Master in Financial Engineering (EPFL)

Course: Financial Econometrics

Case study 1: Modeling a financial time series, cointegration, and pair trading (Statarb)

Exercise 1: Modeling the 5-year interest rate

Using the excel file exercise_1_case_study_1.xls and the 5-year interest rate series (say, X_t) over the period 1960Q1-2013Q3, one would like to propose an ARIMA(p,d,q) specification. In so doing, one proceeds in three steps: (1) (Partial) identification of the autoregressive order (p), the moving average order (q) and the differentiation order (d); (2) Estimation and selection of some ARIMA models; (3) Implementation of some diagnostic tests (especially, for the error terms).

Step 1: Identification of orders

1. Plot the times series (X_t).
2. Identification of d:
   (a) Display the autocorrelation function of (X_t) and $\Delta X_t = X_t - X_{t-1}$. Interpret the results: (1) What do you observe? (2) Using this informal graphical procedure, is there some evidence that the series (X_t) might be non-stationary? Explain carefully.
   (b) To provide more evidence, a unit root test, the so-called Augmented Dickey-Fuller test (`ADF`), is implemented (see Appendix 1) in order to assess whether there exists a root on the unit circle in the autoregressive lag polynomial of an AR(p) specification. The `ADF` test is based on estimating the following test regression:

$$X_t = \sum_{j=1}^{p} \rho_j X_{t-j} + z_t' \delta + \epsilon_t$$

   where (ϵ_t) is a weak white noise, z_t is a set of exogenous regressors $(z_t = \{1, t\}, z_t = \{t\},$ or $z_t = \{1\})$, p is the lag order (that can be determined by differen techniques) and $(\rho_1, \cdots, \rho_p)'$ and δ are parameters to estimate. Using the so-called Beveridge-Nelson decomposition, the test regression can be rewritten as follows

$$X_t = (1 - \Phi(1)) X_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta X_{t-j} + z_t' \delta + \epsilon_t$$

where $\Phi(1) = 1 - \sum_{j=1}^{p} \phi_j$. It can also be written using a first-difference transformation: note however that Matlab uses the first specification (level-based specification). The specification of the test is:

$$H_0 : \text{The series is non-stationary—}\Phi(1) = 1$$
$$H_a : \text{The series is "weakly stationary"—}|\Phi(1)| < 1.$$

Using $z_t = 1$ for all t (i.e., only a constant term), $p = 5$ and the *adf* matlab function, can one conclude that there is some evidence of the presence of unit root?

Remark: As a confirmatory-based analysis, one can conduct a unit root test on (ΔX_t) or make use of the `KPSS` test (the null hypothesis being the one of weak stationarity). Note also that some more powerful unit root tests exist (see Appendix 1).

(c) Which order do you suggest for d?

3. (Partial) Identification of (p, q)

   – Using the result of Question 2, plot the ACF and PACF of the relevant series (X_t or ΔX_t). Provide some upper bounds for p and q, say $p_{\max}$ and $q_{\max}$.
   Hint: (1) Upper bound of the moving average part (respectively, autoregressive part) can be achieved by looking at the `ACF` (respectively, `PACF`); (2) By default, Matlab provides some "conservative" confidence bands (the true ones being larger in the presence of a moving average component).

Step 2: Estimation and selection of ARIMA Models

At this stage, the differentiation order is assumed to be known (Question 2), and $0 \leq p \leq p_{\max}$ and $0 \leq q \leq q_{\max}$. In the sequel, one assumes $p_{\max} = 4$ and $q_{\max} = 4$. Furthermore, AR(p), MA(q) and weak white noise specifications rules out, i.e. one rules out the following (p, q) combinations $(0, 0)$, $(0, q)$ and $(p, 0)$.

4. Compute the `AIC` (Akaike's information criterion) and `BIC` (Bayesian information criterion) for the 16 remaining models—see Appendix 2 for information criteria.

   Hint: Among other methods, this can be done as follows

   – Define the competing models
   – Fit models to the data
   – Compute the `AIC` (or the small-sample size correction of the `AIC`) and `BIC` for each model
   – Example: See Appendix 3.

5. Provide a table with the estimates of your best three specifications.
   Hint: Select the models that minimize each information criteria. Note that:

- The statistical properties of these two information criteria are different (especially, in terms of consistency and efficiency). It turns out that these two criteria might select different models.
- Instead of selecting a unique model (specification) with each criterion, it is often better to select a model that minimizes the corresponding information criterion but also models that are "adjacent" (i.e., models that have a value of the information criterion close to the minimal one).

Step 3: Diagnostic tests

6. Among the different tests of specifications, one would like to test the absence of correlations of the error terms. Using the results of Question 5, conduct a Ljung-Box test for the first 10 lags. What can you conclude?

7. All in all, Which model is your preferred specification?

   Hint: Make use of all previous results as well as the principle of parsimony (i.e., among two competing models, one would generally select the one with less parameters especially with respect to the moving average component).

Exercise 2: The term structure of interest rates

The expectations hypothesis of the term structure of interest rates claims that the following relationship between the m-year and n-year interest rate is satisfied (with $m < n$):

$$y_t^{(n)} = \alpha + \beta y_t^{(m)} + u_t \tag{1}$$

where u_t is the (i.i.d.) error term, and α is the term premium. The pure expectations hypothesis predicts that $\beta = 1$. Using the excel file exercise_2_case_study_1.xls (same data as in Homework 6) and the time period 1960Q1-2013Q4, answer the following questions.

Part I: OLS estimation (without controlling for the order of integration of the two variables)

1. Plot the 1-year interest rate and the 5-year interest rate. What do you observe?

2. Estimate Equation 1 using OLS—The dependent variable being the 5-year interest rate and the explanatory variable being the 1-year interest rate. Interpret the results (statistical significance of the coefficients, sign of the coefficients, the coefficient of determination, and the financial interpretation of the estimated relationship).

Part II: Error correction model

3. Compute the ACF and PACF of the dependent (respectively, explanatory) variable (using the first 10 lags). Discuss the stationarity properties of each series.

4. Test the null of non-stationarity for both series using the augmented Dickey-Fuller with a constant and no time trend and p = 3 lags. Interpret the results.

5. Compute the residuals of the OLS regression (Question 1). Using an augmented Dickey-Fuller test, show that there is a cointegration relationship between the two variables (i.e., the null of non-stationarity is rejected).
   Hint: If one can reject the null hypothesis of non-stationarity, it means that the two variables are cointegrated (taking that the two variables of interest are integrated of order one). In this respect, there exists a linear combination of the two variables which is weakly stationary and there is a long-run (cointegration) relationship between the two variables (given by Equation 1).

6. Given the previous results, estimate an error correction model in which the dependent variable is ($\Delta y_t^{(5y)}$) using the two-step method of Engle and Granger.
   Hint: A simple two-step procedure can be implemented as follows: (1) Estimate Eq. 1 by OLS; (2) Compute the OLS residuals and especially $\hat{u}_{t-1}$; (3) Conduct the following OLS regression:

   $$\Delta y_t^{(5y)} = c + \lambda \hat{u}_{t-1} + \sum_{j=1}^{k_1} a_j \Delta y_{t-j}^{(5y)} + \sum_{j=1}^{k_2} b_j \Delta y_{t-j}^{(1y)} + \epsilon_t$$

   where the parameters are c, λ (< 0), $\{a_j\}_{j=1,\cdots,k_1}$, and $\{b_j\}_{j=1,\cdots,k_2}$, and the lag orders k_1 and k_2 can be determined by using information criteria or a general-to-specific procedure (with $k_1 \leq 3$ and $k_2 \leq 3$).

7. Interpret the results of Question 6 (the speed of adjustment, the short-run parameter estimates, and the financial interpretation of the estimated equation).

Exercise 3: Pair trading or Statarb
Consider a pair of daily stock (log-)prices, say (p_{1t}, p_{2t}). Suppose that these two prices are cointegrated.

- How can we us an error correction model to implement a "simple" financial strategy? Explain carefully.
- Find such a pair among your favorite stocks in a given industry, and estimate the corresponding model.