Fig. 3 having adjusted for the above-mentioned parameters, the external costs estimated in the two studies still differ. The reason for this difference is not related to the models used, but is allocated to the location of the plant. The air quality models predict the level of the emissions in different locations influenced by the emissions from the plant. This level is called the delta concentration. The delta concentration is the only factor that is calculated in the models, and it is different for the involved emissions. Each impact is explained by the delta concentration times the population multiplied by a dose–response function. This is included in the models, but may as well be calculated manually, having estimated the delta concentration in the computer models. The difference in delta concentration and population used in the two models is a result of different locations of the same plant, and will result in different amounts of impacts for the two locations.

For PM_{10} the delta concentration times the population has been found to be a factor 1.75 higher in EXMOD than in EcoSense. This means that the impacts of PM_{10} estimated in EXMOD should be 1.75 larger than the same impacts estimated in EcoSense. However, this is only the case when using the same linear dose–response functions in the two models.

As an example, restricted activity days are estimated in EcoSense by the following function:

$$\mathrm{RAD}_{\mathrm{eco}} = 25 \times \text{delta concentration} \times \text{population} \times \text{adults}/1000,$$

where adults are defined as 57% of the total population.

In EXMOD the function is as follows:

$$\mathrm{RAD_{ex}} = 58.4 \times \text{delta concentration} \times \text{population} \times \text{adults}/1000.$$

Here adults are defined as 83% of the total population.

Giving that Delta Concentration × population is 1.75 times larger using EXMOD than using EcoSense and merging the two functions results in the following:

$$\mathrm{RAD_{eco}} = 0.168 \times \mathrm{RAD_{ex}}$$

The same calculations can be made for other impacts as far as the dose–response functions are linear.

It must be noted that the delta concentration depends on the emission, meaning that the impacts, which are 1.75 larger in EXMOD compared to EcoSense, only relate to PM_{10} emission. For other emissions, like nitrate and sulphate, the situation is different.

The damage costs of chronic bronchitis in Fig. 2 are larger in EcoSense than in EXMOD. This is a result of more cases of chronic bronchitis using EcoSense, although the monetary value is larger in the EXMOD model. Again, the reason for this is the difference in delta concentration times the population in Europe and in US, and is not related to the model used.

Fig. 6 shows the importance of the difference in emissions in the two models. In EcoSense, the secondary emissions sulphate and nitrate have nearly the same weight, while PM_{10} has much smaller weight on the impacts. Comparing this with the weighting factors in EXMOD, nitrate is the most dominant, followed by PM_{10}, while sulphate has a relatively small effect.

Using EXMOD the plant is situated in the New York Capital District, which is a suburban site outside of Albany, while the same plant in EcoSense is situated in Roskilde, Denmark. Differences in the dispersion and impacts of the emissions in the two cases may be caused by differences in background levels of the emissions in the two locations with different surroundings and because of differences in population size.

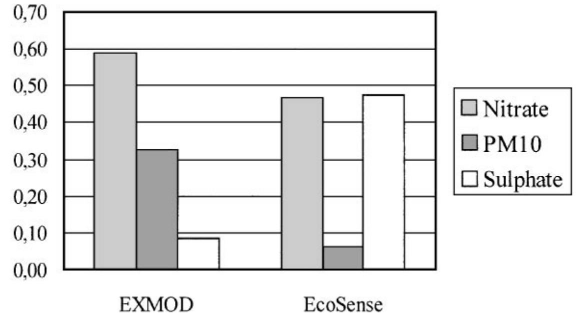


Fig. 6. The relative weighting factors of nitrate, PM_{10} and sulphate on the impacts in the two models.

6. Conclusion

External costs may be used by politicians to assess the importance of different kinds of energy technologies and also by the electricity utilities to choose between different technologies in capacity building. However, it is quite important to note that external costs for power generation technologies may be assessed using different approaches, and therefore the external costs may differ for the same technology depending on the approach used.

In this paper the same approach — the bottom–up approach — has been used, but with two different models. The models are in principle built up in the same way as with air dispersion models and dose–response functions for the calculation of impacts. These impacts are multiplied with monetary values to calculate the external costs. Although the models seem more or less similar, the resulting external costs are five times larger in the ExternE study using the EcoSense model than in the New York study using the EXMOD model for the same power plant.

The above paragraphs have demonstrated the importance of the difference in impacts included, the dose-response functions used and finally the use of different monetary values in different studies. A difference in the above-mentioned parameters may give rise to large differences in the external costs of the energy technologies analysed. The resulting difference after harmonising the models is not related to the model, but to the location of the plant (expressed in delta concentration). Although the parameters have been explained separately, the tables and figures have shown that the parameters have influence on each other.

When politicians use externalities to assess the importance of different kinds of energy technologies, it is therefore quite important that they use external costs for the technologies based on the same approach, i.e. calculating the same impacts and using same monetary values and dose-response functions. This is also the case when externalities are used by the electricity utilities in order to choose between different technologies in capacity building; otherwise the comparison of the technologies may be based on incorrect assumptions.

Acknowledgements

Wolfram Krewitt from IER, Stuttgart University, Germany, has contributed to the study with valuable information and discussion concerning the EcoSense model. Concerning the EXMOD model the estimations have been discussed with Stephen S. Bernow and William W. Dougherty from Tellus Institute in Boston, USA, who have been helpful with information regarding the model. The study has been partly financed by the Danish Energy Agency, Copenhagen, Denmark.

References

Bhattacharyva, S.C., 1997. An estimation of environmental costs of coal-based thermal power generation in India. International Journal of Energy Research 21(3), 289–298.

CEC, 1995. Commission of the European Communities Joule Programme. ExternE: Externalities of Energy, Vol. 1–6. EUR 16520-25, Directorate-General XII, Luxembourg.

ETSU, 1994. Full Fuel Cycle Study on Power Generation Schemes, Incorporating the Capture and Disposal of Carbon Dioxide, Vols. 1–5. IEA Greenhouse Gas R&D Programme, Harwell, UK.

Hohmeyer, O., 1988. Social Costs of Energy Consumption. Springer Verlag, Berlin.

Oak Ridge National Laboratory, 1992. U.S.-EC Fuel Cycle Study: Background Document to the Approach and Issues. Report No. 1 on the External Costs and Benefits of Fuel Cycles: A Study by the US Department of Energy and the Commission of the European Communities, Oak Ridge 11.

Ott, W., 1997. External Costs and External Price Addings in the Swiss Energy Sector, ECONCEPT, Zürich, Switzerland. In Hohmeyer, O. et al. (Eds) Social Costs and Sustainability, Springer Verlag, Berlin. pp. 176–184.

Rowe, R., Lang, C., Chestnut, L., Latimer, D., Rae, D., Bernow, S., White, D., 1995. The New York Electricity Externality Study. Oceana Publications, Dobbs Ferry, NY.

Schleisner, L., Nielsen, P.S., 1997a. External Costs related to Power Production Technologies ExternE National Implementation for Denmark. Risø National Laboratory.

Schleisner, L., Nielsen, P.S. (Eds.), 1997b. External Costs related to Power Production Technologies ExternE National Implementation for Denmark, Appendix 1. Risø National Laboratory.

ELSEVIER

Energy Policy 35 (2007) 828–843

ENERGY POLICY

www.elsevier.com/locate/enpol

Internalisation of external cost in the power generation sector: Analysis with Global Multi-regional MARKAL model

Peter Rafaj*, Socrates Kypreos

Energy Economics Modelling Group, General Energy Department, Paul Scherrer Institute, CH-5232. Villigen-PSI, Switzerland

Available online 17 April 2006

Abstract

The Global MARKAL-Model (GMM), a multi-regional "bottom-up" partial equilibrium model of the global energy system with endogenous technological learning, is used to address impacts of internalisation of external costs from power production. This modelling approach imposes additional charges on electricity generation, which reflect the costs of environmental and health damages from local pollutants (SO_2, NO_x) and climate change, wastes, occupational health, risk of accidents, noise and other burdens. Technologies allowing abatement of pollutants emitted from power plants are rapidly introduced into the energy system, for example, desulphurisation, NO_x removal, and CO_2 scrubbers. The modelling results indicate substantial changes in the electricity production system in favour of natural gas combined cycle, nuclear power and renewables induced by internalisation of external costs and also efficiency loss due to the use of scrubbers. Structural changes and fuel switching in the electricity sector result in significant reduction of emissions of both local pollution and CO_2 over the modelled time period. Strong decarbonisation impact of internalising local externalities suggests that ancillary benefits can be expected from policies directly addressing other issues then CO_2 mitigation. Finally, the detailed analysis of the total generation cost of different technologies points out that inclusion of external cost in the price of electricity increases competitiveness of non-fossil generation sources and fossil power plants with emission control.

Keywords: External cost; Technology learning; Air pollution; Carbon capture; Total generation cost

Abbreviations: ¢, cent (10^{-2}\$); ASIA, developing Asian countries: Centrally Planned Asia, India, South East Asia, Pacific Asia; C, carbon; CHP, combined heat and power (cogeneration); CI, carbon intensity (tonne CO2/GJ); CNG, compressed natural gas; CO_2, carbon dioxide; $DeNO_x$, nitrogen oxides abatement, denitrification; $DeSO_x$, sulphur oxides abatement, desulphurisation; EC, european Commission; ED, elastic demands; EEFSU, eastern Europe and Former Soviet Union; ETL, endogenous technological learning; ExternE, externalities of energy; FC, fuel cell; FGD, flue gas desulphurisation; GDP, gross domestic product (T\$/yr); GFC, gas fuel cell (based on natural gas); GHG, greenhouse gas; GMM, global multi-regional Markal model; GtC, giga tonnes carbon (10^9 ton); H_2FC, hydrogen fuel cell; IGCC, integrated coal gasification combined cycle; IPCC, intergovernmental panel on climate change; LAFM, latin America, Africa, and Middle East region; LBD, learning-by-doing; LWR, light water reactor; MARKAL, market allocation model; mill, mills (10^{-3}\$); Mt, mega ton (10^6 ton); NAME, North American region; NCCR, The National Centre of Competence in Research; NGCC, natural gas combined cycle; NNU, New (design of) nuclear power plant; NO_x, nitrogen oxides; O&M Cost, operation and maintenance cost; OECD, organization for Economic Cooperation and Development; OOECD, other OECD region: Western Europe, Japan, Australia, and New Zealand; PFBC, pressurised fluidised bed combustion; ppmv, parts per million by volume; pr, progress Ratio; PSI, Paul Scherrer Institut; RD&D, Research, development and demonstration; RES, reference energy system; SO_2, sulphur dioxide; SPV, solar photovoltaic system; SRES, special report on emission scenarios; T&D, transport and distribution; WTP, willingness to pay; η, conversion efficiency

*Corresponding author. Tel.: +41 56 310 2675; fax: +41 56 310 2199.

E-mail addresses: Peter.Rafaj@psi.ch (P. Rafaj), Socrates.Kypreos@psi.ch (S. Kypreos).

0301-4215/\$ - see front matter
doi:10.1016/j.enpol.2006.03.003

1. Introduction

Internalisation of external costs into the full energy production cost is considered an efficient policy instrument for reducing negative impacts of energy supply and use. Approach of merging production (or generation) cost with external cost into a total specific cost serves as a comparative indicator for evaluation of economic and environmental performance of optional energy technologies. Consideration of externalities, where quantified or quantifiable, might be useful for providing an indication of damages/benefits associated with different energy options, for assessing trade-offs between different energy options, and for ranking energy options, as well as serving as a basis for the introduction of economic instruments to reflect better the social costs of energy (Fouquet et al., 2001).

Although such an instrument omits other important aspects of the policy- and decision-making processes, for instance the political and social acceptance of certain energy systems (Hirschberg et al., 2000), it is important to analyse the possible impacts of internalising externalities in the energy system.

This study has been performed with the Global Multi-regional MARKAL (GMM) model with endogenous technological learning (ETL) developed at Paul Scherrer Institute. Since the GMM model has a rich representation of the power generation sector (including ETL specification of selected technologies), and because the assumptions on the external cost from the electricity production were provided (EC, 1998a), this analysis focuses primarily on the electric power sector. This paper describes the economic, environmental and structural impacts of a full internalisation of external costs in the electricity generation sector, which are based on results of the EC ExternE project. However, no attempt has been made to verify these costs as fully representing the environmental and health damages.

Three scenarios were analysed with the research objective as specified above—the *Baseline scenario* without inclusion of the external cost; *Local externality scenario* with internalised external costs resulting from local air pollution (SO_2, NO_x); and finally the *Global externality scenario* where the external costs comprise both local air pollutants (SO_2, NO_x) and emittants causing global climate change (CO_2).[1]

2. Description of the modelling framework

The analysis presented in this paper has been executed by using the GMM model with ETL, developed by Barreto (2001), and further upgraded by the authors. MARKAL is a dynamic linear programming, ''bottom-up'', energy planning model allowing detailed representation of energy technology options on both demand and supply side of the energy system (Fishbone and Abilock, 1981).

Fig. 1. Definition of the world regions in the GMM model.

Five world regions are considered in the GMM model, as shown in Fig. 1. Two regions shape industrialised countries of North America (NAME) and the rest of the OECD (OOECD). One region covers transition-economies of Central & Eastern Europe and the Former Soviet Union (EEFSU). Two additional regions represent the developing world: the developing countries in Asia (ASIA) and Latin America, Africa and the Middle East (LAFM).

There are six end-use demand sectors in the GMM model. Industrial and residential & commercial sectors are divided into thermal and electric (specific) uses. The transportation sector merges together passenger and freight transport. Finally, the non-commercial use of biomass (i.e., fuel-wood) and non-energy feedstock is represented. In each of the demand sectors, a set of generic end-use devices is defined (e.g., coal-based heating in industry, oil-based transport).

The supply sector and energy conversion processes are represented with some detail. Technologies for the production of electricity, heat and a variety of final fuels (e.g., oil products, alcohol, hydrogen, natural gas) from different fossil and non-fossil sources are included, as well as the corresponding transport and distribution (T&D) chains. Investment, fixed O&M and variable O&M costs are specified for all supply technologies considered. A schematic representation of the standard reference energy system (RES) used for all the regions, containing all the possible energy chains that can be chosen by the model is shown in Fig. 2.

Levels of power generation based on renewable and nuclear energy sources are controlled in GMM through the imposition of exogenous bounds and annual growth/declination rates for each technology. Bounds applied for renewable resources reflect the regional technologically achievable potential of each type of source and is provided by IEA (2001), UNDP (2000), and Riahi and Roehrl (2000). As indicated in Table 1, except for hydropower, only upper bounds are applied in 2050 for renewable power generation; the level of actual generation, therefore, is not forced, but is left free for determination through

[1]In both scenarios with internalised external cost, impacts of the following burdens are always included, besides the air emissions: solid wastes, liquid wastes, risk of accidents, occupational exposure to hazardous substances, noise, others, e.g., exposure to electro-magnetic fields, emissions of heat (EC, 1998a).

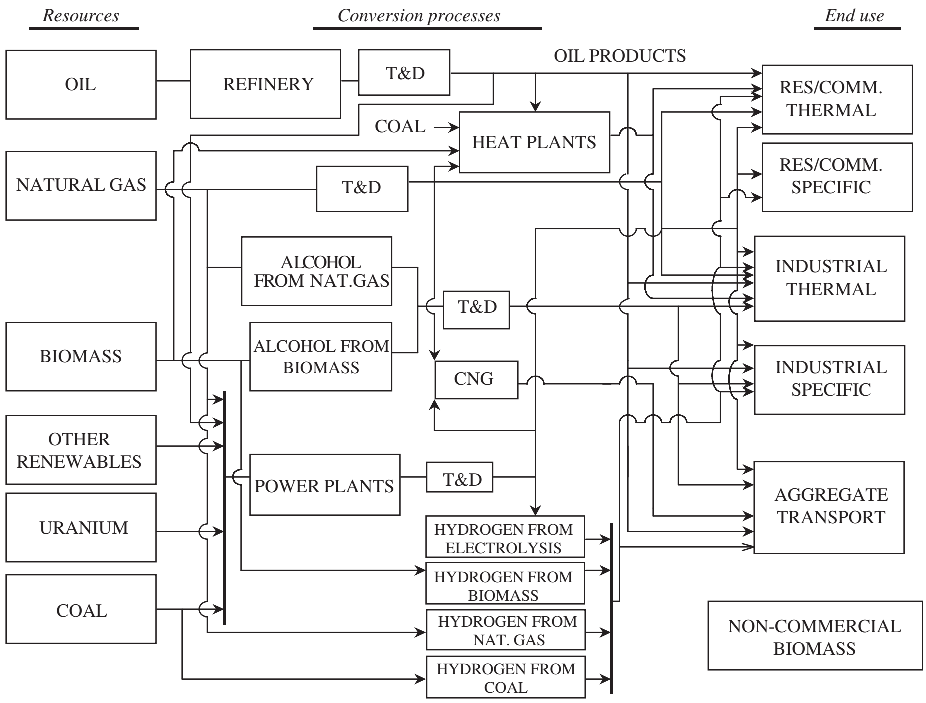


Fig. 2. Reference energy system applied in GMM. Figure taken from Barreto (2001).

Table 1
Assumptions for renewable and nuclear electricity sources applied in GMM

Bounds for renewable electricity sources in 2050 (EJ)						
Regions	NAME	OOECD	EEFSU	ASIA	LAFM	WORLD
Hydro max	2.8	3.4	5.8	7.6	8.5	28.1
Hydro min	2.2	2.0	1.1	1.2	2.4	8.9
Wind max	9.4	12.0	9.3	9.9	9.8	50.4
Solar PV max	3.6	2.2	1.6	14.6	5.2	27.3
Biomass max[a]	8.4	3.3	10.8	53.5	112.4	188.4
Geothermal max	1.0	0.8	2.0	5.0	2.0	10.8
Bounds for nuclear power in 2050 (EJ)						
Nuclear max	18.0	18.0	9.5	20.0	18.0	83.5
Nuclear min	2.0	2.9	0.9	1.5	0.1	7.4

[a]Biomass potential refers to both electricity and heat production.

competition. Power-network stability aspects are taken into account by assuming a maximum penetration fraction of intermittent power generation (e.g., wind power, solar photovoltaic) of 25% of total electricity production. In the case of nuclear power, the lower bound in 2050 corresponds to the present global level of generation. No limit is provided for CO_2 that can be stored in any type of reservoirs. The level of carbon sequestration, however, is controlled by annual growth rates of technologies being operated with CO_2 emissions removal.

An important characteristic of the GMM model is its ability to treat technology dynamics in energy-system development through the incorporation of learning curves of selected technologies within the model. Endogenisation of the technological learning (ETL) enables the modeller to analyse how the specific investment cost of a ‘‘learning'' technology declines with accumulated installed capacity of the respective technology (Messner, 1997). No learning spillovers across technology clusters are defined in the model, instead spatial spillovers of separate learning technologies are assumed. The detailed description and mathematical formulation of the ‘learning-by-doing' (LBD) modelling approach applied in GMM can be found in Barreto and Kypreos (2002). The technological, cost, and learning specification of electricity generation technologies represented in GMM are given in Table 2.

The GMM model version used for this analysis applies the ETL option in combination with a partial equilibrium algorithm, that adjusts demands for energy services to the increased marginal cost of services that can result from the

Table 2
Cost and performance of power generation technologies in GMM

Technology	Start year	Life time	Load factor (max.)		Electric Efficiency		Investment cost	Fixed O&M cost	Variable O&M cost	Progress ratio
			Start	2050	Start	2050	($/kW)	($/kW/yr)	($/GJ)	
Fossi-fuel based power plants										
Coal conventional electric	1990	30	0.65	0.75	0.370	0.380	1050	38	0.72	
Coal conventional electric with $DeSO_x/DeNO_x$	2000	30	0.65	0.75	0.360	0.370	1150	48	1.22	
Coal conv. with $DeSO_x/DeNO_x$ and CO_2 seq	2010	30	0.65	0.75	0.296	0.304	2090	80	1.53	
Coal cogenation	1990	20	0.65	0.75	0.370	0.380	1155	49	1.5	
Coal advanced electric (Supercritical, PFBC)	1990	30	0.65	0.8	0.429	0.500	1584	4.5	0.75	0.94
Coal advanced electric with CO_2 seq	2010	30	0.65	0.8	0.365	0.425	2060	90	1.13	0.93
Integrated Coal-Gasification Combined Cycle (IGCC)	2000	30	0.85	0.85	0.425	0.500	1401	40	0.88	0.94
Coal IGCC with CO_2 seq	2010	30	0.85	0.85	0.361	0.425	1910	52	1.23	0.93
Natural Gas Combined Cycle (NGCC)	1990	20	0.65	0.75	0.510	0.588	560	36.6	0.63	0.9
NGCC with CO_2 scrubber	2010	20	0.65	0.75	0.459	0.529	1015	50	0.88	0.9
Gas turbine	1990	20	0.2	0.2	0.360	0.360	350	58.5	0.51	
Gas steam conventional	1990	20	0.65	0.65	0.333	0.410	987.7	50.6	0.56	
Cogeneration gas turbine	1990	20	0.4	0.46	0.370	0.370	750	51.6	0.63	
Gas fuel cell (GFC)	2000	20	0.65	0.65	0.599	0.649	2463	43.5	0.63	0.82
Hydrogen fuel cell (CHP) in industry (H_2FC)	2010	20	0.85	0.9	0.4	0.6	3500	20	7.5	0.82
Hydrogen fuel cell (CHP) in res&com. (H_2FC)	2010	20	0.85	0.9	0.4	0.5	3500	20	5.8	0.82
Oil electric	1990	20	0.65	0.8	0.303	0.400	991	63.6	0.57	
Nuclear and renewable power plants										
Nuclear plant—Light Water Reactor (LWR)	1990	30	0.75	0.9	0.327	0.327	1800	80	0.19	
Advanced new nuclear power plant (NNU)	2010	30	0.85	0.9	0.345	0.345	1900	80	0.19	0.96
Hydro-electric plant (small and large)	1990	50	0.42	0.46	0.385	0.471	1850	49.5	0.12	
Solar photovoltaics (SPV)	1990	20	0.2	0.25	0.400	0.400	5000	9	1.25	0.81
Solar thermal electric	2000	20	0.2	0.2	0.400	0.400	2900	9	1.25	
Wind turbine	1990	20	0.3	0.3	0.330	0.330	1150	13.5	0.83	0.9
Biomass power plant	1990	20	0.75	0.75	0.333	0.333	2650	47.8	0.92	
Geothrmal electric	1990	20	0.75	0.75	0.381	0.381	2900	28	0.9	

All costs are given in $(1995). The progress ratio (*pr*) is the rate at which the cost declines each time the cumulative capacity doubles. The data presented derives from various sources and literature reviews, e.g., Lako and Seebregts (1998); EIA (2003a); Wu et al. (2001), etc. Characteristics of technologies with CO_2 removal are adopted from David and Herzog (2000); additional CO_2-storage cost (10 \$/t-CO_2 or 36.7 \$/tC for every ton captured) is charged for these technologies. This cost comprises expenditures for CO_2 transport, injection and disposal. A lower value of *pr* for technologies with CO_2-capture is based on an assumption, that the CO_2-capture device applied to the reference power plant might contribute to the "learning" potential of a reference plant. Technologies equipped with CO_2-capture, therefore, could undergo stronger cost reduction.

imposition of a policy constraint (Loulou and Lavigne, 1996); in the analysis herein it is the external cost. The MARKAL model with elastic demands (referred to as the MARKAL-ED) makes use of a procedure whereby the energy end-use demands are not fixed, but instead are elastic to their own prices, endogenously computed by the model in the baseline, and self-adjusted if modified scenario conditions affect the prices. The model attains partial energy equilibrium when the sum of producer and consumer surpluses is maximised. Consequently, the model objective function is comprised of two terms: the energy/technology costs, and the loss of welfare associated with demand reduction (Kanudia and Loulou, 1999). Internalisation of externalities results in allocation of resources through the integration of externalities in energy prices. Although not modelled explicitly within this study, in the "real-world" situation it can be expected that the extra charge imposed on the power generation is recycled back into economy and used for different purposes. The total amount of externality charges levied on the electricity sector is discussed separately in Section 5.3.2.

Additionally, the GMM model allows simulation of the global trade of selected energy or environmental commodities (e.g., fuels, emission permits, etc.) and defines a shadow price of the commodity globally traded among regions. In the scenarios reported herein, electricity is not traded among regions or intra-regionally.

3. External cost specification

External cost values used in this study have been derived from the outcomes of the European Commission (EC) ExternE Project. The methodology used for this project applies the impact pathway approach (i.e., the pathways of polluting substances are followed from the release source to the point of damage occurrence). The consecutive negative impacts (damage) are quantified using a damage function. Economic valuation of the damage is obtained by the "willingness-to-pay" of the affected individual to avoid a negative impact resulting from energy production from an actual power plant.[2] This 'bottom-up' approach emphasizes detailed site-specific characterization of technologies, thereby enabling consideration of every important stage in different energy chains and comparison between different fuel-cycles and different kinds of burden and impact within a fuel-cycle (EC, 1998a).

[2]The underlying principle of the "Willingness-to-pay" (WTP) concept is to obtain a monetary valuation of preferences of affected individual to avoid a negative impact (EC, 1998b). The main advantage of the WTP approach relies in its foundation on the individual viewpoint of the concerned population. This approach attempts to estimate the demand (or WTP) for an improved environmental quality. WTP is measured by how much the concerned individuals are ready to pay in order to improve their own life-quality or the one of other people. Adding the amounts of all concerned individuals results in a value that a group of individuals attributes to the reduction of environmental impacts.

Table 3
Basic assumptions made for the external cost calculation

Region	Population density	Sulphur content in coal [%]	Starting year of externality charges
NAME	Medium	1	2010
OOECD	High	1	2010
EEFSU	Medium	1	2010
ASIA	High	1	2010
LAFM	Medium	1	2010

For the purpose of internalisation of the external cost within the total electricity cost in different world regions, the ExternE results were adjusted to reflect the GMM level of aggregation. The determinants for scaling the externalities were the population density in regions; fuel quality expressed as the content of the sulphur in coal and oil; technology specification with respect to installation of the emissions control systems; and finally, the possible improvement in conversion efficiency over the modelled time horizon.

Table 3 summarises basic assumptions made for the adjustment of external cost. The world regions are grouped in two population density categories according to present statistical data (GeoHive, 2003). ASIA and OOECD are located within the category of High density of population, and the remaining regions are assumed to have Medium population density. Changes in the population density over time are not considered. Sulphur content in coal is assumed to be 1% in all world regions. Even though standardised statistical data are not available, a literature survey indicates, that this value represents the typical average of all different coal types used for power production (Hinrichs, 1999). An optimistic assumption has been made, that a global policy of imposing the external costs on the electricity production starts from the same period (2010) in all regions. Simultaneously, it is expected that a global spill over of experience and know-how transfer from North to the South takes place.

External cost was further scaled as a function of conversion efficiency so that exogenously given efficiency improvements could be taken into consideration. The following formula has been used:

$$ExtCost_t = ExtCost_{original,t=0} * \frac{\eta_{orig,t=0}}{\eta_t}$$

if

$$\eta_t > \eta_{orig,t=0},$$

where η is the conversion efficiency of respective power plant.[3]

[3]Sample calculation for a pulverised coal power plant with 0% $DeSO_x$, 50% $DeNO_x$, 80% DEDUST, $\eta_{2010} = 37\%$, $\eta_{2050} = 38\%$, sulphur

The resulting external costs applied for five world regions in GMM are displayed in Appendix II. Two different kinds of externalities were considered: external cost associated with local air pollutants (SO_2, NO_x, particulates) but excluding CO_2 impacts, and external cost that merges damages from local pollutants and global climate change (CO_2) (see also Note 1 on p. 2). While evaluation of externalities for local pollutants within the ExternE project is based on detailed bottom-up analysis, valuation of external cost for CO_2 emissions bears much larger level of uncertainty. External cost of global warming used in this study refers to the global warming damage cost of 25$ per tonne of CO_2 (91$/tC). This value lays at the lower end of a range of recommended global-warming damage estimates reported by the ExternE project (EC, 1998b). Ranges in the values of external cost, as shown in Appendix II, represent regional differences resulting from assumptions and scaling, as explained above. For further details on assumptions and external cost calculation, see ACROPOLIS (2003).

3.1. Treatment of external costs in MARKAL model

External costs are implemented in the GMM model by multiplying the amount of electric power generated (i.e., kWh) from each power plant during each time period in each region with corresponding external cost (i.e., ¢/kWh). In this way, it is assured that the matching external costs are directly charged to every unit of output from each power plant. The sum of discounted annual externality charges for every region in GMM is reflected in the total discounted system cost (i.e., objective function used in GMM).

An alternative approach that could be used is to apply the damage costs for different pollutants as an environmental tax levied on the entire energy system. The environmental tax would be charged per unit of pollutant emitted (e.g., 8000$/t$SO_2$), which would affect all emitting technologies in all sectors present in the energy system of each region (i.e., including refineries, demand devices, transport sector, etc.). Because this analysis is explicitly focused on the externality impacts on the power generation sector, the approach explained in the previous paragraph has been chosen.

The GMM model has different response options for the extra charges imposed on the electricity sector with the aim of minimising the total energy system cost: (a) to pay (or not) an external charge on power production from a technology; (b) to install (or not) a (costly) system with DeNO_x, DeSO_x, or CO_2-capture & sequestration; (c) to reduce (or not) the energy/electricity consumption in different demand sectors and to substitute the electricity by other fuels; (d) to apply (or not) the inter-fossil fuel switching and technological change for technologies with lower external cost (renewables and nuclear power plants).

(*footnote continued*)
content = 1%.

Low population density (Adjustment factor $AF = 1$)

$$Ext_{2010} = 9.9\,¢/\text{kW}h * 1 = 9.9\,¢/\text{kWh}$$

$$Ext_{2050} = Ext_{2010} * (0.37/0.38) * 1 = 9.6\,¢/\text{kWh}$$

High population density (Adjustment factor $AF = 1.5$)

$$Ext_{2010} = 9.9\,¢/\text{kWh} * 1.5 = 14.8\,¢/\text{kWh}$$

$$Ext_{2050} = Ext_{2010} * (0.37/0.38) * 1.5 = 14.4\,¢/\text{kWh}$$

Table 4
Scenarios definition

Scenario name	Scenario definition
Baseline	Business-as-usual, no local, no global externalities, ETL
Local externality	External cost from local air pollution (SO_2, NO_x) internalised in the electricity sector, Partial equilibrium, ETL
Global externality	External cost from local air pollution (SO_2, NO_x) and emissions causing global climate change (CO_2) internalised in the electricity sector, Partial equilibrium, ETL

4. Scenarios analysed

As already mentioned in Section 1 and summarised in Table 4, three scenarios were explored in this study, all of which include ETL and partial equilibrium:

The underlying story-line for the Baseline scenario refers to the B2 scenario reported by the IPCC/SRES (Intergovernmental Panel on Climate Change/Special Report on Emission Scenarios) project[4] (IPCC, 2000).

The base year in GMM is 1990. The model time horizon is 1990–2050. Ten-year periods are considered. A discount rate of 5%/year is applied to the calculations. Availability of fossil fuel resources by different cost categories is based on Rogner (1997). The demand projections and potentials for renewable resources correspond to those of the characterisation of the B2 storyline performed with the MESSAGE model (IPCC, 2000; Riahi and Roehrl, 2000). However, no attempt has been undertaken to calibrate the baseline scenario to match the results of the SRES-B2 scenario. In this respect, the baseline development corresponds to a PSI scenario, since the allocation of resources is based on an optimisation performed under conditions of perfect foresight with ETL considerations.

[4]The B2 scenario is a "dynamics-as-usual" scenario where differences in the economic growth across regions are gradually reduced, and concerns for environmental and social sustainability at the local and regional levels rise along the time horizon. Population growth is consistent with the United Nations median projection increasing to 9.4 billion people in 2050, which is a continuation of historical trends. Economic growth is gradual, with world GDP increasing at an average rate of 2.8% per annum between 1990 and 2050. Income per capita grows at a global average of 1.8% per year for the same period reaching an average value of 11700 USD (1990) per capita in the year 2050 (at market exchange rates) (IPCC, 2000).

5. Results

Although the energy system of five world regions is fully modelled, results presented here emphasize the global electric power generation sector (e.g., fuel mix, choice of technologies, costs) and the environmental impacts (SO_2 and NO_x emissions from the power production, and CO_2 emissions from the entire energy system). Changes in the rest of the energy system are summarised and reported in the form of primary and final energy use and the system costs.

5.1. System changes

Based on the analysis of structural changes induced by inclusion of external costs in the electricity sector, the following findings are identified:

- There is a significant change in the electricity generation mix. Conventional coal systems without emission control are eliminated early after policy implementation and are replaced by advanced coal technologies (SO_2/NO_x scrubbers, carbon capture).
- Natural gas combined cycle, nuclear power and renewables increase their shares in the electricity market.
- External charges imposed result in a decrease of total power generation. The reduction in final demand for electricity occurs in both industrial and residential/commercial sectors. The industrial sector shows the greatest ability to switch.
- The role of coal in the primary energy demand is significantly reduced in the externality scenarios. Decrease in coal is balanced with increased importance of nuclear and renewable energy sources.

5.1.1. Electricity production

The power generation in the Baseline scenario is dominated by systems based on coal combustion. Different types of coal power plants contribute by 50% to the total global power generation at the end of the time horizon (2050). From the year 2030, the conventional coal plants are replaced by advanced coal 'learning' systems (i.e., supercritical plants, pressurized fluidised-bed combustion—PFBC), and integrated coal gasification combined cycle (IGCC). The second most competitive system in 2050 is the NGCC, which contributes more than 31% of total power production. Approximately 20% of the electric power is supplied by the carbon-free nuclear and renewable energy sources in the year 2050, as is shown in Fig. 3.

Introduction of external costs into total production cost influences significantly the structure of the power generation mix. In the local externality scenario, coal remains the major contributor to total power production; however, its share is reduced in 2050 by 15% relative to the Baseline (in absolute terms by 11 000 TWh/yr). Moreover, the conventional pulverised coal combustion is steadily being replaced by advanced coal plants and pulverised coal systems with SO_2 and NO_x emissions control, i.e., flue gas desulphurisation-FGD, low-NO_x burners, etc. The NGCC plants with other natural gas based systems increase their relative share in power production to a level of 35% of the total electricity supply by 2050, although in absolute terms reduction in the total power generation from NGCC by 900 TWh/yr is reported. The share of renewables and

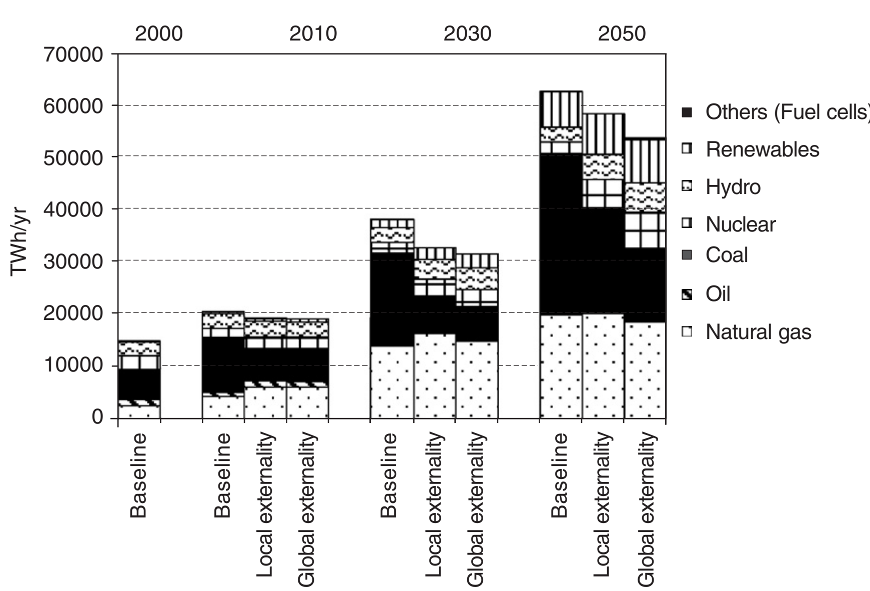


Fig. 3. Development in the electricity production by fuel.

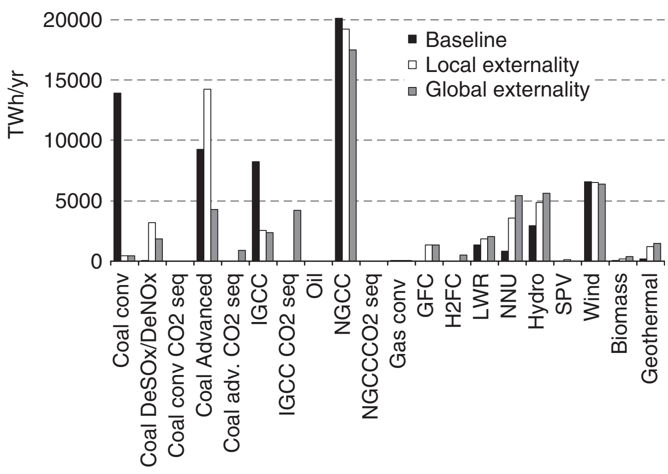


Fig. 4. Power generation profile in 2050 in all scenarios considered.

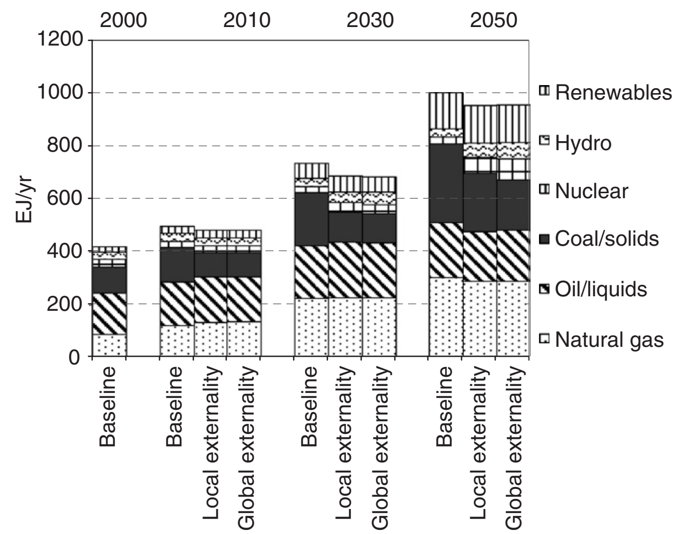


Fig. 5. Development in primary energy consumption.

nuclear plants is increased by 10% relative to the Baseline because of lower external cost charged on these systems.

Changes in electric power mix become more pronounced in the Global externality scenario. The generation from coal reaches only 26% while natural gas fired power stations produce around 32% of total electricity in 2050. The advanced coal and IGCC systems with CO_2-capture become competitive and penetrate the market between 2030 and 2050 at considerable levels. Nuclear energy supplies 14% of electricity in 2050; both the light water reactor (LWR) and advanced nuclear plants (NNU) play more significant role in the power supply during the whole time horizon after implementing external cost as compared to the Baseline. In the Global externality scenario, technologies based on renewable sources contribute 25% of total generation by the year 2050.

Because of rising cost of electricity, the overall power generation in 2050 is decreased by 7% in the Local externality scenario and by 14% in the Global externality scenario, relative to the Baseline (effect of reduced demand for electricity due to partial equilibrium).

Fig. 4 illustrates the power generation profile in 2050 for scenarios considered. Coal technologies with $DeSO_x$/$DeNO_x$ systems produce a considerable amount of electricity in the case where local external cost are incurred, but this amount is almost halved in the Global externality scenario. On the other hand, when global external costs are imposed, the systems with CO_2-capture become competitive, and the IGCC technology with carbon capturing and sequestration is the second largest coal-based power producer at the global level. This finding suggests, that internalised external cost makes the IGCC with CO_2-removal an attractive technological option for carbon mitigation strategies.

In both externality scenarios is the growth in renewable and nuclear electricity sources limited by the exogenous bounds and assumed growth rates, as is described in Section 2. The results of this analysis indicate a substantial increase in generation from advanced nuclear power plants in all regions modelled in GMM. On the other hand, the growth in hydropower production is most pronounced in the ASIA and LAFM regions, where the total generation is close to assumed exploitable limits. Similarly, the power production from wind turbines in externality scenarios is approaching its technical potential specified as an upper bound in GMM. Furthermore, the growth in generation from wind turbines is bounded by the constraint imposed on generation from intermittent sources of electricity, which explains a slightly lower contribution of wind power to the generation profile in 2050 in absolute terms in externality scenarios relative to the Baseline.

5.1.2. Primary energy consumption

Total global primary energy consumption decreases if external costs of power generation are included. In the year 2050, reductions by 4–5% in total primary energy consumption relative to the Baseline in the externality scenarios, are reported. This behaviour is a result of the use of fossil equivalent for calculation of the contribution of non-fossil sources to primary energy consumption and because of the switch to other fuels than electricity in the final energy demand. As shown in Fig. 5, the Local externality scenario is characterised by a large reduction in coal use, which is replaced by mainly by renewable and nuclear energy. This trend becomes even more obvious in the case of global externalities, where coal use (primarily for power generation) is substituted with nuclear energy (partly also by natural gas and oil between 2010 and 2030), and a large increase in renewable electricity consumption at the end of the time horizon is projected. Reductions in natural gas use in the end of modelled horizon under the externality scenarios are associated with a decrease in power generation from the NGCC systems. The consumption of oil is slightly lowered in the Local externality scenario, but its contribution is increased again in the global externalities case, which is again related to inter-fossil fuel substitution.

Changes and fuel switching in the primary energy demand are most significant towards the end of time horizon. This observation is related to a larger penetration of low-emitting technologies induced by externality charges, and further accelerated by cost reducing effects of ETL. Furthermore, the replacement of coal use for power generation goes along with given declination rates assumed for retirement of conventional coal plants without emission control.

5.1.3. Final energy demand

In both externality scenarios the total final energy consumption decreases (4–5%) compared to the Baseline in 2050. Comparison of shares in the final-demand fuel mix summarised in Fig. 6 shows, that the consumption of heat, biomass and other fuels increases towards the end of horizon relative to the Baseline, while the demand share of electricity, natural gas, coal and oil is reduced.

The induced electricity-price increase results in electricity demand reductions and substitution of electricity for other fuels by the end-users. The price elasticity for the attendant end-use demand reductions is −0.30 for all demand sectors represented in GMM. Fig. 7 illustrates changes of the final electricity demand in externality scenarios compared to the Baseline. While the consumption of electricity is reduced in both industrial and residential & commercial sectors, the transport sector is not affected. The largest reduction is projected in the industrial sector, since this sector has the greatest ability to switch from electricity to other fuels. The most significant electricity-demand reductions are observed between 2020 and 2040 and are associated with premature closing of existing electricity sources based of coal combustion during this period. The electricity-demand reductions are lowered in 2050, and represent for the industrial sector a relative decrease over the Baseline of 9% in the Local and 24% in the Global externality scenario.

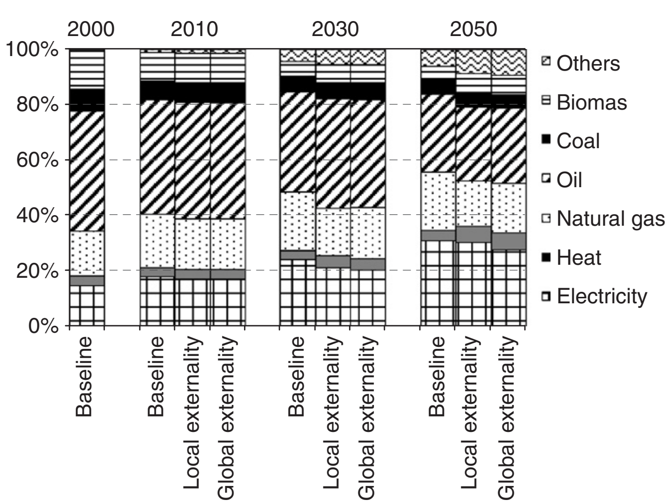


Fig. 6. Relative fuel shares in total final energy demand in 2050 in all scenarios.

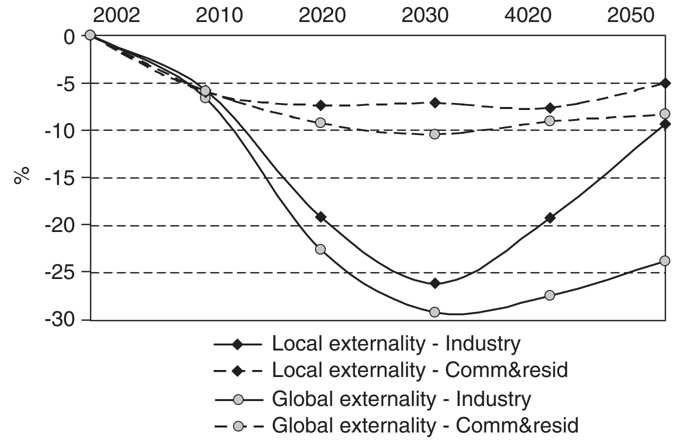


Fig. 7. Change in electricity consumption in the industrial and residential&commercial demand sectors relative to the Baseline.

5.2. Environmental impacts

Internalisation of external cost into the total production cost of electricity leads to rapid emission-reducing effect in both externality scenarios. Fig. 8 represents the relative change of global air emission over the Baseline. For all considered emissions (CO_2, SO_2, NO_x), the most significant reduction occurs within the period 2000–2030 and is associated with a substantial fallback of coal-based power generation implicit to the premature retirement of coal plants without SO_2/NO_x control. Until the year 2040, the emission reduction is partly stabilised. At the end of the time horizon, different developments can be observed in CO_2 emissions and local pollutants. As the (learning) technologies based on fossil fuels coupled with CO_2-removal start to penetrate the market between 2040 and 2050, total CO_2 emissions are reduced by 25% in the Global externality scenario, as compared to the Baseline. On the other hand, substantial decrease in SO_2 and NO_x emissions relative to the Baseline scenario, reported for periods by 2040, is less pronounced in the end of horizon. By 2050, the advanced fossil systems with ETL option (NGCC, advanced coal, IGCC) increase the market share as compared to the earlier periods.

Significant CO_2-emission reduction for the Local externality scenario suggests, that important ancillary benefits can be expected from policies that directly address other environmental issues than CO_2-mitigation.

5.2.1. Emissions of local air pollutants (SO_2/ NO_x)

Fig. 9 shows total SO_2 and NO_x emissions from the power production. To illustrate the effect of external cost on the emission reduction, no local or regional pollution mitigation policies are considered across the world regions in the Baseline scenario. The SO_2 emissions peak in the period 2030–2040 at the rate of 140 Mt SO_2 per year in the Baseline scenario, with region of ASIA being the main contributor to the emissions level. With lowered share of conventional coal plants, the sulphur emissions decrease

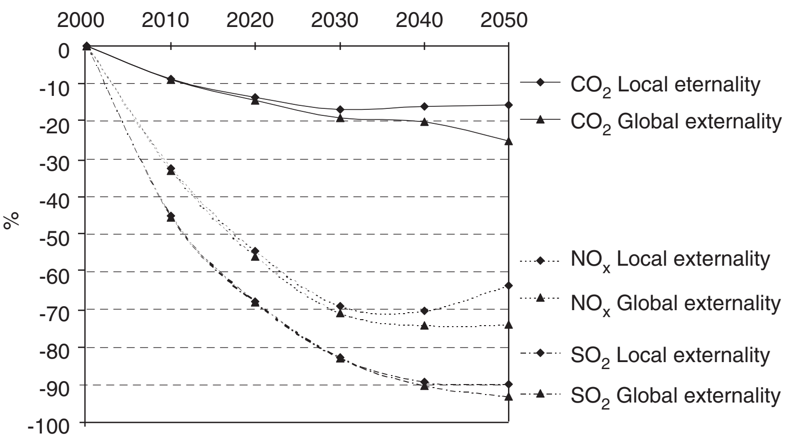


Fig. 8. Change in the global air emissions over the Baseline scenario.

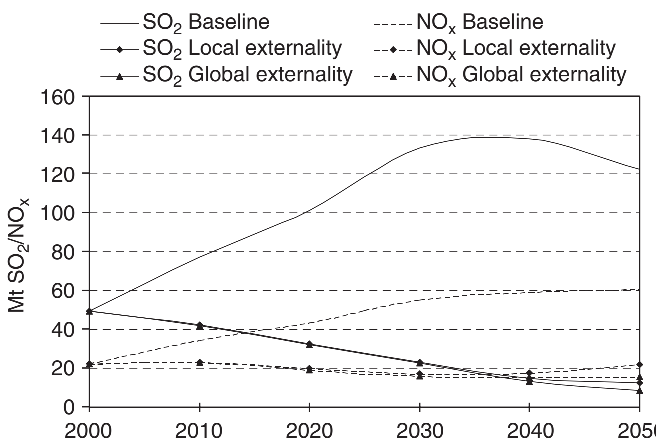


Fig. 9. Development of SO_2 and NO_x emissions from the power generation sector under the Baseline and externality scenarios.

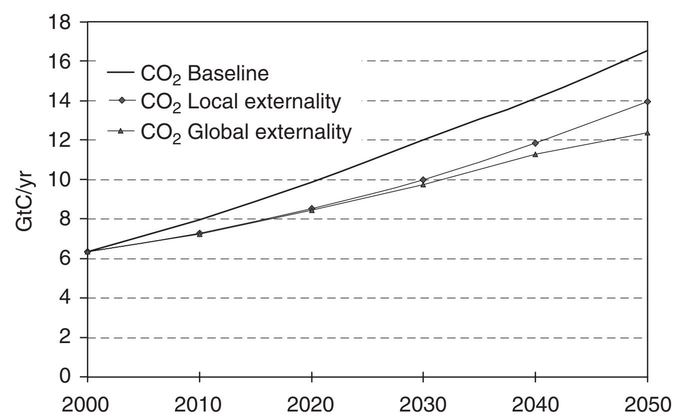


Fig. 10. Development of global CO_2 emissions under the Baseline and externality scenarios.

significantly until 2050. Since the desulphurisation systems together with advanced coal and IGCC displace the conventional coal from the system in the externality scenarios, the reduction effect is considerable. The emissions of NO_x increase in the Baseline until 2030 and then are stabilised at the annual rate around 60 Mt NO_x per year. In the externality scenarios, there is no substantial increase in the NO_x emissions observed until 2040. Then in 2050, the level of NO_x grows by 20% in the Local externality scenario and by 5% in the Global externality scenario relative to 2040, because of increased penetration of new fossil-based technologies.

5.2.2. Global CO_2 emissions

Total global carbon emissions in the Baseline scenario rise during the whole time horizon with an annual rate of 1.8% and reach a level of 16.5 Gt of carbon in 2050. In the Local externality scenario, total emission level is lowered by 15% in 2050, and the annual growth rate is reduced to 1.5%. In the Global externality scenario, the CO_2 emissions annual growth is 1.3% and culminates around the year 2040. As shown in Fig. 10, the carbon emissions growth reduction appears by 2050 and the level of 12.4 Gt of carbon is projected at the end of the time horizon.[5]

The decarbonisation effect of the policy comprising internalisation of external cost can be demonstrated by a break-down of the different CO_2 reduction components, as is shown in Fig. 11. Five carbon reducing components were considered: carbon capture and sequestration, inter-fossil switching (i.e., from coal to natural gas), reduction of fossil fuel fraction resulting from increases in nuclear energy use,

[5]To eliminate the adverse effects of global climate change, the 550 ppmv CO_2-concentration target is frequently used as a precautionary, but attainable level and represent the middle value of stabilisation level identified by Wigley et al. (1996). Global carbon-emission trajectory aimed to achieve the 550 ppmv target in the long run, as indicated by IPCC (2001), implies the maximum energy related CO_2 emissions of 10 GtC/yr by ~2050. The results presented in this section indicate that the policies internalising external cost only to the power sector, as formulated in this modelling exercise, might not be sufficient to reduce global carbon emissions to levels needed for 550 ppmv target, or, that the efforts to curb CO_2 emissions will have to be further accelerated in the second half of the 21st century.

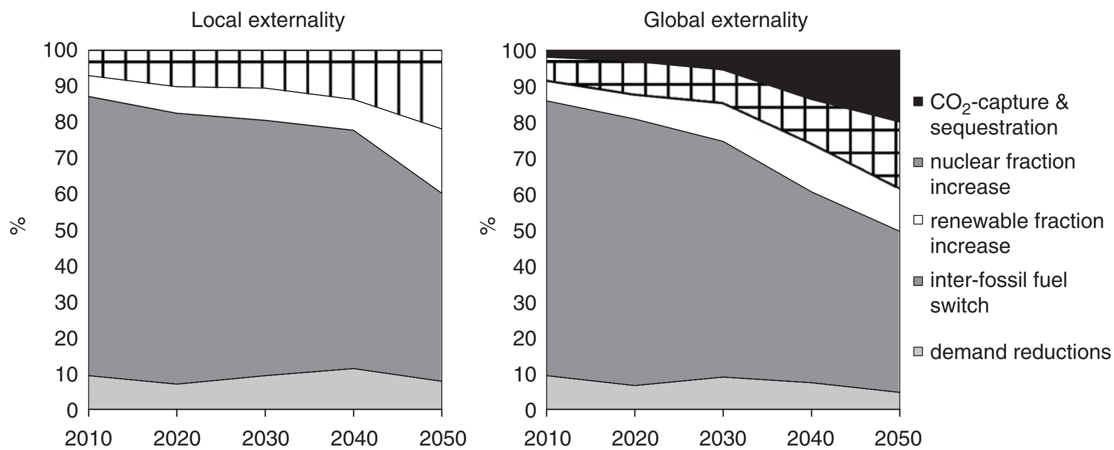


Fig. 11. Break-down of CO_2 reduction components under externality scenarios.

reduction of fossil fuel fraction in favour of renewables, and finally the reduction of end-use demand due to implementation of externality policies (Kypreos, 1990). In both externality scenarios, the inter-fossil fuel switching plays a dominant role in carbon mitigation and contributes 45–53% of CO_2 reduction in 2050. The important role of a larger use of nuclear energy is reflected in the CO_2-emissions reducing effect, as in the time period 2020–2050 the nuclear energy contributes to 10–22% of the total reduction. Carbon removal from fossil fuel combustion plays a significant role in the Global externality scenario. Its share in overall CO_2-mitigation process in 2050 corresponds to 20%.

5.3. Cost impacts

5.3.1. Electricity generation cost analysis

To evaluate the competitiveness of different power generation technologies, a simplified calculation of electricity generation cost has been performed. The calculation assesses the impacts of internalisation of different externality modes on total production cost, as well as the effect of 'learning-by-doing' on the cost development over time. The methodology used for calculation of electricity generation cost is elaborated in Appendix I.

Fig. 12 summarizes results of the total generation cost calculation for the Baseline and externality scenarios for the present situation and cost projection for the year 2050. The ASIA region is taken as an example for the analysis. The Baseline scenario results in 2000 indicate, that without external cost, NGCC, conventional pulverised coal and coal power plants with $DeSO_x$/$DeNO_x$ are the cheapest alternatives at 3.1, 3.4 and 3.9 ¢/kWh, respectively. The projected generation costs in the Baseline scenario in 2050 reflect the change in fuel cost, the impact of ETL towards reduction of investment cost with accumulation of installed capacity by 'learning' technologies in 2050, and expected improvement in the conversion efficiency and a higher average load factor. The least cost systems are wind turbines, IGCC and advanced coal power plants with projected generating cost at the level of 2.4, 2.5 and 2.8 ¢/kWh, respectively. Clearly, the significant cost reduction for technologies undergoing strong "learning" effect in the Baseline scenario is not autonomous, but is related to the specific assumptions about energy-technology dynamics. Cost reduction inherent to LBD-concept will not occur without policy-actions in favour of advanced generation systems.

Applying policies that internalise external cost from the local pollution in the generation cost, the competitiveness of technology portfolio changes towards the end of the time horizon. The least cost options in this case in 2050 are the wind turbines, IGCC and advanced coal plants with total generation cost of 2.6, 3.5 and 4.0 ¢/kWh, respectively. High external cost makes the coal power plant without emission control the most expensive electric-power source among fossil-fuelled systems, which explains the massive elimination of this technology from the generation mix. In the case of internalised global externalities, the most competitive systems are those with low- or zero-emission rates: the wind turbines (2.6 ¢/kWh), followed by advanced nuclear power plants (4.8 ¢/kWh) and IGCC with CO_2-capture (5.1 ¢/kWh). Although the generation costs in both externality scenarios increase as compared to the Baseline scenario, the higher competitiveness of advanced fossil systems, advanced nuclear and renewable energy technologies implies a decreased dependency of the electricity sector on the fossil-fuel supplies.

The regional impacts of the policy instrument that internalise external cost into the power generation cost, as implemented in this analysis, is portrayed by comparing changes in the shadow price of electricity in regions represented in GMM[6]. Table 5 shows that the range of

[6]The shadow price of electricity resulting from the model run is equal to the marginal value of the electricity for the regional energy system as a whole. There are six electricity prices, one for each time-slice defined in GMM (i.e., summer day; summer night; winter day; winter night;

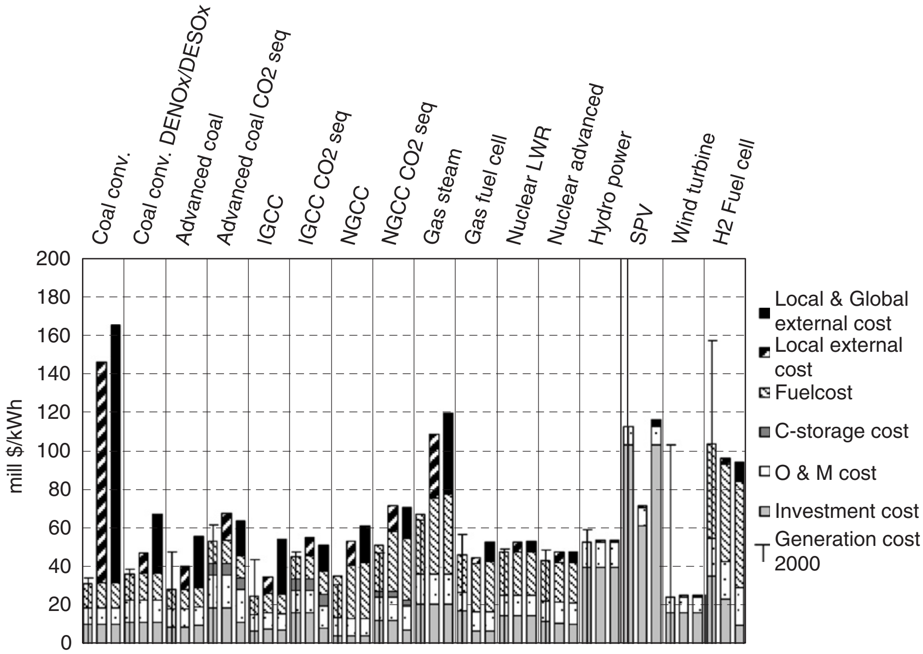


Fig. 12. Break-down of the cost components for power generation by scenarios in 2050 (example region ASIA) at 5% real discount rate. Bars from left to right represent the Baseline, Local externality and Global externality scenarios. [1 \$ = 100 ¢ = 1000 mills].

Table 5
Increase in average shadow price of electricity in externality scenarios relative to the Baseline

Region	Scenario	2010 (¢/kWh)	2020 (¢/kWh)	2030 (¢/kWh)	2040 (¢/kWh)	2050 (¢/kWh)
NAME	Local externality	1.7 (41%)	1.8 (25%)	1.1 (22%)	1.3 (23%)	1.0 (19%)
	Global externality	2.3 (58%)	2.7 (51%)	1.8 (45%)	1.7 (39%)	1.7 (42%)
OECD	Local externality	1.9 (38%)	1.5 (30%)	1.3 (24%)	1.5 (26%)	1.5 (32%)
	Global externality	2.7 (54%)	2.1 (60%)	2.0 (52%)	1.0 (23%)	2.5 (77%)
EEFSU	Local externality	1.6 (36%)	1.5 (24%)	3.2 (37%)	2.2 (33%)	0.8 (16%)
	Global externality	1.7 (40%)	2.5 (52%)	4.0 (72%)	2.4 (51%)	1.7 (40%)
ASIA	Local externality	2.6 (52%)	6.4 (54%)	4.2 (44%)	3.8 (42%)	1.7 (28%)
	Global externality	5.2 (103%)	6.7 (123%)	4.1 (77%)	3.8 (71%)	2.2 (50%)
LAFM	Local externality	0.4 (10%)	0.7 (14%)	1.2 (27%)	1.1 (24%)	1.2 (21%)
	Global externality	0.6 (14%)	0.8 (19%)	1.8 (57%)	1.7 (47%)	1.3 (29%)
WORLD	Local externality	1.6 (36%)	2.4 (51%)	2.2 (50%)	2.0 (45%)	1.2 (30%)
	Global externality	2.5 (55%)	3.0 (64%)	2.7 (62%)	2.1 (48%)	1.9 (46%)

increases in the average shadow price values is rather large (from 0.4 to 6.7 ¢/kWh), depending on the time period and the region. More interesting than the numerical results is the observation, that the price increase is significantly higher in both externality scenarios for regions largely relying on the coal-based electricity production, e.g., the ASIA region. An increment in the shadow price decreases over the time horizon in most of the regions. Large increase in the price in periods 2010–2020 suggests, that the timing of implementation of the policy is particularly important, and a smoother or a gradual introduction of externalities is appropriate for developing regions where fossil fuels burning constitute the main source of energy.

It has to be stressed, that the results presented in this section are indicative and bear all the uncertainties related to the fuel prices development and assumed learning parameters of systems with ETL option (progress ratio, annual growth and declination rates, floor cost; see Table 1). Another policy relevant comment pertinent to the presented values is that the extent of externality charges associated with emission of pollutants influences significantly the level of cumulative installed capacity of power plants. In other words, the technologies with high external cost are introduced into the system at a lower rate

(*footnote continued*)
intermediate day; intermediate night). This provides a composite electricity price which is the average of the 6 electricity shadow prices, and which represents the price of a kWh produced throughout the 6 time-slices (EIA, 2003b).

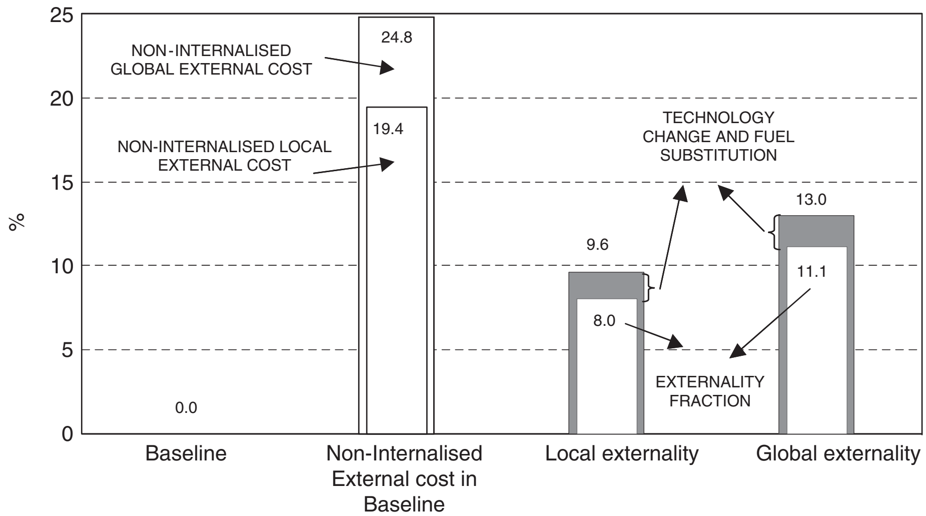


Fig. 13. Change in the cumulative discounted energy system cost relative to the Baseline scenario, including external cost fraction.

and their investment cost reduction because of ETL is impaired. On the contrary, a reverse effect can occur for technologies with low externality charges: their learning performance is accelerated and results in a higher market penetration.

5.3.2. *Total system cost*

As the energy system tries to avoid paying the external costs, new (investment intensive) technologies are being installed and structural changes take place. This leads to significant increase in total system cost over the whole time horizon. The highest contribution of the externality charges to the increase in total system cost, however, occurs in the first period after their introduction (2010). This result confirms again the importance of proper timing of the policy implementation. Fraction of undiscounted externality charges in the total cost is shrinking both in relative and absolute terms towards the end of horizon, reflecting the capability of the energy system to minimize the extra charges through the structural changes and fuel switching.

Model runs indicate a high relative change in the cumulative total discounted system costs, i.e., the objective function used in GMM, due to inclusion of the additional charges on the power generation. This increase over the Baseline totals in externality scenarios to 9.6% and 13%. As is indicated in Fig. 13, the contribution of the external cost itself counts for 85% of the total increase in both externality scenarios. The reminder is attributed to the structural changes and fuel switch occurring within the energy system.

Fig. 13 also shows a ‘‘hypothetical’’, non-internalised external cost associated with the Baseline scenario. The non-internalised external cost approximates the cumulative discounted damage cost produced by the electricity sector. This cost is not taken into account in the price of electricity, but is imposed on the society in a form of environmental and health damages. This analysis indicates that the non-internalised externalities might represent around 25% of the total discounted system cost of the Baseline scenario. On the other hand, the level of energy system cost increase in externality scenarios demonstrates the ability of the energy system to adjust the overall cost well below the environmental damages that occur in the Baseline.

6. Conclusions

Internalisation of external cost in the price of electricity is an important policy instrument towards the sustainable development in the energy use. Modelling the impacts of such policies carries certain limitations and uncertainties, among which the most important are issues of valuing socio-political priorities of future energy sector developments, socio-political acceptance of technological options, income distribution effects, discounting of the future damages to the present value, regional differences in valuing externalities, or the rate of technological change. While these issues were beyond the scope of our analysis, number of conclusions and insights can be derived from the inclusion of externalities into the power generation system, as performed using the Global Multi-regional MARKAL model.

Internalisation of externalities with and without climate change impacts fosters a rapid introduction of emissions control systems and low-emitting power plants. Scenarios analysis reveals substantial changes in the electricity production system, i.e., diffusion of advanced technologies and fuel switching. In the case of the local externalities, the technologies such as coal power plants with emission control, advanced coal power plants and IGCC replace the conventional coal systems. Natural gas combined cycle, nuclear power and renewables increase their share in the power generation mix. The scenario with global externalities further accelerates the structural changes in the power production sector. Contribution of the coal-based generation is strongly reduced, and production from the systems with carbon removal accounts for 36% of total electricity generation from coal power plants. Natural gas combined cycle systems play a dominant role, and a significant increase in the nuclear energy production is reported. Renewable systems, as well as fuel cells, increase their competitiveness. GMM model runs indicate some efficiency loss due to the use of scrubbers ($DeNO_x$, $DeSO_x$, and C-capturing), however, the dependency of the electricity sector on the fossil fuels is considerably lower as compared to the Baseline.

Externality charges on power generation increase the price of electricity for the end-users. Therefore, the reduction in final demand for electricity in industrial and residential & commercial sectors takes place; electricity consumption is partly substituted by other fuels, e.g., heat, biomass.

The inclusion of external costs in the price of electricity has positive global and local environmental impacts due to significant emissions reduction. Emissions of SO_2 and NO_x decrease by 70–85% in 2030 relative to the Baseline scenario, then their elimination slows down with rising installation of new fossil-based systems, such as advanced coal, IGCC, NGCC. The modelling results show a strong decarbonisation effect of policies internalising externalities in the electricity sector. Breakdown of carbon emissions reduction components suggests the major contributions of the inter-fossil switch and increase in nuclear and renewable fraction in the primary energy use. Since the carbon sequestration technologies become competitive in the Global externality scenario, they appear to be an attractive technological option in carbon abatement process.

Significant reduction in CO_2-emissions associated with the Local externality scenario suggests, that synergies and ancillary benefits can be invoked by policies that directly address other sustainability issues than CO_2-mitigation.

Increase in the total energy system cost in the externality scenarios associated with structural changes and fuel substitutions induced by internalisation of externalities represent 1.6% and 1.9% relative to the Baseline. On the other hand, 'learning-by-doing' aids in moderating the level of external cost penalty. While analysis performed with GMM indicates that advanced systems with emission control and carbon capture will undergo significant cost reduction and will become competitive in the long run, policies supporting these technologies is a prerequisite to establish them in the electricity markets, especially in the initial period of their market penetration. This refers to policy measures for the stimulation of technological progress via learning investments and RD&D expenditures that help advanced and carbon-free technologies to follow their learning curves.

Acknowledgment

This study has been conducted within the Swiss NCCR-Climate (grant from the Swiss National Science Foundation) and the EU Research Project ACROPOLIS (grant from the Swiss Federal Office for Education and Science).

We would like to thank to Mr. B. Krakowski, Mr. J. Sager, Mr. D. Krzyzanowski and Mr. L. Gutzwiller for their critique and comments on the analysis. Constructive suggestions from three unknown referees that helped to improve the presentation of this study are highly appreciated.

Appendix I. Calculation of the electricity generation cost

Total electricity generation cost (also referred to as levelized cost of energy or busbar cost) is calculated according this formula adopted from Drennen et al. (2003):

$$TGC = \frac{I * CRF}{Q} + \frac{FIXO\&M}{Q} + \frac{VARO\&M}{Q} + \frac{F}{Q} + \frac{E}{Q},$$

where: I is the capital investment cost, CRF the capital recovery factor, Q the annual plant output (kWh), $FIXO\&M$, the fixed O&M cost, $VARO\&M$ the variable O&M cost (including CO_2-storage cost), F the fuel cost and E the external cost

$$CRF = dr * \frac{(1 + dr)^n}{(1 + dr)^n - 1},$$

where dr is the (real) discount rate, and n the plant life time.

Calculation of specific investment cost for the learning technologies in 2050 follows the approach described by Barreto (2001):

$$I_{2050} = I_0 * \left(\frac{CC_{2050}}{CC_0}\right)^{-b},$$

where I_{2050} is the specific investment cost in 2050, I_0 the specific investment cost at the starting point when technology is introduced into the system, CC_{2050} the cumulative capacity of the technology in 2050, CC_0 the cumulative capacity of the technology at the starting point, b the learning index

$$-b = \frac{\ln pr}{\ln 2} \Rightarrow pr = 2^{-b},$$

where, pr is the progress ratio, or the rate at which the cost declines each time the cumulative production doubles (e.g., a progress ratio of 80% implies that the costs are reduced by 20% relative to the original value when the cumulative capacity is doubled).

Appendix II. Region-specific external cost in ¢/kWh

See Table 6.

Tabel 6
External costs applied for five world regions in GMM ¢/kWh

Technology	ASIA				OOECD				NAME				EEFSU				LAFM			
	excl CO_2		*incl CO_2*		*excl CO_2*		*incl CO_2*		*excl CO_2*		*incl CO_2*		*excl CO_2*		*incl CO_2*		*excl CO_2*		incl CO_2	
Year	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050	2010	2050
Fossi-fuel based power plants																				
Coal conventional electric	15.23	13.73	17.72	15.98	11.12	10.94	13.49	13.28	7.48	7.36	9.86	9.70	10.23	9.22	12.72	11.47	10.23	9.22	12.72	11.47
Coal conventional electric with $DeSO_x/DeNO_x$	1.34	1.26	3.91	3.66	1.28	1.26	3.74	3.66	0.93	0.91	3.39	3.31	0.97	0.91	3.54	3.31	0.97	0.91	3.54	3.31
Coal conv. with $DeSO_x/DeNO_x$ and CO_2 seq	1.75	1.53	2.42	2.12	1.56	1.53	2.16	2.13	1.13	1.11	1.68	1.64	1.26	1.11	1.88	1.64	1.26	1.11	1.88	1.64
Coal advanced electric	1.60	1.44	3.58	3.22	1.60	1.44	3.58	3.22	1.13	1.01	3.10	2.79	1.13	1.01	3.10	2.79	1.13	1.01	3.10	2.79
Coal advanced electric with CO2 seq	1.92	1.69	2.42	2.14	1.92	1.69	2.42	2.13	1.35	1.20	1.78	1.58	1.35	1.20	1.78	1.58	1.35	1.20	1.78	1.58
Coal IGCC	1.14	1.00	3.85	3.38	1.14	1.00	3.85	3.38	0.84	0.74	3.56	3.12	0.84	0.74	3.56	3.12	0.84	0.74	3.56	3.12
Coal IGCC with CO_2 seq	1.34	1.17	1.84	1.61	1.34	1.17	1.84	1.61	0.99	1.00	1.45	1.40	0.99	1.00	1.45	1.40	0.99	1.00	1.45	1.40
Natural Gas Combined Cycle (NGCC)	1.61	1.46	2.51	2.28	0.48	0.44	1.38	1.25	0.39	0.35	1.29	1.17	1.14	1.03	2.04	1.85	1.14	1.03	2.04	1.85
NGCC with CO_2 scrubber	1.82	1.63	2.13	1.90	0.55	0.48	0.72	0.63	0.44	0.39	0.60	0.54	1.29	1.14	1.54	1.37	1.29	1.14	1.54	1.37
Gas turbine	2.38	2.38	3.72	3.72	2.38	2.38	3.72	3.72	1.69	1.69	3.02	3.02	1.69	1.69	3.02	3.02	1.69	1.69	3.03	3.02
Gas steam conventional	4.18	3.92	5.43	5.09	4.08	3.92	5.30	5.09	2.89	2.70	4.14	3.87	2.88	2.70	4.13	3.87	2.88	2.70	4.13	3.87
Cogeneration gas turbine	2.49	2.49	3.88	3.88	2.32	2.32	3.62	3.62	1.76	1.76	3.15	3.15	1.76	1.76	3.15	3.15	1.76	1.76	3.15	3.15
Gas fuel cell (GFC)	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.69	1.17	1.17
Hydrogen fuel cell (CHP) in industry (H_2FC)	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17
Hydrogen fuel cell (CHP) in res&com. (H_2FC)	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17	0.39	0.39	1.17	1.17
Oil electric	3.52	2.67	5.78	4.38	3.52	2.67	5.78	4.38	2.23	1.84	4.31	3.56	2.43	1.84	4.69	3.56	2.43	1.84	4.69	3.56
Nuclear and renewable power plants																				
Nuclear plant—Light Water Reactor (LWR)	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68
Advanced new nuclear power plant (NNU)	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68	0.65	0.65	0.68	0.68
Hydro-electric plant (small and large)	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Solar photovoltaics (SPV)	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39
Solar thermal electric	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39	0.13	0.13	0.39	0.39
Wind turbine	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Biomass power plant	0.39	0.39	0.59	0.59	0.39	0.39	0.59	0.59	0.39	0.39	0.59	0.59	0.39	0.39	0.59	0.59	0.39	0.39	0.59	0.59
Geothrmal electric	0.20	0.20	0.59	0.59	0.20	0.20	0.59	0.59	0.20	0.20	0.59	0.59	0.20	0.20	0.59	0.59	0.20	0.20	0.59	0.59

Costs are given in US$ (1995) Original values given in €(1995) have been converted by using conversion factor 1.3.

References

ACROPOLIS (Assessing Climate Response Options: POLIcy Simulations), 2003. Final report, April 2001–September 2003. Prepared by IER, Germany and IPTS-JRC, Spain, Contract No: ENK6-CT-2000-00443, Framework Programme V (1998–2002).

Barreto, L., Kypreos, S., 2002. Multi-regional Technological Learning in the Energy-Systems MARKAL model. International Journal of Global Energy Issues 17 (3), 189–213.

Barreto, L., 2001. Technological learning in energy optimisation models and deployment of emerging technologies. Ph.D. Dissertation No 14151, Swiss Federal Institute of Technology, Zurich, Switzerland.

David, J., Herzog, H., 2000. The cost of Carbon Capture, presented at the Fifth International Conference on GHG Control Technologies, Cairns, Australia.

Drennen, T.E., Baker, A.B, Kamery, W., 2003. Electricity generation cost simulation model (GenSim), In: second Quarter 2003 Newsletter of The International Association for Energy Economics, pp. 18–21.

EC (European Commission), 1998a. ExternE—externalities of energy, JOULE III non nuclear energy programme, externe final report. European Commission, Directorate-General XII, Science, Research and Development, Brussels. See also: http://externe.jrc.es.

EC (European Commission), 1998b. ExternE—externalities of energy – methodology annexes, JOULE III non nuclear energy programme. European Commission, Directorate-General XII, Science, Research and Development, Brussels.

EIA (Energy Information Administration), 2003a. Assumptions for the annual energy outlook 2003 with projections to 2050. Energy Information Administration, Office of Integrated Analysis and Forecasting, U.S. Department of Energy (DoE), Washington, DC.

EIA (Energy Information Administration), 2003b. System for the analysis of global energy markets (SAGE) – Model Documentation. Energy Information Administration, Office of Integrated Analysis and Forecasting, U.S. Department of Energy (DoE), Washington, DC.

Fishbone, L.G., Abilock, H., 1981. MARKAL, a linear-programming model for energy systems analysis: technical description of the BNL version. International Journal of Energy Research 5, 353–375.

Fouquet, R., Slade, R., Karakoussis, V., Gross, R., Bauen, A., Anderson, D., 2001. External costs and environmental policy in the United Kingdom and the European Union. Occasional Paper 3. Centre for Energy Policy and Technology, Imperial College, London.

GeoHive, 2003. Global Statistics. Electronic database of global statistical information maintained by J. van der Heyden. See also: http://www.geohive.com/.

Hinrichs, W., 1999. Stoffmengenflüsse und Energiebedarf bei der Gewinnung ausgewählter mineralischer Rohstoffe, Teilstudie Steinkohle. Reihe H, Wirtschaftsgeologie, Berichte zur Rohstoffwirtschaft, Heft SH 4, Geologisches Jahrbuch, Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, Germany.

Hirschberg, S., Dones, R., Gantner, U., 2000. Use of external cost assessment and multi-criteria decision analysis for comparative evaluation of options for electricity supply. In: PSI Annual Report 2000—Annex IV, Villigen, pp. 19-24.

IEA (International Energy Agency), 2001. World Energy Outlook—2001 Insights, Assessing Today's Supplies to Fuel Tomorrow's Growth. Economic Analysis Division—OECD/IEA Publications, Paris, France.

IPCC (Intergovernmental Panel on Climate Change), 2000. Special report on emission scenarios. A Special Report of the working group III of intergovernmental panel on climate change, Cambridge University Press, Cambridge.

IPCC (Intergovernmental Panel on Climate Change), 2001. Climate Change 2001: Mitigation. Contribution of Working Group III to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge.

Kanudia, A., Loulou, R., 1999. Advanced bottom-up modelling for national and regional energy planning in response to climate change. International Journal of Environment and Pollution 12 (2/3), 191–216.

Kypreos, S., 1990. Energy Scenarios for Switzerland and Emission Control, Estimated with a Normative Model. PSI Bericht Nr. 90-70. Paul Scherrer Institute, Villigen.

Lako, P., Seebregts, A.J., 1998. Characterisation of power generation options for the 21st century: report on behalf of Macro task E1. ECN Report, ECN-C–98-085, Petten, The Netherlands.

Loulou, R., Lavigne, D., 1996. MARKAL model with elastic demands: application to greenhouse gas emission control. In: Carraro, C., Haurie, A. (Eds.), Operations Research and Environmental Management. Kluwer Academic Publishers, The Netherlands, pp. 201–220.

Messner, S., 1997. Endogenised technological learning in an energy systems model. Journal of Evolutionary Economics 7, 291–313.

Riahi, K., Roehrl, R.A., 2000. Greenhouse gas emissions in a dynamics-as-usual scenario of economic and energy development. Technological Forecasting and Social Change 63, 175–205.

Rogner, H.H., 1997. An assessment of world hydrocarbon resources. Annual Review of Energy and the Environment 22, 217–262.

UNDP (United Nations Development Programme), 2000. World energy assessment: Energy and the challenge of sustainability. In: Goldemberg, J. (Ed.), United Nations Development Programme. United Nations Department of Economic and Social Affairs, World Energy Council, New York, USA.

Wigley, T.M.L., Richels, R., Edmonds, J.A., 1996. Economic and environmental choices in the stabilization of atmospheric CO2 concentrations. Nature 379, 240–243.

Wu, Z., DeLaquil, P., Larson, E.D., Chen, W., Gao, P., 2001. Future Implications of China's Energy-Technology Choices, prepared for the Working Group on Energy Strategies and Technologies (WGEST) of the China Council for International Cooperation on Environment and Development (CCICED), 24 July 2001 (available from the authors).

ECOLOGICAL ECONOMICS 67 (2008) 140–152

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/ecolecon

ANALYSIS

The internalization of externalities in the production of electricity: Willingness to pay for the attributes of a policy for renewable energy☆

Alberto Longo[a,*], *Anil Markandya*[b,c], *Marta Petrucci*[d]

[a]*Queen's University Belfast, UK*
[b]*Department of Economics and International Development, University of Bath, UK*
[c]*Fondazione Eni Enrico Mattei, Italy*
[d]*NERA Economic Consulting, Italy*

ARTICLE INFO

Article history:
Received 20 June 2007
Received in revised form
2 November 2007
Accepted 6 December 2007
Available online 30 January 2008

Keywords:
Non-market valuation
Choice experiments
Willingness to pay
Renewable energy
Energy security
Greenhouse gases emissions

JEL classification:
Q42; Q48; Q51

ABSTRACT

This paper investigates the willingness to pay of a sample of residents of Bath, England, for a hypothetical program that promotes the production of renewable energy. Using choice experiments, we assess the preferences of respondents for a policy for the promotion of renewable energy that: (i) contributes to the internalization of the external costs caused by fossil fuel technologies; (ii) affects the short-term security of energy supply; (iii) has an impact on the employment in the energy sector; and (iv) leads to an increase in the electricity bill. Responses to the choice questions show that our respondents are in favour of a policy for renewable energy and that they attach a high value to a policy that brings private and public benefits in terms of climate change and energy security benefits. Our results therefore suggest that consumers are willing to pay a higher price for electricity in order to internalize the external costs in terms of energy security, climate change and air pollution caused by the production of electricity.

1. Introduction and motivation

Over the last fifteen years there has been a significant research effort in measuring the external costs caused by electricity production (European Commission, 1998; Friedrich and Bickel, 2001; Krewitt, 2002; European Commission, 2003; Markandya, 2003; NewExt, 2004; ExternE-Pol, 2005, European Commission, 2005). It is well established that air pollution, acid deposition, and accidents caused by the production of electricity have negative effects both on human health and on the environment. For example, human health is affected in terms of reduced life expectancy and increased respiratory hospital admissions, while the environment is affected through yield change of crops and global warming (European

☆ The opinions expressed in the paper reflect the authors' views and do not necessarily reflect the views of their respective institutions.

* *Corresponding author.* Institute of Agri-Food and Land Use, School of Biological Sciences, Queen's University Belfast, David Keir Building, Stranmillis Road, BT9 5AG, Belfast, Northern Ireland, UK. Tel.: +44 28 90976537; fax: +44 28 90976513.
E-mail address: a.longo@qub.ac.uk (A. Longo).

0921-8009/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolecon.2007.12.006

Commission, 2003). Using a non demand[1] based bottom-up impact pathway approach,[2] the team of researchers of ExternE has quantified in monetary terms most of the damages to human health and the environment caused by the different fuels and technologies that generate electricity. The external costs estimates are substantial; for example, ExternE-Pol (2005) has estimated that they are in the range of 1.6–5.8 c€/kWh for current fossil fuel based systems, with figures at the lower end for gas based generation technologies and the upper end for traditional coal technologies. The results of the ExternE research also indicate the importance of the effects in terms of human health and global warming: at the end of the 1990s ExternE identified that health impacts comprised 98% of the external costs from SO_2 emissions and almost 100% of those from particulates (European Commission, 1999), with mortality impacts accounting for around 80% of those health impacts. The costs associated specifically with global warming range widely and differ for fuel. The current phase of ExternE uses the *avoidance cost* methodology for valuing the external costs of global warming because, according to ExternE, the monetary valuations to date of global warming externalities have not yet been satisfactory. ExternE suggests to use the value of 19 €/tCO_2 because that is the avoidance cost for achieving the Kyoto targets in the EU (European Commission, 2005).

One reason to support the promotion of renewable energy comes from the internalization of external costs. Economists have shown that when externalities are present, markets are not efficient unless these external costs are internalized and economic agents take into account these costs when making decisions. Market based instruments, such as subsidies have been widely used to address externalities. The recent "Community guidelines on state aid for environmental protection" of the European Commission support the use of subsidies to promote the production of renewable energy: the guidelines allow Member States to "grant operating aid [limited to a maximum of 5 c€/kWh] to new plants producing renewable energy that will be calculated on the basis of external costs avoided" (European Commission, 2001). A second major reason to stimulate the production of renewable energy comes from the increasing demand for electricity, and, moreover, a demand for secure electricity (European Commission, 1995, 1997, 2000).

The UK energy policy aims to comply with the commitment of the Kyoto Protocol, which requires a reduction of the UK greenhouse gases emissions to 12.5% below the 1990 levels during the period 2008–2012. By increasing the share of consumption of renewable energy from 3% of 2002 levels to 10% of UK electricity in 2010, "renewable energy will also play an important part in reducing carbon emissions, while also strengthening energy security" (DTI, 2003, page 11). The electricity supply industry was liberalised in Great Britain in 1999 (Batley et al., 2001); today consumers have the opportunity to decide their supplier and the mix of energy, whether traditional or "green" electricity. This means that the demand for renewable electricity might directly contribute to an increase in its production. As Ek (2005) points out, if people are willing to pay to support the production of renewable energy, we can expect that an increase in the production of renewable energy would be welcome. In particular, it is essential to understand what people think about these changes since they are the ones primarily affected. This change in the supply of electricity can have major effects on the structure of society, including employment in the energy sector.

The focus of this paper is to investigate the perception and the willingness to pay of UK energy users for different characteristics of energy programs that stimulate the production of renewable energy by using choice experiments (Louviere et al., 2000; Alberini et al., 2007). The methodology is applied to a sample of respondents in the city of Bath, England, to elicit their preferences for different hypothetical policies for the promotion of renewable energy that (i) contribute to the internalization of the external costs caused by fossil fuel technologies; (ii) affect the security of short-term energy supply; (iii) have an impact on the employment in the energy sector; (iv) and lead to an increase in the electricity bill. The paper is structured as follows. Section 2 reviews the literature on willingness to pay to avoid energy black-outs and to support renewable energy. Section 3 presents the survey instrument and its administration. Section 4 describes the economic and econometric models aimed at answering our research questions. Sections 5 presents the results of the econometric models and Section 6 presents some concluding remarks.

2. Literature review

Past research has investigated consumers' willingness to pay (WTP) for renewables focusing on environmental effects (Champ and Bishop, 2001; Roe et al., 2001; Alvarez-Farizo and Hanley, 2002; NewExt, 2004; Ek, 2005; Bergmann et al., 2006), on health effects (Johnson and Desvousges, 1997; Bergmann et al., 2006), and on social aspects (Johnson and Desvousges, 1997; Bergmann et al., 2006). Other studies have focused on renewable energy without directly investigating the impacts on environment, health or social aspects (Farhar and Houston, 1996; Farhar, 1999; Farhar and Coburn, 1999; Zarnikau, 2003; Wiser, 2003; Menges et al., 2005; Aravena et al., 2006). A group of studies has focused on the value of security of energy

[1] Turner et al. (1994, p.114) divide the methods for valuing the environment into two groups, demand curve approaches, such as contingent valuation, choice experiments, travel cost method; and non-demand curve approaches, such as dose-response method, avoidance cost method. The latter approach even if it fails to provide 'true' valuation information, has been favoured by ExternE as it provides useful information to policymakers. On the contrary, the approach we use in our study, choice experiments, is a demand curve approach.

[2] "The impact pathway assessment is a bottom-up-approach in which environmental benefits and costs are estimated by following the pathway from source emissions via quality changes of air, soil and water to physical impacts, before being expressed in monetary benefits and costs. The use of such a detailed bottom-up methodology – in contrast to earlier top-down approaches – is necessary, as external costs are highly site-dependent (cf. local effects of pollutants) and as marginal (and not average) costs have to be calculated" (European Commission, 2003, page 8).

Table 1 – Studies on WTP (2005 US$) for renewable energy and energy shortages

A. WTP for improving renewable energy

Study	Goett et al. (2000)	Champ and Bishop (2001)	Roe et al. (2001)	Wiser (2003)	Batley et al. (2001)	Bergmann et al. (2006)
Data year	1999	1997	1997	2001	1997	2003
Stated preference method[a]	CE	CV, SBDC	CE	CV, SBDC	CV, OE	CE
Questionnaire type	Phone–mail–phone	Mail	Intercept	Mail–phone	Mail	Mail
Completed questionnaires	1205	193	835	1574	742	219
Surveyed area	US	Madison, Wisconsin (US)	8 US cities	US	Leicester, England	8 Council Districts in Scotland
Hypothetical scenario	Increase in renewable share (25% of energy supplied by hydro)	WTP for wind energy	Increase in renewable energy of 1% and a decrease of emissions of 1%	Increase renewable energy from 2% to 8%	Increase in renewable sources	Renewable energy projects that have no increase in air pollution
Households WTP/year	98.44[b]	71.79	16.32[c]	39.72[d]	95.20	25.26

B. WTP for avoiding short-term energy shortages (black-outs)

Study	Hartman et al. (1991)	Beenstock et al. (1998)	Layton and Moeltner (2005)	Baarsma et al. (2005)	Carlsson and Martinsson (2004a)
Data year	1988	1990–1991	1998	2003–2004	2004
Stated preference method[a]	CV, OE	CR	CE	CR	CV, OE
Questionnaire type	Mail	In person	Mail	Mail	Mail
Completed questionnaires	1501	2950	1421	12,409	1678
Surveyed area	California, US	Israel	US	The Netherlands	Sweden
Hypothetical scenario	1 h shortage	1 kWh unsupplied electricity	1 h shortage	1 h shortage	1 h shortage
Households WTP/year	65.77	10.46	16.12	78.16	1.29[e]

[a] CV = contingent valuation, SBDC = single bounded dichotomous choice, OE = open ended; CE = choice experiments; CR = contingent ranking.
[b] Goett et al. (2000) report a WTP of 1.46 cUS$/kWh. We multiplied this value for the average consumption of energy per household, adjusted at year 2005 (see http://www.carbonfund.org/assumptions.php, http://minneapolisfed.org/research/data/us/calc/, http://www.xe.com/).
[c] WTP for the median respondent living in the Northeast of the US, with high school degree and no environmental organization membership.
[d] Median WTP.
[e] Carlsson and Martinsson (2004a) investigate the WTP for both planned and unplanned black-outs starting at 6 pm on an evening in January. Only unplanned black-outs are considered here.

supply (Hartman et al., 1991; Doane et al., 1988a,b; Woo et al., 1991; Beenstock et al., 1998; Goett et al., 2000; de Nooij et al., 2005; Baarsma et al., 2005; Layton and Moeltner, 2005). Most of these studies have used the contingent valuation method (Mitchell and Carson, 1989) and more recently have relied on choice experiments (Goett et al., 2000; Roe et al., 2001; Ek, 2005; Baarsma et al., 2005; Bergmann et al., 2006; Aravena et al., 2006). All these studies, despite their differences in study designs, find that consumers generally have a positive WTP for renewable energy policies.

Stated preference studies on WTP for security of energy supply have generally focused on short-term security of supply (black-outs), rather than on price volatility or long-term security of supply (e.g. dependence on Russia). See for example Doane et al. (1988a,b), Woo et al. (1991), Hartman et al. (1991), Beenstock et al. (1998), Goett et al. (2000), Carlsson and Martinsson (2004a,b), Layton and Moeltner (2005), Baarsma et al. (2005). Our study conforms to this literature.[3]

In the UK, the few studies that have been conducted show that people are becoming more supportive of renewable energy: Fouquet (1998) and Batley et al. (2001) find that 20% and 34% of respondents respectively are willing to pay more for electricity generated from renewable sources. In Scotland, Hanley and Nevin (1999) find that local populations support wind energy. Support for renewable energy in Scotland is confirmed by Bergmann et al. (2006) who find that their

[3] People may attach different values to energy shortages depending in part on whether shortages are announced or not (Beenstock et al, 1998; Carlsson and Martinsson, 2004a,b; Baarsma et al, 2005.) Our questionnaire focuses on unannounced short-term energy shortages, as in the UK the public has generally not been informed of forthcoming electricity black-outs.

respondents support renewable energy projects causing no increase in air pollution, but are not willing to pay anything for creating new long-term jobs in the energy industry.

Table 1 summarizes the characteristics and the results of the studies on WTP for renewable energy and on WTP for short-term energy security.

To our knowledge, our study is the first to investigate consumers' WTP for renewable energy that have an impact on the internalization of the external costs of energy production from three sources: reduced greenhouse gases emissions, better security of energy supply, and higher employment level in the energy sector.

3. Structure of the questionnaire and survey administration

3.1. *Selection of the attributes and conjoint choice questions*

In a choice experiments-based survey, respondents are asked to choose between hypothetical public programs or commodities described by a set of attributes (see Hanley et al., 2001); hypothetical programs or commodities differ by the level that two or more attributes take. Respondents trade-off the levels of the attributes of the programs or goods, one of which is usually its cost to the respondent, allowing researchers to infer the willingness to pay for public goods or programs and the implicit value of each attribute (see Hanley et al., 1998). In our choice experiments, the hypothetical policies for the promotion of renewable energy are described by four attributes:[4] (i) annual percentage reduction in greenhouse gases (GHG), (ii) length of shortages of energy supply, (iii) variation in the number of employed in the energy sector, and (iv) increase in the electricity bill. We focused on these four attributes because we were interested in understanding the trade-off between (i) the internalization of the external costs causing damages to human health *and* the environment, (ii) the need of electricity for day to day activities, (iii) a social element always important in political decisions, such as jobs creation/loss (iv) and finally the cost of the policy to understand the willingness to pay for renewable energy.

In a choice experiments-based survey it is essential to present a realistic and clear description of the hypothetical program or good that the respondents are asked to value. This means that the attributes chosen to describe the policy for the promotion of renewable energy presented in the choice sets and their levels have to be realistic and consistent with the government policies, as well as relevant and understandable to respondents.

In choosing the first attribute, the percentage reduction in GHG emissions, we were interested in selecting an attribute that would consider the long-term climate change impacts as well as internalize some of the associated external costs of local pollutants that cause damages to human health and the environment.[5] At first we wanted to use three separate attributes, one for human health effects of local air pollutants, one for damages to the environment, and one for climate change impacts. However, after focus groups and one-on-one interviews, we decided to use only one attribute because participants felt that the three effects, on environment, on human health, and on climate change were too closely correlated. Unfortunately this makes it difficult to separate the local pollution valuation of individuals from that of broader climate change benefits, except on the basis of the share of damages associated with each when GHG emissions are reduced by a given amount. The decision to use the annual percentage change in GHG emissions matches with the recent UK Energy White Paper (DTI, 2003) description of the potential benefits that renewable sources might bring to the internalization of the external costs: the UK has set the target to decrease GHG emissions by 60% below the 1990 levels by 2050. In order to reach this target, it needs to reduce the emissions of CO_2 by at least 15 or 25 MtC before 2020 (DTI, 2003). An increase in the share of renewable sources in the production of energy could bring a reduction of CO_2 emissions of 3–5 MtC (DTI, 2003). This means that renewable energy can contribute to cut GHG emissions by 1% per year compared to 1990 levels. Therefore, the levels chosen for this attribute in the questionnaire are: 1%, 2% or 3% reduction in GHG emissions per year in the UK.[6,7]

The second attribute presented is short-term energy security. Insecurity of energy supply, in the form of sudden physical shortages, can disrupt the economic performance and social welfare of the country in the event of supply interruptions and/or large, unexpected short-term price increases (JESS, 2003). According to the UK Energy White Paper (DTI, 2003), the UK production of oil and gas will strongly decline in the next years and the UK will become a net importer of these resources. As a consequence, the UK will be more vulnerable to price fluctuations and interruptions of supply. DTI (2005b) reports that over the year April 2004 to March 2005 the total number of customer interruptions in the UK was around 22 million. In 2001/02, UK customers suffered on average 86 min of power cuts during the year (JESS, 2003). These figures, combined with previous works on energy security (Hartman et al., 1991; Beenstock et al., 1998; Goett et al., 2000) and focus groups indications, suggested that we set the levels for energy security as follows: 30, 60, 120 min black-out per year, with the business as usual scenario being set at 90 min per year.

[4] Bateman et al. (2002) suggest that not more than 4–5 attributes, including price, should be presented in a choice experiments-based questionnaire.

[5] Local air pollution reduction associated with reductions in GHG is called an ancillary benefit. Studies for the UK and other countries show that such benefits are very policy and location specific, and vary between £2 and £334 per ton of carbon reduced, (DEFRA, 2002), and according to the OECD they could be as much as twice the climate change benefits (OECD, 2000).

[6] Even though 3% might be a well too optimistic scenario, we felt it was necessary to have such a variation among the levels of this attribute so that respondents could better appreciate the different contribution of different hypothetical policies to GHG reductions.

[7] In preparing the questionnaire we were worried whether respondents would understand the differences between GHG reductions of 1, 2, and 3%, but in our focus groups we found that people did understand these differences and deemed them feasible.

The third attribute presented in the questionnaire is the one related to employment. As Bergmann et al. (2006) claim, employment is an essential aspect about changes in the structure of society due to new energy policies. In our study, we assume that the increasing demand for renewable energy might increase the number of jobs in renewable energy sectors, but might decrease the number of jobs in the fossil fuel energy sectors. Focus groups discussions and previous studies (Bergmann et al., 2006) suggested to set the following levels for the attribute "employment": +1000 new jobs, −1000 jobs, and no new jobs in the energy sector in the UK. The values, which are comparable to actual shifts in jobs in the energy sector in the UK, were calculated assuming a hypothetical variation of about 0.5% in the total number of employees in the energy sector.[8,9]

In a choice experiment exercise, when the focus is on the marginal price of attributes and the willingness to pay for a hypothetical program or good, it is necessary to include a payment vehicle among the attributes. Following the literature (Farhar, 1999; Goett et al., 2000; Bergmann et al., 2006), we used an increase in the electricity bill as a payment mechanism for the policy to promote renewables. The levels of the electricity bill chosen are an increase by £6.5, £16, £25 and £38 on the quarterly electricity bill paid by the respondents. These correspond to an increase by 10%, 25%, 40%, and 60% from the average electricity bill in the UK.[10]

In our choice experiments we included the 'status quo' option in each choice set to compare the stated preferences of our respondents with the current situation. Such a comparison is necessary when researchers want to compute the value (WTP) of each alternative policy (Hanley et al., 2001). Table 2 summarizes the attributes and their levels for the present study.

In our conjoint choice questions, respondents are asked to indicate which they prefer between policy A and B and the status quo. To create the pairs of alternative hypothetical policies, we first created the full factorial design, i.e., all of the possible combinations of attribute levels. This gave a total of 108 possible combinations of hypothetical policies. To reduce the number of possible combinations, we opted for a fractional factorial design (Louviere et al., 2000). We then randomly selected two of these alternatives, but discarded pairs containing dominated or identical alternatives.[11] At the end we prepared six different versions of the questionnaire with six choice experiments each. An example of choice experiment question is shown on Fig. 1.

[8] In July 2005 British Gas announced a cut of 2000 jobs as a consequence of a modernization plan of the company (see http://news.bbc.co.uk/2/hi/uk_news/england/4730487.stm accessed October 31st 2007). When we presented the attributes, respondents were made aware that the hypothetical policies might have similar effects.

[9] According to the Office for National Statistics (2005), the total number of employees in the Energy and Water Industry Sector in the UK during the second quarter of 2005 was 177,000.

[10] The average annual electricity bill in the UK according to the National Statistics is equal to £251 (DTI, 2005a; Table 2.2.2). The electricity consumption in 2003 was equal to 337.443 billion kWh (IEA, 2003).

[11] A dominated alternative is one that should obviously be less preferred to the other. For instance, if two projects are identical in every respect except for the price, the project with the higher cost is dominated by the other.

Table 2 – Attributes and their levels for the choice experiments

Attribute	Level 1	Level 2	Level 3	Level 4	Status quo
Annual reduction in GHG emissions due to renewable energy increase (3 levels)	1%	2%	3%	–	No additional greenhouse gases emissions reduction
Annual length of electricity shortages in minutes (3 levels)	30	60	120	–	Current level of black-outs
Change in number of employees in the electricity sector (3 levels)	+1000	−1000	0	–	No employment change in the energy sector
Increase in electricity bill in £ (4 levels)	6.5	16	25	38	No price increase in the electricity bill

3.2. *Structure of the questionnaire and survey administration*

The questionnaire starts by presenting the topic of the survey: people's opinions on hypothetical renewable energy policies. Respondents face a few warm-up questions aimed at investigating the level of knowledge of respondents on the level of externalities caused by different energy fuels. The second part prepares the respondents with the hypothetical policies: one page describes the four attributes that define the possible impacts of a policy for the promotion of renewables. Respondents are asked to focus only on the four attributes we consider and not to think of other elements that might characterize the impacts of a policy for renewable energy. The next section is the central part of the questionnaire with the six choice experiments questions. The fourth section presents some debriefing questions to verify whether the respondents considered all the attributes in their choices, or only one. The fifth part of the questionnaire collects the usual socio-demographic characteristics. At the end, the interviewers took note whether respondents seemed annoyed by the interview or seemed to not understand the choice exercises.[12]

Following Roe et al. (2001), the survey was administered in person to 300 respondents intercepted in shopping areas, public parks and other central areas of Bath, England, in July and August 2005. The surveys were carried out by professional interviewers who were instructed to interview an even number of men and women and to ensure given proportions of respondents in various age groups. To mitigate possible biases in the sample, interviewers were instructed to follow the common practice of stopping every 7th person passing by.[13] We chose to interview people through in-person interviews to

[12] A copy of the questionnaire is available by the authors upon request.

[13] Intercept surveys can be collected in convenient locations, such as shopping malls or street (Champ, 2003, p. 70), or at visitation sites. The sampling strategy we implemented follows the intercept survey described by Davis (2004).

Characteristics	Policy A	Policy B	Neither
Greenhouse Gases emissions	3% reduction per year	1% reduction per year	no additional greenhouse gases emissions reduction
Black-outs	30 min per year	60 min per year	current level of black-outs
Employment	no employment change in the energy sector	-1,000 jobs	no employment change in the energy sector
Increase in electricity bill	£25 per quarter	£6.5 per quarter	no price increase in the electricity bill
Which policy would you choose?	☐	☐	☐

Fig. 1 – Example of choice experiments question.

guarantee a high quality in the answers. The budget constraint of this study limited our analysis to sample residents of Bath and North East Somerset. The results presented in this study should therefore be interpreted with caution: they are not representative of the UK population, but of the residents of a quite wealthy medium sized town of the South of the UK.

4. Economic model and econometric model

4.1. *The willingness to pay for a policy for renewable energy*

Our statistical analysis of the responses to the choice questions is based on the random utility model (RUM), which assumes that a respondent's indirect utility is broken down into two components. The first component is deterministic, and is a function of the attributes of the alternatives, the characteristics of the individuals, and a set of unknown parameters, while the second component is an error term (McFadden, 1974). Formally,

$$V_{ik} = \overline{V}(\mathbf{x}_{ik}, \mathbf{b}) + \varepsilon_{ik} \tag{1}$$

where the subscript i denotes the respondent, the subscript k denotes the alternative, $\mathbf{x}$ is a 1×4 vector comprised of the four attributes: the annual GHG emissions reductions (GHGREDUCTION), the yearly minutes of energy shortages (BLACKOUT), the permanent impact on the energy job market (JOBS), and the increase in the quarterly electricity bill (PRICE), $\boldsymbol{\beta}$ is a vector of unknown coefficients and ε is an error term that captures individual- and alternative-specific factors that influence utility, but are not observable to the researcher. We further assume that the indirect utility function is linear in parameters:

$$V_{ik} = \beta_{0k} + \mathbf{x}_{ik}\boldsymbol{\beta} + \varepsilon_{ik} \tag{2}$$

We further posit that in each of the choice questions the respondent selects the alternative with the highest indirect utility:

$$\pi_{ik} = \Pr(V_{ik} > V_{i1}, V_{ik} > V_{i2}, \ldots, V_{ik} > V_{iK}) = \Pr(V_{ik} > V_{ij}) \quad \forall j \neq k \tag{3}$$

where π_{ik} signifies the probability that option k is chosen by individual i. If the error terms ε are independent and identically distributed and follow the type I extreme value distribution, the probability that the hypothetical policy k is selected out of K policies is:

$$\Pi_{ik} = \Pr(\text{resp. } i \text{ chooses } k) = \exp(\mu V_{ik}) / \sum_{j=1}^{K} \exp(\mu V_{ij}) \tag{4}$$

where μ is the scale parameter which is inversely proportional to the standard deviation of the error terms. Eq. (4) is the contribution to the likelihood in a conditional logit model. In our questionnaire, $K=3$. The log likelihood function of the conditional logit model is:

$$\ln L = \sum_{i=1}^{n} \sum_{k=1}^{K} y_{ik} \cdot \ln \pi_{ik} \tag{5}$$

where y_{ik} takes on a value of 1 if the respondent chooses k, 0 otherwise. The coefficients are estimated using a Maximum Likelihood Estimation Method. The model described by (4) and (5) allows us to estimate the trade-off between any two attributes and the willingness to pay for different policies. The marginal price of attribute k is given by:

$$\text{MP}_k = -\frac{\hat{\beta}_k}{\hat{\beta}_c} \tag{6}$$

where $\hat{\beta}_k$ is the utility from an extra unit of k. Divided by the price coefficient, $\hat{\beta}_c$, it gives us the monetary value of the utility coming from an extra unit of k. Finally we can derive the willingness to pay for a certain policy, formally:

$$\text{WTP}_{ik} = -\frac{x_{ik}\hat{\beta}}{\hat{\beta}_c} \tag{7}$$

Where x is the vector of the levels of attributes of policy k given to individual i.

We use model (2) to test the findings by Bergmann et al. (2006) that, in choosing a policy for the promotion of renewable energy, the impact on the job market does not matter (Hypothesis I). An insignificant sign of the coefficient of (JOBS) would fail to reject Hypothesis I.

The second hypothesis (Hypothesis II) of our model is that respondents value the externalities on human health and the environment more than those arising from energy disruptions. This hypothesis would mirror the actual interest that the ExternE team has had in valuing the externalities caused by energy.[14] To test this hypothesis we compare how much our respondents are willing to pay to avoid energy shortages, considering the average length of energy shortages of 90 min per year, with how much they are willing to pay to decrease

[14] Researchers of the ExternE team have only recently moved their attention also to energy shortages (see the European Commission FP6 funded projects NEEDS and CASES).

GHG emissions from promoting renewables to comply with the DTI (2003) targets of reducing emissions by 60% below the levels of 1990 by 2050, which is roughly given by a GHG reduction of 1% a year.

4.2. *Heterogeneity among respondents and specific hypothesis*

The conditional logit model described by Eqs. (4)–(5) is easily amended to allow for heterogeneity among the respondents. Specifically, one can form interaction terms between individual characteristics, such as age, gender, education, etc., and all or some of the attributes, and enter these interactions in the indirect utility function to test other specific hypotheses.[15]

Our Hypothesis III aims to investigate the internal validity of our responses. Therefore we add interaction terms between respondents' income (INCOME) and GHGREDUCTION and between INCOME and BLACKOUT.

In the literature on non-market valuation, researchers usually try to disentangle the components of the good being estimated into its use and non-use value components (see Freeman, 2003). For our good it is quite difficult to identify the use value component of a policy for the promotion of renewable energy. Such values arise from the direct benefits that respondents receive from the policy, such as the improvement in their own health status, the conservation of the natural environment that they visit, and especially the reduction in energy shortages in their own dwellings. Non-use components are made up of benefits that will emerge in the long run. For example, future generations will be more likely to experience the benefits in terms of health and global warming of the cuts in GHG emissions. If the share of ancillary benefits from GHG reduction is of the order of 10% (OECD, 2000) then we can say, roughly, that 10% of the willingness to pay for renewables is for personal benefits (although even here others also gain from the reduction) and 90% is for the longer term benefits of future generations. Our Hypothesis IV is therefore that respondents who have children are more willing to accept the policy for the promotion of renewables and are willing to pay more than those without children. To test this hypothesis, we add an interaction term between GHGREDUCTION and a dummy variable (CHILD) that takes on a value of 1 if a respondent has children, and 0 otherwise. To further test whether respondents that care for future generations have a higher willingness to pay for renewables, we add an interaction term between BLACKOUT and CHILD.

Finally, we wish to see whether the level of schooling or the membership in an environmental organization influences the WTP for renewables. While being member of an environmental organization should suggest a positive WTP for environmental programs, previous research on the effects of education on WTP has found mixed results. For example, Blomquist and Whitehead (1998), Witzke and Urfei (2001), Li et al. (2004) have found that the level of education has a positive influence on environmental willingness to pay; Danielson et al. (1995), Krupnick et al. (2002), Bergmann et al. (2006) have found a negative effect on WTP; finally Berrens et al. (2004), Veisten et al. (2004), Popp (2001) have found that the level of education has no significant effect on WTP. Therefore we want to investigate how both characteristics, being a member of an environmental organization (ENV_MEMBER) and having a college degree (COLLEGE), affect WTP. Our Hypothesis V is that the coefficients of the interaction terms of ENV_MEMBER*GHGREDUCTION and COLLEGE*GHGREDUCTION to be positive and significant. Table 3 summarizes the hypotheses.

Table 3 – Hypotheses tested with our model

Hypothesis	Description
I	In a policy for the promotion of renewable energy, the number of jobs created or lost does not matter.
II	It is more important to internalize external costs affecting human health and the environment than guaranteeing energy security.
III	Test the internal validity of the responses: WTP increases with income.
IV	WTP is higher for respondents with children.
V	Members of environmental organizations and having a college degree positively affect the WTP for a policy for renewable energy

To further allow for variations in taste among individuals and to relax the IIA[16] hypothesis implicit in the conditional logit model, we also estimate a more complex variant of model (4), which allows for the coefficients β to be random variables and to vary over the population with density $f(\beta)$. In the random-parameter logit model (Train, 2003), the utility function of Eq. (2) is augmented by a vector of parameters $\boldsymbol{\theta}$ that takes into account of individual's preference deviations with respect to the mean preference values expressed by the vector $\boldsymbol{\beta}$:

$$V_{ik} = \beta_0 + \mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{x}_{ik}\boldsymbol{\theta} + \varepsilon_{ik} \tag{8}$$

where $\boldsymbol{\theta}$ is a vector of deviations from the mean β parameters estimated. Clearly, estimation of the likelihood function based on (8) requires that assumptions be made about which coefficients are random, and about the joint distribution of these coefficients.

5. Results

5.1. *Description of the data*

Table 4 reports descriptive statistics for our sample and compares them with those for the population of Bath and North East Somerset, showing that the socio-demographics of our sample are for the most part very similar to those of the population of Bath and North East Somerset. Our sample tends to be slightly richer and younger than the population of Bath and North East Somerset.

[15] Since respondents' characteristics do not vary across alternative hypothetical policies, socio-demographic characteristics must be introduced as interaction terms with the attributes or the alternative specific constants.

[16] The independence of irrelevant alternatives (IIA) states that the relative probability of choosing between any two alternatives is independent of all other alternatives (Haab and McConnell, 2002).

Table 4 – Descriptive statistics

Variable	Observations	Sample average or percent (Standard deviation)	Bath and North East Somerset
Individual characteristics			
Age (in years)	300	35.75 (12.52)	38.4 [a]
Annual income (in £)	299	37,687.29 (26,528.63)	31,000 [b]
Male	300	51.3%	48% [a]
Has a college degree	299	22.6%	25.90% [a]
Has children	300	25.6%	
Member of environmental organizations	296	22.3%	
Electricity bill (in 2005 £)	197	70.86 (38.78)	
Electric heating	300	30.3%	
Choice experiments			
Ranking of the attributes			
GHG reduction ranked as 1st	300	68.3%	
Number of jobs created/lost ranked as 1st	300	16.7%	
Energy shortages ranked as 1st	300	6.3%	
Electricity bill increase ranked as 1st	300	8.7%	
Found the choice experiments difficult (1 = very difficult; 5 = very easy)	300	4.16 (0.88)	
Considered all attributes in the choice questions	300	69.7%	
Attribute mostly considered…			
GHG reductions	300	21.7%	
Number of jobs created/lost	300	4.7%	
Energy shortages	300	1.7%	
Electricity bill increase	300	2.3%	
Interviewer debriefing questions			
Understood the choice questions	300	95.7%	
Annoyed by the questionnaire (1 = very annoyed; 5 = not annoyed at all)	300	4.47 (0.68)	

[a] Source: National Neighbourhood Statistics — http://neighbourhood.statistics.gov.uk/dissemination/home.do?$ph=60.
[b] Gross annul household income in the UK. Source: HMRC CACI Paycheck Model 2005.

Our average respondent is 35 years old, has an annual gross household income of about £37,000, and pays £70 per quarter on electricity bill. The sample is well balanced in terms of gender, with about one quarter of our respondents having one or more children. About 22% of our respondents are members of an environmental organization, and almost 31% have electric heating.

The first task when analysing the data was to look at the initial set of questions (warm-up questions) where respondents

Table 5 – Are the following electricity sources environmentally friendly?

	Our survey			European Commission (2003)
	Yes	No	Don't know	
Biomass	38.00%	36.33%	24.67%	"There are dozens of different biomass technologies, and depending on the care given on gas cleaning technologies, the biomass options can range from low to high external costs."
Nuclear	20.33%	70.33%	9.33%	"Nuclear power in general generates low external costs, although the very low probability of accidents with very high consequences and the fuel cycle impacts are included. It is also a technology with very low greenhouse gas emissions."
Gas	31.00%	52.00%	17.00%	"Gas-fired technologies are quite clean, with respect to classical pollutants, but their impact on climate change depends strongly on the efficiency of the technology."
Hydro	93.67%	3.00%	3.33%	Hydropower exhibits low external costs of all systems, but they may increase on sites were higher direct emission of GHG from the surface of reservoir occurs (ExternE-Pol, 2005).
Oil	3.33%	90.33%	6.33%	Oil has high external costs due to air pollution with impacts on global warming and human health. Introduction of advanced technology (Combined Cycle) substantially reduces the external costs of fossil systems (ExternE-Pol, 2005).
Solar	99.00%	0.67%	0.33%	"Photovoltaics is a very clean technology at the use stage, but has considerable life cycle impacts."
Wind	96.33%	3.33%	0.33%	"Wind technologies are very environmentally friendly with respect to emissions of "classical" pollutants (SO_2, NO_X, dust particles) and with respect to greenhouse gas emissions."
Coal	3.67%	92.67%	3.67%	"Coal technologies carry the burden of their very high CO_2 emissions, even for new, more efficient technologies, and in addition cause quite high impacts due to the primary-secondary aerosols."

were asked to state whether the different electricity fuels presented were environmentally friendly or not. The results are reported in Table 5. Eight different sources were presented. The results can be viewed in the light of the broad qualitative conclusions of ExternE (European Commission, 2003). The two sources with more uncertain answers are biomass and gas, probably due to a lack of knowledge of the sources itself, especially for biomass. Despite the quite positive consideration of nuclear power by ExternE, our respondents consider this source of energy highly hazardous for human health and the environment. The explanation we received the most was related to the risk of accidents and the problems with the nuclear waste.

The perceptions regarding the other sources, namely oil, natural gas and wind power are confirmed by the external costs estimated by ExternE (European Commission, 2003). Hydro, solar and wind are generally considered environmentally friendly by our respondents, while oil and coal are deemed dangerous to human health and the environment by more than 90% of the respondents.

Before running our econometric models, the quality of the responses was checked. In a debriefing question, most respondents considered the choice experiments as easy: on a 1 to 5 Likert scale, where 1 means very difficult and 5 very easy, the average value given by respondents is 4.16. To further check the quality of the responses, at the end of the questionnaire interviewers noted whether they thought that respondents understood the choice exercise or were annoyed during the interview. Only a few respondents seemed annoyed by the questionnaire, and only 13 respondents did not understand the choice experiments. We also check the percentages of respondents who always choose the alternative displayed on the left-hand side of the card (alternative A hereafter), or the alternative displayed on the right-hand side of the card (alternative B hereafter), which may signal the presence of abnormal response patterns (Viscusi et al., 1991). Only 1 respondent selected alternative A for all of the six choice questions included in the questionnaire, and no one selected always either alternative B or the status quo for all of the six choice questions. These preliminary observations suggest that the choice tasks were not prohibitive and were accepted by our respondents.

Table 6 – Conditional logit model estimates

	Model 1		Model 2	
	Coeff.	t-stat	Coeff.	t-stat
A_Alt.1	2.0498	12.90	2.1290	13.12
A_Alt.2	1.9786	12.77	2.0488	12.95
GHGREDUCTION	0.6804	7.29	0.4437	3.78
BLACKOUT	−0.0088	−7.45	−0.0055	−3.09
JOBS	0.0006	6.69	0.0007	4.98
PRICE	−0.0224	−3.34	−0.0244	−3.58
AGE*JOBS			−0.0046[a]	−1.33
INCOME*BLACKOUT			−0.0006[b]	−1.70
INCOME*GHGREDUCTION			0.0453[b]	2.84
ENV_MEMBER* GHGREDUCTION			0.4872	4.83
COLLEGE*GHGREDUCTION			−0.3755	−4.02
CHILD*GHGREDUCTION			0.3380	3.30
CHILD*BLACKOUT			−0.0050	−2.38
Loglikelihood	−1358.72		−1297.47	
Observations	1722		1692	

[a] The coefficient of (AGE*JOBS) has been multiplied by 1000.
[b] The coefficients of (INCOME*BLACKOUT) and (INCOME*GHGREDUCTION) have been multiplied by 10,000.

Table 7 – Implicit prices in British pounds (standard error in parenthesis calculated with the delta method)

	Model 2[a]
GHGREDUCTION	29.65*** (5.50)
BLACKOUT	−0.36*** (0.08)
JOBS	0.02*** (0.00)

***Significant at the 1% level.
[a] Calculated at the mean values of the socio-demographic characteristics of the respondents.

5.2. *Results from the discrete choice models*

In this section, we report the results of the econometric models estimated by dropping the observations of the 13 respondents who did not understand the choice exercises. We began with random-coefficient models, but found no evidence that coefficients are random, and subsequently ran only conditional logit estimators;[17] here we present the results from the latter. The first specification of the model uses only the four attributes as independent variables and the alternative specific constants to take into account of the status quo effect (see Holmes and Adamowicz, 2003). Model 1 of Table 6 shows that all coefficients are significant at the 1% level and have the correct sign. The positive sign in GHGREDUCTION and JOBS implies that our respondents are more likely to favour a policy that reduces the emissions of GHG and supports the creation of new jobs. Model 1 allows us to reject Hypothesis I: contrary to the findings by Bergmann et al. (2006), our respondents are not indifferent to a policy for the promotion of renewables that affects the number of jobs in the energy market. The negative sign of the BLACKOUT coefficient means that our respondents shy away from policies that have longer electricity shortages. Also the negative sign on PRICE suggests that our respondents do not like a policy that entails higher energy prices, with all other characteristics of the policy remaining constant. The positive sign of the alternative specific constants suggests that our respondents do prefer a new policy for the promotion of renewable energy in comparison with the status quo. A Wald test of equality of coefficients for the two alternative specific constants does not reject the null hypothesis of equality of coefficients (Chi-squared = 1.79) and confirms that our respondents did not systematically prefer alternative A over alternative B, or vice versa.

Model 2 of Table 6 controls for socio-demographic characteristics of the respondents by adding interaction terms for

[17] We experimented with log-normal distributions for all the coefficients. Since the coefficients on price and on blackout should be negative, we specified a lognormal distribution for the negative of this coefficient (Train, 2003). In all cases the standard deviation of the coefficient was very small relative to the mean of the coefficient, was insignificant, and the model reduced to a conditional logit.

Table 8 – Willingness to pay for selected hypothetical policies in British Pounds

	Policy A	Policy B	Policy C	Policy D	Policy E
GHGREDUCTION	0.5%	No improvement	0.5%	No improvement	1%
BLACKOUT [a]	45 min	60 min	30 min	0 min	90 min
JOBS	const	+1000	−1000	const	const
WTP	32.19***	34.78***	13.21***	33.16***	29.65***
(Standard error)	(6.29)	(6.47)	(4.00)	(7.15)	(5.50)

[a] The business as usual scenario is 90 min of black-out. Policies A, B, C and D which provide black-outs shorter than 90 min offer improvements compared to the status quo.

age, income, level of education, membership in environmental organizations, and whether the respondent has children or not. A Likelihood ratio test shows that Model 2 outperforms Model 1 at the conventional levels (LR test=76.244). When considering the level of income, we find that respondents with higher income are willing to pay more for the reduction in GHG emissions, as well as for decreasing the shortages of energy. These considerations provide us with reasons to not reject Hypothesis III: our model is internally valid, with WTP increasing with income. AGE is negatively associated with the number of jobs created in the energy market, but is not significant. Model 2 also provides evidence in support of Hypothesis IV: respondents with children have a higher WTP for a policy that stimulates the promotion of renewable energy. The interaction term between CHILD and GHGREDUCTION shows that respondents with children are more responsive to a policy that internalizes a higher percentage of GHG emissions; and the interaction between CHILD and BLACKOUT suggests that this group of respondents is willing to pay more than respondents without children for a policy that minimizes the minutes of energy shortages. These results suggest that our respondents do recognize the importance of a policy for the promotion for renewable energy and are willing to pay for the benefits that such a policy will entail also to future generations.

Model 2 provides little support in favour of Hypothesis V: similar to previous results by Danielson et al. (1995), we find that having a college degree negatively affects the probability of choosing a policy that internalizes a higher percentage of GHG, while members of environmental organizations are more likely to select a policy that internalizes a higher percentage of GHG.

Finally, to evaluate Hypothesis II, we need to look at the marginal prices of the attributes, as we do in the next section.

5.3. *Marginal prices and willingness to pay*

Table 7 reports the implicit prices of the attributes used in the choice experiments estimated from Model 2 and calculated at the mean values of the socio-demographic characteristics of the respondents. This table shows that respondents are on average willing to pay in addition to their electricity bill: (i) £29.65 (s.e. 5.50) to decrease the GHG emissions by 1% a year; (ii) £0.36 (s.e. 0.08) to decrease the shortages of energy by 1 min a year; (iii) £0.02 (s.e. 0.00) to increase the permanent number of jobs in the energy sector by 1.

Model 2 can also be used to assess the marginal prices for different groups of respondents, according to their socio-demographic characteristics. For example, respondents with a college degree, with children and a membership in an environmental organization are willing to pay £45.54 (s.e. 10.49) in addition to their electricity bill for a policy that decreases the GHG emissions by 1%, while respondents with a college degree, no children and no membership in any environmental organization are willing to pay £9.77 (s.e. 3.96) for the same policy. Model 2 can also be used to study the impact of having children in valuing energy shortages: a respondent with children is willing to pay £0.52 (s.e. 0.13) in addition to his electricity bill for a policy that decreases energy shortages by 1 min per year, while a respondent without children is willing to pay only £0.31 (s.e. 0.07) for the same policy.

Results from Model 2 can also be used to estimate the WTP for the effects of specific policies for the promotion of renewable energy. Table 8 reports the WTP for five different policies characterized by different effects on the reduction of GHG emissions, black-outs and employment in the energy sector. For example, our respondents are on average willing to pay about £32 for a hypothetical policy (A) that reduces GHG

Table 9 – Implied society's WTP for reducing emissions by 1 ton of CO_2 per year in 2005 (all the prices are in 2005 US$)

Study	Total WTP/ year [a] (in 2005 US$ million)	Mtons of CO_2 emissions per year [b]	MTons of CO_2 reduction/ year [c]	Implied WTP per ton CO_2 per year per country
Roe et al. (2001)	$870	4778	47.78	$39
Wiser (2003)	$4552	5075	50.75	$89
Goett et al. (2000)	$1,283	4972	49.72	$227
Batley et al. (2001)	$2475	549	5.49	$451
This study	$5368	555	5.55	$967

[a] Total WTP was calculated as (Households' WTP per year)/(number of persons in a household)*(population of the country). For this study we used the results for a reduction of 1% of GHG reported in Model 2. For the other studies the scenarios considered are those reported in Table 1.

[b] Mtons of CO_2 emissions including net CO_2 from land use, land use change and forestry in year of the survey. Source: United Nations Framework Convention on Climate Change, GHG emission profiles for Annex I Parties, available at http://unfccc.int/ghg_emissions_data/items/3954.php.

[c] We assume that each programme provides a reduction of 1% of greenhouse gases emissions.

emissions by 0.5% a year, limits energy shortages to 45 min per year and maintains the current level of employment in the energy sector.

A comparison of Policy D and Policy E investigates how our respondents consider the internalization of the external costs affecting human health and the environment compared to guaranteeing energy security (Hypothesis II). Policy D offers a scenario with no black-outs, while keeping the current level of GHG emissions. Policy E provides a reduction of GHG emissions equal to 1% a year, and keeps the current level of energy security. Both policies maintain the number of jobs in the energy sector constant. The results show that our respondents are willing to pay about £33 for the policy that guarantees energy security, and about £29 for a policy that decreases GHG emissions by 1% per year, a target consistent with the DTI (2003) goals of reducing emissions by 60% below the levels of 1990 by 2050. This result suggests that our respondents consider energy security as an important externality and it supports the recent interest of the European Commission and of national governments to improve energy security. Respondents are willing to pay about £3.5 more for Policy D than for Policy E, suggesting that they are slightly more interested in the current benefits of improving energy security than the long-term benefits of limiting climate change.

6. Discussion and conclusions

It is of considerable interest to policy makers to know how much more individuals are willing to pay for renewable energy than for fossil fuel energy. A number of studies in the UK and US have tried to elicit the additional value of renewable energy and have come up with figures ranging from \$16 a year to as much as \$98 (2005US\$) (see Table 1). Translating this into reductions in CO_2 is an approximate exercise and comes up with estimates of \$39 to \$451 (2005US\$) per year per country per ton. Details are given in Table 9. In our study we find a willingness to pay equal to \$967 for a ton of CO_2. This value represents how much *society* in the UK as a whole is willing to pay every year for reducing carbon emissions by one ton of CO_2. These payments have to be seen as a payment for a public good if individuals make the 'Cournot Nash' assumption that only they are making the payment. In that case the additional benefits at the personal level are insignificant and the WTP is a gesture of social goodwill. This assumption may, however, be suspect. Perhaps individuals are assuming that the programs of shifting to renewable energy apply to society as a whole, in which case there could be important local pollution reduction benefits.

In our study we explicitly assumed that the reduction being paid for was at the national level. Hence there is a public good benefit as well as some personal health benefit. If emissions of GHGs are reduced by 1%, so will associated particles and other health related pollutants and the individual will benefit. Of course, even here, as in the case of the previous studies, there is the potential for free riding — to state a zero or very low WTP because a large part of the benefits go to others. Notwithstanding this possibility, the figures of WTP in both sets of studies appear to be significant. Hence the extent of free riding behaviour appears to be limited.

How can the results of Table 9 be reconciled? There are a number of possible explanations. First, our study took place about more than four years after the latest of the earlier studies. In that time awareness of the climate problem has grown and WTP may have risen considerably. Second, our estimates come from a sample of residents of Bath, a quite wealthy area in the UK and may overstate the WTP of the UK population. Third, the higher willingness to pay for abating emissions in the UK compared to the results from the US studies might further be explained by the different preferences of the two societies for renewable energy programs and for the reduction of GHG emissions, a result that mirrors the positions of the two national governments in climate change negotiations.

Finally, it is interesting to see that our estimates for the value of one ton of CO_2 abatement are considerably higher than the values found in studies that employ the *damage cost* method. Most of the results in the damage cost literature are in the range of 5 to 125 US\$/tCO_2 (Pearce et al., 1996), but these figures are subject to high uncertainty (Tol, 2005). For the UK, in 2002, the Government Economic Service recommended an illustrative estimate for the social costs of carbon of £70/tCO_2, within a range of £35 to £140/tCO_2, for use in policy appraisal across Government. A recent review by DEFRA (2004) suggests to update the estimates range to £12–£260/tCO_2 for emissions abated in 2010, but also states that the current modelling reveals that estimates of the social cost of CO_2 span at least three orders of magnitude, from 0 to over 1000 £/tCO_2, reflecting uncertainties in climate and impacts, coverage of sectors and extremes, and choices of decision variables. Our results, and the results presented on Table 9, indicate that studies that employ the WTP methodology and use a demand curve approach, find estimates for the values of CO_2 emissions much higher than those based on the damage cost method, suggesting that the benefits to society are substantial.

Other major results of interest from our study are the following: (i) despite previous results (Bergmann et al., 2006), people are ready to pay little extra in order to increase renewable energy through policies that increase employment; (ii) the WTP to avoid black-outs is in the range of £22 per hour (£0.37 per minute), is comparable to previous studies (see Table 1), and shows that current governments support for improving energy security is justified by people's preferences.

REFERENCES

Alberini, A., Longo, A., Veronesi, M., 2007. Basic statistical models for conjoint choice experiments. In: Kanninen, B. (Ed.), Valuing Environmental Amenities using Choice Experiments: A Common Sense Guide to Theory and Practice. Springer, Dordrecht, The Netherlands, pp. 203–228.

Alvarez-Farizo, B., Hanley, N., 2002. Using conjoint analysis to quantify public preferences over the environmental impacts of wind farms. An example from Spain. Energy Policy 30 (2), 107–116.

Aravena, C., Martinsson, P., Scarpa, R., 2006. Does money talk? The effect of a monetary attribute on the marginal rate of substitution in a choice experiment. Paper presented at the 3rd World Congress of Environmental and Resource Economists, Kyoto, Japan, 3–7 July, 2006.

Baarsma, B.E., Berkhout, P.H.G., Hop, J.P., 2005. Valuation of the quality of the electricity grid — power outages have a price too. SEO Discussion Paper 41, University of Amsterdam, Amsterdam.
Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D.W., Sugden, R., Swanson, J., 2002. Economic Valuation with Stated Preference Techniques: A Manual. Edward Elgar, Ltd., Cheltenham. 458 pp.
Batley, S.L., Colbourne, D., Fleming, P.D., Urwin, P., 2001. Citizen versus consumer: challenges in the UK green power market. Energy Policy 29 (6), 479–487.
Beenstock, M., Goldin, E., Haitovsky, Y., 1998. Response bias in a conjoint analysis of power outages. Energy Economics 20 (2), 135–156.
Bergmann, A., Hanley, N., Wright, R., 2006. Valuing the attributes of renewable energy investments. Energy Policy 34 (9), 1004–1014.
Berrens, R., Bohara, A., Jenkins-Smith, H., Silva, C., Weimer, D., 2004. Information and effort in contingent valuation surveys: application to global climate change using national internet samples. Journal of Environmental Economics and Management 47 (2), 331–363.
Blomquist, G.C., Whitehead, J.C., 1998. Resource quality information and validity of willingness to pay in contingent valuation. Resource and Energy Economics 20, 179–196.
Carlsson, F., Martinsson, P., 2004a. Willingness to pay among Swedish households to avoid power outages — a random parameter Tobit model approach. Working Papers in Economics no. 154, Department of Economics, Gothenburg University, Gothenburg.
Carlsson, F., Martinsson, P., 2004b. Does it matter when a power outage occurs? — a choice experiment study on the willingness to pay to avoid power outages. Working Papers in Economics no. 155, Department of Economics, Gothenburg University, Gothenburg.
Champ, P.A., 2003. Collecting survey data for nonmarket valuation. In: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecth, The Netherlands, pp. 59–98.
Champ, P.A., Bishop, R.C., 2001. Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. Environmental and Resource Economics 19 (4), 383–402.
Danielson, L., Hoban, T.J., Van Houtven, G., Whitehead, J.C., 1995. Measuring the benefits of local public goods: environmental quality in Gaston County, North Carolina. Applied Economics 27, 1253–1260.
Davis, J., 2004. Assessing community preferences for development projects: are willingness-to-pay studies robust to mode effects? World Development 32 (4), 655–672.
de Nooij, M., Bijvoet, C.C., Koopmans, C.C. 2005. The value of supply security: the costs of power interruptions: economic input for damage reduction and investment in networks. SEO Discussion Paper 35, University of Amsterdam, Amsterdam.
DEFRA, 2002. Ancillary Effects of Green House Gas Mitigation Policies. Department of Environment Food and Rural Affairs, London. October.
DEFRA, 2004. The Social Costs of Carbon (SCC) Review — Methodological Approaches for Using SCC Estimates in Policy Assessment. Department of Environment Food and Rural Affairs, London.
Doane, M.J., Hartman, R.S., Woo, C.K., 1988a. Households' perceived value of service reliability: an analysis of contingent valuation data. The Energy Journal 9, 135–150.
Doane, M.J., Hartman, R.S., Woo, C.K., 1988b. Household preference for interruptible rate options and the revealed value of service reliability. The Energy Journal 9, 121–134.
DTI, 2003. Energy White Paper — Our Energy Future – Creating a Low Carbon Economy. Department of Trade and Industry, London.
DTI, 2005a. Quarterly Energy Prices. National Statistics Publication, Department of Trade and Industry, London.
DTI, 2005b. Secretary of State's First Report to Parliament on Security of Gas and Electricity Supply in Great Britain. Department of Trade and Industry, London.
Ek, K., 2005. Public and private attitudes towards "green" electricity: the case of Swedish wind power. Energy Policy 33 (13), 1677–1689.
European Commission, 1995. White paper of energy policy. Com (95), 682 Final, Brussels.
European Commission, 1997. Energy for the future: Renewable Source of Energy, White Paper for a Community Strategy and Action Plan, Com(97) final (26/11/1997), Brussels.
European Commission, 1998. Externalities of fuel cycles 'ExternE' Project. National Implementation Programme. DGXII, Science, Research and Development, JOULE.
European Commission, 1999. ExternE externalities of energy. Methodology Update. Office of Publications for the European Communities, EC – DG XII, Brussels, vol 7.
European Commission, 2000. Green paper: towards a European strategy for the security of energy supply. Com (769) Final, Brussels.
European Commission, 2001. Community guidelines on state aid for environmental protection. Official Journal C 037, 0003–0015 Brussels.
European Commission, 2003. External costs: research results on socio-environmental damages due to electricity and transport. Directorate-General for Research, Directorate J-Energy, Brussels.
European Commission, 2005. ExternE — externalities of energy: methodology 2005 update. Directorate-General for Research, Sustainable Energy Systems, Brussels.
ExternE-Pol, 2005. Externalities of energy: extension of accounting framework and policy applications. Final Technical Report, August.
Farhar, B.C., 1999. Willingness to pay for electricity from renewable resources: a review of utility market research. NREL/TP-550-26148, National Renewable Energy Laboratory, Golden, CO.
Farhar, B.C., Houston, A., 1996. Willingness to pay for electricity from renewable energy. NREL/TP-460-21216, National Renewable Energy Laboratory, Golden, CO.
Farhar, B.C., Coburn, T., 1999. Colorado homeowner preferences on energy and environmental policy. NREL/TP-550-25285, National Renewable Energy Laboratory, Golden, CO.
Fouquet, R., 1998. The United Kingdom demand for renewable electricity in a liberalized market. Energy Policy 26 (4), 281–293.
Freeman, M.A., 2003. The Measurement of Environmental and Resource Values Theory and Methods, Second Edition. Resources for the Future Press, Washington, DC.
Friedrich, R., Bickel, P. (Eds.), 2001. Environmental External Costs of Transport. Springer-Verlag, New York.
Goett, A.A., Hudson, K., Train, K.E., 2000. Customers' choice among retail energy suppliers: the willingness-to-pay for service attributes. Energy Journal 21 (4), 1–28.
Haab, T.C., McConnell, K.E., 2002. Valuing environmental and natural resources. The Econometrics of Non-market Valuation. New Horizons in Environmental Economics. Edward Elgar, Cheltenham.
Hanley, N., Nevin, C., 1999. Appraising renewable energy developments in remote communities: the case of the North Assynt Estate, Scotland. Energy Policy 27 (9), 527–547.
Hanley, N., Wright, R.E., Adamowicz, W., 1998. Using choice experiments to value the environment: design issues, current experience and future prospects. Environmental and Resource Economics 11, 413–428.
Hanley, N., Mourato, S., Wright, R.E., 2001. Choice modelling approaches: a superior alternative for environmental valuation? Journal of Economic Surveys 15 (3), 435–462.

Hartman, R.S., Doane, M.J., Woo, C.K., 1991. Consumer rationality and the status quo. Quarterly Journal of Business and Economics 106, 141–162.
Holmes, T.P., Adamowicz, W., 2003. Attribute-based methods. In: Champ, P.A., Boyle, K.J., Brown, T.C. (Eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, Dordrecht.
IEA, 2003. Electricity Statistics. International Energy Agency Energy Statistics available at http://www.iea.org/textbase/stats/electricityresult.asp.
JESS, 2003. Joint Energy Security of Supply Working Group — Second Report. Department of Trade and Industry. February http://www.dti.gov.uk/energy/jess/reports.shtml.
Johnson, R.F., Desvousges, W.H., 1997. Estimating stated preferences with rated-pair data: environmental, health, and employment effects of energy programs. Journal of Environmental Economics and Management 34 (1), 79–99.
Krewitt, W., 2002. External costs of energy — do the answers match the questions? Looking back at 10 years of ExternE. Energy Policy 30, 839–848.
Krupnick, A., Alberini, A., Cropper, M.L., Simon, N., O'Brien, B., Goeree, R., Heintzelman, M., 2002. Age, health, and the willingness to pay for mortality risk reductions: a contingent valuation survey of Ontario residents. Journal of Risk and Uncertainty 24, 161–186.
Layton, D.F., Moeltner, K., 2005. The cost of power outages to heterogeneous households. In: Scarpa, R., Alberini, A. (Eds.), Applications of Simulation Methods in Environmental and Resource Economics. Springer, Dordrecht, pp. 35–54.
Li, H., Berrens, R.P., Bohara, A.K., Jenkins-Smith, H.C., Silva, C.L., Weimer, D.L., 2004. Would developing country commitments affect US households support for a modified Kyoto protocol? Ecological Economics 48, 329–343.
Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.
Markandya, A., 2003. Expected external costs and their structure characterizing electricity options and their role in the strategic assessment of these options. Conference on Electric Power Supply in the 21st Century, Warsaw.
McFadden, D.L., 1974. Conditional logit analysis of qualitative choice analysis. In: Zarembka, P. (Ed.), Frontiers in Econometrics. New York Academic Press, New York, pp. 105–142.
Menges, R., Schroeder, C., Traub, S., 2005. Altruism, warm glow and the willingness-to-donate for green electricity: an artefactual field experiment. Environmental and Resource Economics 31 (4), 431–458.
Mitchell, R.C., Carson, R.T., 1989. Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington, DC.
NewExt, 2004. New elements for the assessment of external costs from energy technologies. Final Report to the European Commission, DG Research.
OECD, 2000. Ancillary benefits and costs of greenhouse gas mitigation. Organization for Economic Co-operation and Development, Paris.
Office for National Statistics, 2005. Employee jobs by industry 1978–2005. London. http://www.statistics.gov.uk/STATBASE/tsdataset.asp?vlnk=3412005.
Pearce, D.W., Cline, W.R., Achanta, A.N., Fankhauser, S., Pachauri, R.K., Tol, R.S.J., Vellinga, P., 1996. The social costs of climate change: greenhouse damage and the benefits of control. In: Bruce, J.P., Lee, H., Haites, E.F. (Eds.), Climate Change 1995: Economic and Social Dimensions—Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp. 179–224.
Popp, D., 2001. Altruism and the demand for environmental quality. Land Economics 77 (3), 339–349.
Roe, B., Teisl, M.F., Levy, A., Russell, M., 2001. US consumers' willingness to pay for green electricity. Energy Policy 29 (11), 917–925.
Tol, R.S.J., 2005. The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties. Energy Policy 33, 2064–2074.
Train, K., 2003. Discrete Choice Methods with Simulation Cambridge University Press, Cambridge. 342 pp.
Turner, R.K., Pearce, D., Bateman, I., 1994. Environmental Economics: An Elementary Introduction. Prentice Hall, London. 328 pp.
Veisten, K., Hoen, H.F., Navrud, S., Strand, J., 2004. Scope insensitivity in contingent valuation of complex environmental amenities. Journal of Environmental Management 73, 317–331.
Viscusi, W.K., Magat, W.A., Huber, J., 1991. Pricing environmental health risks: survey assessment of risk-risk and risk-dollar tradeoffs for chronic bronchitis. Journal of Environmental Economics and Management 21, 32–51.
Wiser, R., 2003. Using Contingent Valuation to Explore Willingness to Pay for Renewable Energy: A Comparison of Collective and Voluntary Payment Vehicles. Lawrence Berkeley National Laboratory, LBNL-53239, Berkeley.
Witzke, H.P., Urfei, G., 2001. Willingness to pay for environmental protection in Germany: coping with the regional dimension. Regional Studies 35 (3), 207–214.
Woo, C-K., Pupp, R., Flaim, T., Mango, R., 1991. How much do electric customers want to pay for reliability? New evidence on an old controversy. Energy Systems and Policy 15, 145–159.
Zarnikau, J., 2003. Consumer demand for 'green power' and energy efficiency. Energy Policy 31 (15), 1661–1672.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Energy 33 (2008) 740–746
ENERGY
www.elsevier.com/locate/energy

Internalizing externalities into capacity expansion planning: The case of electricity in Vietnam

Khanh Q. Nguyen[a,b,*]

[a]Department of Energy Economics, Institute of Energy, 6 Ton That Tung, Dong da, Hanoi, Vietnam
[b]GreenTech Ltd., 31/43 Chua Boc, Dong da, Hanoi, Vietnam

Received 4 May 2007

Abstract

This paper examines the impacts of including external costs such as environmental and health damages from power production on power generation expansion planning in Vietnam. Using the MARKAL model and covering a 20-year period to 2025, the study shows that there are substantial changes in the generation structure in favor of renewable energy technologies and other low emitting technologies. These changes lead to a reduction in fossil fuel requirements, and consequently, a reduction of CO_2, NO_x, SO_2, and PM emissions which could be expected to also reduce the associated environmental and human health impacts. The avoided external costs would be equivalent to 4.4 US cent/kWh. However, these gains are not free as the additional electricity production cost would be around 2.6 US cent/kWh higher if the switch to more expensive, but lower emitting technologies were made. The net benefit of internalizing these externalities is thus around 1.8 US cent/kWh.

Keywords: Electricity generation; Externalities; Vietnam

1. Introduction

Power consumption in Vietnam grew at an annual rate of over 15.1% during 1995–2005, far exceeding the annual growth in GDP of 7.2% in the same period. The highest annual growth rate occurred in the residential sector (16.4%), followed by the industrial sector (14.4%). Electricity demand in the coming period (2005–2025) is expected to continue to grow at a significant rate, driven by increasing urbanization and strong population growth, as well as economic growth and industrialization.

The Institute of Energy [1] forecasted that electricity demand in Vietnam would increase from 45,603 GWh in 2005 to 381,163 GWh in 2025, at an average annual growth rate of 11%. Such rapid development raises a number of questions concerning the choice of power generation technologies, particularly given concerns about the impact of particulate and greenhouse gas emissions on human health and climate change. So far, in Vietnam, the choice of technologies, as in many other countries, does not generally consider external costs imposed on society and environment due to unpriced pollutants emitted from electricity generation. The major pollutants are sulfur oxides (SO_x), nitrogen oxides (NO_x), particulate matters (PM), and carbon dioxide (CO_2).

This study looks at the impacts of internalizing these external costs on the choice of electricity technology to meet increasing demand. It proceeds as follows: Section 2 provides an overview of the power sector of Vietnam. Section 3 presents the modeling framework used to examine the impacts of including externalities on Vietnam's future power generation structure. Section 4 discusses the results and Section 5 concludes.

2. The Vietnam power sector

The power sector in Vietnam is governed by The Electricity of Vietnam (EVN), a utility wholly owned by the Government of Vietnam.

At the end of 2005 the total installed generation capacity connected to the network was 10,770 MW, which

*Tel.: +84 4 852 3385; fax: +84 4 852 3311.
E-mail address: mail2knguyen@yahoo.com

0360-5442/$ - see front matter
doi:10.1016/j.energy.2008.01.014

comprised of 61% thermal power plants and 39% hydro power plants. Of the total thermal generation, gas-based power plants had the biggest share (62%), followed by coal- and oil-based power plants with shares of 23% and 14%, respectively. The distribution of power plants in Vietnam is heavily influenced by its natural geography and energy reserves. In the north, hydro and coal power plants dominate while in the south gas turbines represent the major power source. To enable power exchange between the two regions, a 500 kV north–south transmission line was constructed in 1994. The second line was completed in late 2005 [1].

Power generation and network expansion projects in Vietnam follow power development plans, which are approved by the Government. The master plan is renewed every five years and so far six master plans have been made. The latest one looking to 2025 was finished early this year and is being submitted to the Prime Minister for approval.

The electricity tariff is also tightly regulated by the Government. Vietnam maintains uniform national electricity tariffs, across the country. The weighted average retail tariff including 10% value added tax (VAT) in 2005 was 5.5 US cent/kWh [1].

3. Modeling framework

To examine the impacts of internalizing externalities on Vietnam's future electricity generation mix, the Vietnam power sector is modeled using the MARKAL model. The analysis is carried out over a time frame of 20 years with 5-year time steps. The base year is 2005. MARKAL was developed by a consortium of members of the International Energy Agency (IEA) in the early 1980s based on the general algebraic modeling system (GAMS) [2]. Since then the model has evolved and has been applied to a wide range of energy and environmental issues in many countries other than IEA member countries. The issues that MARKAL has been successfully used to examine include [3]:

- Energy security.
- New technology R&D portfolio prioritization.
- Impacts and benefits of environmental regulations.
- Greenhouse gas (GHG) projections and
- GHG project evaluation and estimates of the value of carbon rights.

MARKAL is a dynamic, multi-period, linear programming bottom-up model of a generalized energy system, in which both the energy supply and demand side are depicted, including energy sources, conversion technologies, energy carriers and demand technologies and sectors. The model can be represented by four basic relationships [4]:

(i) the objective function (usually the total discounted cost of the energy system including external costs);
(ii) mass–conservation relationships that balance the flow of fuel masses across processes;
(iii) energy conservation relationship that connects the flow of fuel masses to energy generation; and
(iv) a set of constraints that establish bounds on both energy and mass flows. For example, levels and rates at which specific generation technologies can be deployed, and emission rates.

3.1. Case definition

To facilitate assessment of the impacts, two cases are analysed: the base case (BC) and the externality case (EC). The BC investigates the power system assuming that the current trend in the energy supply system is maintained into the future. That is, there is no consideration of externalities. In the EC, a policy scenario is developed which considers the externalities produced from power generation. In the EC, advanced lower emission technologies, conventional technologies with emission scrubbers and renewable energy technologies are introduced in addition to those available in the base scenario (Table 1).

3.2. External costs

In this paper, external costs are calculated using external cost factors, which are based on international sources due to the absence of specific estimates for Vietnam. The external cost factors, which are commonly expressed in US cent/kWh, vary between and within fossil fuel-based technologies due to variations in efficiency and the use of emission control systems [5]. To represent the varying external cost per unit of electricity produced, external cost factors of fossil fuel based technologies are adopted with per ton pollutants (unit damage cost of pollutant) while other electricity generations are per unit of electricity produced.

Unit damage costs of pollutants are derived from the outcomes of the European Commission ExternE project of Germany [6]. The major types of damages considered by the project were: human health (SO_2, NO_x, and PM), effects on crops and materials (NO_x, SO_2), and damages caused by global warming driven by greenhouse gases (CO_2). In this project the impact pathway approach is used. That is, the pathway of polluting substances is followed from the release source to the point of damage occurrence. The associated negative impacts (damages) are quantified using damage functions. Economic valuation of the damages are estimated by the "willingness to pay" of the affected individual to avoid a negative impact resulting from energy production from an actual power plant. Emissions of SO_2, NO_x, PM_{10}, and CO_2 from fossil fuel consumption in electricity generation technologies are evaluated using emission factors of these airborne pollutants [7].

In order to apply these external costs to Vietnam, two primary adjustments are made: (i) to reflect differences in income and hence, willingness-to-pay (i.e., regarding the

Table 1
Overview of key characteristics of candidate generation technologies in the Vietnam MARKAL model

Technology	Start year	Lifetime	Efficiency (%)	Investment cost \$US/kW	Fixed O&M cost \$US/kW yr	Variable O&M cost \$US/MWh
Conventional coal power plant	2005	30	39.5	1100	33.6	0.15
Conventional coal with DeSul/DeNox	2005	30	38	1200	38.5	0.25
Conventional coal with DeSul/DeNox and CO_2 scrubber	2005	30	30	2000	39.0	0.32
PFBC-based coal power plant	2015	30	42	1580	37.6	0.15
PFBC-based coal power plant with CO_2 scrubber	2015	30	36	1900	50.13	0.23
Coal IGCC	2015	20	42	1580	37.6	0.15
Coal IGCC with CO_2 scrubber	2015	20	36	1900	50.13	0.23
Natural gas combined cycle	2005	20	48	640	23.8	0.99
Natural gas combined cycle with CO_2 scrubber	2015	20	42	1000	32.51	1.38
Diesel (DO) fired gas turbine	2005	20	36	400	15.8	3.95
Natural gas fired gas turbine	2005	20	36	400	15.8	3.95
Fuel oil (FO) power plant	2005	20	38	800	19.5	1.48
Large hydro power plant	2005	40	100	1300	9.1	–
Small hydro power plant	2005	20	100	1200	18.0	–
Biomass steam turbine	2010	20	27.7	2000	73.0	8.5
Solar photovoltaic	2010	20	100	6000	15.4	–
Geothermal power plant	2010	20	100	2000	78.5	–
Wind turbines	2010	20	100	1000	23.7	–

Source: author based on various sources [1,9,11,14].
Notes: wind energy is actually represented in the Vietnam MARKAL model by three grades of technologies to account for different resource scales [11]. Similarly, solar photovoltaic is represented by two grades of technologies to represent different solar conditions in the north and the south of Vietnam [13]. PFBC, pressurized fluidized bed combustion; IGCC, integrated coal gasification combined cycle.

valuation of the damages), Vietnam's GDP per capita is used, and (ii) to reflect differences in the magnitude of the physical damage per ton of pollutant, Vietnam's population density is used [8]. The scaling factor is derived by dividing the magnitude of the parameter for Vietnam by the corresponding value for Germany. As for CO_2, a global external cost of climate change of 50 USD/ton of CO_2 is assumed [9]. Table 2 shows the result of these adjustments.

As for non-fossil fuel generation technologies, similar adjustments are performed to the external cost values of Germany so that they are applicable to Vietnam. Table 3 summarizes the external cost factors used for Vietnam after the adjustments to the external costs of Germany.

The external costs are internalized into the model differently depending on technology categories: Fossil fuel-based technologies or renewable energy technologies due to the reason given above. For fossil fuel-based technologies, the external costs are incorporated as an externality tax. This externality tax would be charged directly per unit of pollutant emitted by relevant generation technologies in the system. For renewable energy technologies, the external costs are added directly to the variable cost of corresponding technologies. In this way, it is assured that the external costs are charged to every unit of output. MARKAL has three response options to react to the extra charges on power generation: (i) to pay an external charge on power production from a technology, (ii) to install a costly emissions reduction system with DeNox, DeSuf, and CO_2 capture ability, or (iii) to change towards technologies with lower external costs (e.g. renewable energy technologies).

Table 2
Estimates of per unit damage costs for Vietnam

Pollutant	Germany (US\$ per ton)	Real GDP scaling factor[a] (%)	Damage scaling factor[b] (%)	Scaled unit damage cost (US\$ per ton)
SO_2	12,350	7.3	112.4	1201
NO_x	7250	7.3	112.4	705
PM	23,670	7.3	112.4	2302
CO_2				50

[a]Based on the following 1995 PPP (purchasing power parity) GDP for Germany and Vietnam in 2004: US\$ 22,361 and US\$ 1633, respectively [10].
[b]Based on the following population density for Germany and Vietnam: 223 and 251 person/km^2, respectively [10].

Table 3
External costs of non-fossil fuel electricity generation in Vietnam

Region	Biomass	Hydro	Geothermal	Solar PV	Wind
Vietnam (US cent/kWh)	0.0046	0.0082	0.0464	0.0556	0.0046

3.3. Other assumptions

The following main constraints and assumptions are implemented in the MARKAL modeling framework:

- *Peak reserve*: The peak reserve capacity is set to decrease gradually, from 35% in 2005 down to 25% in 2025.

- *Fuels*: The maximum annual supplies along with costs of domestic coal and natural gas to the power sector are according to development plans of respective industries [1]. For imported fuels—coal and oil products (DO, FO)—no restriction is set on the import level.
- *Technology and capacity constraints*: For hydro and geothermal power the maximum capacity is assumed to be 17,000 and 400 MW, respectively, in accordance with their respective potentials [11]. Wind energy is represented by three grades of wind turbines to represent a variety of wind resource conditions. The maximum capacity for all these three representative wind turbines is 13,500 MW [11]. As for solar photovoltaic and biomass-fired power plant, a modest level of 1000 MW for each is set. In addition, investment costs are assumed to decline by 1.5% for solar photovoltaic and 0.7% for wind turbine each successive year to account for technology learning effects [12].
- *Power demand*: The study assumes that there is no link between the average power price and total power demand. That means that a change in the generation price due to externality considerations does not reduce total power demand. However, the use of individual electricity generation technologies is influenced by relative prices.
- *Discount rate*: A discount rate of 10% is applied for the present study. This rate is recommended by the World Bank for analysis of the technological choices in Vietnam [1].

4. Results and discussion

4.1. The BC

This case assumes the continuation of the current trend of power development in Vietnam. That is, there is no consideration of externalities and no policy for emission abatement. Accordingly, generation capacity is expected to grow from 10.77 GW in 2005 to 74.04 GW in 2025, i.e., at an average annual growth rate of 10.1% (Table 4). At the same time, generation structure is assumed to change significantly. The share of hydro power plants reduces from 39% in 2005 to 24% (18 GW) in 2025, whereas coal power plants undergo considerable growth, from a 14% share in 2005 to a 69% share (51.3 GW) in 2025. The switch from hydro- to coal-based power plants drives 12.6%/year growth in fossil fuel consumption, from 303.60 PJ in 2005 to 3277.69 PJ in 2025. In order to meet this rapidly growing demand, Vietnam would need to import energy, such as coal, after 2010. The proportion of imported coal in total fuel consumption is expected to increase strongly from 14.1% (92.6 PJ) in 2010 to 78% (2585.5 PJ) by 2025. CO_2 emissions in this period are projected to grow at 14.3% per year, from 21.33 million ton in 2005 to 307.26 million ton by 2025 (Table 6). Per capita, the increase would be from 0.26 million ton in 2005 to 3.04 million ton in 2025, equivalent to a growth rate of 13.2% per year. Emissions of SO_2 are much lower, however, but they are expected to increase at a significant rate, 17.4% per year. Emissions of NO_x are also small in size; however, they are also expected to increase at a considerable rate of 15.2% per year, from 55.76 thousand ton in 2005 to 946.6 thousand ton in 2025. These emissions could impose huge costs on the society and the environment. The total damage from pollutants in 2005 is assessed at about 1225 million USD, equivalent to 2.4% of the real GDP. Damages are projected to grow to 19,656 million USD by 2025. This is equivalent to 7.5% of the projected GDP,[1] i.e., a bigger percentage of a larger GDP. Representing these in terms of US cent per electricity consumed, the increase would be from 2.8 US cent/kWh in 2005 to 5.2 US cent/kWh by 2025, primarily driven by the increasing share of coal.

4.2. The EC

Including external costs in the total production cost of electricity changes the generation mix (Table 5). Even though coal continues to dominate, its share is reduced by 2025 by 21.6% compared with the BC. Specifically, by 2025, 11.08 GW of coal power plants in the BC is replaced by 3.29 GW of gas turbine and 14.63 GW of renewable energy technologies of geothermal, wind, and biomass. Moreover, selected coal-based technologies are those with low emission and/or emission control such as conventional coal power plant with DeSuf/DeNox and CO_2 scrubber and coal IGCC with CO_2 scrubber. This change in generation mix indicates that investments in additional low emitting technologies such as coal-based technologies with emission control and renewable energy technologies (wind, geothermal, and biomass) are more economic than paying taxes. Relative to the BC, this change in generation mix delays the need for coal importation by five years and also reduces CO_2, NO_x, SO_2, and PM emissions. Specifically, emissions are reduced by 253.7 million ton of CO_2, 2.79 million ton of SO_2, 836 thousand ton of NO_x, and 69 thousand ton of PM by 2025, relative to the BC (Table 6). However, this does not bring about a significant reduction in fossil fuel consumption given that coal-powered plants remain economic if emission controls are used. Also, coal IGCC with CO_2 scrubbers generally have a lower efficiency than that of the conventional coal power plants which are selected in the BC (36% versus 39.5%) resulting in higher fossil fuel consumption. These resulting increases in fossil fuel consumption are almost equal to the decreases in fossil fuel consumption brought about by renewable energy technologies.

The reduction in emissions reduces the external costs imposed on society and the environment. By 2025, the external costs in the EC are 2868 million USD or 1.1% of the projected GDP for the same year compared to 19,656 million USD or 7.5% of projected GDP in the BC.

[1] GDP is projected to grow at an annual average rate of 8.5% from 2005 to 2020 and 8% thereafter [1].

Table 4
Future capacity development and fossil fuel requirement in the base case

	2005	2010	2015	2020	2025
Total capacity (GW)	10.77	22.02	35.98	52.63	74.04
Conventional coal power plant	1.50	7.49	15.86	29.91	51.32
Conventional coal with DeSul/DeNox	0	0	0	0	0
Conventional coal with DeSul/DeNox and CO_2 scrubber	0	0	0	0	0
PFBC-based coal power plant	0	0	0	0	0
PFBC-based coal power plant with CO_2 scrubber	0	0	0	0	0
Coal IGCC	0	0	0	0	0
Coal IGCC with CO_2 scrubber	0	0	0	0	0
Natural gas combined cycle	4.09	4.09	4.09	4.09	4.09
Natural gas combined cycle with CO_2 scrubber	0	0	0	0	0
Large hydro power plant	4.15	9.61	15.00	17.00	17.00
Small hydro power plant	0.07	0.07	0.27	1.00	1.00
Geothermal	0	0	0	0	0
Solar photovoltaic	0	0	0	0	0
Wind	0	0	0	0	0
Others	0.96	0.76	0.76	0.63	0.63
Fossil fuel requirement (PJ)	303.60	650.01	1153.85	1995.29	3277.69
Coal	94.30	453.00	953.20	1796.80	3076.20
Domestic	94.30	360.40	453.40	504.60	490.70
Imported	0	92.60	499.80	1292.20	2585.50
Natural gas	182.30	201.50	201.50	201.50	201.50
FO	27.00	0	0	0	0

Others: sum of DO, FO power plants.

Table 5
Future capacity development and fossil fuel requirement in the externality case

	2005	2010	2015	2020	2025
Total capacity (GW)	10.77	22.79	37.79	55.94	79.88
Conventional coal power plant	1.50	1.29	0.85	0.85	0.85
Conventional coal with DeSul/DeNox	0	0	0	0	0
Conventional coal with DeSul/DeNox and CO_2 scrubber	0	2.05	2.05	2.05	2.05
PFBC-based coal power plant	0	0	0	0	0
PFBC-based coal power plant with CO_2 scrubber	0	0	0	0	0
Coal IGCC	0	0	0	0	0
Coal IGCC with CO_2 scrubber	0	0	7.85	19.17	37.34
Natural gas combined cycle	4.09	7.38	7.38	7.38	7.38
Natural gas combined cycle with CO_2 scrubber	0	0	0	0	0
Large hydro power plant	4.15	9.61	15.00	17.00	17.00
Small hydro power plant	0.07	0.20	0.50	1.00	1.00
Geothermal	0	0.10	0.40	0.40	0.40
Solar photovoltaic	0	0	0	0	0
Wind	0	1.40	3.00	7.46	13.23
Others	0.96	0.76	0.76	0.63	0.63
Fossil fuel requirement (PJ)	303.60	606.20	1075.10	1868.80	3142.10
Coal	94.30	242.70	711.60	1505.30	2778.60
Domestic	94.30	242.70	453.40	504.60	490.70
Imported	0.00	0.00	258.20	1000.70	2287.90
Natural gas	182.30	363.50	363.50	363.50	363.50
FO	27.00	0	0	0	0

Others: sum of DO, FO power plants.

Representing in US cent/kWh, the avoided external costs would be equivalent to 4.4 US cent/kWh. These gains are, however, not free as the average generation cost of electricity would be around 7.3 US cent/kWh or about 2.6 US cent/kWh higher under this case than the BC (Table 6).

Table 6
Major emissions and the average generation cost of electricity of both cases

Case study	2010	2015	2020	2025
The base case				
CO_2 (thousand ton per year)	54,838	102,971	184,146	307,260
SO_2 (thousand ton per year)	424	892	1682	2880
NO_x (thousand ton per year)	159	309	563	947
PM (thousand ton per year)	11	22	42	73
External cost per kWh ($ cent/kWh)	3.49	3.93	4.56	5.16
The average generation cost of electricity ($ cent/kWh)	4.57	4.62	4.68	4.79
The externality case				
CO_2 (thousand ton per year)	30,460	29,111	38,462	53,468
SO_2 (thousand ton per year)	114	49	65	91
NO_x (thousand ton per year)	91	75	89	110
PM (thousand ton per year)	5	3	3	4
External cost per kWh ($ cent/kWh)	1.79	0.95	0.81	0.75
The average generation cost of electricity ($ cent/kWh)	9.03	7.26	7.30	7.43

5. Conclusions

While electricity plays a vital role in the socio-economic development of a country, by-products of its production have an undesirable effect on the environment, which imposes external costs to society and to individuals. This study examines the impact of internalizing external costs on the least cost choice of generation mix. Although modeling the impacts of such policies carries certain limitations and uncertainties such as the external cost values, advanced technology options, and the rate of technology change, a number of conclusions can be derived:

- The total damage from pollutants from power generation in 2005 is assessed at about 1225 million USD in Vietnam, equivalent to 2.4% of the nominal GDP. In the absence of price reform and control policies, it is estimated that the total damages will grow to 19,656 million USD by 2025. This is equivalent to 7.5% of the projected GDP, i.e., a larger percentage of a larger GDP.
- The inclusion of external costs in production costs drives not only a change in generation mix from coal-based power plants to renewable energy technologies (in particular, wind energy and geothermal) but also an increase in the capacity of natural gas combined cycle and advanced coal-based technologies. In that sense, including external costs into full energy production cost could be considered as an effective instrument for promoting the fast introduction of low-emitting technologies such as renewable energy technologies.
- In general, domestic energy supplies to the power sector in Vietnam might not keep pace with the strongly growing fuel requirement. Thus, the country would need to import energy after 2010. The internalization of external costs in the power generation sector increases the share of renewable energy technologies in the generation mix. It thus could help lower the dependency on fossil fuels, improve the security of energy supply for the economy and contribute to the reduction of CO_2, NO_x, and SO_2 emissions.
- The reduction of emissions reduces the external costs imposed on the society and the environment. The avoided external costs would be equivalent to 4.4 US cent/kWh. This is higher than the increase in the generation cost of electricity, 2.6 US cent/kWh, which thus indicates that internalization of external costs into capacity expansion planning is an appropriate policy towards emission mitigation.

Acknowledgments

The author would like to thank the editor and three anonymous reviewers for their helpful comments and suggestions on an earlier version of the paper. However, the author only is responsible for any remaining errors.

References

[1] The master plan on power development stage VI. Institute of Energy, Hanoi, Vietnam; 2005.
[2] Fishbone LG, Giesen G, Goldstein G, Hymmen HA, Stock KJ, Vos H, et al. User's guide for MARKAL (BNL/KFA version 2)—a multi-period, linear-programming model for energy system analysis; 1983.
[3] Goldstein GA, Kanudi A, Loulou R. MARKAL—an energy-environment-economic model for sustainable development. See also 〈http://bronze.nescaum.org/projects/ne-markal/MARKAL_BACKGROUND.doc〉.
[4] Loulou R, Shukla PR, Kanudia A. Energy and environment policies for a sustainable future. Allied Publishers Limited; 1997.
[5] European Commission. External costs—research results on socio-environmental damages due to electricity and transport; 2003. See also 〈http://www.externe.info〉.
[6] ExternE. Externalities of fuel cycles 'ExternE' project: aggregation—external costs from electricity generation in Germany and the UK. A draft final report produced for the Commission of the European Communities DGXII; 1997.
[7] Goodwin JWL, Salway AG, Eggleston HS, Murrells TP, Berry JE. UK emissions of air pollutants 1970 to 1996. See

also 〈http://www.aeat.co.uk/netcen/airqual/naei/annreport/annrep96/contents.htm〉 [last accessed August 28, 2007].

[8] Markandya A, Boyd R. Economics of greenhouse gas limitations. Handbook reports. The indirect costs and benefits of greenhouse gas limitations: Mauritius case study. UNEP Collaborating Centre on Energy and Environment, RisØ National Laboratory, Denmark; 1999.

[9] Kypreos S, Krakowski R. Introducing externalities in the power-generation sector of China. Int J Global Energy Issues 2004;22(2–4):131–54.

[10] World Bank. World development indicators 2004; 2005.

[11] Nguyen QK. Impacts of wind power generation and CO_2 emission constraints on the future choice of fuels and technologies in the power sector of Vietnam. Energy Policy 2007;35(4):2305–12.

[12] Messner S. Endogenized technological learning in an energy system model. J Evol Econ 1997;7(3):291–313.

[13] Nguyen QK. Alternatives to grid extension for rural electrification: decentralized renewable energy technologies in Vietnam. Energy Policy 2007;35(4):2579–89.

[14] Rafaj P, Kypreos S. Internalisation of external cost in the power generation sector: analysis with global multi-regional MARKAL model. Energy Policy 2007;35(2):828–43.

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Ecological Economics 46 (2003) 333–350

ECOLOGICAL ECONOMICS

www.elsevier.com/locate/ecolecon

METHODS

Pricing environmental externalities in the power sector: ethical limits and implications for social choice

Patrik Söderholm*, Thomas Sundqvist

Division of Economics, Luleå University of Technology, S-971 87 Luleå, Sweden

Received 17 September 2002; received in revised form 2 May 2003; accepted 26 June 2003

Abstract

During the last decade, a series of valuation studies have made attempts at estimating the external environmental costs of various power generation sources. The purposes of this paper are: (a) to explore some of the ethical limits of the economic valuation of environmental impacts; and (b) to analyze what the implications are of these limits for the social choice between different electric power sources. Environmental valuation based on welfare economic theory builds on restrictive behavioral foundations and can only partly model moral values, although such values are an essential part of people's preference towards the environment. In addition, public preferences are seldom exogenously given as is commonly assumed in economic theory, but are instead formed in public discourse. For this reason, the range of electricity externalities where economic valuation (and thus cost–benefit analysis) should be applied is likely to be narrower than often assumed. After analyzing the scope, methodology and the results of the so-called ExternE project, the paper concludes that many power generation externalities are either inherently 'new' or inherently 'complex'. In these cases, the initial challenge lies not in 'discovering' private preferences, but in specifying the conditions for public discourse over common ways of understanding what the pertinent issues are about. This implies that research on the environmental externalities of power generation must, in addition to refining the theory and the applications of existing non-market valuation techniques, also address the instruments and content of political and moral debate.

Keywords: Power generation; Ethics; Neoclassical welfare economic theory; Environmental valuation; Externalities

1. Introduction

One of the key elements of energy and environmental policies in the western world is to 'get prices right' and to ensure that environmental externalities are accounted for in market mechanisms. Policy makers and economists have particularly targeted the environmental damages arising from power generation. The reasons for focusing especially on the power-generating sector are two-fold. First, power generation generally provides much more flexibility in terms of fuel choices than is the case for other energy sectors (e.g. transport) and the various technologies have significantly different environmental impacts. Second, power

* Corresponding author. Tel.: +46-920-492-078; fax: +46-920-492-035.
E-mail address: patrik.soderholm@ies.luth.se (P. Söderholm).

0921-8009/03/$ - see front matter
doi:10.1016/S0921-8009(03)00185-X

plants are concentrated in relatively few and thus easily identifiable facilities.

A series of valuation studies have made attempts at estimating the environmental costs of various power-generating technologies. Most of these studies were commissioned by governmental authorities, such as the European Commission, the US Department of Energy and the UK Department of Trade and Industry (Ottinger et al., 1990; Pearce et al., 1992; Rowe et al., 1995; European Commission, 1995a, 1999). Stirling (1998) (p. 268) concludes in his review and methodological critique of some of the most important external cost studies that:

> [...], there is little doubt that neoclassical environmental valuation techniques are the approach to environmental appraisal currently preferred by the official bodies responsible for the formulation, implementation, and international coordination of environmental regulation in the electricity supply sector.

In other words, the theoretical support for externality valuation exercises is drawn from the neoclassical welfare economics literature. Within this strand of research, there are a number of valuation methods in use (e.g. abatement cost, contingent valuation, hedonic pricing, etc.), but ultimately they all aim at discovering people's preferences expressed as willingness to pay (WTP) for environmental goods and services (see Sections 2 and 3). The valuation and internalization of externalities is generally deemed necessary for assisting market processes and for making efficient social choices.[1] The implications for energy policy of these external cost assessments are thus essential. For example, in order to improve efficiency in the selection of new power generation sources damage estimates can be used to determine 'adders' to the private production costs (Eyre, 1997). In addition, external cost estimates can be used to evaluate existing pollution taxes and/or tradable permit systems, or help in designing new ones. Taxes and subsidies that reflect the external costs or benefits will then ensure that profit-maximizing firms select the mix of goods and production technologies that best satisfy environmental and economic goals.

However, a number of researchers in the social science field have questioned the use of non-market valuation techniques as the basis for integrating public input into the environmental policy process (e.g. Sagoff, 1988; Spash, 1997). It is argued that these methods rely on overly restrictive assumptions and ethical principles, implying that they often produce poor descriptions of the environmental values people hold and therefore serve as inadequate inputs to policy decisions. So far, though, the validity of these concerns in the empirical context of power generation externalities is only poorly understood (Stirling, 1997).

The purposes of this paper are thus to: (a) explore some of the ethical limits of environmental valuation methods within the welfare economics paradigm; and (b) discuss what the implications of these limits are for the social choice between power-generation technologies. The main thesis of the paper is that the scope of electricity externalities where environmental valuation can be applied from an ethical point of view is probably narrower than commonly assumed. Specifically, many environmental impacts in the power generation sector involve moral concerns for which private preferences are not always readily available, but rather must be formed in public discourse. For this reason, economic valuation provides an insufficient (but not necessarily unnecessary or illegitimate) basis for social choice. Also, since various power sources give rise to different types of externalities—some likely to be less amenable to social cost pricing than others—the choice between different technologies becomes

[1] According to the Coase (1960) theorem, bargaining between the polluter and the affected agent(s) can, under certain circumstances (such as low transaction costs), internalize externalities and achieve an efficient market outcome. However, in most cases, due to the large number of parties involved, such bargaining will be too complex and expensive and government intervention is therefore called for.

more complex than is implied by the welfare economics literature.

Before proceeding it is important to note that one of the most important ethical principles in welfare economics is that 'only' human (subjective) preferences should count; all values in this case are thus anthropocentric in the sense that they lack existence apart from the human valuer. This is the approach taken in this paper. Thus, the possible existence of 'strong' intrinsic values (e.g. Rolston, 1982), implying that the environment has an 'objective' value that is independent of human existence, is brought up neither in economic theory nor in this paper.[2] Our main argument, however, is that in contrast to welfare economics, which assumes a single preference ordering for each individual, there are strong reasons to believe that people possess two or more preference orderings, using different ones in different instances. This implies that the usefulness of economics in making rational choices over limited resources ought to be complemented by other forms of social agreements about what should be the important criteria in energy and environmental policy.

The paper proceeds as follows. In Section 2, we briefly review the methods used to assess the external costs of electricity generation and present some of the results obtained in previous studies. Section 3 discusses the ethical foundations and the limits of environmental valuation techniques as well as alternative philosophical approaches to human preferences and social choice. Section 4 analyzes these ethical limitations in the empirical context of the power generation externalities examined in the European Commission's so-called ExternE project. Finally, Section 5 provides some concluding comments and remarks.

2. The valuation of power generation externalities: methods and results

An externality is an unpriced benefit or cost directly bestowed or imposed upon one agent by the actions of another agent. Externalities cause market failures in the sense that there will exist a difference between the private and the social (private plus external) costs and benefits of an action and the free market's allocation of resources will, as a result, be non-optimal from society's point of view (Varian, 1992). Most electricity externality studies assess the negative externalities (external costs), most importantly the environmental damages, for selected power generation sources. In these cases, the private costs of power production is thus deemed to be lower than the social costs and electricity markets will tend to clear at a price level below the marginal social cost. The social choice between different power generation technologies will be inefficient and biased towards energy sources with low private production costs, but not necessarily low social costs.

Even though externalities are not reflected in market transactions, they do have a direct impact on people's welfare and thus on economic value. The economic valuation of externalities and thus of many environmental impacts, builds on the assumption that people seek to satisfy their preferences, i.e. maximize utility or welfare. The change in the level of individual welfare resulting from a given environmental change is typically measured as the amount of income necessary to maintain a constant level of utility before, and after, the change. In this way, one can elicit welfare changes in monetary terms through willingness-to-pay (or willingness-to-accept) measures (see also Section 3). Externality valuation is thus ultimately concerned with applying different empirical methods to identify these measures. There are two broad methodological approaches employed in

[2] However, we still consider what may be referred to as 'weak' intrinsic values, in the sense that they are non-instrumental (rather than objective) and refer to a situation in which humans consider that something has a value in itself irrespective of whether it has value in attaining something else of value (i.e. they are non-instrumental values). See Stenmark (2002) for a discussion of the distinction between 'weak' and 'strong' intrinsic values.

practice to assess the value of electricity externalities: (a) the abatement cost approach and (b) the damage cost approach.[3]

The abatement cost approach uses the costs of controlling or mitigating damage or the costs of meeting legislated regulations as an implicit value of the damage avoided. The rationale behind this approach is that legislatures are assumed to having considered the willingness of the public to pay for alleviation of the damage and the relevant abatement costs in setting the standard,[4] thus providing a revealed preference damage estimate not necessarily less reliable than the more explicit valuation methods (see below). An example of a study that utilizes the abatement cost methodology is Bernow and Marron (1990).

The damage cost approach, on the other hand, aims at providing an explicit (rather than an implicit) measure of the economic damages arising from a negative externality. Damage costing can be either *top-down* or *bottom-up*. Top-down approaches make use of highly aggregated data to estimate the external costs of, say, particular pollutants. Researchers adopting the top-down approach normally start at the national or the regional level, using estimates of total quantities of a specific pollutant. These physical damages are attributed to power plants and converted to damage costs using available monetary estimates (e.g. US\$ per SO_2 emitted) on the damages arising from the pollutants under study (e.g. Hohmeyer, 1988). In the bottom-up approach, damages from a single source are typically traced, quantified and monetized through damage functions/impact pathways (e.g. European Commission, 1995a). This approach makes use of technology-specific data, combined with dispersion models, information on receptors and dose–response functions to physically quantify the impacts of specific externalities. These physical impacts then need to be converted to damage costs either by using available information or through original valuation studies.

There exist several ways of monetizing these externalities. The first two approaches discussed above—abatement cost and top-down damage cost—directly provide a monetary estimate of the damages associated with the externalities. However, in the third approach—bottom-up damage cost—one needs to translate the identified and physically quantified impacts into monetary terms. Generally, whenever market prices can be used as a basis for valuation, they are used. However, since externalities by definition are external to markets, impacts from externalities are not reflected in market prices. Consequently, any attempt to monetize an externality when making use of the bottom-up damage cost approach need to rely on non-market valuation methods. These methods can in turn be subdivided into (a) direct methods and (b) indirect methods.[5]

The direct methods attempt to create a hypothetical market for the environmental good. These methods are *direct* in the sense that they are based on direct questions to households about willingness to pay. The direct methods possess the advantage that they can assess total economic values, i.e. the use as well as the non-use values (i.e. existence values) associated with the good. Well-known techniques sorting under this approach include contingent valuation and choice experiments. The *indirect* methods take their basis in the actual (rather than the hypothetical) behavior of individuals. Either the welfare effects in terms of willingness to pay show up as changes in costs or revenues in observable markets or in markets closely related to the resource that are affected by the externality. The damage is thus indirectly valued using an existing relation between the externality and some good that is traded in a market. Examples of indirect methods are hedonic pricing and travel costs.

[3] See Sundqvist and Söderholm (2002) for a critical survey of a large number of economic studies focusing on the valuation of environmental externalities in the power generation sector.

[4] Specifically, the public decision makers are assumed to choose the level of abatement at which the marginal damage curve and the marginal abatement cost curve intersects.

[5] There exists an extensive literature on different environmental valuation methods and to review this in detail here would be beyond the scope of this paper. For an excellent overview, however, see Garrod and Willis (1999).

In recent years, policy makers and researchers have given increasing attention to the assessment of external costs in the electricity sector. Several major studies have addressed the issue and examples included in the ExternE-project in Europe (European Commission, 1995a) and in the US, the New York State Environmental Externality Cost Study (Rowe et al., 1995). As noted above, welfare economic theory directs us on how to value externalities and previous electricity externality studies have relied heavily on the methods outlined above. According to the welfare economic theory, the choice of method should not affect results of the externality assessments significantly, i.e. it should not matter for the outcome whether people's willingness to pay has been 'filtered' through the political process or if it has been elicited directly in, for instance, contingent valuation surveys. Still, this presumption builds on the rather strong assumption that politicians make optimal decisions, i.e. they know the true (marginal) abatement and (marginal) damage costs and they aim at maximizing social welfare. In addition, as noted by Joskow (1992), abatement costs will only be representative of damage cost if they are derived from the pollution control strategy that gives the least cost of control.

For the studies that have been completed, the externality estimates produced for each electricity source range from very high effects to more or less insignificant effects. Fig. 1 displays the external cost estimates from 63 different studies carried out during the 1980s and 1990s. For example, looking at coal, the range of external cost estimates is from 0.03 to <1000 US cents per kWh. Similar ambiguities exist for the other electricity sources.

The reported discrepancies in results for similar fuels raise some concerns about the validity and reliability of the conducted valuation studies. Still, it must be made clear that there is no reason to question the general notion that to some extent the numbers *should* differ due to, for instance: (a) the use of different technologies (e.g. implying separate emission factors); (b) the characteristics of the specific site under consideration (e.g. population density, income, transport distances etc.); and (c) differences in scope (e.g. only a fraction of all externalities may be included, the entire fuel cycle rather than only the generation stage has been evaluated etc.). Still, by employing statistical analysis and 132 observations of external cost estimates for a set of different fuels, Sundqvist (2002) shows that one additional and more troubling reason for this disparity is also the choice of externality assessment approach. Most notably, the probability of obtaining a low externality cost value is, *ceteris paribus*, lower when the abatement cost or top-down damage cost approaches are used while the opposite is true for the bottom-up damage cost approach. One reason for the difference in results between the abatement cost approach and bottom-up damage costs is that many analysts tend to base their calculations on existing regulations (rather than the least-cost regulation) when estimating the abatement cost (e.g. Joskow, 1992).[6] However, the analysis in this paper also adds a new perspective to the observed differences in reported externality estimates between the abatement cost approach and the damage cost approach, i.e. between implicit and explicit valuation. Policy makers are in their formulation of regulations likely to base their decisions also on additional ethical foundations and the implicit values reported in abatement cost studies may thus reflect a different reasoning process than that outlined in the welfare economics literature.

Fig. 1 also displays that the ranges intertwine across fuels making the ranking of various fuels with respect to externality impacts a difficult task. Still, some tentative conclusions can be drawn. For instance, the results suggest that fossil fuel fired power, in particular coal and oil, gives rise to the highest external costs, while some of the renewable energy sources, solar, wind and also hydropower, tend to have the lowest.

[6] The reason why the top-down approach also tends to produce relatively high external damage is that there may arise practical problems in attributing the 'exact' damage to each individual source, which may force researchers to rationalize and use standardized rules for the attribution-process. These rules may fail to ascribe the aggregate damage to each and every individual source, especially smaller sources, thus producing estimates for larger power plants that are positively biased since these latter plants, normally, are easily identifiable as well as significant sources of pollution.

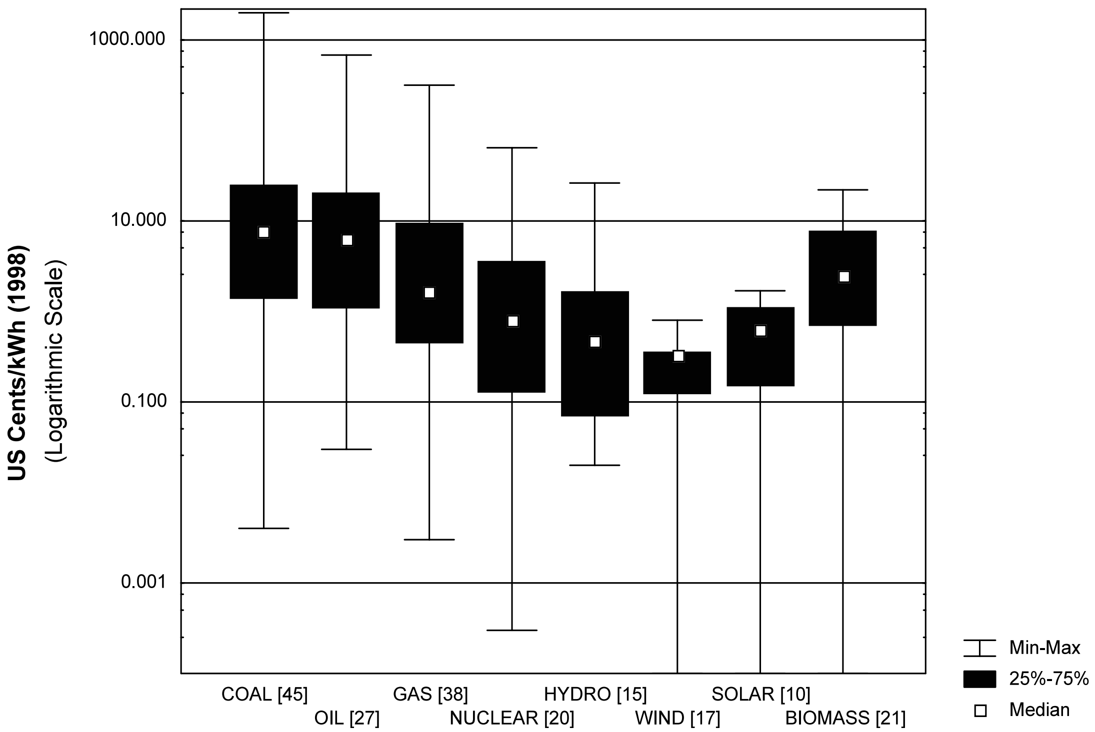


Fig. 1. Range of external cost estimates in power generation. Sources: Sundqvist (2002) and Sundqvist and Söderholm (2002).

According to Stirling (1997) (p. 531), "[t]his ambiguity in the comparison of different options is a serious defect in technique which aspires to present a robust and systematic representation of environmental performance." He argues that one of the most important defects of these studies is that they fail to address the multi-dimensional nature of power generation externalities. The different dimensions relate to, for example, the distribution of effects in terms of space, time and people, the particular forms they take (e.g. in terms of severity, reversibility etc.) and the degree of autonomy of those affected (Ibid.). Thus, according to Stirling, most of the existing valuation studies are still 'immature' and very preliminary; more realism in the treatment of the multi-dimensional nature of the external effects is therefore needed.

While previous critics, such as Stirling, address many of the practical and the methodological problems associated with assessing the externalities arising from power generation, the analysis in this paper is of a more fundamental nature. We argue that the behavioral and ethical foundations of environmental valuation, as applied to the valuation of external effects, are likely to be too restrictive for serving as the sole basis for social choice.

3. Ethical limits of welfare economics and the implications for social choice

Since the basic thesis of this paper is that the economic valuation of environmental externalities relies on specific behavioral assumptions and ethical foundations, it is useful to briefly review these before discussing alternative ethical bases for social choice and their consequences.

In the welfare economics discipline, human beings are treated as autonomous individuals who seek to satisfy their private preferences, which are complete, ethically unchallengeable (i.e. subjective) and exogenously determined. This implies that individuals have given preferences ('indifference maps') for public goods and are willing to consider tradeoffs in relation to the quantity or quality of these goods (Pearman et al., 1999). The objective of the analysis is to elicit from each individual his/her personal valuation of given environmental 'goods', measured in willingness to pay (WTP) terms. For example, within this theoretical framework each individual i's welfare is often expressed as:

$$U_i = (X, Z), \quad (1)$$

where U is the utility of individual i, X is a vector of the quantity private goods and Z represents the quantity of the public environmental good (e.g. air quality). The maximum WTP of individual i for increased provision of the public good is given by the solution to:

$$U_i(X^0, Z^0) = U_i(X^0 - \text{WTP}, Z^1), \quad (2)$$

which is equivalent to the compensating variation associated with the move from Z^0 to Z^1 at the initial level of private-good consumption level, X^0. Thus, if individuals are utility maximizers, welfare may be interpreted as units of measure of the maximum WTP for a given outcome (or reversing the property rights aspect, as a measure of the compensation an individual would require giving up some existing good, i.e. the minimum willingness to accept, WTA).[7]

Generally the welfare economics literature suggests that these welfare measures should be aggregated into the overall preference (utility) of society. The policy that maximizes total preference satisfaction needs to be chosen. The fundamental philosophical positions guiding social choice are thus that the net utility (benefits over costs) from the *consequences* of an action determines whether that action is right or wrong and a sense of society as the sum of the preferences (utilities) of its individual members. It should be noted, however, that this choice of ethical principle for *social* choices does not follow logically from the fact that utility maximization is assumed to constitute the behavioral foundation for *individual* choices. However, in practice they are likely to be closely related. The use of WTP as a welfare measure builds entirely on the assumption of utility maximizing behavior and there would probably be few reasons to estimate WTP as such if these estimates are not intended to form part of, say, social cost–benefit analyses that in turn are important (but not necessarily the only) input into the political decision process. Thus, as many philosophers point out, the development of societal ethical guidelines is largely an empirical question about individual's behavior and values.

While the standard environmental valuation techniques build on the assumption of utility maximizing behavior, the environmental stance of individuals is in many cases likely based on a deontological or rights-based approach to decision-making (e.g. Brennan, 1995). In this context, decisions are made based on whether the act itself is right or wrong regardless of its consequences, i.e. this approach recognizes the priority of the right over the good. For example, people may believe that aspects of the environment, such as wildlife threatened by a hydropower development project, have an absolute right to protection. They are thus willing to defend the existence or the well being of the environment apart from any instrumental value it provides. This is in line with the Sen (1977) distinction between *sympathy*, where concern for others forms one part of the utility function and *commitment*, where acts of altruism are chosen, even though they may result in lower utility for the individual. In other words, deontology denies the rationality attributed to making tradeoffs, whatever the commodity and therefore

[7] In Eq. (2) the unit of WTP is the quantity of private goods. However, by employing so-called indirect utility functions one can express WTP as a money metric measure. See, for instance, Freeman (1993). The choice between WTP and WTA as a welfare measure depends on the assumed property rights situation. For instance, in the case where the individual can be assumed to be the property right owner it can (theoretically) be valid to ask for the WTA in the case of a deterioration of the resource. Otherwise the individual may find the question (scenario) illegitimate and may choose to refuse to respond or can provide a protest bid.

suggests the existence of so-called lexicographic preferences. In this case, the axiom of continuity is violated, and the utility function in Eq. (1) is indefinable for an individual.[8] Thus, the indifference curves collapse to single points, denying the principle of substitution.

Spash and Hanley (1995) present empirical support for the existence of a deontological ethics, and conclude that standard valuation methods that elicit bids for biodiversity preservation fail as measures of welfare changes due to the existence of lexicographic preferences. Stevens et al. (1991) performed a contingent valuation study of species preservation in New England. A majority of the respondents (79%) agreed with the statement that: "all species of wildlife have a right to live independent of any benefit or harm to people." Still, when confronted with the WTP question, most of the respondents refused to pay. In other words, they were reluctant to choose between something of instrumental value (private goods) and a true moral position and in this way they applied a decision-making process inconsistent with the welfare economics paradigm.[9]

The motivation for the existence of a rights-based ethics, however, need not rely solely on empirical evidence. It is equally important to recognize that utilitarianism (and consequentialism) will not in itself be a sufficient moral theory for social choice. Since we cannot evaluate the net utility of an infinite number of alternatives, pure utilitarianism becomes a tautology. Some options simply have to be ruled out and this selection cannot be justified in utilitarian terms; instead we need to choose among options that we regard as morally or politically worth considering.

This does not imply that we should abandon the utilitarian approach to social choice. It merely points to the simple fact that people may approach the same issue in different ways, i.e. with different ethical standpoints. Environmental values often have a broad ethical content and since ethics are a matter for discussion[w1], environmental valuation ought to be endogenous to the political process and ultimately rely on social agreements. In other words, "the collective choice problem is, first of all, about advancing common ways of understanding what the pertinent issues are about. Only then can we develop a basis for collective choice predicated upon the elicitation of individual choice," (Vatn and Bromley, 1994, p. 142). Reasoned political argument among citizens does not exclude utilitarian (or indeed any other) belief systems but contextualizes them and helps us reflect upon our own arguments. We may not agree on the importance of different fundamental moral values but may still be able to come to a consensus on how to deal with moral aspects of practical issues. This consensus on the principles for social choice may (or is even likely to) involve a reliance on social-cost benefit analyses in some—but as indicated above not in all—instances.

This line of reasoning mirrors the work of Sagoff (1988, 1998). He suggests that individuals have two distinct roles; they act both as consumers with private preferences and as citizens with public preferences. Private preferences reflect what the individual thinks is good from a pure utility maximizing perspective, e.g. he or she prefers Coke to Pepsi. Public preferences, in contrast, state what a person believes is best or right for the community as a whole, e.g. 'society should not legalize drugs'. For instance, some people may regard environmental pollution as something inherently wrong, and what Sagoff rejects is the view of such moral objections as constituting just another kind of external cost that can and should enter a cost–benefit analysis.

Although the distinction between private and public preferences often is hard to operationalize, the consequences of not understanding the difference can lead to results that we would normally like to avoid. For example, economists usually argue that for the purpose of cost–benefit analyses it does not matter *why* people value environmental goods. As such, economists assume that all preferences are private and they grant equal credibility to every motive that underlies these preferences. To base social choice on this approach, Sagoff argues, is the equivalent of trying to decide whether a person on trial is guilty by

[8] The seminal work in this area is Georgescu-Roegen (1936).

[9] See also Common et al. (1997), who survey the empirical evidence on this issue and Russell et al. (2001).

discovering, before any evidence has been heard, what the preferences of the jury are in this regard and then calculating the net benefits of the two possible verdicts. It thus involves "an underlying confusion between preferences that may be priced and values that are to be heard, considered, criticized, and understood" (Sagoff, 1988, p. 95).[10]

This suggests therefore that, apart from simply 'speaking out' their given private preferences, individuals engage in a social process in which they form a collective understanding as citizens about what is appropriate, right or good, and in this way construct a *basis* for social choice. In other words, public preferences are endogenous rather than exogenous. For this reason, public values are also context relative, i.e. they are determined by social processes that play important roles in internalizing norms and beliefs about what is right and wrong.[11] Preferences are also likely to change over time due to the influence of education and cultural variations (Norton et al., 1998).

Private preferences towards private goods may of course also be endogenous and thus change over time, but normally this does not call for broader public deliberations about fundamental values. The social learning process however does become particularly important when individuals are confronted with public goods that: (a) they have little past experience of (i.e. preferences normally do not exist until we find a need to build them); (b) involve ethical dilemmas; or (c) have very complex characteristics. This is often the case when environmental goods are involved. The myriad of different classes of environmental effects, the many cross-cutting dimensions of these effects and the different risk characteristics involved cannot be casually separated in many cases. In addition, the conventional way of learning about the attributes of a good—learning by doing—becomes difficult and indeed often risky. It is one thing to choose between Pepsi and Coke, but another to choose between the preservation of an entire ecosystem and the development of a hydropower plant.

In summary, in this section we suggest that environmental goods and services embody characteristics that present serious ethical complications when social choices are to be made on the basis of recommendations derived from standard environmental valuation techniques. Preferences toward public goods are often endogenous to the political process and there is thus an important distinction between private and public preferences. The latter includes not only utility maximizing motives, but also other ethical positions, such as a deontological approach to decision making. In many cases, therefore, the initial challenge lies not in 'discovering' private preferences, but in specifying the conditions for public discourse over what is worth valuing and for what reason.[12] This becomes particularly important for many environmental goods, which are often both 'new' (e.g. global warming) and 'complex' (e.g. ecosystems).

4. The ExternE study as a basis for social choice in the power generation sector

In this section we discuss the relevance of the above theoretical discussion for social choice in the empirical context of power generation externalities

[10] Still, one important limitation of Sagoff's analysis is that even though he stresses the importance of public participation and public discourse for environmental issues, he does not attempt at characterizing this public sphere in a theoretically compelling way. See, however, Fiske (1991, 1992) for an interesting and systematic account of social interaction in which market pricing is only one of four relational models.

[11] This so-called deliberative approach to environmental valuation also lends support from the normative political theory of deliberative democracy, which recognizes that it is no less rational to focus on the procedure of the political decision-making process than on its outcome. See, for instance, Jacobs (1997) and Sagoff (1998) for reviews of this literature.

[12] See also the seminal work by Kapp (1978) who concludes: "Indeed the really important problems of economics are questions of collective decision-making which cannot be dealt with in terms of calculus deductively derived from a formal concept of individual rationality under hypothetically assumed and transparent conditions" (p. 288). Thus, for Kapp environmental policy was a question of political economy rather than a technical issue to be decided by cost–benefit analysis. Of course, in practice valuation based on cost–benefit analysis may not necessarily differ much from that provided by public deliberations. See Page (1992) for some empirical evidence on this latter point.

as addressed in the so-called ExternE project (European Commission, 1995a, 1999). This project aimed at evaluating the external costs of the different power generation fuel cycles in the EU. The results and methods of the studies have been utilized as inputs in important modeling work and have served as vehicles in developing additional methodological work in the environment and energy field .[13] As the ExternE project represents one of the most ambitious and internationally recognized attempts at coming up with 'true' external cost estimates for the different power technologies (Krewitt, 2002), it serves well as a case study of the ethical limits of environmental valuation in the power sector. Tables 1 and 2 present the different power generation externalities quantified and priced within the ExternE core project (European Commission, 1995a).[14]

All studies that form part of the project primarily use the bottom-up damage cost approach. The analyses begin by identifying the range of the burdens and impacts that result from the different fuel chains. Only impacts deemed to have 'significant' effects are included in the final assessment. These are quantified and monetized based on WTP measures, using methodologies appropriate for each specific externality. This implies that the ExternE project is not at all entirely comprehensive in its assessment of environmental externalities (e.g. it omits ozone impacts from gas-fired power generation).[15]

When inspecting Tables 1 and 2 we first note that most of the fuel cycles involve significant impacts on the health and deaths of humans ('public and occupational health'). In the ExternE project considerable attention was put on evaluating these impacts and much was learnt, especially about the importance of fine particles emissions for public health (Krewitt, 2002). In the core project, the value of a statistical life was used to calculate the external costs of mortality[16] and chronic and acute morbidity effects from air emissions were monetized using previous estimates of WTP to avoid different symptoms. However, according to a deontological ethics, human beings are moral ends in themselves and an infinite amount would be required to compensate for the death of a human being. This comes into direct conflict with the ethical basis of the ExternE project, which (implicitly) aims at maximizing society's total utility.

This does not imply that we should spend the entire public budget on saving lives and preventing morbidity impacts; it simply points to the fact that such impacts involve a moral dilemma. To what extent should we treat humans as means to an end (utility) or as ends in themselves? This question cannot be resolved with the help of cost–benefit analyses, but rather within the realms of public discourse.[17] It is not enough in this instance to make the remark that we do already reveal our preferences against health and death risks by our daily risk-taking behavior. "Precisely because we fail, [...], to give life-saving the value in everyday personal decisions [...], we may wish our social decisions to provide us the occasion to display the reverence for life that we espouse but do not always show," (Kelman, 1981, p. 38). This suggests also that, in contrast to the postulations of welfare economic theory, in social choices involving less than perfect information about risks it may be sensible to make a distinction between

[13] See, for instance, Bigano et al. (2000) and Vennemo and Halseth (2001).

[14] In 1999, the ExternE core project was followed up by the so-called national implementation projects (European Commission, 1999), whose aim has been to develop an EU-wide set of external cost data for the different fuel cycles and countries, utilizing the methodology developed within the core project.

[15] In addition, the focus is on environmental externalities, and externalities attributable to, for instance, fuel supply security are beyond the scope of the analysis. See, however, Bohi and Toman (1996) for an overview of the existence of energy security externalities.

[16] In the national implementation part of the ExternE project the decision was made to introduce an alternative measure on which to base the valuation of mortality impacts due to air pollution. This is the so-called years of life lost (YOLL) approach, which essentially assigns a WTP to the risk of reducing life expectancy rather than to the risk of death.

[17] Of course, public deliberations do not guarantee wise or viable decisions. Still, for the resolving of moral issues they should provide an appropriate (if not entirely sufficient) starting point.

Table 1
Externalities priced within the ExternE core project: coal, oil and gas

Externality	Coal	Oil	Gas
Public health	PM, ozone, and accidents: Mortality, morbidity, and transport impacts	PM and ozone: Mortality, morbidity, and transport impacts	PM: Mortality, morbidity, and transport impacts
Occupational health	Diseases from mining and accidents during mining, transport, construction, and dismantling	Accidents: death and injury impacts	Accidents: death and injury impacts
Agriculture	Sulfur, acidification, and ozone: crop and soil impacts	Sulfur, acidification, and ozone: crop and soil impacts	
Forests	Sulfur, acidification, and ozone damages	Sulfur, acidification, and ozone damages	
Marine	Acidification impacts	Accidents with oil tankers	Fishery: extraction impacts
Materials	Sulfur and acidification damages on surfaces	Sulfur and acidification damages on surfaces	Sulfur and acidification damages on surfaces
Amenity	Noise: operational road and rail traffic impacts		Noise: operational impacts
Global warming	CO2, CH4 and N2O damages	CO2, CH4 and N2O damages	CO2, CH4, and N2O damages
Total estimate (US cents/kWh)	2.8–4.1*	2.7–2.9*	1.7*

Source: European Commission (1995a).
* The global warming impacts constitute roughly half of the reported external cost estimates for coal-, oil- and gas-fired power. In the ExternE core project, the global warming estimates were drawn from Cline (1992).

Table 2
Externalities priced within the ExternE core project: nuclear, hydro and wind

Externality	Nuclear	Hydro	Wind
Public health	Radiation and non-radiation: mortality and transport impacts from operations and accidents		Accidents: travel to and from work
Occupational health	Radiation and non-radiation: mortality and transport impacts from operations and accidents	Accidents during construction and operation	Accidents during manufacturing, construction, and operation of turbine
Agriculture		Loss of grazing land	Acidification: damage on crops
Forests		Forest production loss due to flooding and land use	Acidification damages
Marine		Water supply and ferry traffic	Acidification damages
Materials			Acidification damages
Amenity		Visual amenity loss	Noise and visual amenity loss: operational impacts
Global warming			CO_2, CH_4, and N_2O damages
Recreation		Fishing and hunting	
Cultural objects		Objects of cultural and archeological interest	
Biodiversity		Terrestrial and aquatic ecosystems	
Total estimate (US cents/kWh)	0.0003–0.01	0.3	0.1–0.3

Source: European Commission (1995a).

preferences, in terms of individual choices made, and welfare, which is a broader measure of well-being (Johansson-Stenman, 2002).

Since different power-generation sources differ in terms of their relative impact on mortality and morbidity, the above concerns may have a direct impact on the actual choice between fuels. For example, the risks presented by nuclear power are generally more dominated by disease impacts than those of, say, gas and hydropower. In addition, the aggregation of effects of different severity (e.g. morbidity versus mortality) into a single monetary value also raises the ethical question of how society should weigh the importance of each of these impacts.

From an ethical point of view mortality and morbidity impacts are likely to differ from those externalities affecting *materials*, such as corrosion caused by acidic deposition. In the latter case, the implicit trade-off is between higher electricity production and less material damages. Parts of the natural environment (including humans) are (for all practical reasons) never at stake here and for this reason, private preferences may well serve as an appropriate basis for social choice.

Another ambiguity in how to deal with novel social choice problems when one considers the fundamental differences in the nature of risk between the different electricity alternatives. With nuclear power, an option with very low probabilities of very large negative impacts were introduced in the electric power arena. This is in contrast to fossil-fueled power generation, which gives rise to continuous but also comparably modest impacts. The Krewitt (2002) (p. 844) review of the ExternE project concludes that:

> The instruments for the assessment of consequences from beyond design accidents in nuclear power plant are well established, and the message from the use of such models is rather clear and non-ambiguous: the impacts from a single event can be very large, resulting in up to several ten thousand cases of fatal cancers, and in monetary terms they could amount to billions of Euro. Normalized to the probability of the event, and to the electricity generation over the power plant's lifetime, the expected value of risk (i.e., the probability times consequences) is low, a fact which is even robust against uncertainties in the accident probability.

Many experts claim that laypeople in general tend to overestimate the very low probabilities of nuclear accidents, but people are often unimpressed by arguments stating that the *expected* damages of nuclear are lower than those of other alternatives. An extended research tradition (e.g. Slovic, 1987) attempts to explain such behavior. In particular, it is noted that the public finds it especially hard to accept risks that are hard to identify because they arise from novel circumstances or technologies or have a catastrophic potential and may constitute a threat to future generations. Laypersons also rank as serious, risks that are involuntary, uncontrollable or having an uncertain and inequitable distribution of consequences, and for many power generation possesses a large number of these risk profiles (Ibid).

There is thus a large degree of 'catastrophe aversion' among the public. This is far from an indication of 'irrational' behavior; instead, it expresses that the willingness to accept a certain risk is related to the capacity to deal with the consequences should they arise.[18] For example, nuclear waste management risks are essentially irreversible after the plant has been commissioned, while the visual amenity and noise impacts from wind power are more or less reversible since the plant can be removed. Such differences are likely to affect the public preferences toward power-generating technologies. In sum, most people are not willing to engage in a trade-off discussion

[18] Of course, a neoclassical counter-argument would be that catastrophe aversion simply reflects the fact that the insurance market is insufficient and unable to correctly pool risks in the case of a catastrophic incident (e.g. Radetzki and Radetzki, 2000). For this reason, the government has to cover these additional risks and provide a de facto subsidy to the nuclear industry. Nevertheless, we argue that even in the presence of perfectly functioning insurance markets the moral dilemma would still be there, and it is unlikely that compensation for future accidents would make the perceived catastrophe aversion problem disappear.

regarding events that may lead to disastrous effects (for present or future generations) even though the probability of that disaster is extremely low. Thus, in such cases there simply exists no well-defined private utility function on which to base external cost estimates.

Furthermore, in trying to evaluate and assess the problems of nuclear waste management and radiation, individuals need to rely heavily on the statements of scientists. These can provide important information about the main physical relationships and may be able to present different scenarios and discuss the outcome of each. Still, the perceptions of what constitutes a significant risk are essentially socially constructed. The question of how risk should be evaluated therefore requires a broader public discussion in which, of course, scientists ought to take an active part.[19] This would enable the public to revise their perceptions of different risk profiles by considering the arguments of researchers as well as of other laypeople. In this way, any 'overestimations' of the risks involved could be removed, not by informing people about the 'true', 'objective' risks, but by encouraging them to reflect and engage in deliberations with others.

In the 1980s, many countries (e.g. Austria, Germany, Italy, Switzerland, etc.) put a moratorium on further nuclear expansion. These political outcomes do not only reflect the fact that nuclear energy was found economically inefficient. They also express ethical commitments towards future generations and unwillingness to accept the associated risks. Thus, nuclear energy is essentially a new 'good' with complex and far-reaching risk characteristics, for which most societies have not yet found an overall ethical position on which to base public and private decisions.[20]

This latter argument applies to the impacts of global warming (primarily caused by carbon dioxide emissions) as well. If it were not for reports from scientists, people in general would know nothing of their existence. The effects of global warming are inherently global, irreversible, long-term and asymmetrically distributed over time. This is in heavy contrast with other emissions from the power sector (e.g. sulfur dioxide), whose impacts are more tangible and directly connected to present human (dis)utility.

Again, society (and in this case countries) need to establish the conditions on which to base social choices in this matter. For example, an ethical position about the claims of future generations needs to be chosen. In this process, the need for *actual* compensation when there are damages to future generations,[21] as well as any inviolable rights of coming generations would have to be considered (Spash, 1993). Related to this, one has also to decide what is the relevant degree of risk acceptance, and within which limits of risk should pure cost–benefit be applied. A first attempt to agree upon a global climate policy was made by the rich market economies at the Kyoto conference in 1997. After making some necessary simplifying assumptions, Radetzki (2000) concludes that the *implicit* marginal price set on carbon dioxide emissions by the political process in Kyoto is somewhere in the range five to 25 times higher than the more explicit marginal damage cost estimates employed in the ExternE project (see Table 1).[22]

What do we make of this discrepancy? According to welfare economic theory, it suggests that the outcome of the Kyoto process was highly ineffi-

[19] The problem is complicated further by the fact that most scientists tend to transform genuine *uncertainty* into *risk*, where risk reflects a situation where the probabilities of different outcomes are known. In other words, they make an implicit assumption that their understanding of causal effects and overall system behavior (e.g. the nuclear power process) is more or less correct (Shackley and Wynne, 1996).

[20] In addition, the opposition towards nuclear has not only been directed towards environmental and risk-related issues. It has also been a struggle between the local and the national level of the political life, where local communities often see no benefits in nuclear development and resist to accept decisions exclusively taken at the national level.

[21] This differs from the ethical approach in welfare economics, which normally builds on a *potential* compensation criterion.

[22] This implicit price equals the carbon price, which would have to prevail in order to fulfill the emissions reductions agreed to at Kyoto. Krewitt (2002) also reports the existence of substantial differences between this implicit price and marginal damage costs for Europe.

cient, this since the constraints on carbon emissions agreed to are not motivated by generally accepted external cost calculations. However, if one accepts the ethical approach discussed in the present paper one should note that the 'Kyoto price' and the 'ExternE price' reflect different reasoning processes and are therefore not directly comparable. Within the ExternE project, hypothetical prices are established in *advance* as one of the raw materials for calculating the 'total' cost of energy. Thus, these prices together determine whether a specific energy source is better than another. The 'Kyoto price', on the other hand, did not play a causal role in the decision made at Kyoto but at most merely reflects the economic results of the political process. In this latter case, it is therefore the process that defines the legitimacy of choice, not the result. Accordingly, any inadequacies of the outcome arrived at under this process are essentially inadequacies of the process that produced them and cannot be attributed to the fact that the 'in effect' price put on carbon emissions is much higher than the 'true', or 'total' price presented in the ExternE project. As was suggested above (Section 2), this implies that there is a fundamental ethical difference between the abatement cost (regulatory revealed preference) approach and the damage cost approach.

In order to evaluate the legitimacy of the Kyoto process we need to know how ordinary citizens frame their discussions on global warming and develop preferences about climate policy. A major research project has investigated what dimensions of climate change are important to the European public (Kasemir et al., 2000). Focus groups, covering ≈ 600 people in seven densely populated areas in Western Europe were convened. The researchers conclude that the participants usually favored a two-stage policy process. First governments need to set limits—'tolerable windows'—on the behavior of firms and individuals, especially in terms of overall energy use. These limits reflected primarily ethical (and not economic) considerations expressed as safe minimum standards. In a second stage, however, cost considerations become highly important. Climate policy should find cost-efficient ways to stay within these 'windows'. Thus, the deliberations of the groups indicated clearly that value for money rather than monetary valuation—i.e. cost efficiency rather than cost–benefit analysis—appears to be the relevant issue for laypeople in Europe in attempting to reach a judgment on climate policy.

Finally, the ExternE project includes a contingent valuation (CVM) study of some of the impacts of hydropower development in Norway (European Commission, 1995b). These impacts comprise three basic damage components: losses of recreation, cultural objects and ecosystems/nature. The respondents were asked how much they were willing to pay to avoid the above impacts.[23] This sub-study, we argue, implicitly raises many of the ethical dilemmas posed in this paper.

First, the complexity of the three 'goods' differs much. Recreation is essentially a private good, and a hypothetical bid for, say, hunting or fishing permits may be as trustworthy as any market price. However, as soon as the valuation range is broadened to include entire ecosystems, the problem of what is actually valued—and for what reason—becomes apparent. In a CVM study, ecosystems are described in a manner that renders them commodity-like (with a use value and an existence value) and there may be little room for what we would normally claim is the most important aspect of an ecosystem—its functional aspects (e.g. its life-supporting mechanisms and the role of ecological diversity) (Vatn and Bromley, 1994). In addition, the site dependent impacts on local ecosystems may be hard to quantify.

Second, the complexity of ecosystems is also related to the moral philosophies held by individuals. If we believe that a particular ecosystem is essential for life to be worthwhile, there is an indirect moral commitment to the system itself. According to this view, there would be no substitute means for achieving human satisfaction,

[23] It is worth noting that the hypothetical price derived from this CVM study basically equals the total external hydropower cost of 0.3 U.S. cents per kWh reported in Table 2. The remaining external costs are, in other words, comparably small and range between 0.0004 and 0.001 U.S. cents per kWh (European Commission, 1995b).

and this invalidates a contingent pricing analysis.[24] A similar argument can be made for some cultural objects. People wish to see some pattern to their lives and they want their lives to be set in some larger context. In many instances, cultural phenomena provide exactly that desired context. This is in contrast to a pure recreation good; it can normally be replaced by something else with an equivalent value. For the above reasons, it is probably fair to conclude that the values derived from this study, although competently conducted, are likely to serve as an insufficient guide toward an informed choice between preservation and hydropower development.

5. Concluding remarks

The pricing of power-generation externalities, it is argued, is necessary for making consistent and meaningful comparisons between technologies. Tradeoffs (however unfair they may seem) must always be made, and it is best to make them explicit in a cost–benefit analysis. Our main argument in this paper, however, is that this argument is based on restrictive behavioral assumptions and ethical principles outlined in the welfare economics literature. We do not claim that one has to choose this philosophy or reject it; we simply point to the fact that choosing this particular perspective gives us only partial insight into many environmental issues. All policies that attempt to reflect human preferences have to be sensitive to the actual behavior of humans; they cannot simply assume that all humans possess a single well-defined utility function, which is 'employed' in all situations. Since environmental issues often have a broad ethical content and people tend to possess different preference structures, there are no simple answers to the question of how decision-makers should collect public preferences and integrate them into the environmental policy process. What is clear, however, is that any meaningful policy process should aim at incorporating these different modes of articulating preferences towards the environment.

In practice, most societies adopt a two-step approach to achieving environmental goals, and—as we have tried to show in this paper—probably for good reasons. Take the example of the US sulfur allowance system. First, the government sets limits on the behavior of firms and individuals. For example, the Environmental Protection Agency (EPA) sets a cap on overall sulfur dioxide emissions. Ideally, these limits (or minimum standards) reflect not only the social costs and benefits of the policy, but also society's attitude toward risk and its ethical commitments towards the rights of natural amenities and ecosystems. This first step therefore requires a broad political dialogue among citizens and experts in order to illuminate and address the dilemmas and the underlying value conflicts. In a second step, the EPA encourages electric utilities to buy and sell emission allowances. The utilities will do so only when benefits exceed costs. Thus, within the overall emission limit, pure cost–benefit principles are allowed to dominate choices. There is, in other words, a fundamental ethical difference between a tradable permit system (which to some extent represents the solution to a cost-effectiveness analysis), and a pure cost–benefit analysis (that forms the sole basis of the policy decision).

It may well be that the American government, in some sense, allows too much or too little sulfur emissions. Put differently, one may argue that the implicit price on sulfur emissions is too low or too high. However, it is hard to see in what way a cost–benefit analysis of the 'full' cost of electricity would help us resolve this. Environmental valuation based on the welfare economics theory is primarily a tool for aggregation of private preferences and not for public discussion. In a democratic society, however, the discussion itself is important, since ethical positions and public preferences tend to be endogenous to the political process. Our analysis of the ExternE study's evaluation of a number of electricity externalities shows that the understanding of people's preferences towards many environmental impacts in this

[24] This dilemma is probably best illustrated by the building of China's Three Gorges dam. It leads to the flooding of large tropical forests and to the displacement of millions of people (e.g. The Economist, 1999).

sector requires a stronger focus on the instruments and the content of political and moral debate. The ExternE project may very well have provided a nice starting point for such a discussion, but it will not be able to substitute for it. Any talk of the 'full' cost of electricity has thus to be understood as at best metaphorical.

We do not suggest in this paper that standard non-market valuation exercises are fundamentally flawed. Under some circumstances (e.g. private goods, few ethical conflicts, a lot of prior experience on the part of the valuer etc.), they provide very relevant and reliable information for policy makers. What we suggest, however, is that in other cases, e.g. for 'new', 'complex' goods, researchers need to take two issues more seriously than has been the case in the past: (a) the process of preference formation; and (b) the distinction between public and private preferences. Researchers must increasingly help people build preferences (rather than assume them as given).[25] In general, there is a need for combining analyzes based on intensive value structuring, involving small numbers of people in focus groups, with more extensive value information gathered via surveys from large numbers of people. Such studies may also involve monetary valuation (e.g. WTP elicitation), but should also include a strong focus on the ethical values held by the respondent.

Both public and private preferences are important for informed social choices. However, a common problem is that people often express public preferences in surveys designed to elicit private preferences. Put differently, people's view of the issues presented in the scenarios presented to them in CVM surveys is often not compatible with the theoretical framework used to interpret the responses. To some extent this is of course a practical problem, and one may, for instance, alter the scenario preceding the WTP question so as to only trigger private preferences (e.g. Russell et al., 2001). However, in order to trigger also the public preferences one would need to adopt a broader theoretical framework when analyzing people's responses/arguments in focus groups as well as in surveys. The usefulness of economics in making rational choices over limited resources is vital, but in the environment and energy field it must be complemented by other forms of social intelligence about what should be the important criteria in social choice.

[25] See also Johansson-Stenman (2002) and Gregory et al. (1993) for more on these issues.

Acknowledgements

An earlier version of this paper was presented at the Twenty-third IAEE International Conference, Energy Markets and the New Millennium: Economics, Environment, Security of Supply, Sydney, Australia, 7–10, June, 2000. Financial support from Vattenfall and the Kempe Foundations is gratefully acknowledged, as are valuable comments from Christian Azar, Fredrik Carlsson, Marian Radetzki, John Tilton and two anonymous referees. We also acknowledge the support of Pat Wagner at the International Institute for Applied Systems Analysis (IIASA) for careful editing of the paper. Any remaining errors, however, reside solely with the authors.

References

Bigano, A., Proost, S., Van Rompuy, J., 2000. Alternative environmental regulation schemes for the Belgian power generation sector. Environmental and Resource Economics 16, 121–160.

Bernow, S., Marron, D., 1990. Valuation of Environmental Externalities for Energy Planning and Operations, May 1990 Update, Report 900-SB01. Tellus Institute, Boston.

Bohi, D.R., Toman, M.A., 1996. The Economics of Energy Security. Kluwer Academic, Boston.

Brennan, A., 1995. Ethics, ecology and economics. Biodiversity and Conservation 4, 798–811.

Cline, W.R., 1992. The Economics of Global Warming. Institute for International Economics, Washington DC.

Coase, R.H., 1960. The problem of social cost. Journal of Law and Economics 3, 1–44.

Common, M., Reid, I., Blamey, R., 1997. Do existence values for cost benefit analysis exist? Environmental and Resource Economics 9, 225–238.

European Commission, 1995. ExternE: Externalities of Energy. Volume 1, Summary, EUR 16521 EN, DGXII, Science, Research and Development. Office for Official Publications of the European Communities, Luxembourg.

European Commission, 1995. ExternE: Externalities of Energy. Volume 6, Wind and Hydro. EUR 16525 EN, DGXII, Science, Research and Development. Office for Official Publications of the European Communities, Luxembourg.

European Commission, 1999. ExternE: Externalities of Energy. Volume 10, National Implementation. EUR 18528 EN DGXII, Science, Research and Development. Office for Official Publications of the European Communities, Luxembourg.

Eyre, N., 1997. External costs—what do they mean for energy policy? Energy Policy 25, 85–95.

Fiske, A.P., 1991. Structures of Social Life: the Four Elementary Forms of Human Relations. Free Press New York.

Fiske, A.P., 1992. The four elementary forms of sociality: framework for a unified theory of social relations. Psychological Review 99, 689–723.

Freeman, A.M., 1993. The Measurement of Environmental and Resource Values: Theory and Methods. Resources for the Future, Washington, DC.

Garrod, G.D., Willis, K.G., 1999. Economic Valuation and the Environment—Methods and Case Studies. Edward Elgar, Cheltenham.

Georgescu-Roegen, N., 1936. The pure theory of consumer behavior. Quarterly Journal of Economics 50, 545–593.

Gregory, R., Lichtenstein, S., Slovic, P., 1993. Valuing environmental resources: a constructive approach. Journal of Risk and Uncertainty 7, 177–197.

Hohmeyer, O., 1988. Social Costs of Energy Consumption. Springer-Verlag, Berlin.

Jacobs, M., 1997. Environmental valuation, deliberative democracy and public decision-making institutions. In: Foster, J. (Ed.), Valuing Nature? Economics, Ethics and Environment. Routledge, London, pp. 211–231.

Johansson-Stenman, O., 2002. Evironmental policy when people's preferences are inconsistent, non-welfaristic, or simply not developed. In: Bromley, D., Paavola, J. (Eds.), Economics, Ethics and Environmental Policy: Contested Choices. Blackwell, London, pp. 103–119.

Joskow, P.L., 1992. Weighing environmental externalities: let's do it right. Electricity Journal May 5, 53–67.

Kapp, K.W., 1978. The Social Cost of Business Enterprise, third ed. Spokesman, Nottingham.

Kasemir, B., Schibli, D., Stoll, S., Jaeger, C.C., 2000. Involving the public in climate and energy decisions. Environment 42, 32–42.

Kelman, S., 1981. Cost-benefit analysis. An ethical critique. Regulation January: 33–40.

Krewitt, W., 2002. External costs of energy—do the answers match the questions? Looking back at 10 years of ExternE. Energy Policy 30, 839–848.

Norton, B., Costanza, R., Bishop, R.C., 1998. The evolution of preferences. Why 'sovereign' preferences may not lead to sustainable policies and what to do about it. Ecological Economics 24, 193–211.

Ottinger, R.L., Wooley, D.R., Robinson, N.A., Hodas, D.R., Babb, S.E., 1990. Environmental Costs of Electricity. Oceana, New York.

Page, T., 1992. Environmental existentialism. In: Costanza, R., Norton, B.G., Haskill, B.D. (Eds.), Ecosystem Health: New Goals for Environmental Management. Island Press, Covelo, CA.

Pearce, D.W., Bann, C., Georgiou, S., 1992. The Social Costs of Fuel Cycles. HMSO, London.

Pearman, R., Ma, Y., McGilvray, J., Common, M., 1999. Natural Resource and Environmental Economics, second ed. Longman, London.

Radetzki, M., 2000. Coal or nuclear in new power stations: the political economy of an undesirable but necessary choice. The Energy Journal 21, 135–147.

Radetzki, M., Radetzki, M., 2000. Private arrangements to cover large-scale liabilities caused by nuclear and other industrial catastrophes. The Geneva Papers on Risk and Insurance: Issues and Practice 25, 180–195.

Rolston, H., III, 1982. Are values in nature subjective or objective. Environmental Ethics 4, 125–151.

Rowe, R.D., Lang, C.M., Chestnut, L.G., Latimer, D.A., Rae, D.A., Bernow, S.M., White, D.E., 1995. New York State Environmental Externalities Cost Study. Empire State Electric Energy Research Corporation (ESEERCO). Oceana, New York.

Russell, C.S., Bue Bjørner, T., Clark, C.D., 2001. Searching for Evidence of Alternative Preferences: Public as Opposed to Private. Working Paper No. 1-W01, Department of Economics, Vanderbilt University, Nashville.

Sagoff, M., 1988. The Economy of the Earth. Cambridge University Press, Cambridge.

Sagoff, M., 1998. Aggregation and deliberation in valuing environmental public goods: a look beyond contingent pricing. Ecological Economics 24, 213–230.

Sen, A.K., 1977. Rational fools: a critique of the behavioral assumptions of economic theory. Philosophy and Public Affairs 6, 317–344.

Shackley, S., Wynne, B., 1996. Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority. Science, Technology and Human Values 21, 275–302.

Slovic, P., 1987. Perception of risk. Science 236, 280-285.

Spash, C.L., 1993. Economics, ethics, and long-term environmental damages. Environmental Ethics 15, 117–132.

Spash, C.L., 1997. Ethics and environmental attitudes with implications for economic valuation. Journal of Environmental Management 50, 403–416.

Spash, C.L., Hanley, N., 1995. Preferences, information and biodiversity preservation. Ecological Economics 12, 191–208.

Stenmark, M., 2002. Environmental Ethics and Policy-Making. Ashgate, Aldershot.

Stevens, T., Echevarria, J., Glass, R., Hager, T., More, T., 1991. Measuring the existence value of wildlife: what do CVM estimates really show? Land Economics 67, 390–400.

Stirling, A., 1997. Limits to the value of external costs. Energy Policy 25, 517–540.

Stirling, A., 1998. Valuing the environmental impacts of electricity production: a critical review of some 'first-generation' studies. Energy Sources 20, 267–300.

Sundqvist, T., 2002. Power Generation Choice in the Presence of Environmental Externalities, Ph.D. Dissertation, Division of Economics, Luleå University of Technology, Sweden.

Sundqvist, T., Söderholm, P., 2002. Valuing the environmental impacts of electricity generation: a critical survey. Journal of Energy Literature 8, 3–41.

The Economist, 1999. Water Power in Asia: The Dry Facts about Dams. November 20: 82.

Varian, H.R., 1992. Microeconomic Analysis, third ed. Norton, New York.

Vatn, A., Bromley, D.W., 1994. Choices without prices without apologies. Journal of Environmental Economics and Management 26, 129–148.

Vennemo, H., Halseth, A., 2001. Environmental regulation of a power investment in an international market. Resource and Energy Economics 23, 157–173.

Energy Risk management

ELSEVIER

Energy Policy 34 (2006) 3367–3373

ENERGY POLICY

www.elsevier.com/locate/enpol

Viewpoint

Energy risk management and value at risk modeling

Mehdi Sadeghi*, Saeed Shavvalpour

Economics department, Imam Sadiq University, P.B. 14655-159, Tehran-Iran

Available online 26 August 2005

Abstract

The value of energy trades can change over time with market conditions and underlying price variables. The rise of competition and deregulation in energy markets has led to relatively free energy markets that are characterized by high price shifts. Within oil markets the volatile oil price environment after OPEC agreements in the 1970s requires a risk quantification.'' Value-at-risk'' has become an essential tool for this end when quantifying market risk. There are various methods for calculating value-at-risk. The methods we introduced in this paper are Historical Simulation ARMA Forecasting and Variance–Covariance based on GARCH modeling approaches. The results show that among various approaches the HSAF methodology presents more efficient results, so that if the level of confidence is 99%, the value-at-risk calculated through HSAF methodology is greater than actual price changes in almost 97.6 percent of the forecasting period.

Keywords: Energy price risk; Price risk management; Value-at-risk; HSAF methodology; Garch model

1. Introduction

Risk Management embodies the process and the tools used for evaluating, measuring and managing the various risks within a Company's portfolio of financial, commodity and other assets. The value of energy trades can change over time as market conditions and underlying price variables change. A price forecast is the foundation for determining a firm's risk in managing their energy supply and their forward contracts for energy trades.

In energy markets, proper risk management depends not only upon proper portfolio analysis tools but also on a solid foundation of forward price.

Calls for competition in the power and gas industry have made deregulation an attractive option around the world. The rise of competition and deregulation in turn has led to relatively free energy markets that are characterized by high price shifts. Within oil markets the volatile oil price environment after OPEC agreements in the 1970s requires risk quantification. Value-at-Risk has become an essential tool for this end, when quantifying market risk. Within oil markets, value-at-risk (VaR) can be used to quantify the maximum oil price changes associated with a likelihood level. This quantification constitutes a fundamental point when designing risk management strategies. This paper aims at addressing the importance of oil price risk in managing price risk in energy markets and introducing the application of VaR in quantifying oil price risk.

The rest of this paper is set as follows: in Section 2 we put forward the fundamental of managing price risk in energy markets and the importance of price volatility in managing energy risk. Section 3 introduces the VaR modelling procedure and analyzes the main methodologies and models that can be used to determine VaR. Section 4 is devoted to addressing the proposed methodology. The final two sections present the empirical analysis and the main conclusions of this paper, respectively.

*Corresponding author. Tel.: +98 218081404; fax: +98 218093484.
E-mail addresses: sadeghi@isu.ac.ir (M. Sadeghi), shavalpoor@isu.ac.ir (S. Shavvalpour).

0301-4215/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enpol.2005.07.004

2. Price volatility and price risk management in energy markets

Risk management embodies measuring and managing the various risks within a company's portfolio of financial, commodity and other assets. Wengler (2001) argues that in the energy market, producers and providers enter into trading contracts that help match supply with demand. Energy firms buy or sell contracts on the open market to

- meet contracted deliveries when demand exceeds production capacity,
- sell excess capacity when demand is less than supply, and
- speculate to increase earnings through futures contracts

The value of energy trades can change over time as market conditions and underlying price variables change. A firm's portfolio risk is measured by evaluating the risk exposure from changes in any of the variables that affect existing contracts or the firm's projections from demand, supply and prices (Kaushik and Pirrong, 1999). A price forecast is the foundation for determining a firm's risk in managing their energy supply and their forward contracts for energy trades. Accurate price forecasting can therefore help reduce portfolio risk (Kaushik and Pirrong, 1999).

Analysis of expected return on assets based on "Value at-Risk" measures allows the firm to optimize the use of both physical and financial assets. Analysts can then determine the best use of physical and financial capital in order to maximize earnings (Wengler, 2001). As Parsons (1998) suggests, comprehensive risk management strategy that addresses both portfolio and operational risk, allows firms to

- avoid big losses due to price fluctuations or changing energy consumption patterns,
- reduce volatility in earnings while maximizing return on investment, and
- meet regulatory requirements that limit exposure to risk.

2.1. Price volatility and managing energy risk

Price volatility is at the heart of risk; yet it is an elusive concept that is hard to master and model. Volatility is usually defined as a measure for the magnitude of percentage changes in prices over time (Lintner, 1965).

According to the EIA report (2002) calls for competition in the power and gas industry, from the wholesale level to the retail level, have made deregulation an attractive option around the world. New market structures have been studied to search for a good one that can ultimately satisfy regulatory bodies, customers and suppliers.

The rise of competition and deregulation in power and gas markets has had a significant effect on prices so that the new market is relatively free and characterized by high price shifts. An unpredictable, volatile and risky environment has arisen and protection against market risk has become an essential issue.

In resource-based economies, such as those dependent on oil, exports and government revenues are uncertain and highly volatile. Uncertainty means that a variable, say, the oil price for the coming years, is simply unpredictable. In these economies oil price fluctuations not only affect the government budget considerably but also have strong effects on macroeconomic variables and even the stock market (Sadorsky, 1999). Given the effects of oil price volatility and the uncertainty, which is accompanied by these price movements, there is a great need for oil price risk quantification in these countries.

3. Value-at-Risk (VaR)

3.1. Definition

The term VaR did not enter the financial lexicon until the early 1990s, but the origins of VaR measures go further back. These can be traced to capital requirement for US security firms of the early 20th century. Starting with an informal capital test, the New York Stock Exchange (NYSE) first applied to member firms around 1922 (Hilton, 2003).

As Hendricks (1996) implies VaR is the maximum amount of money that may be lost on a portfolio over a given period of time, with a given level of confidence (Fig. 1). VaR describes the loss that can occur over a given period at a given confidence level, due to exposure to market risk (Hilton, 2003). The wide usage of the

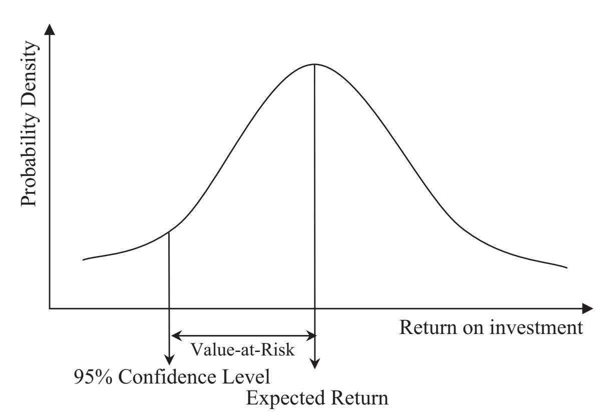


Fig. 1. VaR quantification using the probability density function of returns.

VaR-based Risk Management (VaR-RM) by financial as well as nonfinancial firms stems from the fact that VaR is an easily interpretable summary measure of risk and also has an appealing rationale as it allows its users to focus attention on "normal market conditions" in their routine operations (Basek and Shapiro, 2001).

Cabedo and Moya (2003) suggest that within oil markets, Value-at-Risk can be used to quantify the maximum oil price changes associated with a likelihood level. This quantification is fundamental when designing risk management strategies.

3.2. VaR quantification methods

There are several methods for calculating VaR. among them some methods are based on historical information that can be classified into three groups:

- Historical simulation Approach.
- Monte Carlo Simulation Method.
- Variance–Covariance methods (Hull and White, 1998).

In the Historical Simulation approach, an empirical distribution must be derived for the price changes over a period prior to the time of calculation. In the same way, for the Monte Carlo simulation method, an empirical distribution must be derived for the price changes. In this method some series of pseudo-random variables must be generated assuming that they follow a determined statistical distribution. Finally, within the Variance–Covariance methods it is assumed that potential loss is proportional to return standard deviation. Within the Variance–Covariance method VaR is estimated through:

$$\mathrm{VaR}_t = \lambda\sqrt{\theta}\,\mathrm{SDV}_{tp}, \tag{1}$$

where λ is the likelihood parameter, SDV_{tp} is the return standard deviation for time t; and θ is a parameter used when we calculate VaR for a time period with a length different from that used to estimate the standard deviation. Within the Variance–Covariance methods several methodologies can be used to calculate the VaR; among them Autoregressive Conditional Heteroskedasticity (ARCH) models are now very popular.

The original ARCH models were introduced by Engle (1982) and generalized by Bollerslev (1986) only few years later. Bollerslev's model characterizes the error term (ε_t) distribution in a general regression model conditional on the realized values of the set of exogenous variables (ϕ_{t-1}) as follows:

$$\varepsilon_t|\phi_{t-1}\sim N(o,h_t), \tag{3}$$

where normal distribution variance (h_t) can be expressed through

$$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p}. \tag{4}$$

This model is known as a generalized ARCH model or GARCH (p,q) model, where p denotes the number of considered lagged variance values and q determines this number for the squared deviations.

4. The historical simulation approach

The Historical Simulation approach for VaR quantification contains two methods. One is the Historical Simulation Standard approach and the other the Historical Simulation ARMA Forecasting approach. What makes HSAF methodology different from the historical simulation standard approach is that the first does not directly use the distribution of past returns but rather the distribution of forecasting errors, derived from an estimated ARMA model.(Cabedo and Moya, 2001)

HSAF methodology, introduced and developed in this paper, requires a four-stage procedure (Fig. 2).

In the first stage the past returns are calculated and their stationary behavior analyzed. There are various methods for testing the stationary of series. Dicky Fuller and Augmented Dicky Fuller tests are now the most relevant tests for this end. If the results confirm the stationary behavior of the series, then the procedure should be continued by testing the autocorrelation behavior of the original series. If the stationary hypothesis is rejected, then the consecutive differences over the original series are required.

Whether the original series is stationary or not, the next stage is to test the autocorrelation behavior of the series. The Ljung–Box test calculation is then advisable at this point. If autocorrelation is not statistically significant, then the HSAF methodology is equivalent to the historical simulation standard approach. On the other hand, only when the analysis of the series determines a statistically significant autocorrelation level can the second stage of the procedure be implemented.

In the second stage, by applying Box–Jenkin's methodology and using past returns, a model for past returns behavior can be estimated. Ljung–Box autocorrelation tests are used again in this stage in order to determine the necessary number of lags to consider in order to remove the autocorrelation.

During the third stage, using the coefficients estimated in the second stage, forecasts are made for price returns. Using these forecasts the forecasting errors can be obtained. The statistical distribution of these errors is analyzed and the percentile associated with the desired likelihood level is calculated.

The final stage involves forecasting future returns using the model estimated in the second stage of the procedure. These forecasts are corrected by the percentile obtained in the previous stage. These corrected

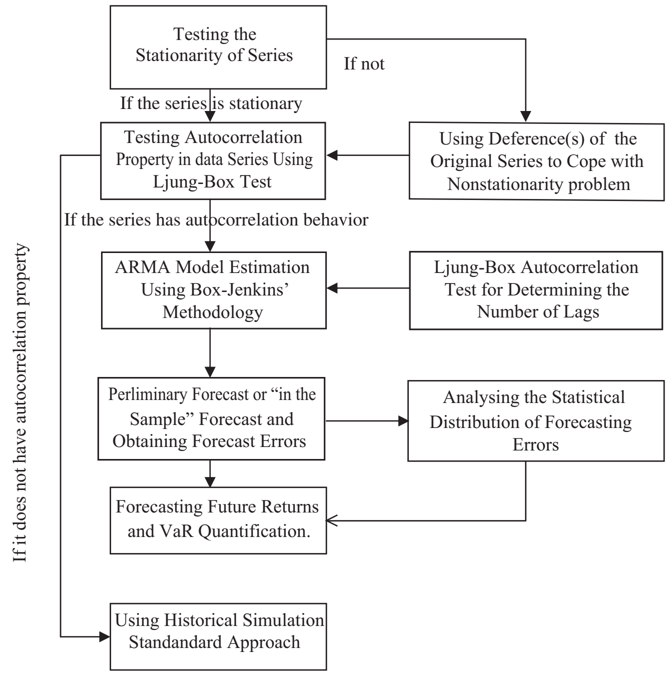


Fig. 2. The procedure for HSAF methodology implementation.

forecasts provide the value-at-Risk associated with a statistical likelihood level equivalent to the percentile used in the third stage.

5. Empirical results

5.1. Data

We used weekly OPEC prices from January 1997 to December 2003 and divided them into two periods: one from 1997 to 2002 which was used to estimate the model coefficients, and the other the year 2003, which was used for forecasting purposes (Fig. 3).

5.2. Historical simulation ARMA forecasting (HSAF) approach

As illustrated in Fig. 1, to apply the HSAF methodology we should follow a five-stage procedure. In the first stage we test the stationary of oil price series by applying ADF tests. Table 1 shows the result obtained. As results show the series is not stationary at conventional significant levels. To cope with this problem we used the first difference of series. Again we applied ADF tests. Table 2 shows the test results. As can be seen in this table, the first difference of the series is stationary at 99% level of confidence.

In the second stage we analyzed the autocorrelation functions of price returns by applying the Ljung–Box test. As can be seen in Table 3, the series show a statistically significant autocorrelation.

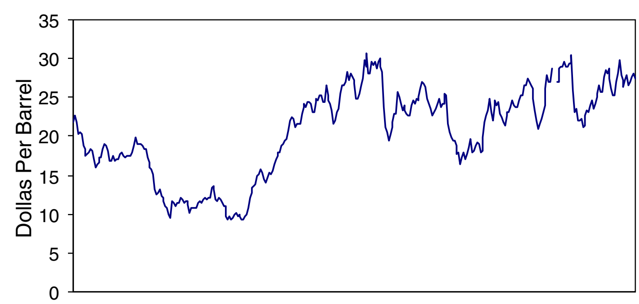


Fig. 3. Opec weekly oil prices January 1997–December 2003.

Table 1
ADF test statistics and critical values for the original series

1% Critical value	5% Critical value	10% Critical value
−3.4536	−2.8712	−2.5719
ADF test statistic:	−1.353785	

Table 2
ADF test statistics for the first- differenced series

1% Critical value	5% Critical value	10% Critical value
−3.4537	−2.8712	−2.5719
ADF Test Statistic:	−7.577756	

Stage 3 is devoted to the ARMA Model estimation. The method we used to estimate an ARMA model is Box–Jenkins. In this stage we estimated an AR(1) model. This estimation is according to the results obtained from analyzing Autocorrelation and Partial Autocorrelation functions. The estimation results are summarized in Exhibit 1.

Exhibit 1 Estimation results for AR(1) model

Variable	Coefficient	Std. error	*t*-Statistic
C	0.29553	0.140021	2.110612
AR(1)	0.222403	0.056164	3.959878

We also analyzed residuals ACF and PACF. The result indicated that there is no statistically significant autocorrelation in residuals (Table 4).

In stage 4, forecasts are made using the coefficients estimated in the third stage. We made these forecasts using the data provided by the "in the sample period" (1997–2002). Using these forecasts, we estimated the forecasting errors without any assumption about the skewness of the statistical distribution of the forecasting errors. We analyzed positive and negative forecasting errors separately and obtained the 99th percentile from their cumulative density function.

Table 3
Ljung–Box *Q*-statistics for the original series

Number of lags	*Q*-stat
12	35.442*
24	44.485*
36	48.181*

* Significant under 95% level of confidence

Table 4
Ljung–Box *Q*-statistics for the residuals of AR(1) models

Number of lags	*Q*-stat	Probability
12	15.154	0.233
24	25.232	0.393
36	32.776	0.623

In the final stage, we used the model coefficients obtained in the third stage to forecast the future value of oil price changes. Actually this is an ex ante forecast.

Using the 99th percentile obtained in the previous stage, we corrected the future price changes. These corrected forecasts are the VaR estimations. The result of this VaR quantification together with actual price changes is shown in Fig. 4. As can be seen, the estimated VaR is greater than actual price changes for 97.6 percent of the forecast period. This is a similar percentage to the 99th likelihood level, which was expected before estimating VaR.

5.3. *The variance–covariance approach for VaR estimation*

Among the various Variance–Covariance-based models for VaR quantification, Autoregressive Conditional Heteroskedasticity (ARCH) models are relatively the most advanced models. Using these models, we can forecast future variance values by combining past deviations and past values.

In applying HSAF methodology in Section 5.2 we analyzed the stationary and the autocorrelation behavior of the oil price series. We concluded there that although the original series is not stationary, its first difference is stationary. Also we found that price changes show an autocorrelation behavior, so we estimated an AR(1) model.

To estimate VaR through an ARCH scheme it is necessary to determine whether the price changes are suitable for this scheme. Fig. 5 illustrates oil price changes during 1992–2003. As can be seen in this graph, large oil price changes are followed by large changes and small changes are followed by small changes.

Although this suggests an ARCH scheme, we cannot rely only upon this criterion. So the suggested behavior was tested with the use of statistical tools. As

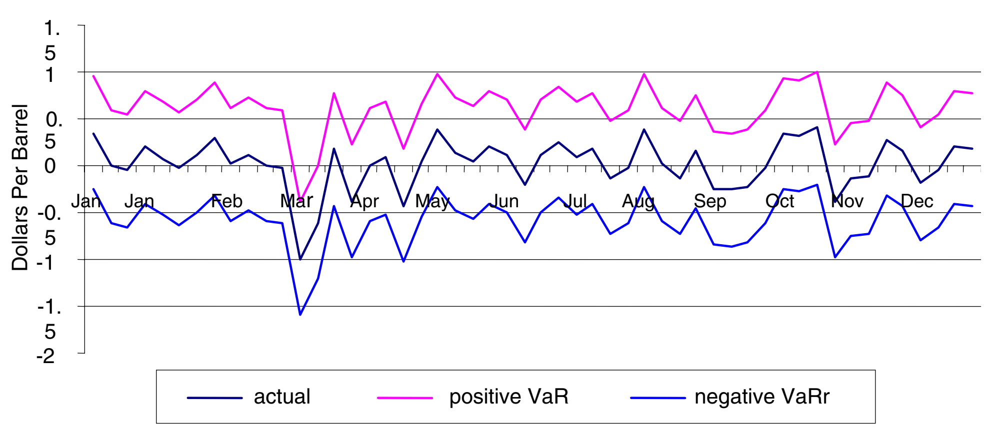


Fig. 4. Estimated VaR through HSAF methodology.

recommended by Enders (2004), using Ljung–Box statistic, we analyzed the autocorrelation behavior of the squared residuals of AR(1) model. Table 5 summarized the Q-Statistic values and their significant level. Results show that autocorrelation is statistically significant. So an ARCH scheme can be used to model the series behavior.

Several ARCH(p) and/or GARCH(p,q) models can be estimated for the analyzed behavior. To determine the best model, we used AIC and SBC model selection criteria. Table 6 reports the calculated values of AIC and SBC criteria for various models. Among them GARCH(1,1) presents the minimum values for both criteria. As Enders (2004) suggests, this model can be selected as the best model among others.

Using the estimated parameters of the GARCH (1,1) model, we forecasted variance values for the "out of sample" period. Also, the forecast obtained from the AR(1) model was used as the price changes of the year 2003.

Assuming that the values of standard deviation have a normal distribution, the corresponding value of the normal standard function for the assumed level of confidence was determined. Then we multiplied this value (2.33) by the forecasted standard deviations for the out of sample period. Finally, VaR was calculated by adding (for the positive returns) and subtracting (for the negative returns) the multiplication results to the return forecasts.

Fig. 6 shows the results of VaR estimation calculated through GARCH and HSAF methodologies. As shown the VaR calculated through HSAF methodology is more efficient. In other words, although the VaR estimated through Variance–Covariance methodology is more than actual price changes in 100% of the forecast period, due to its high variation from actual changes is less reliable than what is estimated through HSAF methodology.

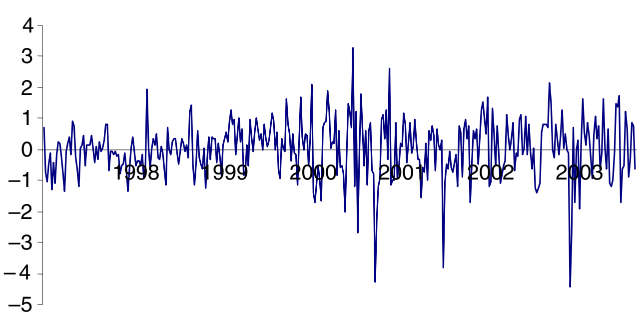


Fig. 5. OPEC weekly oil price changes 1997–2003.

Table 5
Ljung–Box Q-Statistics for the squared residuals of AR(1) model

Number of Lags	Q-stat
12	36.544*
24	60.472*
36	72.175*

* Significant at 95% level of confidence

Table 6
AIC and SBC model selection criteria for various ARCH/GARCH models

Model	AIC	SBC
GARCH(1,1)	2.498	2.546
ARCH(1)	2.584	2.621
ARCH(2)	2.58	2.628
ARCH(3)	2.581	2.64
ARCH(4)	2.579	2.65

6. Conclusions

Risk Management embodies the process and the tools used for evaluating, measuring and managing the various risks within a company's portfolio of financial, commodity and other assets. In energy markets, proper risk management depends not only upon proper portfolio analysis tools but also upon a solid foundation of forward price, volatility and option analysis.

Calls for competition in the power and gas industry, have made deregulation an attractive option around the world. The rise of competition and deregulation in turn has led to relatively free energy markets that are characterized by high price shifts. Within oil markets the volatile oil price environment after OPEC agreements in the seventies requires risk quantification.

Within oil markets, Value-at-Risk can be used to quantify the maximum oil price changes associated with a likelihood level. This quantification constitutes a fundamental point when designing risk management strategies. For this end, the paper proposes to quantify OPEC oil price VaR through various methodologies and to compare the result of VaR calculation through each. We used OPEC weekly oil prices from January 1997 to December 2003 for VaR calculation.OPEC oil price Value-at-Risk is calculated in this paper through Historical Simulation based on ARMA Forecasting (HSAF) and also Variance-Covariance based on GARCH modeling approaches. Results show that if the level of confidence is 99 percent, then the VaR calculated through HSAF methodology is greater than actual price changes in almost 97.6 percent of the forecasting period.We also concluded that although the estimated VaR through Variance–Covariance approach is greater than actual price changes in the whole forecasting period, it is not as efficient as what is calculated through HSAF methodology. Finally the

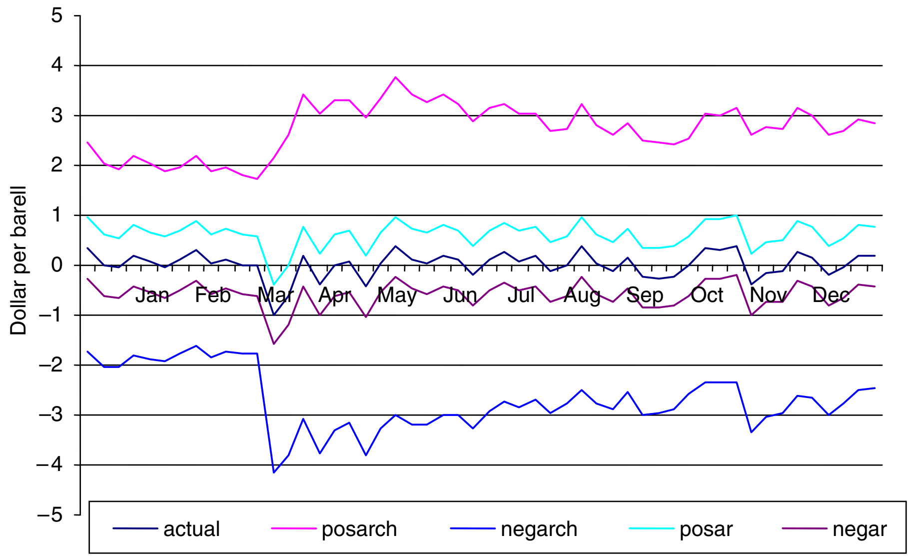


Fig. 6. OPEC oil price value at risk estimation through HSAF and GARCH methodologies.

conclusion is that Value-at-Risk, calculated by any method, is a reliable measure of oil price risk for whoever is concerned with oil price volatility, whether he (she) is a firm manager or a policy maker in the government body.

References

Basek, S., Shapiro, A., 2001. Value-at-risk based risk management optimal policies and asset prices. The Review of Financial Studies 14 (2), 371–405.

Bollerslev, T., 1986. Generalized autoregressive conditional hetero-skedastisity. Journal of Econometrics 31.

Cabedo J. David, Moya Ismael, 2003. Estimating oil price value-at-risk using historical simulation approach. Energy Economics 25, 239–253.

Enders, 2004. Applied Econometric Time Series, second ed. New York.

Energy Information Administration (EIA), 2002. Derivatives and risk management in the petroleum, natural gas and electricity industries.

Engle, R.F., 1982. Autoregressive conditional heteros-kedastisity with estimates of the variance of United Kingdom inflation. Econometrics 50 (4).

Hendricks, D., 1996. Evaluation of value at risk modeling using historical data. Federal Reserve Bank of New York. Economic Policy Review.

Hilton, G.A., 2003. Value-at-risk, Theory and Practice. New York.

Hull, J., White, A., 1998. Incorporating volatility updating into the historical simulation method for VaR. The Journal of Risk 1, 5–19.

Kaushik, V.Ng., Pirrong, C., 1999. Arbitrage-Free Valuation of Energy Derivatives in Managing Energy Price Risk. 2nd ed. Risk Books, London, UK.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Parsons, J., 1998. Alternative Models of Uncertain Commodity Prices for Use with Modern Asset Pricing Methods. Energy Journal 19 (1).

Sadorsky, P., 1999. Oil price shocks and stock market activity. Energy Economics 21 (5).

Wengler John, W., 2001. Managing Energy Risk:A Nontechnical Guide to Markets and Trading. Penn Well Publishing Company.

ELSEVIER

Energy Policy 36 (2008) 627–641

ENERGY POLICY

www.elsevier.com/locate/enpol

A portfolio risk analysis on electricity supply planning

Yun-Hsun Huang[a], Jung-Hua Wu[b,*]

[a]Energy and Environment Laboratory, Industrial Technology Research Institute, Hsinchu 310, Taiwan
[b]Department of Resources Engineering, National Cheng Kung University, Tainan 701, Taiwan

Received 2 May 2007; accepted 1 October 2007
Available online 26 November 2007

Abstract

Conventional electricity planning selects from a range of alternative technologies based on the least-cost method without assessing cost-related risks. The current approach to determining energy generation portfolios creates a preference for fossil fuel. Consequently, this preference results in increased exposure to recent fluctuations in fossil fuel prices, particularly for countries heavily depend on imported energy.

This paper applies portfolio theory in conventional electricity planning with Taiwan as a case study. The model objective is to minimize the "risk-weighted present value of total generation cost". Both the present value of generating cost and risk (variance of the generating cost) are considered. Risk of generating cost is introduced for volatile fuel prices and uncertainty of technological change and capital cost reduction. The impact of risk levels on the portfolio of power generation technologies is also examined to provide some valuable policy suggestions. Study results indicate that replacing fossil fuel with renewable energy helps reduce generating cost risk. However, due to limited renewable development potential in Taiwan, there is an upper bound of 15% on the maximum share of renewable energy in the generating portfolio. In the meantime, reevaluating the current nuclear energy policy for reduced exposure to fossil fuel price fluctuations is worthwhile.

Keywords: Conventional electricity planning; Portfolio theory; Risk-weighted

1. Introduction

Recent escalating global energy prices have caused worldwide concern. The benchmark West Texas Intermediate crude oil price skyrocketed to US$77.23 per barrel in July 2006. This marked a US$46 rise, or a growth of 149.1%, compared to the US$31 per barrel in July 2003. Historically, a stable energy supply has predominated over price. However, rising world energy price has shifted the focus from quantity to fluctuating price risk management. In terms of energy policy, security of supply is enhanced if one relies on several sources whose prices are uncorrelated or negatively correlated. Hence a diverse energy portfolio contributes to energy security. An efficient electricity generation mix helps to mitigate the impact of fossil fuel price shocks and minimize societal risk. Consequently, since renewable energy derives from indigenous sources and does not correlate with fossil fuel price movements, appropriate generation portfolio diversification to include renewable energy would result in reduced exposure to fossil fuel price fluctuations.

The conventional approach to electricity planning applies the least-cost method to select from a range of alternative technologies without assessing cost-related risks. The result is inherent in the bias favoring fossil fuel generating technology. Recent fossil fuel price volatility underlines the potential advantage of including renewable energy sources in the generating portfolio. Furthermore, rapid technological progress has lowered renewable energy cost.

This paper applies portfolio theory in conventional electricity planning with Taiwan as a case study. The overall model objective is to minimize the "risk-weighted present value of total generation cost". This work both considers the present value of generating cost and risk (generating cost variance). Risk of generating cost is introduced for volatile fuel prices and uncertainty of

*Corresponding author. Tel.: +886 6 2096174; fax: +886 6 2380421.
E-mail address: hwaa@mail.ncku.edu.tw (J.-H. Wu).

0301-4215/$ - see front matter
doi:10.1016/j.enpol.2007.10.004

technological change and capital cost reduction. The impact of risk levels on the portfolio of power generation technologies is also examined. Finally, policy suggestions are proposed based on the simulation results for Taiwan electricity sector.

The remainder of the paper is organized as follows. Section 2 provides the literature review; Section 3 introduces Taiwan electricity sector status. Section 4 contains the model description; Section 5 describes data sources and adjustments. Section 6 presents the simulation results from applying the model to the Taiwan electricity sector, and Section 7 is the conclusion.

2. Literature review

Portfolio theory describes how rational investors base their investment decisions in the financial market on optimizing portfolio return for a given risk level (Bodie et al., 2005; Markowitz, 1952). Researchers have recently applied the theory to risk analysis in the energy market. This section provides a short literature review on researches closely related to risk analysis of the energy market, summarized as follows.

Humphreys and McClain (1998) use financial portfolio theory to demonstrate how energy mix consumed in the United States could be chosen given the national goal to reduce risks to the domestic macro-economy of unanticipated energy price shocks. An efficient portfolio frontier is constructed using time-varying variances and covariances estimated with generalized autoregressive conditional heteroskedastic (GARCH)[1] models. Results indicate that while the electric utility industry is operating close to the minimum variance portfolio, a shift towards coal would still reduce overall price volatility for US energy consumption.

Awerbuch and Berger (2003) introduce mean–variance portfolio theory and evaluate its potential application to the development of efficient European Union (EU-15) generating portfolios that enhance energy security and diversification objectives. The portfolio model reflects relevant generating cost stream risk: fuel, operation and maintenance (O&M), and construction period costs. The model illustrates the portfolio effects of different generating mixes. Findings indicate that existing and projected EU-generating mixes are sub-optimal from a risk-return perspective. The analysis suggests that including more renewables in portfolios creates lower cost and risk.

Lesbirel (2004) uses portfolio theory to conceptualize energy security as an insurance mechanism against disruptions to energy import markets. The study provides quantitative measures of systematic and specific risks associated with Japanese energy imports during the period 1970–99. Results conclude that, despite their limitations, portfolio measures provide a much more theoretically and methodologically robust indicator of energy import security than traditional measures.

Zon and Fuss (2005) present an optimum portfolio selection approach integrated in a clay–clay vintage model. The model focuses on fuel price and technological uncertainty as the most important cost volatility determinants in the electricity sector. Results indicate that the cumulative nature of embodied technical change gives rise to uncertain investment responses between the standard results of optimum portfolio theory and real option theory.

Krey and Zweifel (2005) use financial portfolio theory to investigate energy mixes of Switzerland. The efficient frontier is constructed estimating time-varying variances and covariances in energy prices using GARCH models. Additionally, the TGARCH[2] variant serves to control for excess kurtosis. Results suggest that a shift towards nuclear power and away from natural gas and gasoline would reduce both expected increase in the Swiss energy bill and its price volatility.

Research on risk analysis in the energy market is still rare in Taiwan. Only one such study on electricity market has been conducted. Chang and Chen (2005) apply mean–variance portfolio theory to examine electricity generation options currently available to the Taiwan electricity sector and related issues. The model only focuses on fuel price volatility. Model results propose an optimal generation portfolio with significantly fewer oil and natural gas power plants than that planned by the stated owned Taiwan Power Company (Taipower). Oil and natural gas prices are generally more volatile than other energies. Reducing the proportion of these two types of energy in the generation portfolio contributes to energy security.

Summing up this review, the conventional least-cost approach for determining generation mix tends to undermine renewable energy benefits. Portfolio theory was applied to energy portfolio analysis in the late 1990s. Research results show that a higher proportion of renewable energy or a lower share of oil and gas reduces portfolio exposure to fuel price fluctuations. Unfortunately, conventional planning models tend to overlook the risk issue.

Different from the above literature, our model integrates portfolio theory into a vintage electricity planning framework. In this model, risk of generating cost is introduced for volatile fuel prices and uncertainty of technological change and capital cost reduction. In addition to risk consideration, the model also accommodates components of conventional electricity planning approach including unit characteristics, load segments, capacity constraints and power output constraints. This allows us to analyze the investment decisions taken for different vintages of power generation technologies based on different energy sources. In short, the model explicitly takes many characteristics of

[1]The GARCH process introduced by Engle (1982) and Bollerslev (1986) allows the error variance to respond to shocks.

[2]The assumption of conditionally t-distributed errors in combination with the GARCH model is called TGARCH model because the t-distribution has higher kurtosis than the normal distribution.

power generation technologies and the risks that accrue from electricity investment into account. The application is based on data for Taiwan electricity sector.

3. Taiwan electricity sector status

3.1. Installed capacity

Taiwan relies heavily on imported energy. Imported energy dependence was 97.85% in 2005 while imported oil was an even higher 99.94%. The Taiwan electricity sector capacity in 2005 is shown in Table 1. Total installed capacity by year-end 2005 was 43,142.3 MW with 28,643.7 MW from Taipower, 7234.9 MW from independent power producers (IPP), 7013.6 MW from self-generating power systems (including co-generation systems and biomass generation systems), and 250.1 MW from commissioned hydro power plants. Installed capacity distributable by the Taipower electricity system is 36,128.7 MW (excluding self-generating sources), accounting for 83.8% of total installed capacity.

Among those, 16,870.1 MW, or 39.1%, was from coal-fired; 11,099.7 MW, or 25.7%, was from LNG-fired; 5144 MW, or 11.9%, was from nuclear; 4749.1 MW, or 11.0%, was oil-fired; 2602 MW, or 6.1%, from pumped storage hydro; 2538.8 MW, or 5.9%, from renewable energy (including conventional hydro power plant), and 138.6 MW, or 0.3% from waste heat recovery and other energy sources.

3.2. Power generation

The total power generation of Taiwan's electricity sector in 2005 is shown in Table 2. Total power generation in 2005 was 219,460 GWh with 140,480 GWh produced by Taipower, accounting for 64.0% of total output, followed by 42,460 GWh from self-generating power systems (19.4%), 35,850 GWh from IPP (16.3%), and 670 GWh from commissioned hydro power plants (0.3%).

Among those, 115,870 GWh was from coal-fired, accounting for 52.8% of total electricity generation, 38,400 GWh from nuclear (17.5%), 37,450 GWh from LNG-fired (17.1%), 15,760 GWh from oil-fired (7.2%), 7560 GWh from renewable energy (3.4%), 3830 GWh from pumped storage hydro (1.7%), and 590 GWh from waste heat recovery and other energy sources (0.3%).

Thermal power and nuclear power are the current dominant sources of electricity supply in Taiwan, accounting for 87.7% of total installed capacity. However, further expansion of capacity in nuclear is restricted by the "Nuclear-Free Home" policy[3] and concerns regarding the disposal of radioactive waste from nuclear power plants in Taiwan. Economic growth continues to drive power consumption; therefore, power expansions must meet growing demand. The correlation between renewable energy costs and fossil fuel price is relatively low. Hence in addition to increase power supply from thermal power plants, extending the proportion of renewable energy technologies in the generation portfolio will help diversify price fluctuation risk of imported fossil fuel.

4. Model description

4.1. Electricity demand and load duration curve

Load demand is total power generation output less transmission and distribution loss, or the total sum of electricity consumption. The electricity system load curve represents changing output or load values as a time function. Load curve shape is dependent on many factors, such as economic development, climatic conditions, and electricity usage habits. The load curve can convert into the load duration curve by reordering the load data in descending order of magnitude. The load duration curve provides a useful yearly summary (or one period) of hourly fluctuations in electricity demand (Madlener et al., 2005). A discretized load duration curve shown in Fig. 1 divides the load demand into base-load, medium-load, and peak-load demand in the study.

P_1, P_2, P_3, and P_4 represent power technologies allocated to meet certain load sections and average actual power output during the designated time period, and θ_1, θ_2, θ_3 are the length of time the power plants are utilized. Base-load power plants are generally suitable for long time operating and are in use during the entire period (i.e. $\theta_1+\theta_2+\theta_3$), contributing P_1 of power. Peak-load power plants can be easily adjusted to unexpected fluctuations in electricity demand but are more expensive technologies. They operate during the period θ_1 (for low utilization rate), with contribution P_4 and locate at the top of the load duration curve. The load duration curve also illustrates utilization of varying generation power plants. The load duration curve represents electricity demand and power plant utilization in the model.

4.2. Objective function

The "minimization of the risk-weighted present value of total generation cost" is the objective of this model. Both the present value of generating cost and risk (variance of the generating cost) are considered. Generating cost includes energy cost (fuel cost and variable O&M cost) and capacity cost (fixed O&M cost and capital investment cost).[4] Technological change in the model is technological progress embodied in more recent and more efficient capital investment. In other words, production efficiency is driven solely by additional capital investment. Capital is not homogenous and includes investments made in

[3]The "Nuclear-Free Home" policy is the ultimate goal in achieving a non-nuclear homeland in Taiwan. The implementing strategies include banning the development of nuclear weapons, gradually phasing out nuclear power and developing renewable energy to meet future needs.

[4]Investments are viewed as sunk cost; that is, they cannot be recovered. Similarly, the investment costs of existing plants are sunk and thus irrelevant for the present decision.

Table 1
Total installed capacity in Taiwan in 2005

Item	Taipower	IPP	Self-generating power systems	Commissioned power plant	Total
Coal-fired	8650.0	3097.1	5123.0	–	16,870.1 (39.1%)
LNG-fired	6972.4	4120.0	7.3	–	11,099.7 (25.7%)
Nuclear	5144.0	–	–	–	5144 (11.9%)
Oil-fired	3608.7	–	1140.4	–	4749.1 (11.0%)
Pumped storage hydro	2602.0	–	–	–	2602 (6.1%)
Renewable energy	1666.6	17.8	604.3	250.1	2538.8 (5.9%)
Waste heat recovery and other energy	–	–	138.6	–	138.6 (0.3%)
Total	28,643.7 (66.4%)	7234.9 (16.8%)	7013.6 (16.2%)	250.1 (0.6%)	43,142.3

Source: Taiwan Power Company (Taipower) (2006).
Unit: MW.

Table 2
Total electricity generation in Taiwan in 2005

Item	Taipower	IPP	Self-generating power systems	Commissioned power plant	Total
Coal-fired	62,090	20,600	33,180	–	115,870 (52.8%)
Nuclear	38,400	–	–	–	38,400 (17.5%)
LNG-fired	22,310	15,110	30	–	37,450 (17.1%)
Oil-fired	10,430	–	5330	–	15,760 (7.2%)
Pumped storage hydro	3830	–	–	–	3830 (1.7%)
Renewable energy	3420	140	3330	670	7560 (3.4%)
Waste heat recovery and other energy	–	–	590	–	590 (0.3%)
Total	140,480 (64.0%)	35,850 (16.3%)	42,460 (19.4%)	670 (0.3%)	219,460

Source: Taiwan Power Company (Taipower) (2006).
Unit: GWh.

different vintages. Power plants are categorized by periods to highlight differences in embodied technological change. Mathematical models for generating cost and risk of generating cost are developed in the Appendix.

The "risk-weighted present value of total generation cost" is a result of calculating the present value of total generating cost and the risk of generating cost

$$\mathrm{RW(PVTC)} = \mathrm{PVTC} + \lambda \mathrm{Var(PVTC)},$$

λ is the risk-averse parameter. It also stands for the relative contribution of the variance of total generating cost in the objective function. If λ is zero then the risk of generating cost is excluded from the technology portfolio selection. The higher the value, the more risk-averse the investor. The objective function, based on the above, minimizes "risk-weighted present value of total generation cost": Min RW(PVTC).

4.3. Constraints

The following constraints, together with objective function, complete the model formulation.

Constraint 1:

$$\sum_{j=1}^{J}\sum_{v=0}^{t} P_{j,t,v,s}(1-\mathrm{Loss}_t) \geqslant D_{t,s}, \quad s = 1\ldots S, \; t = 1,\ldots,T,$$

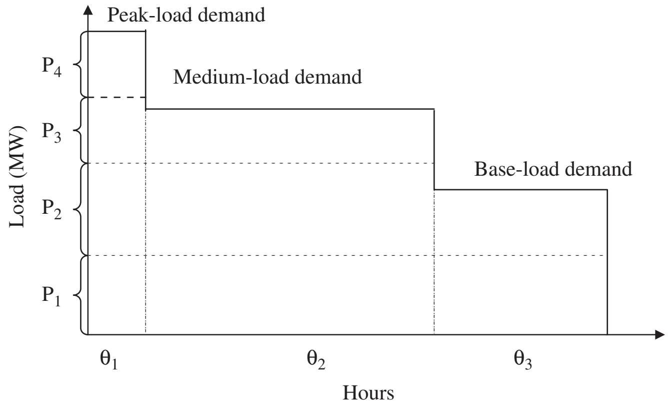


Fig. 1. A discretized load duration curve.

where $D_{t,s}$ is the electricity demand for block s at period t. Loss_t is the distribution and transmission loss at period t. The constraint requires that total generation in each block must be more than or equal to electricity demand during the planning horizon.

Constraint 2:

If technology j is coal or nuclear and s is peak load block (for $s=1$), then:

$$P_{\text{coal},t,v,1}=0, \quad P_{\text{nuclear},t,v,1}=0.$$

Power output is limited by unit characteristics. Nuclear and coal power plants take longer time to start up and shut down and are thus less responsive to sudden power demand changes. The plants are more effective when used continuously to meet base load demand. Therefore, the power output of these two types of plants during peak load block is zero.

Constraint 3:

$$\text{CAP}_{j,t}=\text{CAP}_{j,t-1}+X_{j,t}-\text{Retire}_{j,t}, \quad j=1,\ldots,J, \quad t=1,\ldots,T.$$

$\text{Retire}_{j,t}$ is the amount of installed capacity of technology j retired at period t. This formula calculates cumulative installed capacity of different technology types during each period. Cumulative installed capacity is the sum of cumulative installed capacity of the previous period plus new installed capacity less retired installed capacity.

Constraint 4:

$$\sum_{j=1}^{J}\text{CAP}_{j,t}\geqslant D_{t,1}(1+m), \quad s=1, \quad t=1,\ldots,T,$$

where $D_{t,1}$ is the peak demand (for $s=1$) at period t and m is the reserve margin. The constraint requires that total capacity must meet peak demand (electricity consumption plus reserve margin).

Constraint 5:

(a)

$$P_{j,t,v,s}\leqslant a_j X_{j,v} \quad s=1,\ldots,S, \quad v=0,\ldots,t, \quad t=1,\ldots,T, \quad j=1,\ldots,J$$

(b)

$$\sum_{s=1}^{S}P_{j,t,v,s}\theta_s\leqslant 8760 b_j X_{j,v}, \quad v=0,\ldots,t, \quad t=1,\ldots,T, \quad j=1,\ldots,J,$$

where a_j is availability of a power plant, which refers to the percentage of time that the plant can be used, i.e., it is not out of service due to repairs or maintenance (Kumbaroğlu et al., in press). b_j is the capacity factor, which refers to the ratio between average electricity generation and installed capacity of a given power plant during a specific period of time (e.g. a year). It measures power plant average utilization rate. The constraint requires that output from each plant cannot exceed available capacity.

Constraint 6:

$$\sum_{v=0}^{t}\sum_{s=1}^{S}r_{\text{gas},v}P_{\text{gas},t,v,s}\theta_s\leqslant \text{gasimport}_t, \quad j\in\text{gas}, \quad t=1,\ldots,T,$$

where gasimport_t is the annual supply of liquefied natural gas (LNG) for electricity generation. The amount of LNG-fired to generate electricity shall be less than the amount supplied. Natural gas largely imported in Taiwan implies that the amount of LNG imported cannot exceed the receiving capacity of LNG reception terminals.

Constraint 7:

$$\text{CAP}_{j,T}\leqslant \text{Max} \quad X_j, \quad j\in\text{renewable energy},$$

where $\text{Max}X_j$ is renewable energy development potential. Despite an indigenous energy resource, the capacity of renewable energy technologies is also subject to geographical conditions (Wu and Huang, 2006). Renewable energy capacity must be less than its development potential.

Constraint 8:

$$P_{j,t,v,s}\geqslant 0 \quad s=1,\ldots,S, \;\; v=0,\ldots,V, \;\; t=1,\ldots,T, \quad j=1,\ldots,J \text{ (non-negative constraint).}$$

The full model consists of the objective function (RW (PVTC)) that needs to be minimized, subject to constraints 1–8. Risk of generating cost is focused on volatile fuel prices and uncertainty of technological change and capital cost reduction ($\sigma^2_{\text{FP}},\sigma^2_r,\sigma^2_C$) as the most important cost volatility determinants in the electricity sector. There are no country-specific constraints included in the model formulation. It means that this model could be universally applied to other countries or regions, preparing the necessary data for the model are available. The simulation analysis is based on data for Taiwan electricity sector. In the following, we describe the data sources and adjustments of the model for Taiwan electricity sector. The results associated with these simulations are presented in Section 6, which show how increasing the level of risk aversion would influence the technology portfolio and generating cost.

5. Data sources and adjustments

5.1. Technical data for power plants

5.1.1. Generating cost data

In the model application, we explore electricity planning in Taiwan power sector for the period 2006–2025, differentiating between three types of thermal power plants (i.e. coal-fired, oil-fired, and LNG-fired), nuclear power plants, and six types of renewable energy technologies (i.e. conventional hydro, wind, solar PV, Municipal Solid Waste (MSW), other biomass,[5] and geothermal) of different vintages.

Generating cost, availability, capacity factor, and thermal efficiency data for ten types of power plants are shown in Table 3. The other technical parameters of power plants are summarized in Table 4. Fixed cost includes capital investment cost and fixed O&M cost. Variable cost covers variable O&M cost and fuel cost. Fuel purchase price is not made public as it is power company classified information. Hence, fuel price is based on data released by the Bureau of Energy (average purchase prices for imported energy, 1990–2005). The historical data on fuel price are deflated to year 2001 currency value over the designated period and used to estimate growth rate of the theoretical model function (Table 4). Fuel price estimation in future planned years is made on the basis of fuel price growth rate.

5.1.2. Fuel/output ratio (or fuel consumption rate)

Fuel/output ratio is developed from data released by the Bureau of Energy. The amount of fuel required per unit of power output is calculated by dividing power plant fuel consumption by total electricity generation (Bureau of Energy, Ministry of Economic Affairs, 2006). Fuel/output ratio of MSW is calculated from data released by the Environmental Protection Administration (Environmental Protection Administration (EPA), Executive Yuan, 2006). The fuel/output ratio rate for MSW also applies to other biomass energy due to a lack of data regarding fuel consumption and electricity generation.

The fuel/output ratio for coal-fired power plants in 2005 was 0.389 tons/MWh, oil-fired power plants 0.241 kl/MWh, LNG power plants 0.202 thousand cubic meters/MWh, nuclear power plants 2.502 g/MWh, and biomass power plants 1.969 tons/MWh. Taking LNG for example, Fig. 2 illustrates the change of LNG consumption rate (LNG/power output ratio) between 1990 and 2005. The historical data on fuel/output ratio have been used to estimate growth rate of the theoretical model function (Table 4). Fuel/output ratio estimation in future planned years is made on the basis of fuel consumption growth rate (a negative value implies fuel-saving technological progress).

[5]Other biomass includes industrial waste and agricultural waste.

5.1.3. Capital cost

The learning curve[6] forms the basis of estimated future capital cost in the model. Cumulative installed capacity determines reduction rate in capital cost per unit of installed capital. Only renewable energy technologies are assumed to exhibit learning effects. However, due to cost and cumulative installed capacity data solely available for wind and solar PV technology in Taiwan, estimations are made on the two technologies (Table 4). Taking wind power for example, capital cost is calculated using cumulative installed capacity (x) and unit investment cost (y) (including wind turbines, electric facilities, construction facilities, and grid facilities) data from the Taipower for years 2001–2006. Fig. 3 shows the relationship between cumulative installed capacity and unit investment cost, deflated by the GDP deflator over the designated period. Estimated future target year capital cost is made on the basis of learning rate.

5.1.4. The variance and covariance

The risk of changes in fuel price, fuel/output ratio and capital cost is calculated based on data from the Bureau of Energy and Environmental Protection Administration from 1990 to 2005 (Table 4). In terms of fuel price fluctuation risk, oil is the most risky resource followed by LNG, coal, and nuclear. The fuel price fluctuation risk for conventional hydro, wind, Solar PV, and geothermal technology are zero as no fuel is required. The fuel price fluctuation risk for MSW and other biomass are zero as only waste is involved.

5.2. Expected long-term load and line loss rate

Power load historical data and expected future peak load demand are shown in Table 5. Peak load demand is expected to grow from 30,942.9 MW in 2005 to 63,390.3 MW in 2025 with an average growth rate of 3.65%. The Taipower Department of Planning estimates that line loss rate is projected to fall from 5.6% in 2005 to 5.3% in 2025 thanks to line improvement plans and completion of the Sixth Power Transmission and Substation Project by the end of 2006.

[6]In energy and climate models, the learning curve has been employed with increasing frequency to account for cost reductions due to technology related learning and for endogenizing technological change (Barreto, 2001; Barreto and Kypreos, 2004; Berglund and Söderholm, 2006; IEA, 2000; McDonald and Schrattenholzer, 2001). However, some weakness associated with use of the learning curve should also be acknowledged. The choice of both model specification and estimation technique can have a strong influence on the learning rate estimates (Söderholm and Sundqvist, 2007). Despite uncertainties surrounding the estimated learning curve that stem from model misspecification issues, the learning curve is still a useful tool in modeling cost reductions in energy supply systems and is applied in our model.

Table 3
Cost data, availability, capacity factor and thermal efficiency of power generation technologies

Cost item							
Technologies	Capacity cost		Energy cost		Others		
	Investment cost (NT$/KW)	Annual fixed O&M cost (NT$/KW)	Variable O&M cost (NT$/KWh)	Fuel cost (NT$/KWh)	Availability	Capacity factors	Thermal efficiency
Coal-fired	34,300	694	0.0759	0.75	0.9021	0.9021	0.34
Oil-fired	10,700	292	0.6927	1.93	0.7390	0.7390	0.38
LNG-fired	16,000	254	0.3774	2.03	0.8947	0.8947	0.43
Nuclear	77,100	1716	0.2006	0.15	0.83	0.83	–
Conventional hydro	80,000	1200	0.182	0	0.8	0.5	–
Wind	47,000	650	0	0	0.9	0.31	–
Solar PV	150,000	310	0	0	0.9	0.15	–
MSW	127,400	0	0.41	0	0.8	0.8	–
Other biomass	56,200	0	0.24	0	0.8	0.8	–
Geothermal	70,000	4122	0	0	0.75	0.75	–

Note: (1) 1 NT$ is equivalent to 0.03US$.
(2) The costs are the currency value of year 2001.
(3) Fuel cost is based on year 2005; estimation of fuel costs in other years is based on the growth rate.
(4) Due to lack of data regarding the benefit of burning waste (e.g. disposal fees), the fuel cost of MSW and other biomass is assumed to be zero.
Sources: (1) Taiwan Power Company (Taipower) (2005), (2) Energy and Environment Laboratories (EEL), Industrial Technology Research Institute, 2006.

Table 4
Technical parameters of power generation technologies

Parameters									
Technologies	Initial fuel price	Initial fuel/output ratio	Initial capital cost	Growth rate of fuel price	Growth rate of fuel/output ratio	Learning rate	Variance for fuel price growth (σ_{FP}^2)	Variance for fuel consumption rate (σ_r^2)	Variance for capital cost reduction (σ_C^2)
Coal-fired	1928.56	0.3894	34.3	0.0115	−0.00109	0	0.023241	0.003249	0
Oil-fired	8018	0.2410	10.7	0.0145	−0.00263	0	0.041806	0.000363	0
LNG-fired	10,050	0.2023	16	0.0084	−0.01883	0	0.025784	0.000966	0
Nuclear	56.87	2.5018	77.1	0.0072	−0.01374	0	0.010600	0.002303	0
Conventional hydro	0	0	80	0	0	0	0	0	0
Wind	0	0	47	0	0	0.05	0	0	0.000704
Solar PV	0	0	150	0	0	0.15	0	0	0.009409
MSW	0	1.9686	127.4	0	−0.05675	0	0	0.001662	0
Other biomass	0	1.9686	56.2	0	−0.05675	0	0	0.001662	0
Geothermal	0	0	70	0	0	0	0	0	0

Note: (1) Initial represents year 2005.
(2) The unit of initial fuel price: coal (NT$/tons), fuel oil (NT$/kl), LNG (NT$/thousand cubic meters), uranium (NT$/g).
(3) The unit of initial fuel/output ratio: coal (tons/MWh), fuel oil (kl/MWh), LNG (thousand cubic meters/MWh), uranium (g/MWh).
(4) The unit of initial capital cost: millions NT$/MW.
Sources: (1) Bureau of Energy, Ministry of Economic Affairs (2006), (2) Environmental Protection Administration (EPA), Executive Yuan (2006).

5.3. Retirement schedule of existing power plants

The Taipower expects its nuclear power plants and coal-fired power plants to run for 40 years, oil-fired and LNG-fired power plants for 30 years, and hydro power plants for 70 years. Total installed retirement capacity between 2006 and 2025 is expected to reach 11,245 MW.

5.4. LNG import limitations

The LNG reception terminal receiving capacity is shown in Table 6. Yong-an is currently the only LNG reception terminal in Taiwan with a receiving capacity of 7440 thousand tons per annum. Given the implementation of scheduled addition and expansion projects, its receiving

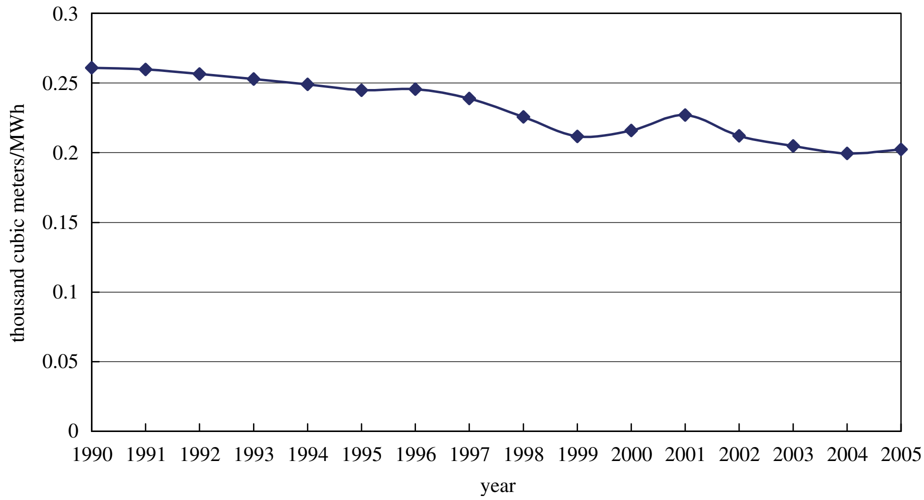


Fig. 2. The change of LNG consumption rate between 1990 and 2005.

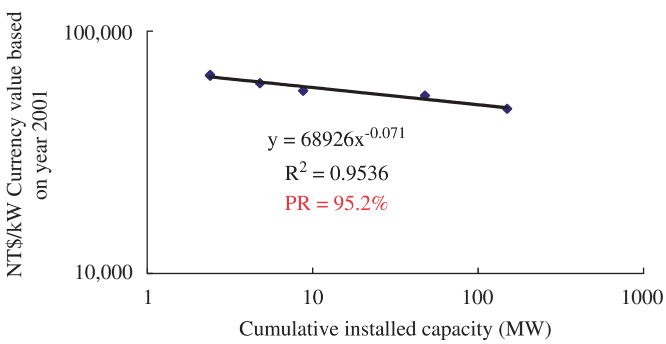


Fig. 3. Learning curve of wind power in Taiwan.

Table 5
Expected long-term load and line loss rate

Year	Peak load demand (MW)	Growth rate (%)	Line loss rate (%)
2005	30,942.9	–	5.6
2006	31,615.8	2.17	5.6
2007	32,833.9	4.01	5.6
2008	34,331.2	4.40	5.5
2009	35,954.4	4.73	5.5
2010	37,724.1	4.92	5.5
2011	39,472.7	4.64	5.5
2012	41,229.9	4.45	5.5
2013	42,925.3	4.11	5.5
2014	44,643.1	4.00	5.5
2015	46,315.9	3.75	5.5
2016	48,016.7	3.67	5.5
2017	49,698.2	3.50	5.4
2018	51,406.5	3.44	5.4
2019	53,139.5	3.37	5.4
2020	54,837.0	3.19	5.4
2021	56,530.1	3.09	5.4
2022	58,230.9	3.01	5.4
2023	59,914.4	2.89	5.3
2024	61,632.8	2.87	5.3
2025	63,390.3	2.85	5.3

Source: Taiwan Power Company (Taipower) (2005).

Table 6
Receiving capacity of LNG reception terminal

Year	Receiving capacity			LNG consumption (Policy target)
	Yong-an terminal	Taichung terminal	Total	
2005	7440	–	7440	–
2006	9000	–	9000	–
2010	9000	3000	12,000	13,000
2020	9000	9000	18,000	16,000–18,000
2025	–	–	–	20,000–22,000

Source: Bureau of Energy, Ministry of Economic Affairs (2005).
Unit: thousand tons per annum.

capacity should reach 12,000 thousand tons per annum in 2010, 18,000 thousand tons in 2020, and 20,000–22,000 thousand tons in 2025.

5.5. *Development potential of renewable energy*

The capacity of renewable energy is subject to geographical conditions despite being an indigenous energy resource. Renewable energy development potential estimation in Taiwan is made based on data from the New Energy and Clean Energy Research and Development Planning Report (Bureau of Energy, Ministry of Economic Affairs, 1999) shown in Table 7.

5.6. *Other parameters*

Reserve margin rate is set at 15%; discount rate is constant at 5%, and planned period is 20 years, from 2006 to 2025.

6. Results

The model is programmed in GAMS (Brooke et al., 2005), and results have been obtained with the solver MINOS.

Table 7
The development potential of renewable energy in Taiwan

Technologies	The development potential of renewable energy
Conventional hydro	Total development potential is estimated to be 4.12 GW (excluding the 0.82 GW development potential in national parks)
Wind	Total development potential for onshore wind power is estimated to be 4 GW. However, after considering the issues such as the high cost of land acquisition only one-fourth, or 1.0 GW, of this potential is expected to be utilized to generate electricity. Together with 2.0 GW from offshore sources total development potential for wind is 3 GW
Solar PV	Total development potential is estimated to be 12 GWp including those installed at residential installations, industrial and commercial installations, public facilities, and other places
MSW	Total development potential is estimated to be 0.9 GW based on the amount of urban waste
Other biomass	Total development potential is estimated to be 0.8 GW based on the amount of industrial waste and agricultural waste
Geothermal	Total development potential is estimated to be 0.5 GW based on data from the 25 non-volcanic geothermal areas

Source: Bureau of Energy, Ministry of Economic Affairs (1999).

6.1. Scenario design

The simulation scenarios in this paper are as follows:

Scenario 0 (C0) is the baseline scenario that examines annual generating portfolio changes under the least-cost principle and relevant constraints without considering risk impact. Scenario 1–scenario 4 (C1–C4) explore behavioral differences of investment exhibiting varying levels of risk-aversion by setting the risk-averse parameter λ at 0.001, 0.0025, 0.005, and 0.0075. The impact on technology portfolio and generating cost is also examined. As Taiwan has limited energy resources, 97.85% of its energy supply is imported. The prices of import energies are highly related to international energy markets. Energy security is reduced when holding inefficient energy portfolios that are exposed to fossil fuel price fluctuations. Hence, scenario 5 (C5) primarily exploits the impact of drastic price fluctuation (high variance) on power generating portfolio. Scenario 5 sets fossil fuel price variance at twice its original value when the risk-averse parameter (λ) is 0.001 to illustrate fossil price fluctuation impact on generating portfolio.

Nuclear power plant capacity expansion is limited given the "Nuclear-Free Home" policy, except for the fourth nuclear power plant currently under construction. Taiwan electricity planning models largely set the upper bounds of generating capacity for nuclear energy at 7844 MW (installed capacity of the fourth nuclear power plant included). However, this paper applies an upper bound of 20% of total installed capacity to nuclear energy to assess nuclear energy significance in a generating portfolio that accounts for risk of generating cost. In order to avoid unrealistically high growth rates for power generation technologies, the maximum annual capacity addition limit for thermal power plants (coal-fired, oil-fired, and LNG-fired) is set at 2000 MW and for renewable energy is set at 500 MW.

6.2. Simulation results

Fig. 4 illustrates the proportion of installed capacity by technology during the planning horizon. Minimizing generating cost is the objective in this scenario. Nuclear power plants gained precedence over other types of generating technology due to low cost. However, there is an upper bound of 20% to nuclear energy share as a proportion of total installed capacity. Hence base load demand is mainly handled by coal-fired power plants suitable for operating for a long time. Coal-fired power plant installed capacity maintains at the 40% level. Oil-fired plants have high generating costs, and their installed capacity fell as no new additions were made to existing plants during the simulation period. LNG-fired power plants are more responsive and have lower fixed cost, and are typically used to meet peak power demand. Installed capacity proportion of LNG-fired plants initially fell as no new additions were made to existing plants. Capacity subsequently rose as new plants were added to meet increasing peak power demand during the simulation period.

Wind power, other biomass, and hydropower are currently the only economically viable renewable energy technologies. New additions have been made to these three types of power plants. Nevertheless, given the objective of minimizing generating cost, more installed capacity will be added in the future should power demand and fossil fuel cost continue to rise. Fig. 4 also shows that new wind power plants will only be added post period 5. Solar PV is uncompetitive due to high generating cost. No new additions were made during the simulation period. As for geothermal power plants, a small amount of installed capacity is added towards the end of the simulation period. Yet its share of total installed capacity was negligible due to only 500 MW development potential in Taiwan. In short, the decision to add new power plants largely depends on generating cost and unit characteristics, without accounting for risk. Coal-fired, LNG-fired, and nuclear power plants take up the lion's share of installed capacity.

Fig. 5 depicts the relationship between total generating cost and risk-averse parameter. Portfolio theory asserts that the more risk- averse an investor, the higher cost to be paid. Fig. 5 also shows that the more risk- averse the generating portfolio, the higher the generating cost involved. The simulation results correspond to those obtained with portfolio theory.

Figs. 6–9 illustrate the impact of different risk-averse parameters (C1–C4) on generating portfolio using coal-fired power plants, LNG-fired power plants, wind power plants, and solar PV power plants as examples. The results

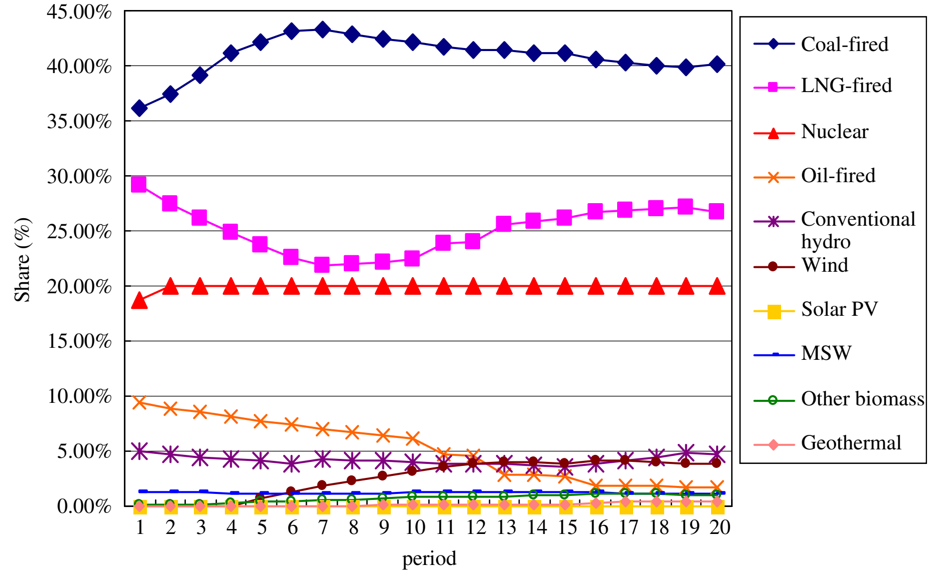


Fig. 4. Share of installation capacity by technology in baseline scenario.

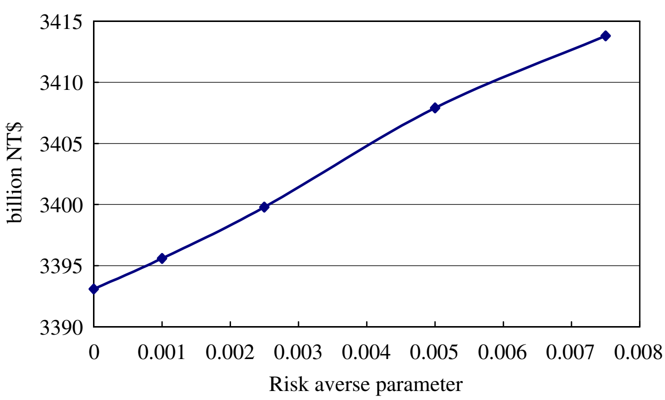


Fig. 5. The relationship between risk-averse parameter and total generating cost.

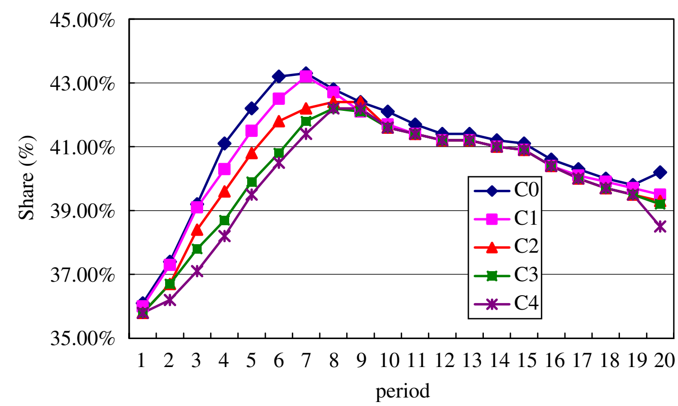


Fig. 6. Changes in the share of installed capacity of coal-fired power plant at different risk levels.

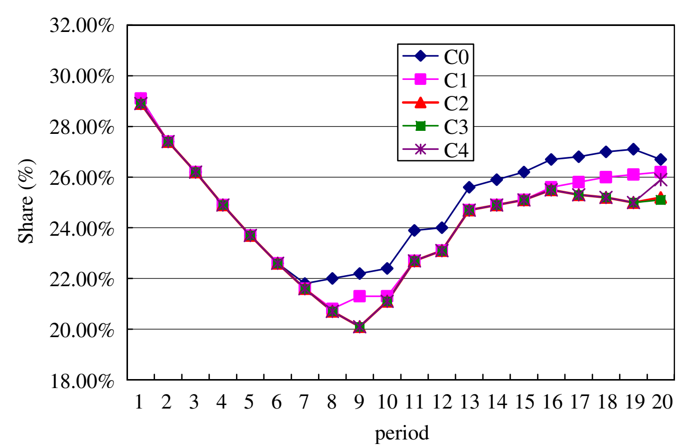


Fig. 7. Changes in the share of installed capacity of LNG-fired power plant at different risk levels.

show while considering generating cost risk, the proportion of high-risk fossil fuel technology reduces and is more likely replaced by less risky renewable alternatives. The effect is particularly pronounced with risk-aversion level increase.

In terms of fuel price fluctuation risk, oil is the most risky energy followed by LNG, coal, and nuclear. Fig. 6 shows that the larger the risk-averse parameter during the initial simulation period (period 1–8), the lower the share of coal-fired power plants. No additional capacity was made to LNG-fired power plants during this period, so new renewable energy plants replaced coal-fired power plants at this stage. Replacement ratio rises with risk-aversion level. However, replacement ratio is limited as geographical conditions dictate renewable energy development capacity. Because the fuel price fluctuations associated with LNG-fired are more than coal-fired, in the latter half of the simulation period (after period 9) LNG-fired power plants are replaced by new renewable energy plants. Fig. 7 also shows that the larger the risk-averse parameter, the lower the LNG power plant proportion in the generating portfolio.

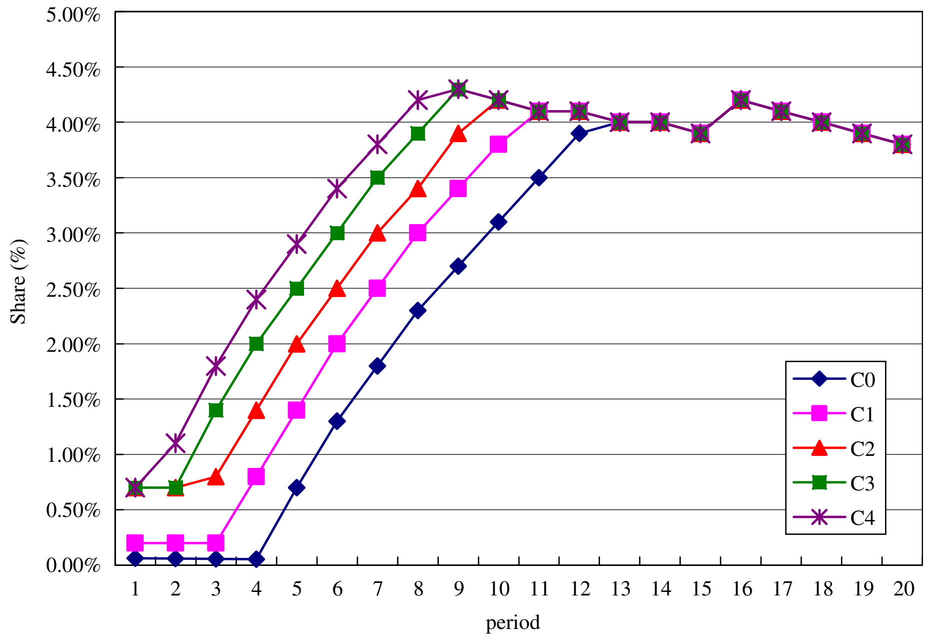


Fig. 8. Changes in the share of installed capacity of wind power plant at different risk levels.

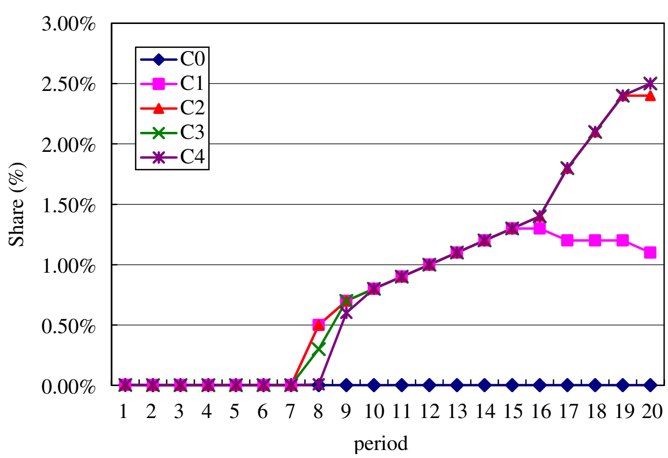


Fig. 9. Changes in the share of installed capacity of solar PV power plant at different risk levels.

Fig. 8 shows that the share of wind power plants increases with the risk-aversion level. Consequently, additional installed wind power capacity made in advance serves as a fossil fuel technology substitute. The additional installed capacity occurs in the baseline scenario in simulation period 5, while this in C4 takes place in period 1. The model sets an upper bound of 3000 MW, given that wind energy development capacity is subject to geographical conditions in Taiwan. Wind power share is identical across scenarios once the upper bound (3000 MW) is reached. Fig. 9 illustrates proportion changes of installed capacity from solar PV technologies. Solar PV as a generating technology is not economically under least-cost principle. With risk consideration, solar PV should be employed when the level of risk-aversion is becoming higher and all other forms of renewable energy have been exhausted. Similarly, the higher the risk averse parameter level, the higher the share of solar PV.

Fig. 10 shows percentage deviation of installed capacity between C5 and C1. The risk-averse parameter is 0.001 in C5 and C1, but fossil fuel price variance of C5 is twice C1's value. Installed capacity proportion of coal-fired and LNG-fired power plants is drop when fuel price fluctuates dramatically. Wind power and hydropower supplies make up the shortfall. The solar PV proportion rose as a fossil fuel substitute toward the end of the simulation period as other renewable energy sources were exhausted.

Oil-fired power plants have higher generation costs and are thus subject to a much higher level of fuel price fluctuation risk. Possible new additions to existing power plants are zero while considering cost-related risk. Therefore, the proportion of installed capacity of oil-fired power plants will decrease each year as illustrated in the baseline scenario. Fuel price volatility of nuclear is less frequent than fossil fuel and less affected by global energy price fluctuations. Nuclear energy could be regarded as a ‘‘quasi-indigenous’’ type of energy due to the relatively small size of fuel material and transportation convenience. Simulation results show that installed capacity proportion generated by nuclear energy be maintained at its upper bound despite the risk involved. Nuclear energy proportion in the generating portfolio does not decrease with rising risk aversion. Nuclear can also be averse to fossil fuel price fluctuations such as LNG and coal. Therefore, reducing the proportion of installed capacity from coal-fired and LNG-fired plants and adding renewable energy and nuclear power will lead to fewer generating cost fluctuations.

In conclusion, taking risk of generation cost into account in electricity supply planning creates a preference for more

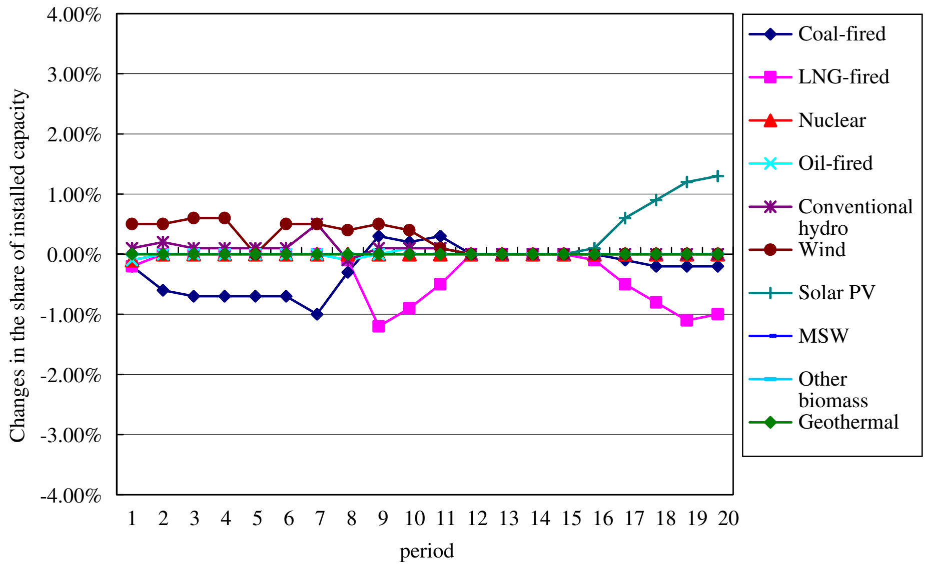


Fig. 10. Percentage deviation of installed capacity between C5 and C1.

risk-averse renewable energy in the generating portfolio and also helps reduce portfolio exposure to fossil fuel price fluctuations. However, producing the same power generation level requires more land resources compared to conventional generating technologies given the relatively lower energy density of renewable energy sources. Renewable energy development potential is also subject to geographical condition constraints. Simulation results illustrate that even in highly risk-averse circumstances, the maximum proportion of installed capacity from renewable energy sources of total installed capacity is 15% due to limited development potential in Taiwan. Looking forward, renewable energy must be able to utilize future technological progresses to improve energy density (e.g. more capacity for a given wind turbine or better energy conversion efficiency of solar cell) and reduce the amount of land resources required for each unit of energy generated. Improved energy density will translate into a higher development potential. Only then will renewable energy sources play a greater role in electricity supply. Meanwhile, reevaluating Taiwan's current nuclear energy policy could be worthwhile. The low generating cost and minor fuel price volatility of nuclear power will reduce exposure to price fluctuations in fossil fuel.

7. Conclusion

Conventional electricity planning applies the least-cost method to select from a range of alternative technologies without assessing cost-related risks. The current approach to determining electricity generation portfolios is in the bias favoring fossil fuel. Taiwan relies heavily on imported energy. The dependence level on imported energy was 97.85% in 2005 while the dependence level on imported oil was an even higher 99.94%.

This paper applies portfolio theory in conventional electricity planning with Taiwan as a case study. The overall model objective is to minimize the "risk-weighted present value of total generation cost". Risk of generating cost is focused on volatile fuel prices and uncertainty of technological change and capital cost reduction as the most important cost volatility determinants in the electricity sector. The current work also examines the impact of risk levels on the technology portfolio and generating cost. Simulation results from applying the model to the Taiwan electricity sector find that the more risk-averse a generating portfolio, the higher the total generating cost. Moreover, incorporating the risk concept in the planning model also leads to a lower proportion of electricity generated from fossil fuel technology and a higher proportion from renewable energy source in substitute for fossil fuel. The higher the risk-aversion level, the more pronounced the effect.

Simulation results also show that replacing fossil fuel with renewable energy helps reduce generating cost risk. However, there is an upper bound of 15% on the maximum share of renewable energy in the generating portfolio due to limited renewable development potential in Taiwan. Renewable energy will only play a more significant role in electricity generation should its energy density improve in terms of unit capacity and conversion efficiency. In the meantime, reevaluating the current nuclear energy policy for reduced exposure to fossil fuel price fluctuations could be worthwhile.

The model adopted in this paper is a preliminary research with some meaning results. Model results could be made more feasible and practical given more relevant data for detailed parameters examination. Furthermore, this study explores the expected cost and risk of electricity generation without considering greenhouse gas emission constraint. Therefore, model results exhibit a preference for coal-fired energy sources to replace more risky technology such as LNG-fired power plants. The Kyoto Protocol commits signatory countries to cut back on greenhouses gas emissions. Future research could cover global warming issues and give consideration to greenhouse gas emissions in this model.

Acknowledgments

We wish to thank the National Science Council (Project NSC 96-2415-H-006-014) for financial support. The authors also appreciate Dr. J.Y. Hsu (Professor, Department of Applied Economics, National Chung Hsing University) for his valuable comments on the earlier version of this paper. Appreciation is extended to the editor and anonymous referees for their valuable suggestions regarding the manuscript.

Appendix

Mathematical models for generating cost and risk of generating cost are derived as follows:

A.1. Generating cost

A.1.1. Fuel cost

$$\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{FP}_{j,t}r_{j,v}P_{j,t,v,s}\theta_s,$$

where parameter j is generation technology $(1,\ldots,J)$ or type of fuel used to generate electricity since technology in this paper is categorized by type of fuel; t is planning period $(1,\ldots,T)$; v is vintage $(0,\ldots,V)$ with 0 for existing power plants and $1,\ldots,V$ for newly installed power plants during the planning period; s is the block $(1,\ldots,S)$ formed by the time axis on the load duration curve.

$\mathrm{FP}_{j,t}$: unit price of fuel j at period t;
$r_{j,v}$: amount of fuel j required for a given level of power output from power plants installed in vintage v;
$P_{j,t,v,s}$: power output of technology j from power plants installed in vintage v during period t in block s;
θ_s: duration time of blocks.

where $r_{j,v}$ is the fuel/output ratio, and it means the amount of fuel required per unit of power output. Assuming that fuel consumption rate per unit of power output is influenced by vintage only, the higher the generation efficiency of new power plants, the less fuel required. Thus the technological change of fuel consumption rate $(r_{j,v})$ takes place at a given exponential growth rate and with a residual:

$$r_{j,v} = r_{j,v-1}\mathrm{e}^{\hat{r}_j+\varepsilon_v^{r,j}},$$

$$r_{j,v-1} = r_{j,v-2}\mathrm{e}^{\hat{r}_j+\varepsilon_{v-1}^{r,j}},$$

$$\vdots$$

$$r_{j,1} = r_{j,0}\mathrm{e}^{\hat{r}_j+\varepsilon_1^{r,j}}.$$

Through continuous iteration obtains $r_{j,v} = r_{j,0}\mathrm{e}^{\hat{r}_j\cdot v+\sum_{v=1}^{v}\varepsilon_v^{r,j}}$, where $r_{j,0}$ is the initial fuel/output ratio of fuel j; $\varepsilon^{r,\,j}$ is the residual; $\hat{r}_j$ is constant growth rate. The negative value of the growth rate means that fuel/output ratio decreases as vintage increases (fuel-saving technological progress).

By analogy, if $\mathrm{FP}_{j,t}$ is the unit price of fuel during different periods (t), $\mathrm{FP}_{j,t}$ also takes place at a given exponential growth rate with a residual:

$$\mathrm{FP}_{j,t} = \mathrm{FP}_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j\cdot t+\sum_{t=1}^{t}\varepsilon_t^{\mathrm{FP},j}},$$

where $\mathrm{FP}_{j,0}$ is the initial price of fuel j; $\varepsilon^{\mathrm{FP},j}$ is the residual; $\widehat{\mathrm{FP}}_j$ is constant growth rate. The positive value of the growth rate means that fuel price will increase over time.

A.1.2. Variable O&M cost

$$\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{VAROM}_{j,t,v}P_{j,t,v,s}\theta_s,$$

where $\mathrm{VAROM}_{j,t,v}$: variable O&M cost of technology j of power plants installed in vintage v during period t.

A.1.3. Fixed O&M cost

$$\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{FIXOM}_{j,t}\mathrm{CAP}_{j,t},$$

where $\mathrm{FIXOM}_{j,t}$ is the fixed O&M cost of technology j of power plants installed in vintage v during period t; $\mathrm{CAP}_{j,t}$ the cumulative installed capacity of technology j at period t.

A.1.4. Capital investment cost

$$\sum_{j=1}^{J}\sum_{t=1}^{T}C_{j,t}X_{j,t},$$

where $C_{j,t}$ is the capital cost per unit of capacity of technology j at period t; $X_{j,t}$ is the new installed capacity[7] of technology j at period t.

Future capital cost estimates are established by the one-way learning curve. Investment cost reductions per unit installed capacity is dependent on cumulative installed capacity. The residual is also included in the learning curve and expressed as follows:

$$C_{j,t} = C_{j,0}(n_t)^{\alpha}\mathrm{e}^{\varepsilon_t^{C,j}} = C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}\mathrm{e}^{\varepsilon_t^{C,j}},$$

[7]For simplicity, we assume that capacity is perfectly divisible.

where $C_{j,0}$ is the initial capital cost for technology j; ε^{C_j} is the residual; n_t is the cumulative installed capacity; and α is the learning parameter.

All error terms ($\varepsilon^{r,j}$, $\varepsilon^{\mathrm{FP},j}$, $\varepsilon^{C,j}$) are assumed to have zero expectation, constant variance, and serially uncorrelated. By assumptions, we have ($E(\varepsilon^{x,j})=0$), $E((\varepsilon^{x,j})^2)=\sigma_x^2$, $E(\varepsilon_w^{x,j}\varepsilon_{w-1}^{x,j})=0$ for $x=r$, FP, C; $w=v$, t.

Based on the above, the present value of total cost (PVTC) of generation is expressed as follows:

$$\begin{aligned}\mathrm{PVTC} =&\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,t}r_{j,v}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{VAROM}_{j,t,v}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)\mathrm{FIXOM}_{j,t}\mathrm{CAP}_{j,t}+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)C_{j,t}X_{j,t},\end{aligned}$$

where $\mathrm{DF}(t)$ is the discount factor, $\mathrm{DF}(t)=(1+R)^{-t}$.

A.2. The risk of generating cost (the variance of generating cost)

By adding all residuals into the model, then PVTC can be rewritten as

$$\begin{aligned}\mathrm{PVTC} =& \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j t+\sum_{t=1}^{t}\varepsilon_t^{\mathrm{FP},j}}\\&\times r_{j,0}\mathrm{e}^{\hat{r}_j v+\sum_{v=1}^{v}\varepsilon_v^{r,j}}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{VAROM}_{j,t,v}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)\mathrm{FIXOM}_{j,t}\mathrm{CAP}_{j,t}\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}\mathrm{e}^{\varepsilon_t^{C,j}}X_{j,t}\end{aligned}$$

$$\mathrm{e}^u=1+u+\frac{u^2}{2!}+\frac{u^3}{3!}+\cdots,\quad \text{when } u \text{ is minimal},\quad \mathrm{e}^u\approx 1+u.$$

A first-order approximation of PVTC is given by

$$\begin{aligned}\mathrm{PVTC} \approx& \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j t}\left(1+\sum_{t=1}^{t}\varepsilon_t^{\mathrm{FP},j}\right)\\&\times r_{j,0}\mathrm{e}^{\hat{r}_j v}\left(1+\sum_{v=1}^{v}\varepsilon_v^{r,j}\right)P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{VAROM}_{j,t,v}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)\mathrm{FIXOM}_{j,t}\mathrm{CAP}_{j,t}\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}(1+\varepsilon_t^{C,j})X_{j,t}.\end{aligned}$$

Suppose $R_t^{1,j}=\sum_{t=1}^{t}\varepsilon_t^{\mathrm{FP},j}$, $R_v^{2,j}=\sum_{v=1}^{v}\varepsilon_v^{r,j}$ (where $\varepsilon_0^{r,j}=0$ (error term for the initial period is zero), hence $R_0^{2,j}=0$), $R_t^{3,j}=\varepsilon_t^{C,j}$.

Then

$$\begin{aligned}\mathrm{PVTC} \approx& \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,0}r_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j t+\hat{r}_j v}\\&\times P_{j,t,v,s}\theta_s(1+R_t^{1,j}+R_v^{2,j})\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{VAROM}_{j,t,v}P_{j,t,v,s}\theta_s\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)\mathrm{FIXOM}_{j,t}\mathrm{CAP}_{j,t}\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}X_{j,t}(1+R_t^{3,j})\\&(\text{Note}: R_t^{1,j}R_v^{2,j}\approx 0).\end{aligned}$$

Because $\mathrm{Var}(\mathrm{PVTC})=E[\mathrm{PVTC}-E(\mathrm{PVTC})]^2=E(\varepsilon^{\mathrm{PVTC}}\varepsilon^{\mathrm{PVTC}})$ Where $\varepsilon^{\mathrm{PVTC}}=\mathrm{PVTC}-E(\mathrm{PVTC})$.

From the above assumption ($E(\varepsilon^{\mathrm{FP},j})=0$, $E(\varepsilon^{r,j})=0$, $E(\varepsilon^{C,j})=0$), we can calculate $E(\mathrm{PVTC})$, using the PVTC and $E(\mathrm{PVTC})$, then:

$$\begin{aligned}\varepsilon^{\mathrm{PVTC}} \approx& \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,0}r_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j t+\hat{r}_j v}\\&\times P_{j,t,v,s}\theta_s(R_t^{1,j}+R_v^{2,j})\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}\mathrm{DF}(t)C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}X_{j,t}R_t^{3,j}.\end{aligned}$$

Suppose

$$B_t^{j,v}=\sum_{s=1}^{S}\mathrm{DF}(t)\mathrm{FP}_{j,0}r_{j,0}\mathrm{e}^{\widehat{\mathrm{FP}}_j t+\hat{r}_j v}P_{j,t,v,s}\theta_s,$$

$$B_t^{3,j}=\mathrm{DF}(t)C_{j,0}\left(\sum_{v=1}^{t}X_{j,v}\right)^{\alpha}X_{j,t},$$

$$\begin{aligned}\varepsilon^{\mathrm{PVTC}} \approx& \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}B_t^{j,v}(R_t^{1,j}+R_v^{2,j})+\sum_{j=1}^{J}\sum_{t=1}^{T}B_t^{3,j}R_t^{3,j}\\=&\sum_{j=1}^{J}\sum_{t=1}^{T}R_t^{1,j}\sum_{v=0}^{t}B_t^{j,v}+\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{v=0}^{t}B_t^{j,v}R_v^{2,j}\\&+\sum_{j=1}^{J}\sum_{t=1}^{T}B_t^{3,j}R_t^{3,j}\\=&\sum_{j=1}^{J}\sum_{t=1}^{T}R_t^{1,j}\sum_{v=0}^{t}B_t^{j,v}+\sum_{j=1}^{J}\sum_{t=1}^{T}R_t^{2,j}\sum_{v=t}^{T}B_v^{j,t}\end{aligned}$$

$$+\sum_{j=1}^{J}\sum_{t=1}^{T}B_t^{3,j}R_t^{3,j}.$$

And suppose $B_t^{1,j}=\sum_{v=0}^{t}B_t^{j,v}$, $B_t^{2,j}=\sum_{v=t}^{T}B_v^{j,t}$

$$\begin{aligned}\varepsilon^{\mathrm{PVTC}} &\approx \sum_{j=1}^{J}\sum_{t=1}^{T}(B_t^{1,j}R_t^{1,j}+B_t^{2,j}R_t^{2,j}+B_t^{3,j}R_t^{3,j})\\ &=\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{k=1}^{3}(B_t^{k,j}R_t^{k,j}),\end{aligned}$$

where k refers to the different types of risks considered in the model: FP is the risk (fluctuation) of fuel price growth, $k=1$; r is the risk of technological progresses of fuel consumption rate, $k=2$; and C is the risk of capital cost reduction, $k=3$.

$$\begin{aligned}\mathrm{Var(PVTC)} &= E(\varepsilon^{\mathrm{PVTC}}\varepsilon^{\mathrm{PVTC}})\\ &= E\left[\sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{k_1=1}^{3}(B_{t_1}^{k_1,j_1}R_{t_1}^{k_1,j_1}) \times\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}\sum_{k_2=1}^{3}(B_{t_2}^{k_2,j_2}R_{t_2}^{k_2,j_2})\right]\\ &= E\left[\sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{k_1=1}^{3}\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}\sum_{k_2=1}^{3}(B_{t_1}^{k_1,j_1}R_{t_1}^{k_1,j_1}) \times(B_{t_2}^{k_2,j_2}R_{t_2}^{k_2,j_2})\right]\\ &= \sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{k_1=1}^{3}\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}\sum_{k_2=1}^{3}B_{t_1}^{k_1,j_1}B_{t_2}^{k_2,j_2} \times E(R_{t_1}^{k_1,j_1}R_{t_2}^{k_2,j_2}).\end{aligned}$$

Suppose the covariances between the three types of risks are zero, i.e., they are uncorrelated to each other in terms of risk. The above formula has a value only when $k_1=k_2$. Hence, Var(PVTC) can be rewritten as

$$\begin{aligned}&\sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}\min(t_1,t_2)B_{t_1}^{1,j_1}B_{t_2}^{1,j_2}\sigma_{t_1,t_2}^{1,j_1,j_2}\quad(k=1)\\ &+\sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}\min(t_1,t_2)B_{t_1}^{2,j_1}B_{t_2}^{2,j_2}\sigma_{t_1,t_2}^{2,j_1,j_2}\quad(k=2)\\ &+\sum_{j_1=1}^{J}\sum_{t_1=1}^{T}\sum_{j_2=1}^{J}\sum_{t_2=1}^{T}B_{t_1}^{3,j_1}B_{t_2}^{3,j_2}\sigma_{t_1,t_2}^{3,j_1,j_2}\quad(k=3),\end{aligned}$$

where $\min(t_1,t_2)$ represents the minimum of t_1 and t_2. If $j_1=j_2$ then $\sigma_{t_1,t_2}^{j_1,j_2}$ is the variance; if $j_1\neq j_2$ then $\sigma_{t_1,t_2}^{j_1,j_2}$ is the covariance of j_1 and j_2.

References

Awerbuch, S., Berger, M., 2003. Applying portfolio theory to EU electricity planning and policy-making. IEA/EET Working Paper EET/2003/03, International Energy Agency (IEA), Paris.

Barreto, L., 2001. Technological learning in energy optimisation models and deployment of Emerging Technologies. Ph.D. dissertation, Swiss Federal Institute of Technology Zurich.

Barreto, L., Kypreos, S., 2004. Endogenizing R&D and market experience in the "bottom-up" energy-systems ERIS model. Technovation 24 (8), 615–629.

Berglund, C., Söderholm, P., 2006. Modeling technical change in energy system analysis: analyzing the introduction of learning-by-doing in bottom-up energy models. Energy Policy 34 (12), 1344–1356.

Bodie, Z., Kane, A., Marcus, A.J., 2005. Investments, sixth ed. McGraw-Hill Irwin, New York.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Brooke, A., Kendrick, D., Meeraus, A., Raman, R., 2005. GAMS—A Users Guide. GAMS Development Corporation.

Bureau of Energy, Ministry of Economic Affairs, R.O.C., 1999. New Energy and Clean Energy Research and Development Planning Report.

Bureau of Energy, Ministry of Economic Affairs, R.O.C., 2005. The Conclusions of 2nd National Energy Conference.

Bureau of Energy, Ministry of Economic Affairs, R.O.C., 2006. Taiwan Energy Statistical Hand Book 2005.

Chang, S.L., Chen, Y.Y., 2005. Evaluation on the strategies of renewable energy development and energy security in Taiwan. The Cross-Strait Conference on Sustainable Development.

Energy and Environment Laboratories (EEL), Industrial Technology Research Institute, 2006. A Study on Long-Term Energy Supply-Demand Planning and Analysis in Taiwan.

Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4), 987–1007.

Environmental Protection Administration (EPA), Executive Yuan, R.O.C., 2006. Taiwan Environmental Conservation Annual Report for 2005.

Humphreys, H.B., McClain, K.T., 1998. Reducing the impact of energy price volatility through dynamic portfolio selection. The Energy Journal 19 (3), 107–131.

IEA, 2000. Experience Curves for Energy Technology Policy. OECD/IEA, Paris.

Krey, B., Zweifel, P., 2005. An efficient energy portfolio for Switzerland. In: 28th Annual IAEE International Conference, Taipei/Taiwan, 3–6 June.

Kumbaroğlu, G., Madlener, R., Demirel, M., in press. A real options evaluation model for the diffusion prospects of new renewable power generation technologies. Energy Economics, doi:10.1016/j.eneco.2006.10.009.

Lesbirel, S.H., 2004. Diversification and energy security risks: the Japanese case. Japanese Journal of Political Science 5 (1), 1–22.

Madlener, R., Kumbaroğlu, G., Ediger, V., 2005. Modeling technology adoption as an irreversible investment under uncertainty: the case of the Turkish electricity supply industry. Energy Economics 27 (1), 139–163.

Markowitz, H.M., 1952. Portfolio selection. Journal of Finance 7 (1), 77–91.

McDonald, A., Schrattenholzer, L., 2001. Learning rates for energy technologies. Energy Policy 29 (4), 255–261.

Söderholm, P., Sundqvist, T., 2007. Empirical challenges in the use of learning curves for assessing the economic prospects of renewable energy technologies. Renewable Energy 32 (15), 2559–2578.

Taiwan Power Company (Taipower), 2005. Long-term load forecasting and long-term power development programming in Taiwan.

Taiwan Power Company (Taipower), 2006. Statistics of Taiwan power company. Annual Report for 2005.

Wu, R.H., Huang, Y.H., 2006. Renewable energy perspectives and support mechanisms in Taiwan. Renewable Energy 31 (11), 1718–1732.

Zon, A. van, Fuss, S., 2005. Investing in energy conversion technologies—an optimum vintage portfolio selection approach. Working Paper 2005-013, Maastricht Economic Research Institute on Innovation and Technology (MERIT).

Available online at www.sciencedirect.com
ScienceDirect
Electrical Power and Energy Systems 29 (2007) 690–697
ELECTRICAL POWER & ENERGY SYSTEMS
www.elsevier.com/locate/ijepes

Risk management in a competitive electricity market

Min Liu [a,*], Felix F. Wu [b]

[a] *Faculty of Electrical Engineering, Guizhou University, Guiyang, Guizhou 550003, PR China*
[b] *Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong*

Received 19 August 2006; accepted 24 May 2007

Abstract

In a competitive electricity market, it is necessary and important to develop an appropriate risk management scheme for trade with full utilization of the multi-market environment in order to maximize participants' benefits and minimize the corresponding risks. Based on the analyses to trading environments and risks in the electricity market, a layered framework of risk management for electric energy trading is proposed in this paper. Simulation results confirmed that trading among multiple markets is helpful to reduce the complete risk, and VaR provides a useful approach to judge whether the formed risk-control scheme is acceptable.

Keywords: Risk management; Mean-variance portfolio theory; Value at risk

1. Introduction

Global deregulation in the electrical power industry has introduced the concept of a competitive electricity market. In this new environment, electricity is traded the same way as other commodities. However, electricity prices are substantially more volatile than any other commodity price since electricity cannot be stored and its transmission is limited by physical and reliability constraints. Confronted with this severe price volatility, market participants need to find ways to protect their benefits (quantified in profits in this paper), i.e., to manage risks involved in the market.

Risk refers to the possibility of suffering harm or loss; danger or hazard. Risk result from uncertainty. In the electricity market, a trader's profit is influenced by many uncertain factors: unit outage, other trader's bidding strategy, congestion in transmission, demand change, etc. These uncertainties bring about risks in electricity pricing and delivery. From mathematics point of view, a trader's profit is a random variable. According to the modern theory of choice under uncertainty [1], the expected profit is an indication of expected profitability, while the variance or the standard deviation of the profit can be used as an indication of the risk involved.

Risk management is the process of achieving a desired profit, taking the risks into considerations, through a particular strategy. In the financial field, there are two means to control risk. One is through risk financing by using hedging to offset losses that can occur and the other is through risk reduction using diversification to reduce exposure to risks. Instruments for risk management include forward contracts, futures contracts, options, swaps, etc. [2]. Forward contracts are agreements to buy/sell an agreed amount of the commodity at a specified price at a designated time. Futures contracts are standardized forward contracts that are traded on exchange and no physical delivery is necessary. Options are contracts that provide the holder the right but not the obligation to buy/sell the commodity at a designated time at specified price. A swap contract is an agreement between two parties to exchange a series of cash flows generated by underlying assets without physical transferring of the commodity between the buyer and seller. Hedging is to use these financial instruments with the payoff patterns to offset the market risks. Diversification is to engage in a wide variety of markets so that the exposure to the risk of any particular market is limited.

* Corresponding author. Tel.: +86 851 4732915; fax: +86 851 4730394.
E-mail address: minliu@graduate.hku.hk (M. Liu).

0142-0615/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijepes.2007.05.003

Applying this concept to energy trading in an electricity market, diversification means to trade electric energy through different physical trading approaches[1]. In the energy market, both physical trading approaches (e.g., spot market, contract market) and financial trading approaches (e.g., futures, options and swaps, etc.) are available. A combination of these trading approaches is defined as a portfolio and the corresponding risk-control methodology is called portfolio optimization. A commonly adopted measure for risk assessment, i.e., assessing risk exposure of financial portfolios, is the value at risk (VaR) [3].

Various aspects of risk management have been applied to electricity markets [4]. For example, different forward contracts that can provide hedging to the risk of spot prices for market participants are proposed in [5,6]. The usefulness of the application of futures contracts in an electricity market is demonstrated in [7,8]. Valuation of different contracts is considered in [9,10]. Monte Carlo simulation and decision analysis have been applied to find the optimal contract combination [11–13]. Approaches of portfolio optimization in Mean-variance portfolio theory have been adopted in the trading scheduling for a Genco [14]. VaR has been applied to risk assessment in electricity markets [15]. Concepts from financial option theory have been utilized in the valuation of generation assets [16].

We are addressing the problem of establishing a framework for risk management in a multi-market environment, i.e., how to make an optimal trading schedule from the point of view of risk control and make an assessment on this trading schedule. We apply the methodologies of the Mean-variance portfolio theory to risk control and VaR to risk assessment on the associated trading portfolio. Following explanation is made from the point of view of a generation company (Genco). We first introduce the background of a competitive electricity market; then give an overview about the framework of the risk management; explain the methodologies of the risk control and risk assessment, respectively; give a numerical example to demonstrate the proposed approach; and finally draw a brief conclusion.

2. Electricity markets and pricing systems

In the electricity market, energy-trading markets can be divided into two categories: physical markets and financial markets. In the physical market, energy is physically traded. While financial markets only operate as hedging instruments and no physical energy transactions are involved.

Most of the electricity markets provide two types of physical markets: the spot market and the contract market. The term "spot market" in the electricity market typically refers to a market in which trades covers a short period in the very near future. In this paper, we adopt FERC-USA definition in its standard market design [17] that all the energy traded in the real-time and day-ahead market as spot energy. From a Genco's point of view, selling energy in the spot market means to submit a bid (price and quantity) to the exchange (Power Pool/ISO) and get either of the two alternative results: (1) the exchange accepts the Genco's bid and pays the Genco the market clearing price (MCP) for its actual energy output; or (2) the exchange rejects the Genco's bid, i.e., the Genco sells nothing in the spot market. The MCP depends on everybody's bids, as well as the load demand, and is therefore uncertain. The risk of the price fluctuation is therefore, the most important risk in the spot market.

In the contract market, a Genco trades energy by way of signing contacts, which are referred to as physical forward contracts, with its counterparters (e.g., energy consumers). Specific details such as trading quantity (MW), trading duration (hours), trading price ($/MWh) and delivery point are bilaterally negotiated between Gencos and consumers or their agents. Bilateral contracts are signed before the actual trading period. In other words, trading quantity and price are set in advance. The main risk in the contract market is the risk of the congestion charge. Congestion charges depend on the specific pricing system of an electricity market. There are three pricing systems currently adopted in the electricity market: uniform marginal pricing, zonal pricing and locational marginal pricing (LMP) [18]. In a market with uniform marginal pricing system, only one price is used for ex post settlement for each trading interval. A Genco can make certain of its revenue by signing bilateral contracts with its customers at fixed prices. In other words, there is no risk of the congestion charge in a uniform pricing market. In a zonal pricing market or LMP market, bilateral contracts face the risk of congestion charge since the marginal prices will vary from zones or locations when there is a congestion between the zones (zonal pricing market) or in the transaction system (LMP market). The congestion charge is the product of the price difference between the zones or locations and the trading amount involved.

In financial markets, several financial instruments are provided to offset the particular sources of risk in the electricity market. For example, futures contracts, options and swaps are used to hedge the risk of the price volatility in the spot market. While financial transmission rights[2] are adopted to protect market participants from the risk of congestion charge in the contract market.

[1] Physical trading approach refers to the trading approach in which actual physical energy are traded while financial trading approach only involves financial settlement, and no physical delivery is necessary.

[2] Financial transmission rights (FTRs) [19] are contracts that exist between a market participant – in fact, any individual or organization – and the system operator. FTRs do not entitle their holders to an exclusive right to use the transmission system but to be paid the transmission price on a given path (multiplied by the number of rights the owner has), or, in a nodal market, the price difference between two nodes.

3. Overview of the framework

In order to provide a clear hierarchy of the risk management process, this paper proposes to establish a risk management framework in which four steps are involved: (1) determination of the trading objective of a Genco; (2) identification to the associated trading constraints such as trading environments, market rules and trading horizons, etc.; (3) translating the objectives and constrains into risk-control strategies, and making a trading schedule/trading portfolio under a specific strategy; and (4) risk assessment on the formed trading portfolio. The risk management process is completed if the assessment result is acceptable to the Genco. Otherwise, modifications to the risk-control strategy and the corresponding trading portfolio are needed.

A Genco's objective, in a competitive electricity market, is characterized by the benefit-and-risk trade-off between the expected benefits that the Genco wants (benefit requirements) and how much risk it is willing to assume (risk tolerance). Different benefit requirements (e.g., variable or constant) and risk tolerance (e.g., conservative or variable) result in different objectives. Rather than enumerating all the benefit-risk combinations, which is an impossible task, we divide Gencos' objectives roughly into three types: normal conservative, more conservative and less conservative. A Genco with a normal conservative objective would like to accept variable benefits (of course as high as possible) provided that the corresponding risk may be reduced with risk-control instruments. A more conservative Genco concerns about the risk more so than the benefit and therefore, requires expected benefit as close to constant as possible. While a less conservative Genco focuses its attention on transactions with potentially high benefits rather than try to search for the optimal benefit-risk profile for the entire portfolio.

Trading constraints includes trading environments, market rules, and trading horizons, etc. Trading environments refer to the types of physical and financial trading approaches provided by the specific electricity market (e.g., spot market, forward contract market, futures market, FTRs market, etc.). Market rules vary from one market to another. From risk management point of view, two aspects of the market rules are concerned. The first concern is the pricing method adopted in the spot market, i.e., uniform pricing or zonal pricing or locational marginal pricing or others. The second concern is the specific rules on trading proportion of each trading market. That is, the maximum and minimum trading proportions of each trading market. For example, some markets require at least 80% of a Genco's energy should be traded through forward contract market and the remainders be traded in the spot market. Trading horizon can be divided into three levels: (1) short-term trading schedule (e.g., weekly or monthly schedule); (2) mid-term trading schedule (e.g., quarterly or annually schedule); and (3) long-term trading schedule (e.g., a schedule for several years).

Risk control has two levels. The first level is diversification through portfolio optimization. That is, trying to find a portfolio with a reasonable trade-off between the expected benefit and risk. The second level of risk control is hedging specific risks with specific financial instruments. Different Gencos with different objectives and different constraints would adopt different risk-control strategies. Detailed discussions are made in the following.

Risk assessment is made to give a picture of the risk of a trading portfolio and let a decision-maker decide whether the trading schedule is acceptable more intuitively. VaR is adopted to value the trading portfolio in the following.

4. Risk control

4.1. General case: risk-control strategy for a normal conservative Genco

Assume that all existing physical and financial trading approaches are available in an electricity market. The risk-control strategy of a normal conservative Genco includes two aspects: (1) diversification among multiple physical trading markets; and (2) hedging with specific financial instruments. That is, physical energy is allocated between spot markets and contract markets, while specific risks of the spot market (i.e., spot-price risk) and contract market (i.e., congestion-charge risk) are hedged, respectively with specific financial instruments. Detailed risk-control scheme is subject to the specific pricing method adopted in the spot market[3].

In a uniform pricing market, a Genco would sign contracts with customers who offer the highest price without considering congestion cost. As for the spot market, the price risk can be hedged with financial instruments such as futures, options, and swaps. Among these instruments, futures are suitable for mid-term trading schedule since they are generally traded monthly and up to 18 months. Options can be traded in the exchange or over the counter (OTC)[4] market and therefore, can be used in short-term, mid-term and long-term trading schedule. But it tends to be more expensive than futures and swaps due to the premium payment. Swaps are OTC derivatives and therefore, suitable for short-term, mid-term and long-term trading schedules provided that counterparties are available.

In a zonal pricing/LMP market, a Genco would sign bilateral contracts with customers located in different pricing areas, as well as trade energy in the spot market, aiming at reducing the total risk of the trading portfolio. Spot-price risk can be hedged with futures, options or swaps, while congestion-charge risk can be hedged with FTRs.

---

[3] Due to the limitation of the paper size, this paper only gives the associated conclusions about risk-control schemes in different pricing markets. Please refer [20] for more explanation and demonstration.

[4] OTC is a kind of derivatives market in which non-standard products (e.g., contracts) are traded. Trades on the OTC market are negotiated directly with dealers.

Making a trading schedule/portfolio refers to the determination of energy allocation ratios of the markets including both physical trading markets (i.e., spot market and forward contract market) and financial trading markets (i.e., futures market and FTRs market, etc.). There are three steps to achieve an optimal trading schedule. Firstly, calculate the optimal hedge-ratio[5] for each physical transaction; then calculate the optimal energy allocation ratio to each physical trading approach; finally, calculate the optimal allocation ratios to each financial trading approach.

4.1.1. Step 1: Hedging with financial instruments

For each physical trading approach, assuming that all energy is traded in this approach, calculate the optimal hedge-ratio denoted by x_i^* (i is the index of the physical trading approach and the associated financial hedging market). Mathematically, the optimal hedge amount can be achieved by minimizing the variance of the profit on the hedged physical trading approach with respect to the trading amount in the financial hedging market. For example, for the transaction in the spot market, if futures contracts are used to hedge the spot-price risk, the optimal hedge-ratio can be obtained by minimizing the total risk[6] with respect to the amount traded in the futures market. That is, suppose that a Genco sells α MWh energy in futures market at futures price f (\$/MWh) before the beginning of the trading period. Upon the date of delivery (i.e., the beginning of trading period) the Genco settles the futures contracts by buying back α MWh energy in the futures market at futures price f^* (\$/MWh). Let π_O be the profit on the spot market. If the transaction costs of futures contracts are ignored, the Genco's profit from both spot market and futures market is π_N, where $\pi_N = \pi_O + \alpha(f\text{-}f^*)$. The optimal quantity of energy sold in the futures market can be achieved by minimizing the total risk ($\mathrm{Var}(\pi_N)$) with respect to α, i.e.,

$$\min_{\alpha} \mathrm{Var}(\pi_N) = \mathrm{Var}(\pi_O) + \alpha^2 \mathrm{Var}(f^*) - 2\alpha \cdot \mathrm{Cov}(\pi_O, f^*)$$

where $\mathrm{Var}(\pi_O)$, $\mathrm{Var}(\pi_N)$, $\mathrm{Var}(f^*)$ are variances of π_O, π_N, f^*, respectively; $\mathrm{Cov}(\pi_O, f^*)$ is the covariance between π_O and f^*. Solving this optimization problem results in the optimal selling quantity α^*, where $\alpha^* = \frac{\mathrm{Cov}(\pi_O, f^*)}{\mathrm{Var}(f^*)}$. And the optimal hedge-ratio is $x^* = \alpha^*/E$, where E is the quantity of energy sold in the spot market.

4.1.2. Step 2: Energy allocation among physical trading approaches

The risk preference of a risk-averse Genco can be described with a utility function which combines the benefit (expected profit) and risk (variance of profit) into a simple relation, e.g., $U(\pi) = E\ (\pi)\text{-}B \cdot \mathrm{Var}(\pi)$ [14], where $U(\pi)$ is the utility value; B is the risk penalty factor which indicates the extent that a Genco "penalizes" the expected profit considering the risk of obtaining the corresponding profit[7]. According to the utility theory [1], a trading portfolio with the highest utility value is preferred. The trading objective of a normal conservative Genco, i.e., maximizing benefit and minimizing the associated risk, is then achieved by maximizing the utility function.

For each hedged trade (with the optimal hedge-ratio x_i^*) determined from Step 1, calculate its expected profit and risk (i.e., $E(\pi_i)$, $\mathrm{Var}(\pi_i)$). For any two trades that have been hedged, calculate their covariance (i.e., σ_{ij} where i, j are the indexes of trading approaches). The optimal energy allocation among different hedged trades can be achieved by maximizing the Genco's utility value with respect to the proportions allocated to the trades. Assuming there are n physical trading approaches available in the electricity market, this optimization problem can be described as follows:

$$\begin{aligned}
&\max_{w_i} U(w_1, \ldots, w_i, \ldots, w_n) = \sum_{i=1}^{n} w_i E(\pi_i) - B \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \\
&\text{s.t.} \quad \sum_{i=1}^{n} w_i = 1 \\
&\quad w_i^{\min} \leqslant w_i \leqslant w_i^{\max}
\end{aligned} \tag{1}$$

where w_i is the proportion allocated to the ith trade; $w_i^{\min}$ and $w_i^{\max}$ are the upper limit and lower limit of the trading proportion for the ith physical trading approach, respectively, which are specified by a specific electricity market. Solutions to this optimization problem are the optimal allocation ratios to the physical trading approaches denoted by w_i^* $(i = 1 - n)$.

4.1.3. Step 3: Optimal hedging proportions

The proportion of total energy traded in a financial instrument market aiming at hedging the risk of the ith physical trading approach is called the optimal hedging proportion for the ith physical trading approach. This ratio is calculated as $w_i^* x_i^*$.

To summarize, the optimal proportion of total energy allocated to the physical trading market is $w_i^* (i = 1 - n)$; the optimal proportion of total energy allocated to the financial trading market is $w_i^* x_i^* (i = 1 - n)$.

4.2. Discussion: Risk-control strategies for more conservative Gencos and less conservative Gencos

A more conservative Genco would like to trade physical energy both in spot and contract markets. Spot-price risk is hedged with swaps since it can be negotiated

[5] Hedge-ratio refers to the ratio of the energy quantity traded in the financial hedging instrument to the quantity traded in the underlying physical trading approach.

[6] Here, the total risk is the risk of trading in a physical trading market (i.e., spot market) with the corresponding hedge effect (i.e., hedging with futures contracts) into consideration.

[7] The value of B can be calculated with formula $B = A/2C$ [14], where C is the production cost of a Genco, A is an index of the decision-maker's risk-aversion. The moderate value of A is 3, $A > 3$ means more risk averse and $A < 3$ indicates less risk averse [21].

bilaterally and no transaction cost is involved. For the contract trade, the Genco would sign contracts with customers who offer the highest price in the uniform pricing market since no congestion cost is charged to the bilateral transaction. While in a zonal pricing/LMP market, the Genco would sign contract with local customers. The reason is that contracts signed with non-local customers face potential congestion charges and whether the corresponding financial transmission rights can be obtained and completely hedge the congestion risk is quite uncertain.

A less conservative Genco, in a uniform pricing market, would like to trade physical energy only in the spot market and hedge the spot-price risk with options. Because purchasing a put option with a physical sale of electric energy to the spot market let the Genco avoid the risk of lower prices and benefit from increase in spot price although a premium is needed. But in the zonal pricing/LMP market, the Genco would like to trade energy in the spot market and with customers who offer the highest contract price in the form of forward contracts. It prefers to hedge the spot-price risk and congestion-charge risk with options and FTRs, respectively.

5. Risk assessment

5.1. Risk-assessment technique

Value at risk (VaR) is a risk management concept developed and promoted in the banking industry to provide a common measurement for the risk exposure of financial portfolios. It is defined, in the financial literature, as a monetary value that the portfolio will lose less than that amount over a specified period of time with a specified probability. For example, a one-day 95% VaR of \$500,000 indicates that the portfolio is expected to lose 95 days out of 100 days an amount less than \$500,000.

There are numerous methods to calculate VaR, which use different assumptions and techniques. Since VaR calculations are very sensitive to assumptions and data, quantitative results will differ when the same techniques are applied using different assumptions or different data sets. Ref. [3] distinguishes four separate routes to measuring VaR: delta-normal method, historical simulation, stress-testing method and Monte Carlo approach.

5.2. VaR application in trading scheduling

In this paper, the VaR of a trading portfolio is defined as the expected minimum profit (a monetary value) of the portfolio over a target horizon within a given confidence interval. The target horizon is the trading horizon. The confidence level depends on the extent of the Genco's risk-aversion. Normally, a Genco with moderate risk-aversion adopts 95% confidence level; a more risk-averse Genco may require 99% confidence level and a less risk-averse Genco could use 92.5% confidence level.

In the most general form, VaR can be derived from the probability distribution of the future portfolio value $f(\pi)$ (where π denotes the profit on the trading portfolio). At a given confidence level c, we wish to find the lowest possible realization $\hat{\pi}$ such that the probability of exceeding this value is c:

$$c = \int_{\hat{\pi}}^{\infty} f(\pi)\mathrm{d}\pi$$

Or such that the probability of a value lower than $\hat{\pi}$, $p = \mathrm{Prob}(\pi \leqslant \hat{\pi})$, is $1-c$:

$$1 - c = \int_{-\infty}^{\hat{\pi}} f(\pi)\mathrm{d}\pi = \mathrm{Prob}(\pi \leqslant \hat{\pi}) = p$$

The number $\hat{\pi}$ is called the sample quantile of the distribution. In the simulation approaches to measuring VaR such as the historical simulation, stress testing and Monte Carlo simulation, this sample quantile can be found out from the simulation results of samples. For example, suppose 100 samples are used to simulate the expected profit on a trading portfolio. Simulation results are ranged from the highest profit to the lowest profit. If the confidence level is 95%, then $\hat{\pi}$ is equal to the 95th simulation result in the simulation result list.

In this paper, VaR is used to measure the risk extent of the scheduled trading portfolio aiming at providing a rough figure that helps the decision-maker judge whether the scheduled portfolio is acceptable. Therefore, we may use a relatively rough but simplified method to calculate the value of VaR, i.e., the delta-normal method. Suppose the prices of electricity and fuel are normally distributed. The profit on the trading portfolio, which is a linear combination of normal random variables, is then also normally distributed. Under this condition, the VaR figure can be derived directly from the portfolio's standard deviation, using a multiplicative factor that depends on the confidence level.

First, the general normal distribution $f(\pi)$ is translated into a standard normal distribution $\phi(u)$, where u has mean of zero and standard deviation of unity. That is, $\hat{\pi}$ is associated with a standard normal deviate α ($\alpha > 0$) by setting

$$-\alpha = \frac{\hat{\pi} - \mu}{\sigma}$$

where, $\mu = E(\pi^*) = \sum_{i=1}^{n} w_i^* E(\pi_i), \sigma = \sqrt{\mathrm{Var}(\pi^*)} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i^* w_j^* \sigma_{ij}}$. w_i^* is the optimal allocation ratio of the ith physical trading approach, which is the calculation result in the process of risk control. It is equivalent to set

$$1 - c = \int_{-\infty}^{\hat{\pi}} f(\pi)\mathrm{d}\pi = \int_{-\infty}^{-\alpha} \phi(u)\mathrm{d}u = p$$

Thus, the problem of finding a value at risk is equivalent to finding the deviate α such that the area to the left of it is equal to $1 - c$. It is made possible by turning to tables of the cumulative stand normal distribution function, which is the area to the left of a standard normal variable with value equal to c. To find the VaR of a standard normal variable, select the desired confidence level in the table, say 95%. This corresponds to a value of $\alpha = 1.65$. Then the VaR of the portfolio (i.e., the cutoff profit $\hat{\pi}$) can be calculated as follows:

$$\mathrm{Var} = \hat{\pi} = \mu - \alpha\sigma \tag{2}$$

6. Example

The PJM electricity market is a LMP market with several pricing zones such as PSEG, PECO, PENELEC, etc. [22]. Suppose a Genco is located in PENELEC and is normal conservative. The Genco is making a monthly trading plan of August, i.e., determining the trading amount/trading ratio of each market (e.g., spot market, contract market, etc.). Before applying the methodologies proposed in this paper, the profit characteristics (i.e., expected value, variance and covariance) of each trading approach is clarified firstly in the following.

6.1. Profit characteristics

Suppose that there are M trading intervals during planning period (one trading interval can be one hour, one day, one week, one month or even one year depending on the planning horizontal). Trading time for each trading interval is t (hour). The following notation will be used:

i, j	the index of the trading area or pricing node
k	the index of the trading interval
$\lambda^{\mathrm{B}}_{i,k}$	the kth trading interval's electricity contract price signed with customers of area i
$\lambda^{\mathrm{S}}_{i,k}$	the kth trading interval's electricity spot price of area i
λ^{F}_{k}	the kth trading interval's fuel spot price
π_i	profit on the ith trade, $i = 0$ denotes local contract; $i = 1 - n$ denotes non-local contract; $i = n + 1$ denotes spot transaction
e_k	the kth trading interval's trading energy
$E(\cdot)$	expectation
$Var(\cdot)$	variance
$\sigma_{i,j}$	covariance between profits on transaction i and j
b	heat rate (or consumption coefficient) of a unit

Assume that the Genco's production exhibits constant returns to scale. Production cost is a function of energy output and fuel price, i.e., $c(\cdot) = be\lambda^F$. For the local contract and spot transaction, the associated cost only involves the production cost; for the non-local contract, the associated costs include congestion charge as well as production cost. Generally, congestion charges should be paid by the associated bilateral transaction. But who (Gencos or energy purchasers) should pay how many percentage of the involved congestion charges depends on the specific market rules. In this paper, a factor β ($0 \leqslant \beta \leqslant 1$), is used to denote the payment proportion of the Genco. According to the methodologies of probability, the expectation, variance and covariance of profits on each transaction can be derived as follows[8]:

$$E(\pi_0) = \sum_{k=1}^{M} \left[\lambda^{\mathrm{B}}_{0,k} - bE(\lambda^{\mathrm{F}}_k)\right] \cdot e_k \tag{3}$$

$$\mathrm{Var}(\pi_0) = \sum_{k=1}^{M} (be_k)^2 \cdot \mathrm{Var}(\lambda^{\mathrm{F}}_k) \tag{4}$$

$$E(\pi_i) = \sum_{k=1}^{M} \left[\lambda^{\mathrm{B}}_{i,k} + \beta E(\lambda^{\mathrm{S}}_{0,k}) - \beta E(\lambda^{\mathrm{S}}_{i,k}) - bE(\lambda^{\mathrm{F}}_k)\right] \cdot e_k \quad (i = 1\text{–}n) \tag{5}$$

$$\mathrm{Var}(\pi_i) = \sum_{k=1}^{M} e_k^2 \begin{bmatrix} \beta^2\mathrm{Var}(\lambda^{\mathrm{S}}_{0,k}) + \beta^2\mathrm{Var}(\lambda^{\mathrm{S}}_{i,k}) + b^2\mathrm{Var}(\lambda^{\mathrm{F}}_k) \\ -2\beta^2\mathrm{Cov}(\lambda^{\mathrm{S}}_{0,k}, \lambda^{\mathrm{S}}_{i,k}) - 2b\beta\mathrm{Cov}(\lambda^{\mathrm{S}}_{0,k}, \lambda^{\mathrm{F}}_k) \\ +2b\beta\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{F}}_k) \end{bmatrix} \quad (i = 1\text{–}n) \tag{6}$$

$$E(\pi_{n+1}) = \sum_{k=1}^{M} \left[E(\lambda^{\mathrm{S}}_{0,k}) - bE(\lambda^{\mathrm{F}}_k)\right] \cdot e_k \tag{7}$$

$$\mathrm{Var}(\pi_{n+1}) = \sum_{k=1}^{M} e_k^2 \left[\mathrm{Var}(\lambda^{\mathrm{S}}_{0,k}) + b^2\mathrm{Var}(\lambda^{\mathrm{F}}_k) - 2b\mathrm{Cov}(\lambda^{\mathrm{S}}_{0,k}, \lambda^{\mathrm{F}}_k)\right] \tag{8}$$

$$\sigma_{0i} = -\sum b\beta e_k^2\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{0,k}) + \sum b\beta e_k^2\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{i,k}) + \sum b^2 e_k^2\mathrm{Var}(\lambda^{\mathrm{F}}_k)(i = 1\text{–}n) \tag{9}$$

$$\sigma_{0,n+1} = -\sum be_k^2\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{0,k}) + \sum b^2e_k^2\mathrm{Var}(\lambda^{\mathrm{F}}_k) \tag{10}$$

$$\sigma_{i,n+1} = \sum e_k^2 \begin{bmatrix} \beta\mathrm{Var}(\lambda^{\mathrm{S}}_{0,k}) - \beta\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{S}}_{0,k}) - b\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{0,k}) - \\ b\beta\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{0,k}) + b\beta\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{i,k}) + b^2\mathrm{Var}(\lambda^{\mathrm{F}}_k) \end{bmatrix} \quad (i = 1\text{–}n) \tag{11}$$

$$\sigma_{i,j} = \sum e_k^2 \begin{bmatrix} \beta^2\mathrm{Var}(\lambda^{\mathrm{S}}_{0,k}) - \beta^2\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{S}}_{0,k}) - 2b\beta\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{0,k}) \\ -\beta^2\mathrm{Cov}(\lambda^{\mathrm{S}}_{j,k}, \lambda^{\mathrm{S}}_{0,k}) + \beta^2\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{S}}_{j,k}) \\ +b\beta\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{i,k}) - b\beta\mathrm{Cov}(\lambda^{\mathrm{F}}_k, \lambda^{\mathrm{S}}_{j,k}) + b^2\mathrm{Var}(\lambda^{\mathrm{F}}_k) \end{bmatrix} \quad (i, j = 1\text{–}n) \tag{12}$$

The statistics of prices involved in above equations, i.e., $E(\lambda^{\mathrm{S}}_{i,k})$, $\mathrm{Var}(\lambda^{\mathrm{S}}_{i,k})$, $E(\lambda^{\mathrm{F}}_k)$, $Var(\lambda^{\mathrm{F}}_k)$, $\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{S}}_{j,k})$ and $\mathrm{Cov}(\lambda^{\mathrm{S}}_{i,k}, \lambda^{\mathrm{F}}_k)$, can be estimated based on historical data according to the statistical method.

6.2. Simulation results

Following numerical simulation is performed based on the historical data of daily electricity prices in the PJM electricity market [22]. The statistical characteristics (i.e., expectation, variance, and covariance) of prices in each trading interval (i.e., one day 24 h) of 10 pricing zones are calculated. The average value (i.e., the statistical characteristics of monthly price) is shown in Table 1. The Genco owns a coal-fired generation unit with 600 MW capacity and 8.9 MBtu/MWh heat rates [23]. The coal price

[8] Please refer [14] for details.

Table 1
Statistical characteristics of prices in the PJM market (August)

Zone	Expectation of price ($/MWh)	Standard deviation of price (in percentage)	Zone	Expectation of price ($/MWh)	Standard deviation of price (in percentage)
PSEG	47.738	84.554%	METED	46.715	84.643%
PECO	46.351	88.662%	PEPCO	47.097	79.188%
PPL	44.301	87.376%	AECO	49.058	86.245%
BGE	47.157	84.294%	DPL	48.148	91.941%
JCPL	46.044	87.095%	PENELEC	40.653	69.177%

is given as $\lambda_k^{\mathrm{F}} = 2.0\$/\mathrm{MBtu}$ ($k = 1$–31). The unit is located in PENELEC.

If the Genco only trades its energy in the spot market, with formulae Eqs. (7) and (8), the expected profit and corresponding variance are calculated and shown in the second column of Table 2. If the Genco is normal conservative, i.e., $c = 95\%$, the VaR value is calculated as $6,481,300 (see Table 2). The simulation results indicate that, in the spot market, the expected profit is $10,202,000; within 95% confidence level, the minimum profit is $6,481,300 during the trading month. If the Genco think the VaR value is quite lower and would like to reduce the risk of spot transaction, it can control the risk by trading energy both in spot market and contract market.

Table 2
Simulation results

Characteristic	Spot trading	Trading portfolio (1)	Trading portfolio (2)
$E(\pi^*)$ ($)	1.0202×10^7	1.0055×10^7	1.0063×10^7
$Var(\pi^*)$ ($\$^2$)	5.084×10^{12}	3.8263×10^{11}	4.0435×10^{11}
VaR($) ($c = 95\%$)	6.4813×10^6	9.0339×10^6	9.0135×10^6

Table 3
Contract prices

Location of consumers	Contract price ($/MWh)	Location of consumers	Contract price ($/MWh)
PSEG	47.2	METED	46.2
PECO	45.9	PEPCO	46.6
PPL	43.8	AECO	48.6
BGE	46.7	DPL	47.6
JCPL	45.5	PENELEC	40.0

Suppose the Genco can sign bilateral contracts with consumers located in different pricing zones with contract prices shown in Table 3. Assume that the Genco pays all the congestion charges involved in the bilateral transaction, i.e., $\beta = 1$. The upper and lower limit of the trading proportion of each trading market are 100% and 0%, respectively, i.e., $w_i^{\max} = 1, w_i^{\min} = 0$. The index of the Genco's risk-aversion is set to 3 (i.e., $A = 3$) since the Genco is normal conservative. Then the optimal allocation ratios w_i^* can be obtained by solving problem (1) and shown in Table 4; VaR value is calculated with formula (2) and shown in the third column of Table 2. Under this trading portfolio (i.e., PECO 19.6%, BGE 32.6%, PENELEC 14.1% and spot market 33.7%), the one-month 95% VaR is $9,033,900. That is, under the same confidential level, the minimum profit of the trading portfolio is increased 39.4% compared to that of the spot trading. Of course, the values of allocation ratio and VaR depend on the price of each bilateral contract. For example, if the consumers located in DPL would like to offer a higher price, say 47.8$/MWh, the allocation ratio to the contract signed with DPL consumers increases and other allocation ratios change accordingly (see Table 5). The expected profit and the associated risk also change as it is shown in the forth column of Table 2. Trading portfolio 1 and 2 both demonstrate the effect of diversification, i.e., trading among multiple trading approaches can reduce the risk of the complete portfolio.

If the minimum profit with 95% confidential level of the trading portfolio 1 or 2 is still not accepted by the Genco, modification to the risk-control strategy is needed. For example, hedging can be adopted to hedge the specific risks of spot transaction and non-local contract. Firstly, spot-price risk could be hedged with available financial

Table 4
Allocation ratio of each transaction (1)

Transaction	Spot	Bilateral contract									
		PSEG	PECO	PPL	BGE	JCPL	METED	PEPCO	AECO	DPL	PENELEC
Ratio	0.337	0	0.196	0	0.326	0	0	0	0	0	0.141

Table 5
Allocation ratio of each transaction (2)

Transaction	Spot	Bilateral contract									
		PSEG	PECO	PPL	BGE	JCPL	METED	PEPCO	AECO	DPL	PENELEC
Ratio	0.341	0	0	0	0.294	0	0	0	0	0.201	0.163

instrument such as futures, options and swaps, etc. Physical energy is then allocated among the hedged spot transaction, local contract and non-local contract. The risk of the modified trading portfolio is expected to be reduced. The risk management process is completed if the corresponding VaR value is acceptable. Otherwise, hedging the risk of non-local contract (i.e., congestion-charge risk) with FTRs can be considered.

7. Conclusions

This paper developed an overall framework of risk management for Gencos' trading in a competitive electricity market. That is, firstly, a Genco's objective and trading constraints are identified. Then the identified objective and constraints are translated into a reasonable and feasible risk-control strategy under which a specific trading schedule could be made. Finally, the formed trading portfolio is assessed with standard risk measurement technique – VaR. If the risk-assessment result is not acceptable, readjusting the risk-control strategy and the associated trading schedule until the Genco accepts it.

Risk-control strategy varies from different trading objective and constraint. With trading constraints assumed general, this paper discussed the risk-control strategies for Gencos with different objectives. Simply stated, a normal conservative Genco would like to control risk with all available trading approaches. That is, trading physical energy between the spot market and contract market; hedging spot-price risk with futures or swaps according to the specific trading horizon, and hedging the congestion-charge risk with financial transmission rights. A more conservative Genco prefers to control risk through diversification, i.e., trading energy between spot and contract markets. While a less conservative Genco tends to trade physical energy in spot markets only and hedge spot-price risk with options.

The risk-management process of a Genco with normal conservative objective was demonstrated based on the historical data of electricity prices in the PJM market. Simulation results confirmed that diversification, i.e., trading among multiple physical approaches, is helpful to reduce the complete trading risk, and VaR provides a useful approach to judge whether the formed trading portfolio is acceptable.

To summarise, the proposed framework of risk management provides a clear hierarchy of the risk management process, which should help a Genco to identify its objective and achieve an optimal trading portfolio in markets involving risks. It is also applicable to other market participants such as energy purchasers, with little modification.

Acknowledgements

This work has been supported by the Research Grant Council, Hong Kong SAR, China, through Grant HKU7174/04E, and Guizhou University, Guizhou, China, under Grant GUT2004-014.

References

[1] Biswas T. Decision-making under uncertainty. New York: St. Martin's Press, Inc.; 1997.
[2] Luenberger DG. Investment science. New York: Oxford University Press; 1997.
[3] Jorion P. Value at risk. New York: McGraw-Hill; 1997.
[4] Liu M, Wu FF, Ni YX. A survey on risk management in electricity markets. In: The proceedings of 2006 IEEE PES General Meeting, Montreal, 2006.
[5] Kaye RJ, Outhred HR, Bannister CH. Forward contracts for the operation of an electricity industry under spot pricing. IEEE Trans Power Sys 1990;5:46–52.
[6] Gedra TW. Optional forward contracts for electric power markets. IEEE Trans Power Sys 1994;9:1766–73.
[7] Collins RA. The economics of electricity hedging and a proposed modification for the futures contract for electricity. IEEE Trans Power Sys 2002;17:100–7.
[8] Tanlapco E, Lawarree J, Liu CC. Hedging with futures contracts in a deregulated electricity industry. IEEE Trans Power Sys 2002;17:577–82.
[9] Bjorgan R, Song H, Liu CC, Dahlgren R. Pricing flexible electricity contracts. IEEE Trans Power Sys 2000;15:477–82.
[10] Deng SJ. Pricing electricity derivatives under alternative stochastic spot price models. In: Proceedings of the thirty-third annual Hawaii international conference on system sciences. 2000; p. 1313–22.
[11] Vehviläinen I, Keppo J. Managing electricity market price risk. Eur J Operat Res 2003;145:136–47.
[12] Sheble GB. Decision analysis tools for GENCO dispatchers. IEEE Trans Power Sys 1999;14:745–50.
[13] Siddiqi SN. Project valuation and power portfolio management in a competitive market. IEEE Trans Power Sys 2000;15:116–21.
[14] Liu M, Wu FF. Managing price risk in a multi-market environment. IEEE Trans Power Sys 2006;21:1512–9.
[15] Dahlgren R, Liu CC, Lawarree J. Risk assessment in energy trading. IEEE Trans Power Sys 2003;18:503–11.
[16] Deng SJ, Johnson B, Sogomonian A. Spark spread options and the valuation of electricity generation assets. In: Proceedings of the thirty-second annual Hawaii international conference on system sciences. 1999.
[17] Notice of proposed rulemaking. United States of America Federal Energy Regulatory Commission Docket No.RM01-12-000, 2002.
[18] Ma XW, Sun DI, Cheung KW. Evolution toward standardized market design. IEEE Trans Power Sys 2003;18:460–9.
[19] Lyons K, Fraser H, Parmesano H. An introduction to financial transmission rights. The Elect J 2000;13:31–7.
[20] Liu M. Energy allocation with risk management in electricity markets. Ph.D. Thesis, The University of Hong Kong, Hong Kong, 2004.
[21] Bodie Z, Kane A, Marcus AJ. Investments. 4th ed. Chicago: Irwin/McGraw-Hill; 1999.
[22] PJM website:<http://www.pjm.com>.
[23] Wood AJ, Wollenberg BF. Power generation, operation, and control. 2nd ed. New York: Wiley and Sons; 1996.

Available online at www.sciencedirect.com
SCIENCE DIRECT
ENERGY
Energy 31 (2006) 940–953
www.elsevier.com/locate/energy

Electricity derivatives and risk management

S.J. Deng[a,*], S.S. Oren[b]

[a]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA
[b]Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA

Abstract

Electricity spot prices in the emerging power markets are volatile, a consequence of the unique physical attributes of electricity production and distribution. Uncontrolled exposure to market price risks can lead to devastating consequences for market participants in the restructured electricity industry. Lessons learned from the financial markets suggest that financial derivatives, when well understood and properly utilized, are beneficial to the sharing and controlling of undesired risks through properly structured hedging strategies. We review different types of electricity financial instruments and the general methodology for utilizing and pricing such instruments. In particular, we highlight the roles of these electricity derivatives in mitigating market risks and structuring hedging strategies for generators, load serving entities, and power marketers in various risk management applications. Finally, we conclude by pointing out the existing challenges in current electricity markets for increasing the breadth, liquidity and use of electricity derivatives for achieving economic efficiency.

1. Introduction

Electricity spot prices are volatile due to the unique physical attributes of electricity such as non-storability, uncertain and inelastic demand and a steep supply function. Uncontrolled exposure to market price risks could lead to devastating consequences. During the summer of 1998, wholesale power prices in the Midwest of US surged to a stunning \$7000 per MWh from the ormal price range of \$30–\$60 per MWh, causing the defaults of two power marketers in the east coast. In February 2004, persistent high prices in Texas during a 3-day ice storm led to the bankruptcy of a retail energy provider that was exposed to spot market prices. And of course, the California electricity crisis of 2000/2001 and its devastating economic consequences are largely attributed to the fact that the major utilities were not properly hedged through long-term supply contracts. Such expensive lessons have raised the awareness

* Corresponding author. Tel.: +1 404 894 6519; fax: +1 404 894 2301.
E-mail address: deng@isye.gatech.edu (S.J. Deng).

0360-5442/\$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2005.02.015

of market participants to the importance and necessity of risk management practices in competitive electricity market.

Hedging of risk by a corporation should in principle be motivated by the goal of maximizing firm's value. Hedging achieves value enhancement by reducing the likelihood of financial distress and its ensuing costs, or by reducing the variance of taxable incomes and its associated present value of future tax liabilities. Regulatory rules also play an important role in hedging practices. In California, for instance, the regulators granted the incumbent investor-owned utilities (IOUs) a fixed time frame to recover their stranded generation costs through the Competition Transition Charge. Fearing adverse market conditions causing insufficient recovery of the stranded costs, one major utility company hired investment bankers to structure and implement an extensive hedging strategy for its stranded-cost recovery. On the other hand, the reluctance of the regulators in California to immunize the IOUs against ex-post prudence review of long-term supply contracts discouraged the adoption of such contracts, resulting in over-reliance of the IOUs on the spot market for electricity procurement. This excessive exposure led to the near collapse of the California utility industry in 2001, with devastating economic losses due to prolonged outages and substantial rate increases.

As the competitive but volatile electricity markets mature, generation companies, power marketers and load serving entities (LSEs) seek certainty in their costs and revenues through hedging practices and contracting and active trading. Such activities involve quantifying, monitoring and controlling trading risks in the wholesale and retail power markets, which in turn require appropriate risk management tools and methodology.

On the supply side, managing risk associated with long-term investment in generation and transmission requires methods and tools for planning under uncertainty and for asset valuation. Much of the demands for generation asset valuation methods were spurred by the mandatory divestiture of generation assets already owned by major utility companies in various jurisdictions. For example, in California, most of the fossil-fuel plants held by the three IOUs, which account for about 60% of the total installed capacity in California by 2000, have been or will be divested to other parties. The need for asset valuation also rises from analysis of investment in new generation capacity and from efforts by regulators in the US and abroad to develop incentives for investment in generation capacity to meet supply adequacy and system reliability objectives.

A fundamental vision underlying the worldwide movement toward a competitive electricity industry has been that most of the efficiency gains from restructuring come from long-run investments in generating capacity. Under the state-ownership or required rate-of-return regulatory regime, utility companies were allowed to earn a regulated rate of return above their cost of capital. Once regulators approved the construction costs of a power generating plant, the costs would be passed onto consumers through regulated electricity prices over the life of the investment, independent of the fluctuation in market value of the investment over time due to changing energy prices, improving technology, and evolving supply and demand conditions. Most of the investment risks in generating capacity were allocated to consumers rather than producers. Firms, therefore, had little incentives to avoid excessive cost of investment and they focused on improving and maintaining quality of service rather than on developing and adopting new generation technology.

Electricity market reforms around the world have shifted much of the investment risk from consumers to producers. Under the ideal theoretical paradigm, shareholders bear all the investment risk and consumers bear the price risk, with competitive entry pushing generation capacity toward desired long-term equilibrium. In such an ideal market environment, suppliers and consumers are free to choose their desired level of risk exposure, achieved through voluntary risk management practices. Unfortunately, this idealized vision of a competitive electricity market is not working as expected, primarily due to such market imperfections as lack of demand response, abuse of locational market power, and political resistance to high prices reflecting scarcity rents and shortages.

With few exceptions such as Australia (where electricity spot prices are allowed to rise to $10,000 per MWh), most restructured electricity markets in the US and around the world have backed away from the idealized economic market models and instituted price caps and various capacity payment mechanisms. Such regulatory interventions allocate risks between consumers and producers by limiting price volatility for consumers and assuring investment cost recovery for generators. From a risk management perspective, these intervention schemes are mandatory backstop hedging that limits the exposures of consumers and producers. The proper design of such schemes requires the same pricing and asset valuation tools as voluntary risk management practices in a competitive market. For instance, a price cap of $1000/MWh can be viewed as a mandatory call option imposed on all produced electricity with a strike price of $1000/MWh, with the option premium being the proper capacity payment for generators abiding by the cap.

The organization of the rest of the paper is as follows. Section 2 describes the institutional features of several types of commonly traded electricity instruments. Section 3 highlights the essential elements in electricity derivative pricing and introduces the pricing methodologies. Section 4 illustrates the roles of these electricity instruments in risk management applications. Section 5 concludes.

2. Different types of electricity financial and physical instruments

This section reviews various electricity financial/physical instruments traded on the exchanges and over the counters. Most of the electricity futures and options on futures are traded on the New York Mercantile Exchange (NYMEX) [1]. However, the trading volume of electricity futures is less than electricity forwards traded in the over-the-counter (OTC) markets. A large variety of electricity derivatives are traded among market participants in the OTC markets, including forward contracts, swaps, plain vanilla options, and exotic (i.e. non-standard) options like spark spread options, swing options and swaptions [2–6]. Other important trading vehicles for hedging the price risk of long-term revenue streams and service obligations are termed as structured transactions, including tolling agreements [7,8] and load-serving full requirement contracts. The institutional details of these instruments are given below.

2.1. Electricity forwards, futures and swaps

The plainest forms of electricity derivatives are forwards, futures and swaps. Being traded either on the exchanges or over the counters, these power contracts play the primary roles in offering future price discovery and price certainty to generators and LSEs.

2.1.1. Electricity forwards

Electricity forward contracts represent the obligation to buy or sell a fixed amount of electricity at a pre-specified contract price, known as the forward price, at certain time in the future (called maturity or expiration time). In other words, electricity forwards are custom-tailored supply contracts between a buyer and a seller, where the buyer is obligated to take power and the seller is obligated to supply. The payoff of a forward contract promising to deliver one unit of electricity at price F at a future time T is

$$\text{Payoff of a Forward Contract} = (S_T - F) \tag{1}$$

where S_T is the electricity spot price at time T. Although the payoff function (1) appears to be the same as for any financial forwards, electricity forwards differ from other financial and commodity forward contracts in that the underlying electricity is a different commodity at different times. The settlement price S_T is usually calculated based on the average price of electricity over the delivery period at the maturity time T.

Consider a forward contract for the on-peak electricity on day T. 'On-peak electricity' refers to the electricity delivered over the daily peak-period, traditionally defined by the industry as 06:00–22:00. The daily 'off-peak' period is the remaining hours of the day. In this case, S_T is obtained by averaging the 16 hourly prices from 06:00 to 22:00 on day T.

Based on the delivery period during a day, electricity forwards can be categorized as forwards on on-peak electricity, off-peak electricity, or 'around-the-clock' (24 h per day) electricity. As almost all electricity derivatives have such categorization based on the delivery time of a day, we will not repeat this point.

Generators such as independent power producers (IPPs) are the natural sellers (or, short-side) of electricity forwards while LSEs such as utility companies often appear as the buyers (or, long-side). The maturity of an electricity forward contract ranges from hours to years although contracts with maturity beyond two years are not liquidly traded. Some electricity forwards are purely financial contracts, which are settled through financial payments based on certain market price index at maturity, while the rest are physical contracts as they are settled through physical delivery of underlying electricity. Examples of financially settled electricity forwards include the Contract for Differences in the United Kingdom and Australian power markets.

Electricity forwards with short maturity like 1 h or 1 day are often physical contracts, traded in the physical electricity markets such as the Pennsylvania–New Jersey–Maryland (PJM) power pool market and the energy balancing market operated by the California Independent System Operator (CAISO) in US. Those with maturity of weeks or months can be either physical contracts or financial contracts and they are mostly traded through brokers or directly among market participants (namely, traded in the OTC markets).

Electricity forward contracts are the primary instruments used in electricity price risk management. LSEs (e.g. local distribution companies) typically combine several months of forward/futures contracts to form a close match to the long-term load shape of their customers. Other power marketers usually use forwards to hedge their positions in electricity options and other complex electricity derivatives.

2.1.2. Electricity futures

First traded on the NYMEX in March 1996, electricity futures contracts have the same payoff structure as electricity forwards. However, electricity futures contracts, like other financial futures contracts, are highly standardized in contract specifications, trading locations, transaction requirements, and settlement procedures. The most notable difference between the specifications of electricity futures and those of forwards is the quantity of power to be delivered. The delivery quantity specified in electricity futures contracts is often significantly smaller than that in forward contracts.

For example, a Mid-Columbia electricity futures traded on the NYMEX specifies a delivery quantity of 432 MWh of firm electricity, delivered to the Mid-Columbia hub at a rate of 1 MW per hour, 16 on-peak hours per day during delivery month, while a corresponding forward contract has a delivery rate of 25 MW per hour for the same delivery periods in a month.

Electricity futures are exclusively traded on the organized exchanges, while electricity forwards are usually traded over-the-counter in the form of bilateral transactions. This fact makes the futures prices more reflective of higher market consensus and transparency than the forward prices. The majority of electricity futures contracts are settled by financial payments rather than physical delivery, which lower the transaction costs. In addition, credit risks and monitoring costs in trading futures are much lower than those in trading forwards, since exchanges implement strict margin requirements to ensure financial performance of all trading parties. The OTC transactions are vulnerable to financial non-performance due to counterparty defaults. The fact that the gains and losses of electricity futures are paid out daily, as opposed to being cumulated and paid out in a lump sum at maturity time, as in trading forwards, also reduces the credit risks in futures trading.

In summary, as compared to electricity forwards, the advantages of electricity futures lie in market consensus, price transparency, trading liquidity, and reduced transaction and monitoring costs while the limitations stem from the various basis risks associated with the rigidity in futures specification and the limited transaction quantities specified in the contracts.

2.1.3. Electricity swap

Electricity swaps are financial contracts that enable their holders to pay a fixed price for underlying electricity, regardless of the floating electricity price, or vice versa, over the contracted time period. They are typically established for a fixed quantity of power referenced to a variable spot price at either a generator's or a consumer's location. Electricity swaps are widely used in providing short- to medium-term price certainty up to a couple of years. They can be viewed as a strip of electricity forwards with multiple settlement dates and identical forward price for each settlement.

Electricity locational basis swaps are also commonly used to lock in a fixed price at a geographic location that is different from the delivery point of a futures contract. That is, a holder of an electricity locational basis swap agrees to either pay or receive the difference between a specified futures contract price and another locational spot price of interest for a fixed constant cash flow at the time of the transaction. These swaps are effective financial instruments for hedging the basis risk on the price difference between power prices at two different physical locations.

2.2. Electricity options

The power industry had been utilizing the idea of options through embedded terms and conditions in various supply and purchase contracts for decades, without explicitly recognizing and valuing the options until the beginning of the electricity industry restructuring in UK, US and the Nordic countries in the 1990s. The emergence of the electricity wholesale markets and the dissemination of option pricing and risk management techniques have created electricity options not only based on the underlying price attribute (as in the case with plain vanilla electricity call and put options), but also other attributes like volume, delivery location and timing, quality, and fuel type.

Basically, a counterpart of each financial option can be created in the domain of electricity options by replacing the underlying of a financial option with electricity (see [9] for introduction to various kinds of financial options). Here, we describe a sample of electricity options that are commonly utilized in risk management applications in generation and distribution sectors. These options usually have short- to medium maturity times such as months or a couple of years. Options with maturity times longer than 3 years are usually embedded in long-term supply or purchase contracts, which are termed as structured transactions.

2.2.1. Plain call and put options

Electricity call and put options offer their purchasers the right, but not the obligation, to buy or sell a fixed amount of underlying electricity at a pre-specified strike price by the option expiration time. They have similar payoff structures as those of regular call and put options on financial securities and other commodities. The payoff of an electricity call option is

$$\text{Payoff of an electricity call option} = \max(S_T - K, 0) \tag{2}$$

where S_T is the electricity spot price at time T and K is the strike price.

The underlying of electricity call and put options can be exchange-traded electricity futures or physical electricity delivered at major power transmission inter-ties, like the ones located at California–Oregon Border and Palo Verde in the Western US power grid. The majority of the transactions for electricity call and put options occur in the OTC markets. Electricity call and put options are the most effective tools available to merchant power plants and power marketers for hedging price risk because electricity generation capacities can be essentially viewed as call options on electricity, particularly when generation costs are fixed.

2.2.2. Spark spread options

An important class of non-standard electricity options is the spark spread option (or, spark spread). Spark spreads are cross-commodity options paying out the difference between the price of electricity sold by generators and the price of the fuels used to generate it. The amount of fuel that a generation asset requires to produce one unit of electricity depends on the asset's fuel efficiency or heat rate (Btu/kWh). The holder of a European- spark spread call option written on fuel G at a fixed heat rate K_{H} has the right, but not the obligation, to pay at the option's maturity K_{H} times the fuel price at maturity time T and receive the price of one unit of electricity. Thus, the payoff at maturity time T is

$$\text{Payoff of a spark spread call} = \max(S_T - K_{\mathrm{H}} \times G_T, 0) \tag{3}$$

where S_T and G_T are the electricity and fuel prices at time T, respectively.

Abstracting away the operational characteristics of a fossil fueled power generator (e.g. startup cost and ramping constraints), the per kW benefit of owning the right to use the generator is equivalent to having 1 kW spark spread call option with a strike heat rate matching the generator's operating heat rate. Based on this observation, it is clear that spark spread call options play important roles in hedging the price risk of the output electricity of fossil fueled power plants and further serve as key instruments in valuing those generation assets [10,11].

2.2.3. Callable and putable forwards

Two interesting types of electricity derivatives termed as callable forward and putable forward are introduced in Refs. [12,13] to mimic the interruptible supply contracts and the dispatchable independent power producer contracts. In a callable forward contract, the purchaser of the contract longs one forward contract and shorts one call option with a purchaser-selected strike price. The seller of the forward contract holds opposite positions and can exercise the call option if the electricity price exceeds the strike price, effectively canceling the forward contract at the time of delivery. The purchaser gets an 'interruptibility' discount on the forward price, which is equal to the option premium at the time of contracting continuously compounded to the delivery time.

In a putable forward, the purchaser longs one forward contract and one put option with a seller-selected strike price. The seller holds the corresponding short positions. The purchaser exercises the put option if the electricity price drops below the strike price at the maturity time, effectively canceling the forward contract. At the time of contracting, the purchaser needs to pay a 'capacity availability' premium over the forward energy price, which equals the put option price at that time, continuously compounded to the maturity time.

One variation of the callable forwards is proposed by adding an earlier notification date for exercising the call option in a callable forward before the contract matures [14,15]. This emulates an interruptible service contract with early notification [16].

2.2.4. Swing options

Electricity swing options are adopted from their well-known counterparts in the natural gas industry [5]. Also known as flexible nomination options, swing options have the following defining features. First, these options may be exercised daily or up to a limited number of days during the period in which exercise is allowed. Second, when exercising a swing option, the daily quantity may vary (or, swing) between a minimum daily volume and a maximum volume. However, the total quantity taken during a time period such as a week or a month needs to be within certain minimum and maximum volume levels. Third, the strike price of a swing option may be either fixed throughout its life or set at the beginning of each time period based on some pre-specified formula. Last, if the minimum-take quantity of any contract period is missed by the buyer, then a lump sum penalty or a payment making up the seller's revenue shortfall needs to be paid (i.e. take-or-pay).

2.3. Structured transactions

Structured bilateral transactions are powerful tools for power market participants to share and control a variety of risks including price and quantity risks over a potentially long time horizon.

2.3.1. Tolling contracts

Tolling is one of the most innovative structured transactions embraced by the power industry. A tolling agreement is similar to a common electricity supply contract signed between a buyer (e.g. a power marketer) and an owner of a power plant (e.g. an IPP) but with notable differences. For an upfront premium paid to the plant owner, it gives the buyer the right to either operate and control the scheduling the power plant with the ISO or simply take the output electricity during pre-specified time periods subject to certain constraints. In addition to inherent operational constraints of the underlying power plant, there are often other contractual limitations in the contract on how the buyer may operate the power plant or take the output electricity. For instance, a tolling contract almost always has a clause on the maximum allowable number of power plant restarts. These constraints make the pricing of tolling contracts a very challenging task. The analogy between holding a tolling contract and owning the underlying merchant power plant, however, leads to a numerical approach for valuing and hedging tolling contracts [7]. Alternatively, one may use a statistical approach for benchmarking the price reasonableness of tolling contracts based on historical electricity price and fuel costs [8].

2.3.2. Load-serving full-requirement contracts

Most large electricity consumers prefer a power supply contract with flexible consumption terms. Specifically, they desire to pay a fixed rate per unit of energy for the actual consumption quantity, regardless of the quantity being high or low. Such a contract is termed as a load-serving full-requirement contract.

Suppose an electricity supplier (or, LSE) signs a full-requirement contract with a customer and then utilizes futures contracts to lock in a fixed quantity of electricity supply at a fixed cost for hedging the expected energy consumption of the customer [17,18]. The LSE is then at the risk of either under- or over-hedging, as the consumption quantity of the customer will almost surely deviate from the amount hedged by the futures contracts. When the electricity spot price is high (low), the total demand for electricity is likely to be high (low) as well. A case in point is the periods of unusual cooling/heating needs. Hence, if the market price of electricity is higher than the fixed contract rate for serving electricity, chances are that the customer's energy consumption level is significantly higher than the hedged quantity. As a result, the LSE is under-hedged relative to its load obligation and must purchase electricity in the open market to serve its customer at a loss because the wholesale spot price most likely exceeds the contracted price paid by consumers. Conversely, when the electricity spot price is low, the LSE faces the risk of being over-hedged and having to sell the surplus in the spot market or settle it financially at a price below its long-term contract price.

The above illustrates the under- and over-hedging exposures faced by an LSE due to the volumetric uncertainty in customers' load and the positive price-load correlation. To hedge the volumetric risk, the LSE would need to buy an electricity option on the consumption quantity of its customers. Unfortunately, such an option is usually unavailable in the marketplace. Although perfect hedging may not be possible, weather derivatives [19,20] that exploit the correlation between load and temperature can be used. Section 4.4 describes another approach based on an optimal hedging portfolio of standard derivatives that exploits the positive correlation between power prices and consumption quantity [21].

2.4. Financial derivatives on electricity transmission capacity

Open access to, efficient utilization of, and adequate investment in transmission networks are critical for the electricity wholesale markets and retail competitions to be workable and efficient. Intuitively, rights are required for using transmission networks and rules are needed for rationing transmission usage when networks become congested. There are two major proposals for using financial instruments as transmission rights in US: (a) the point-to-point financial transmission rights (FTRs) [22–24]; and (b) the flowgate rights (FGRs) [25,26], as outlined in the Standard Market Design (SMD) put forth by the Federal Energy Regulatory Commission (FERC). FTRs and FGRs are electricity derivatives, with their values derived from the network transmission capacity.

2.4.1. FTR and FTR options

In an electricity market such as the PJM that employs locational market price (LMP), a point-to-point FTR is specified over any two locations in the power transmission grid. An FTR entitles its holder to receive compensation (or pay) for transmission congestion charges that arise when the grid is congested. The congestion charge/payment (or, payoff) associated with one unit of FTR is equal to the difference between the two locational prices of one unit of electricity resulting from the re-dispatch of generators out of merit order to relieve transmission congestion. The primary markets for the FTR trading are auctions held by the independent system operators (ISOs) of power markets.

An FTR option offers the right to the FTR settlement without the obligation to pay when that settlement is negative. Hence, the settlement of an FTR option equals to the positive part of the corresponding two-sided point-to-point FTR.

2.4.2. FGRs

Flowgates are defined over all transmission elements such as lines, transformers, or linear combinations of them. Each transmission element has two elemental flowgates, one in each direction. An elemental flowgate has a rated capacity in megawatts in its pre-specified direction corresponding to the capacity of an underlying transmission element. Thus, flowgate rights are link-based transmission rights for hedging transmission risks. The values of flowgate rights can be established through auctions conducted by the ISOs. The spot price upon which the settlement of flowgate rights is based is given by the real time shadow price on the corresponding constrained element, determined by the security constrained economic dispatch algorithm employed by an ISO. Since these shadow prices are nonnegative, FGRs are inherently defined as options.

3. Pricing electricity derivatives

Since the value of electricity derivatives are based on the underlying electricity prices, modeling electricity price is the most critical component in pricing electricity derivatives. Due to the unique physical and operational characteristics of electricity production and transmission processes, electricity price exhibits different behaviors than other financial prices which can be often described by Geometric Brownian Motion. There has been a growing literature addressing mainly two competing approaches to the problem of modeling electricity price processes:

(a) ‘Fundamental approach’ that relies on simulation of system and market operation to arrive at market prices; and
(b) ‘Technical approach’ that attempts to model directly the stochastic behavior of market prices from historical data and statistical analysis.

While the first approach provides more realistic system and transmission network modeling under specific scenarios, it is computationally prohibitive due to the large number of scenarios that must be considered. Such analysis may be necessary for pricing financial transmission rights (in particular, flowgate rights) but not for the other electricity derivatives. Therefore, we shall focus our attentions on the second approach and review the corresponding methodologies for pricing electricity derivatives.

Approaches to characterize market prices include discrete-time time series models such as GARCH and its variants [27–32], Markov regime-switching models [33], continuous-time diffusion models such as mean-reversion [11,34,35], jump-diffusion [2,3,36], and other diffusion models [37,38]. There are also models proposed for direct modeling of electricity forward curves [39,40].

While a straightforward application of the maximum likelihood estimation (MLE) method yields the parameter estimates of a discrete-time time series model, it does not yield analytic expressions for derivative prices. In fact, Monte Carlo simulation and lattice-based approaches are the only feasible derivative pricing methods under time-series price models. For continuous-time diffusion models, model parameters can be estimated by applying moment-based methods, such as the generalized method of moments, which may not be as efficient as the MLE method. Nonetheless, more option pricing methods (e.g. the analytic solution approach and the partial differential equation (PDE) approach) become applicable under the diffusion price models.

Deng [3] was the first to employ a multifactor affine jump diffusion (AJD) processes to model electricity spot prices under several specifications, including regime switching and stochastic volatility. Under the assumption that electricity prices follow AJD processes, an extended Fourier transform technique developed in Ref. [41] can be applied to derive analytic expressions (up to Fourier inversion) for a variety of derivative prices. Specifically, prices of forwards, calls/puts and spark spreads were derived in Ref. [3] under three different electricity price models, and prices of callable forwards with an early notification were obtained in Ref. [14].

When there is a large set of market data available, the most appropriate approach to pricing electricity options is to infer the risk-neutral distribution of the underlying electricity price from the market data and then obtain the prices of the electricity derivatives based on the premise of no-arbitrage. If there is not enough forward-looking market information for implementing a no-arbitrage pricing model, then equilibrium models can be applied to obtain derivative prices, as in Refs. [31,34,40,42,43] for forward prices and [44] for spark spreads. In certain cases, statistical benchmark analysis based on historical data can provide a sense of the reasonableness on the electricity options prices [8].

The binomial/multinomial lattice and Monte Carlo simulation methods are powerful numerical tools for pricing electricity options with complex structures and/or under a complicated model for the electricity price process. For instance, given the complex structure of a swing option or a tolling contract, it is impossible to obtain prices of such contracts either in closed-forms or through PDEs. Thus, swing options are priced by lattice models [45,46], or by approximation methods for obtaining price lower bounds [47]. The pricing of tolling contracts requires a combination of Monte Carlo simulation with dynamic programming [7].

4. Risk management applications

4.1. Hedging a generator's output

Albeit having simple payoff structures, forwards, swaps, and call options are effective tools for a generator with fixed per unit cost to lock in profits by selling forwards, fixed-price swaps, and call options on electricity. When the forward/swap rate or the strike price of the call options is higher than the fixed cost, the generator's profits are guaranteed.

However, if the generating costs are market-based (e.g. a natural gas fired merchant power plant that burns natural gas at market price), the selling forwards, swaps and calls will expose the generator to potential fuel cost increases. In such a case, a properly constructed portfolio of spark spread calls would be the right tool for hedging a generator's revenue stream over a given time period.

The operational efficiency of a natural gas fired power plant is characterized by its operating heat rate. Therefore, the financial benefit of owning a portfolio of spark spread calls with strike heat rates identical to the operating heat rate of the plant is the same as owning the power plant during the time period of the options' maturity times. This observation leads to the valuation and hedging method for generation capacity proposed in Refs. [10,11]. When taking into account the operational characteristics, lattice-based method [48] and simulation method [35] are necessary to determine pricing and hedging strategies of generation capacity.

In the case, where the electricity forward market at the generator's location is not liquidly traded, electricity forwards from adjacent trading hubs or even forwards on the input fuel, which are liquidly traded, can be utilized to cross-hedge the electricity output price [49,50].

4.2. Ensuring generation adequacy

Oren [51,52] and Chao and Wilson [53] propose a new role for options with long maturity to address the resource adequacy problem. They propose a scheme for ensuring generation adequacy via call options as obligations imposed on the LSEs. Call options provide an attractive alternative to artificial capacity products such as installed capacity (ICAP) employed in New York, New England, and PJM, whose demand is based only on administrative requirements and which have no intrinsic value. By requiring LSEs to purchase a proper portfolio of options, a regulator can achieve spot price volatility reduction by implementing price insurance while using the premium to stabilize generators' income and enhance investment incentives.

4.3. Callable forwards and interruptible service contracts

The restructured electricity markets have shown little demand response to price spikes. The enormous price volatility affirms the need for demand responsiveness to make these markets workable. As load curtailment can provide an efficient substitute for generation capacity in meeting balancing energy and reserves needs, flexible loads are viable and valuable resources in taming price volatility.

Consider the traditional utility interruptible service contracts utilized in demand-side management (DSM) to mitigate supply shortages. These interruptible contracts are readily implementable through standard electricity derivatives [12–14]. For instance, a synthetic interruptible service contract offered by an LSE is a callable forward under which the LSE sells a forward to and buys a call option from its customer. Furthermore, with a liquid electricity derivative market, the discounts offered to the interrupted services would be set through market trading instead of bilateral negotiations thus making the pricing of the interruptible services more transparent and efficient.

4.4. Hedging congestion risk of bilateral transactions

From the perspective of new power network transmission users, FTRs can be viewed as an instrument for hedging their exposure to congestion cost risk. A 1-MW bilateral transaction between two points in a transmission network is charged (or credited) the nodal price difference between the point of withdrawal and the point of injection. At the same time (assuming that transmission rights are fully funded), a 1 MW FTR between two points is an entitlement (or obligation) for the difference between the nodal prices at the withdrawal node and the injection node. Thus regardless of how the system is dispatched, a 1 MW FTR between two nodes is a perfect hedge against the uncertain congestion charge between the same two nodes.

The hedging properties of FTRs make them ideal instruments for converting historical entitlements to firm transmission capacity into tradable entitlements that hold the owners of such entitlements harmless, while enabling them to cash out when someone else can make more efficient use of the transmission capacity covered by these entitlements. In other words, FTRs make it relatively easy to preserve the status quo while opening up the transmission system to new and more efficient use. A word of caution is that the hedging function of FTRs may not be perfect due to changing network operating conditions and potential inherent trading inefficiency [54]. Some ISOs derate FTR settlements in order to cover congestion revenue shortfalls due to transmission contingencies not accounted for in the FTR auction. In such cases, depending on the derating approach, FTRs may not provide perfect hedges either.

4.5. Hedging volumetric risks

LSEs providing electricity service at regulated prices in restructured electricity markets are wary of both price and quantity risks [17,18]. As the electricity markets are inherently incomplete, the quantity risk cannot be perfectly hedged. Commonly proposed hedging alternatives include the implementation of a minimal variance hedge through purchasing electricity forwards [18] and the utilization of weather derivatives.

Recent work reported in Ref. [21] addresses the problem of hedging volumetric risks by risk-averse LSEs, whose hedging objective is to maximize a concave utility function. Exploiting the correlation between consumption quantities and spot prices, the authors developed an optimal, zero-cost hedging function described by a payoff function of spot price. They also demonstrate how such a hedging strategy can be implemented through a portfolio of standard forwards and a spectrum of call and put options with various strike prices.

5. Conclusion

In electricity market restructuring, electricity derivatives play an important role in establishing price signals, providing price discovery, facilitating effective risk management, inducing capacity investments in generation and transmission, and enabling capital formation. Custom design of electricity financial instruments and structured transactions can provide energy price certainty, hedge volumetric risk, synthesize generation and transmission capacity, and implement interruptible service contracts.

Admittedly, many exotic forms of electricity options can meet specific needs for hedging and speculation. However, we emphasize the importance of standardization. Future research should focus on identifying standardized electricity derivatives and utilization of financial engineering tools to synthesize and replicate alternative contracts using standardized instruments. Such standardization will reduce transaction costs and produce liquidity, which in turn will improve the efficiency of risk management practices.

Acknowledgements

This research is supported in part by the NSF Grant ECS-0134210 (Deng), National Science Foundation Grant ECS-0224779 (Oren) and a Power System Engineering Research Center (PSERC) grant (Deng and Oren).

References

[1] Johnson B, Sogomonian A. Electricity futures. The US power market. London: Risk Publications; 1997 p. 83–98.

[2] Clewlow L, Strickland C. Energy derivatives: pricing and risk management. London: Lacima Publications; 2000.

[3] Deng SJ. Stochastic models of energy commodity prices and their applications: mean-reversion with jumps and spikes. POWER working paper, University of California Energy Institute, Berkeley; 1999.

[4] Eydeland A, Wolyniec K. Energy and power risk management. Chichester: Wiley; 2003.

[5] Kaminski V. The challenge of pricing and risk managing electricity derivatives. The US power market. London: Risk publications; 1997.

[6] Pilipovic D. Energy risk: valuing and managing energy derivatives. New York: McGraw-Hill; 1998.

[7] Deng SJ, Xia ZD. Pricing and hedging power supply contracts: the case with tolling agreements. Working paper, Georgia Institute of Technology; 2003.

[8] Woo CK, Olson A, Orans R. Benchmarking the price reasonableness of an electricity tolling agreement. Electricity J 2004; 17(5):65–75.

[9] Hull J. Options, futures, and other derivative. 4th ed. Englewood, NJ: Prentice-Hall; 2000.

[10] Deng SJ. Valuation of investment and the opportunity to invest in power generation assets with spikes in power prices. Manage Financ 2005;31(6):95–115.

[11] Deng SJ, Johnson B, Sogomonian A. Exotic electricity options and the valuation of electricity generation and transmission assets. Decis Support Syst 2001;30(3):383–92.

[12] Gedra TW. Optional forward contracts for electric power service contracts. PhD Thesis, University of California, Berkeley; 1991.

[13] Gedra TW, Pravin PV. Markets and pricing for interruptible electric power transactions. IEEE Trans Power Syst 1992;92 [WM 169–3 PWRS].

[14] Kamat R, Oren SS. Exotic options for interruptible electricity supply contracts. Oper Res 2002;50(5):835–50.

[15] Oren SS. Integrating real and financial options in demand-side electricity contracts. Decis Support Syst 2001;V30:279–88.

[16] Strauss TP, Oren SS. Priority pricing of interruptible electric power with an early notification option. Energy J 1993;14(2): 175–95.

[17] Alexander BR. Managing default service to provide consumer benefits in restructured states: avoiding short-term price volatility. Butte, Montana: National Center for Appropriate Technology; 2003.

[18] Woo CK, Karimov R, Horowitz I. Managing electricity procurement cost and risk by a local distribution company. Energy Policy 2004;32(5):635–45.

[19] Cao M, Wei J. Pricing the weather. Risk Mag 2000;67–70.

[20] Pirrong C, Jermakyan M. The price of power: The valuation of power and weather derivatives. Working paper, Oklahoma State University; 2001.

[21] Oum Y, Deng SJ, Oren SS. Volumetric hedging in electricity procurement. Working paper, UC Berkeley; October 2004.

[22] Harvey SM, Hogan WW, Pope SL. Transmission capacity reservations and transmission congestion contracts. Cambridge, MA: John F. Kennedy School of Government, Harvard University; 1997.

[23] Hogan WW. Contract networks for electric power transmission. J Regul Econ 1992;4:211–42.

[24] Hogan WW. Financial transmission right formulations. Cambridge, MA: Center for Business and Government, John F. Kennedy School of Government, Harvard University; 2002.

[25] Chao H-P, Peck S. A market mechanism for electric power transmission. J Regul Econ 1996;10(1):25–60.

[26] Chao H-P, Peck S, Oren SS, Wilson R. Flow-based transmission rights and congestion management. Electricity J 2000; 13(8):38–58.

[27] Bunn DW, editor. Modelling prices in competitive electricity markets. Chichester: Wiley; 2004.
[28] Duffie D, Gray S. Volatility in energy prices. Manage energy price risk. 2nd ed. London: Risk Publications; 1998.
[29] Goto M, Karolyi GA. Understanding electricity price volatility within and across markets. Working paper, Ohio State University; 2003.
[30] Knittel C, Roberts M. An empirical examination of deregulated electricity prices. Working paper, University of California Energy Institute; 2001.
[31] Longstaff FA, Wang AW. Electricity forward prices: a high-frequency empirical analysis. J Finance 2004;59(4): 1877–900.
[32] Worthington AC, Kay-Spratley A, Higgs H. Transmission of prices and price volatility in Australian electricity spot markets: a multivariate GARCH analysis. Energy Econ 2003.
[33] Mount T, Ethier R. Estimating the volatility of spot prices in restructured electricity markets and the implications for option values. PSerc Working paper, Cornell University; 1998.
[34] Lucia JJ, Schwartz ES. Electricity prices and power derivatives: evidence from the Nordic power exchange. Rev Deriv Res 2002;5(1):5–50.
[35] Tseng CL, Barz G. Short-term generation asset valuation: a real options approach. Oper Res 2002;50(2):297–310.
[36] Barz G, Johnson B. Modeling the prices of commodities that are costly to store: the case of electricity. Proceedings of the Chicago Risk Management Conference, Chicago, IL 1998.
[37] Barlow M. A diffusion model for electricity prices. Math Finance 2002;12(4):287–98.
[38] Deng SJ, Jiang WJ. Levy process driven mean-reverting electricity price model: a marginal distribution analysis. Working paper, Georgia Institute of Technology; 2002.
[39] Audet N, Heiskanen P, Keppo J, Vehvilainen I. Moleling electricity forward curve dynamics in the Nordic market. Working paper, University of Michigan; 2003.
[40] Bjork T, Landen, C. On the term structure of futures and forward prices. SSE/EFI working paper series in economics and finance no. 417; 2000.
[41] Duffie D, Pan J, Singleton K. Transform analysis and asset pricing for affine jump-diffusion. Econometrica 2000;68: 1343–76.
[42] Bessembinder H, Lemmon ML. Equilibrium pricing and optimal hedging in electricity forward markets. J Finance 2002; 57:1347–82.
[43] Kellerhals BP. Pricing electricity forwards under stochastic volatility. Working paper, POWER, Department of Finance, College of Economics and Business Administration, Eberhard Karls University, Tubingen; 2001.
[44] Routledge B, Seppi D, Spatt CS. Equilibrium forward curves for commodities. J Finance 2000;55:1297–338.
[45] Jaillet P, Ronn E, Tompaidis S. Valuation commodity based swing options. Working paper, University of Texas, Austin; 2001.
[46] Thompson AC. Valuation of path-dependent contingent claims with multiple exercise decisions over time: the case of take-or-pay. J Financial Quant Anal 1995;30:271–93.
[47] Keppo J. Pricing of electricity swing options. Preprint, University of Michigan 2002.
[48] Deng SJ, Oren SS. Incorporating operational characteristics and startup costs in option-based valuation of power generation capacity. Probab Eng Inform Sci 2003;17(2):155–81.
[49] Emery G, Liu QW. An analysis of the relationship between electricity and natural gas futures prices. J Futures Markets 2002;22(2):95–122.
[50] Woo CK, Horowitz I, Hoang K. Cross hedging and forward-contract pricing of electricity. Energy Econ 2001;23:1–15.
[51] Oren SS. Capacity payments and generation adequacy in competitive electricity markets. Proceedings of SEPOPE IIV conference, Curitiba, Brazil, May 2000;22–6.
[52] Oren SS. Ensuring generation adequacy in competitive electricity markets. In press, University of Chicago Press.
[53] Chao H-P, Wilson R. Resource adequacy and market power mitigation via option contracts. Proceeding of the POWER Conference, University of California Energy Institute, Berkeley, CA 2004.
[54] Deng SJ, Meliopoulos SA, Oren SS. The inherent inefficiency of simultaneously feasible financial transmission rights auctions. Working paper, Georgia Institute of Technology; 2004.