Master in Financial Engineering (EPFL)

Course: Financial Econometrics

A Primer on Time Series Analysis

Professor: F. Pelgrin

These lecture notes briefly summarize some important concepts in time series analysis.[1]

1 Definition of a time series

Suppose that one is interested in a variable (say, a price) or a transformation of that variable (say, a return or log-return) over a given time period: Asset price, asset return, interest rates, exchange rates, bond yields, dividend returns, earnings returns, etc. This variable of interest (the realizations) is observed at different (regular) periods ($t = t_1, \cdots, t_k$) and the time elapsed between two realizations is constant (say, daily, monthly, quarterly, yearly observations).[2]

Taking the frequency of the data, the sequence of realizations is a time series, i.e. the time-ordered sequence of observations, say $x_1, x_2, \cdots, x_T$, is a time series. Moreover, the data generating process underlying these realizations is called a stochastic process, which is the building block of time series theory. More specifically, a real (univariate) stochastic process can be defined as follows.

Definition 1. *A real-valued (discrete-time) stochastic process is a sequence of random variables indexed by $t \in \mathbb{Z}$ on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$:*

$$\begin{aligned} &\mathbb{Z} \times \Omega \to \mathbb{R} \\ (t, \omega) \quad &\mapsto X(t, \omega) = X_t(\omega). \end{aligned}$$

where Ω is the sample space, $\omega \in \Omega$ is a state of the nature such that $x_t = X_t(\omega)$, $\mathcal{A}$ is a σ-algebra, and $\mathbb{P}$ is a probability measure.

For instance, one can define the following stochastic processes (for a given ω):

- For all $t \in \mathbb{Z}$, $X_t = \phi X_{t-1} + \epsilon_t$ where ϵ_t is a weak white noise;
- For all $t \in \mathbb{Z}$, $X_t = \epsilon_t + \theta \epsilon_{t-1}$ where ϵ_t is a weak white noise;
- For all $t \geq 0$, $X_t = X_{t-1} + \epsilon_t$ where ϵ_t is a weak white noise.

[1] For a more advanced overview, see Tsay (2002) or Hamilton (2005).

[2] In practice, this is not always the case (e.g., tick-by-tick data or transaction time).

It is worth noticing that one might interpret such a (discrete-time) stochastic process as being a "kind" of stochastic difference equation.

Using Definition (1), one can define a realization of a stochastic process for a given initial state of the world (or a given event ω) and thus the concept of time series (or chronological series).

Definition 2. *A realization of the stochastic process $(X_t)_{t\in\mathbb{Z}}$ for a given $\omega \in \Omega$ is the mapping defined by :*

$$\begin{aligned}&\mathbb{Z} \to \mathbb{R}\\ &t \mapsto x_t(\omega).\end{aligned}$$

The realization of a stochastic process is said to be a time series or a chronological series.
□

2 Stationarity

In order to propose some statistical models, a time series must be "enough regular", meaning that some of its features can be explained by a proper modeling. Among these properties, the concept of stationary is essential.

Briefly speaking, stationarity "means" that **some essential properties of a time series remain constant over time** (as for instance, the mean, the variance, and the autocovariances). Depending on whether all characteristics (as for instance, all moments) or only some particular ones (as for instance, the first and second moments, and the covariances) are of interest, different types of stationarity can be defined: strong stationarity, weak stationarity, etc.

To provide some intuition, Figure 1 plots the S&P500 composite index.[3] The series is clearly not mean-reverting and it might display a (stochastic) trend. This series looks like a non-stationary series. On the other, using the first-difference of the initial series in Figure 2, i.e. $\Delta X_t = X_t - X_{t-1}$ where $X_t = \log(P_t)$, the (log-) return is mean-reverting and looks like a (weakly) stationary series.

[3] A second intuition is provided in Appendix 1.

Figure 1: Time series of the S&P500 (January 1957 - December 2012)

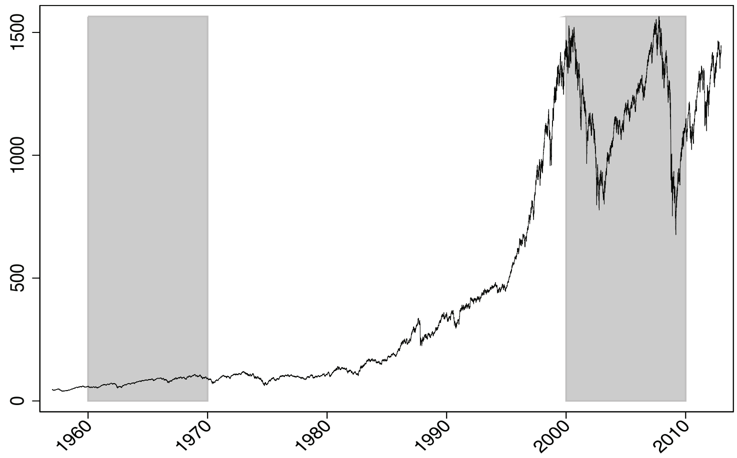
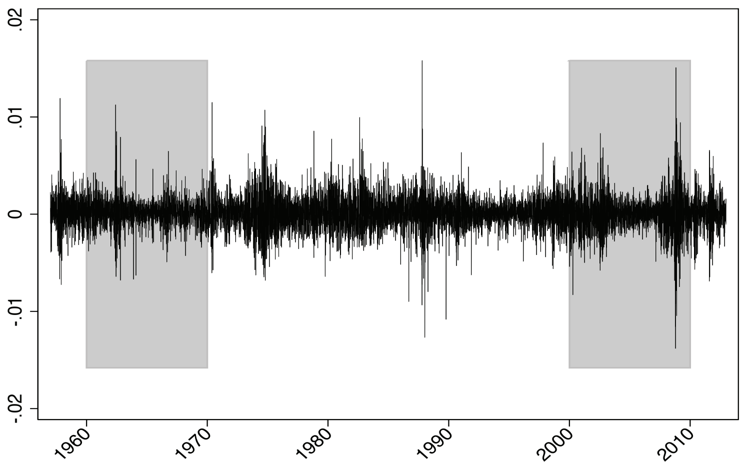
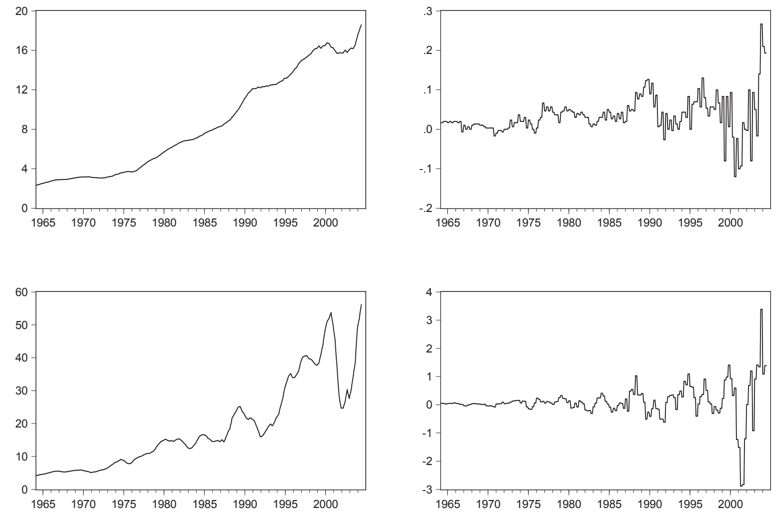


Figure 2: Time series of the S&P500 log returns (January 1957 - December 2012)



The same patterns are observed when comparing the dynamics of dividends (top panel) or of earnings (bottom panel). The left panel provides the initial series (i.e., in level) whereas the right panel displays the first-difference of the initial series.

Figure 3: Dividend and earnings (1965M1 - 2005M12)



In both examples, the statistical properties of the initial series and the transformed series are rather different. This has some key implications regarding the modeling of financial time series, estimation methods, etc. Therefore, it is critical to disentangle the nature of the time series, and especially whether it is weakly stationary or non-stationary. In this respect, weak (second-order or covariance) stationarity can be defined as follows.[4]

Definition 4 (Weak stationarity)**.** *A stochastic process (X_t) is weakly stationary, covariance or second-order stationary if it satisfies the following properties:*

1. $\mathbb{E}(X_t) = m$ *is independent of* t*;*
2. $\mathbb{V}(X_t)$ *is time-invariant.*
3. $\mathbb{C}ov(X_t, X_{t+h}) = \mathbb{E}\left[(X_t - m)(X_{t+h} - m)\right] = \gamma_X(h)$ *is time-invariant.*

Say differently, a process is weakly stationary if both the population mean and the population variance are constant over time and if the covariance between two realizations (observations) is a function only of the distance between them and thus does not depend on time. At the same time, weak stationarity exploits the "stability" of the first two moments whereas strong stationarity implies the stability of all the moments (among others). This means that strong stationarity implies weak stationarity as long as the first two moments exist. The converse is not true in general.[5]

Example 1: One prominent example of a second-order stationary series is a weak white noise. More specifically, a weak white noise process (see Figure 4), (ϵ_t), satisfies the following properties

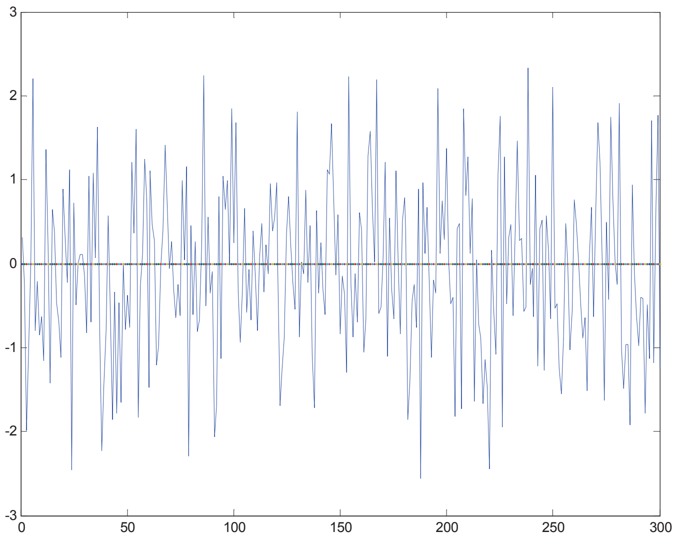
[4] Weak stationarity is often opposed to strong stationary.

Definition 3 (Strong stationarity)**.** *A stochastic process (or a time series) $(X_t)_{t\in\mathbb{Z}}$ is said to be strongly or strictly stationary if the distribution of $(X_t)_{t\in\mathbb{Z}}$ is identical to that of $(Y_t)_{t\in\mathbb{Z}}$ with $Y_t = X_{t+h}$.*
□

[5] It is worth noticing that if (X_t) is a Gaussian stochastic process, weak stationarity is equivalent to strong stationarity.

1. $\mathbb{E}(\epsilon_t) = 0$ for all t;
2. $\mathbb{V}(\epsilon_t) = \sigma_\epsilon^2 < +\infty$ for all t.
3. $\mathbb{C}ov(\epsilon_t, \epsilon_{t-h}) = 0$ for $h \neq 0$

Figure 4: White noise process



A quick visual inspection reveals that the series looks like a mean-reverting series (fluctuations around a zero-mean here). Moreover, there is no specific pattern regarding the autocorrelation and fluctuations behave "regularly" (i.e., the variance might not depend on time).

Example 2: Any (financial) time series that can be written as a linear combination of a weak white noise process is also weakly stationary :

$$X_t = \mu + \sum_{k\geq 1} a_k \epsilon_{t-k}.$$

Example 3: A fundamental (discrete-time) stochastic process in finance is the so-called random walk with or without drift (i.e., a constant):

$$\begin{aligned} X_t &= \mu + X_{t-1} + \epsilon_t \quad \text{(random walk with a drift)} \\ X_t &= X_{t-1} + \epsilon_t \quad \text{(random walk without a drift).} \end{aligned}$$

Such a process is a non-stationary. In the case of random walk without a drift, one has (by backward induction):

$$\begin{aligned} X_t &= X_{t-2} + \epsilon_{t-1} + \epsilon_t = \cdots \\ &= X_0 + \epsilon_1 + \cdots + \epsilon_t = X_0 + \sum_{j=0}^{t-1} \epsilon_{t-j}. \end{aligned}$$

Such a process is non-stationary since (assuming that the initial realization of X_0 is fixed):

$$\begin{aligned} \mathbb{V}(X_t) &= \mathbb{V}\left(X_0 + \sum_{j=0}^{t-1} \epsilon_{t-j}\right) = \mathbb{V}\left(\sum_{j=0}^{t-1} \epsilon_{t-j}\right) = \sum_{j=0}^{t-1} \mathbb{V}(\epsilon_{t-j}) \\ &= \sigma_\epsilon^2 \sum_{j=0}^{t-1} 1 = t\sigma_\epsilon^2, \end{aligned}$$

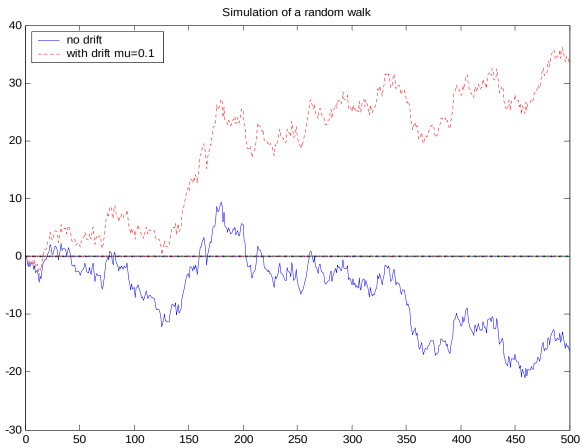
and thus the variance does depend on t!

In the case of a random walk with a drift, one has:

$$\begin{aligned} X_t &= \mu + X_{t-1} + \epsilon_t \\ &= \mu + (\mu + X_{t-2} + \epsilon_{t-1}) + \epsilon_t = \cdots \\ &= X_0 + \mu \times t + \epsilon_1 + \cdots + \epsilon_t = X_0 + \mu \times t + \sum_{j=0}^{t-1} \epsilon_{t-j}. \end{aligned}$$

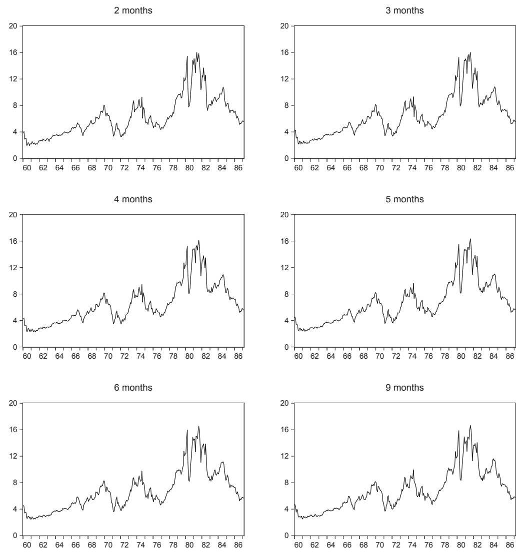
It is then straightforward to show that the three conditions of weakly stationary do not hold anymore. Interestingly, one can see that there is a "kind" of "deterministic trend", "$\mu \times t$", after using a backward induction: this is a direct consequence of the presence of a so-called unit-root—the coefficient of the first lag, X_{t-1}, is one! This is rather a so-called stochastic trend. Figure 5 displays two examples of random walks. Note that the random walk with a drift depicts a "trend" (a stochastic trend) and it might justify why random walks might be used to describe dynamics of stock prices (among others).

Figure 5: Examples of random walks



This "proximity" of the first autoregressive coefficient (the one of X_{t-1}) is often considered as being a "signal" of non-stationary. For instance, consider data on US zero-coupon bonds with different maturities (Figure 6).

Figure 6: Data on US zero-coupon bonds with different maturities



Using a simple `OLS`-based estimation (for different maturities), one obtains the results of Table 1.[6] Irrespective of the maturity, the autoregressive coefficient estimate is "close" to one : there might be some issues regarding the non-stationarity of these series. Note that a formal testing procedure can be implement using the so-called "unit root tests".[7]

Table 1: OLS estimate of AR(1) process

Maturity	Intercept		slope	
(months)	$\hat{\mu}$	s_μ	$\hat{\phi}$	s_ϕ
2	0.164	0.086	0.974	0.012
3	0.156	0.084	0.976	0.012
4	0.149	0.084	0.977	0.012
5	0.149	0.084	0.978	0.012
6	0.154	0.086	0.977	0.012
9	0.158	0.088	0.977	0.012

3 Identification tools of weakly stationary (linear) time series

Taking Lecture 1, a covariance stationary (financial) time series can be characterized by some descriptive descriptive statistics (including some distribution tests), and especially the autocovariance function, the autocorrelation function (ACF) and their sample counterparts (e.g., the sample autocorrelation function—SACF). In the case of a linear time series, some further statistical properties can be captured by the partial autocorrelation function PACF (respectively, the sample partial autocorrelation function—SPACF).[8] In

[6] It is worth noting that the asymptotic distribution of the (normalized) OLS estimator is not "standard" when the autoregressive coefficient is greater than or equal to one.

[7] For an introduction to unit root tests, see Hamilton (2005).

[8] For a definition of linear time series, see Definition 10.

the case of (weakly stationary) linear time series, it can be shown that most of the relevant features of a time series are captured by the sample autocorrelation function and the sample partial correlation function.

Autocovariance and autocorrelation functions Taking a (financial) time series or more generally a stochastic process, the autocovariance is a function that yields the covariance of the time series (stochastic process) with itself at pairs of time points (i.e., at different lags).

Definition 5. *The autocovariance function of a stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ is defined to be :*

$$\begin{aligned} \gamma: \quad & \mathbb{Z} \to \mathbb{R} \\ & h \mapsto \gamma_X(h) = \mathbb{C}ov(X_t, X_{t-h}). \end{aligned}$$

with:

$$\gamma_X(h) = \gamma_X(-h).$$

On the other hand, the autocorrelation function captures the (unconditional) linear dependency of the (financial) time series at different lags, i.e. the linear dependency between the two "series" (X_t) and (X_{t-h}) for $h \geq 0$.[9]

Definition 6. *The autocorrelation function of a stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ is defined to be:*

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_t, X_{t-h})$$

$\forall h \in \mathbb{Z}$.

Properties:

1. $\rho_X(-h) = \rho_X(h)\ \forall h$
2. $\rho_X(0) = 1$
3. The range of ρ_X is $[-1; 1]$.

The autocorrelation function is often used to first assess the stationarity or non-stationarity of a (financial) time series.[10] In addition, different patterns of the autocorrelation function can be observed and it often provides some useful information regarding the choice of certain models.[11] Notably, the autocorrelation function, which captures the intrinsic

[9]Scatter plots, (x_t, x_{t-h}), can be used to visualize the linear dependency. See also Lecture 1.

[10]This is not a formal testing procedure.

[11] Taking the autocorrelation function, if one observes the following patterns, some specifications are preferable (see further):

- Exponential decay to zero: Autoregressive model (use the partial autocorrelation plot to identify the order p);
- Damped oscillations decaying (exponentially) to zero: Autoregressive model;

persistence of the (financial) time series, is often used to (partially) characterize an important class of time series models: the so-called (mixing) AutoRegressive and Moving Average models (ARMA).

Using Definition 6, the sample autocorrelation function (see Lecture 1) can be expressed as follows.

Definition 7. *Given a sample of* T *observations,* $x_1, \cdots, x_T$*, the sample autocorrelation function, denoted by* $(\hat{\rho}_X(h))$*, is computed by:*

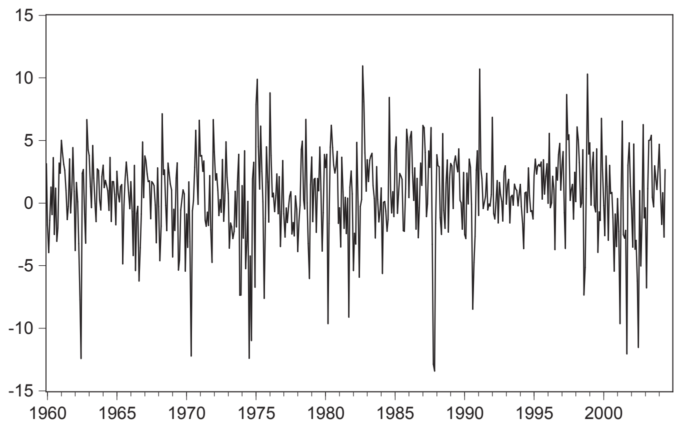
$$\hat{\rho}_X(h) = \frac{\sum_{t=h+1}^{T} (x_t - \hat{\mu})(x_{t-h} - \hat{\mu})}{\sum_{t=1}^{T} (x_t - \hat{\mu})^2}$$

where $\hat{\mu}$ *is the sample mean:*

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} x_t.$$

Figure 7 displays US real equity returns (in %) with monthly data. A quick eye inspection suggests that the series is mean-reverting and does not display any stochastic and/or deterministic trend (among possible alternative models). This can be further supported by looking at the autocorrelation function.

Figure 7: US monthly data on real equity returns (in %)



- One or more spikes, the rest is essentially zero: Moving average model (order q identified by where autocorrelation plot becomes zero);
- Exponential decay starting after a few lags: Mixed autoregressive and moving average model;
- No significant autocorrelations (zero or close to zero): White noise;
- High values at fixed intervals: Include seasonal autoregressive terms;
- No decay to zero or very slow decay: Non-stationarity or long-memory effects...

In particular, instead of computing the sample autocorrelations (Definition 7) , it is worth noting that the sample autocorrelations can be determined by using the following sequence of linear regressons and by estimating ϕ_h with the ordinary least squares method:

$$\texttt{rer}_t = \mu_h + \phi_h \texttt{rer}_{t-h} + u_t$$

Notably,

- For $h = 1$,[12]

$$\texttt{rer}_t = \underset{(0.150)}{0.419} + \underset{(0.042)}{0.241} \texttt{rer}_{t-1}$$

  i.e. $\hat{\rho}_X(1) = 0.241$.

- For $h = 2$,

$$\texttt{rer}_t = \underset{(0.155)}{0.549} + \underset{(0.043)}{0.008} \texttt{rer}_{t-2}$$

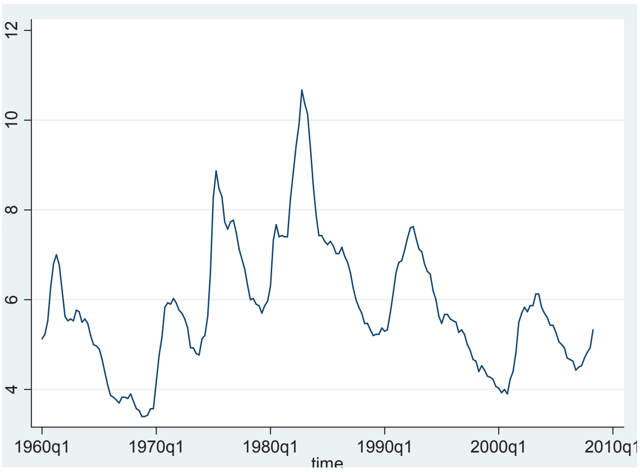
  i.e. $\hat{\rho}_X(2) = 0.008$.

- $\cdots$

After iterating this procedure and computing (say) the first 10 to 20 lags, there is some evidence that the autocorrelation function decreases (exponentially) toward zero, i.e. the series is weakly stationary. Note also that the intrinsic persistence of the series is quite low, i.e. a few lags might be sufficient to take into consideration the autocorrelation properties of the series.

As a second example, consider the following financial time series (Figure 8) in the US over the period 1960Q1-2010Q1 (quarterly data)

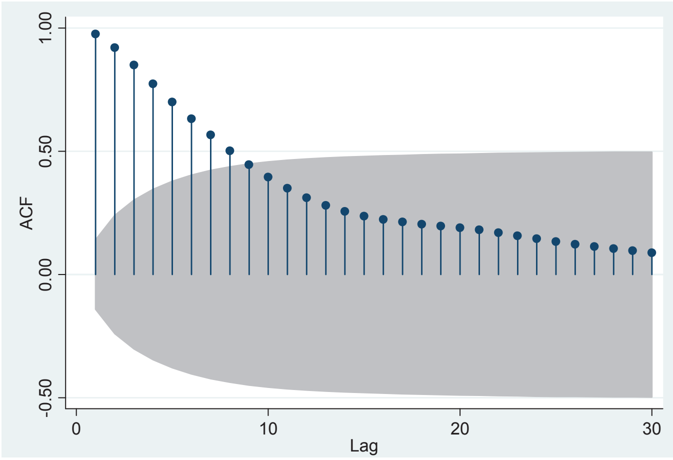
Figure 8: Financial time series (interest rate)



The autocorrelation function (with some confidence bands in grey) are reported in Figure 9. The financial time series has some (intrinsic) persistence, the first autocorrelation (i.e., the first autoregressive coefficient) is close to one, and the autocorrelation function decreases toward zero. This suggests that the series might be weakly stationary with a strong linear dependency for the first lags or that it might be a (nearly) non-stationary process.

[12]Standard errors are in parentheses.

Figure 9: Autocorrelation function



Partial correlations function A partial correlation coefficient measures the correlation between two random variables at different lags after adjusting for the correlation this pair may have with the intervening lags: the (sample) PACF thus represents the sequence of **conditional correlations**. On the other hand, a (sample) correlation coefficient between two random variables at different lags does not adjust for the influence of the intervening lags: the (sample) ACF thus represents the sequence of **unconditional correlations**.

Definition 8. *The partial autocorrelation function of a stationary stochastic process* $(X_t)_{t\in\mathbb{Z}}$ *is defined to be:*

$$a_X(h) = Corr(X_t, X_{t-h} \mid X_{t-1}, \cdots, X_{t-h+1})$$

$\forall h > 0$.

As stated in Definition 9, the partial autocorrelation function can be determined by using a sequence of multiple linear regressions (see Lecture 2).

Definition 9. *The partial correlation function can be viewed as the sequence of the h-th autoregressive coefficients in a h-th order autoregression. Let* $a_{h\ell}$ *denote the* ℓ*-th autoregressive coefficient of an AR(h) process:*

$$X_t = a_{h1}X_{t-1} + a_{h2}X_{t-2} + \cdots + a_{hh}X_{t-h} + \epsilon_t.$$

Then

$$a_X(h) = a_{hh}$$

for $h = 1, 2, \cdots$

Indeed, the autoregressive coefficient a_{hh} (in a multiple linear regression with the first h lags) can be interpreted (in an OLS sense) as the effect of X_{t-h} onto X_t after controling for the (partial) effects of $X_{t-1}, \cdots, X_{t-h+1}$. This is precisely the idea behind a conditional correlation and then a partial correlation of order h. The information provided by the (sample) partial correlation function is also useful to (partially) identify some linear time series models, as for instance `ARMA` models.

Application: As a first example, using the US real equity returns series (Figure 7), one can compute the (sample) partial correlation function as follows.

- For $h = 1$,

$$\texttt{rer}_t = \underset{(0.150)}{0.419} + \underset{(0.042)}{0.241}\texttt{rer}_{t-1}.$$

Therefore the first partial correlation is given by $\hat{a}_X(1) = 0.241$. Note that this value is also the one of the first autocorrelation, i.e. $\hat{a}_X(1) = \hat{\rho}_X(1)$.

- For $h = 2$,

$$\texttt{rer}_t = \underset{(0.151)}{0.443} + \underset{(0.043)}{0.255}\texttt{rer}_{t-1} - \underset{(0.041)}{0.053}\texttt{rer}_{t-2}$$

Consequently, the second partial correlation is given by $\hat{a}_X(2) = -0.053$.

- For $h = 3$,

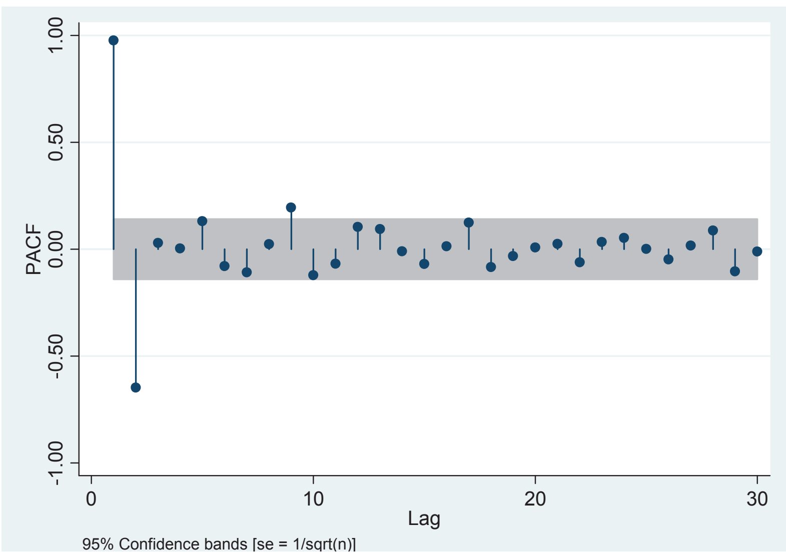
$$\texttt{rer}_t = \underset{(0.152)}{0.431} + \underset{(0.043)}{0.256}\texttt{rer}_{t-1} - \underset{(0.045)}{0.062}\texttt{rer}_{t-2} + \underset{(0.043)}{0.026}\texttt{rer}_{t-3}$$

In the same respect, the third partial correlation is $\hat{a}_X(3) = 0.026$.

- ...

As a second example, consider the financial time series (interest rate) in Figure 8. The corresponding sample partial correlation (with some confidence bands) is reported in Figure 10. After comparing Figure 9 and Figure 10, one can see that the information provided by the sample autocorrelation function and the sample partial correlation function substantially differ.

Figure 10: Sample partial correlation function



4 Some linear time series models

Some linear time series models can be proposed to capture the dynamics of a (financial) time series.[13] In the case of weakly stationary processes, the starting point is the so-called Wold's theorem.

Theorem 1 (Wold's decomposition). *Any covariance (weak) stationary time series (X_t) can be represented in the form:*

$$X_t = \mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}$$

where μ is the mean of X_t, $\theta_0 = 1$, and (ϵ_t) is a white noise process with $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{V}(\epsilon_t) = \sigma^2_\epsilon < +\infty$, and $\mathbb{C}ov(\epsilon_t, \epsilon_{t-k}) = 0$, for all $k \neq 0$.

This representation only exploits the covariance stationary property. In particular, this decomposition requires neither any distributional assumption nor any independence of the error terms. While being useful from a theoretical perspective, this decomposition is not so useful from an applied perspective since it is characterized by an infinite number of parameters and the error terms are not observable.

In this respect, linear time series models can be viewed as an approximation of this decomposition. A fundamental class of linear time series model is the so-called family of **A**uto**R**egressive and **M**oving **A**verage (ARMA) models (Box and Jenkins, 1976).

Definition 11 (ARMA(p,q)). *A stochastic process is said to be an* **ARMA**(p, q) *process if it has the following representation:*

$$\begin{aligned} X_t &= \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ &\quad + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ &= \mu + \sum_{k=1}^{p} \phi_k X_{t-k} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}. \end{aligned}$$

where $\phi_p \neq 0$, **p** *is the order of the* ***autoregressive part*** *(the past values of X_t), $\theta_0 = 1$ and $\theta_q \neq 0$,* **q** *is the order of the* ***moving average part*** *(the contemporaneous and past value of the error term ϵ_t), and ϵ_t is a weak white noise process.*

[13] A linear time series can be defined as follows.

Definition 10 (Linear time series). *A time series (X_t) is said to be linear if it can be written as:*

$$X_t = \mu + \sum_{k=0}^{\infty} \theta_k \epsilon_{t-k}$$

where μ is the mean of X_t, $\theta_0 = 1$, and (ϵ_t) is a white noise process with $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{V}(\epsilon_t) = \sigma^2_\epsilon < +\infty$, and $\mathbb{C}ov(\epsilon_t, \epsilon_{t-k}) = 0$, for all $k \neq 0$.

More generally, a linear stochastic process has the following form:

$$X_t = \mu + \sum_{k=-\infty}^{+\infty} \theta_k \epsilon_{t-k}.$$

Several points are worth commenting. First, the autoregressive part captures the (intrinsic) persistence of the series whereas the moving average part captures the noisy component. Second, some restrictions (stationarity conditions) are commonly imposed on the autoregressive coefficients such that the stochastic process (e.g., the financial time series) is weakly (second-order or covariance) stationary—see Appendix 3. Third, some further restrictions must be considered in order to interpret the weak white noise process, (ϵ_t), as being the innovation process. Indeed, one should expect for estimation purposes that there is no correlation between ϵ_t and past values of the stochastic process, $X_{t-1}, X_{t-2}, \cdots$. In particular, if ϵ_t is the **innovation** at time t, it can be interpreted as **the new information that appears at time t and that was not predictable at time $t-1$.**[14] In this case, given that $\mathbb{E}(\epsilon_t) = 0$, one has (kind of "exogeneity" assumption—see Lecture 2):

$$\mathbb{E}(\epsilon_t X_{t-k}) = \texttt{Cov}(\epsilon_t, X_{t-k}) = 0 \ \forall k > 0.$$

Fourth, the autoregressive component and the moving average component can be identified to some extent by using the (sample) autocorrelation function and the (sample) partial correlation function. For instance, using some lags of the dependent variable only makes sense if there is some persistence or some information content of past values. Fifth, when $q = 0$, one gets an AutoRegressive (`AR`) process of order p. In contrast, when $p = 0$, one obtains a Moving Average (`MA`) process of order q.

Definition 13 (`AR`(p))**.** *A stochastic process (X_t) is said to be an* `AR`*(p), if it has the following representation:*

$$X_t \ = \ \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t \ \forall t$$

where $\phi_p \neq 0$, μ is a constant term, and (ϵ_t) is a weak white noise process.

Definition 14 (`MA`(q))**.** *A stochastic process (X_t) is said to be a* `MA`*(q), if if it has the following representation:*

$$X_t \ = \ \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \ \forall t$$

where $\theta_q \neq 0$, μ is a constant term, and (ϵ_t) is a weak white noise process.

[14] The innovation process can be defined as follows.

Definition 12. *Let (X_t) denote a covariance stationary process. The innovation process of (X_t), denoted (ϵ_t), is defined to be:*

$$\epsilon_t = X_t - X_t^* \equiv X_t = X_t - \mathbb{EL}\left(X_t \mid \underline{X}_{t-1}\right)$$

where the best linear forecast of X_t given the available information at time $t-1$, denoted $I_t = \{X_{t-1}, X_{t-2}, \cdots\} \equiv \left\{\underline{X}_{t-1}\right\}$, is defined to be:

$$X_t^* = \mathbb{EL}\left(X_t \mid I_{t-1}\right) \equiv \mathbb{EL}\left(X_t \mid \underline{X}_{t-1}\right).$$

It is worth noticing that the innovation process is a weak white noise process, but the converse is generally not true.

Definitions 11, 13, and 14 are useful for estimating some macroeconomic time series as well as some financial time series (exchange rates, interest rates). However, this is not sufficient to capture the dynamics of log-return (see Lecture 1).

As a final remark, when the series is non-stationary and the source of non-stationarity is the presence of one stochastic trend (e.g., in the case of a random walk), one can still use a first-difference transformation so that the resulting series will be weakly stationary. In this case, the order of integration of the initial series is one, denoted $X_t \sim I(1)$, and the order of integration of the first-differenced series is zero, denoted $\Delta X_t = X_t - X_{t-1} \sim I(0)$. More specifically, any weakly stationary series is $I(0)$ and a series integrated of order one needs a first-difference transformation to be weakly stationary.[15] In this respect, the dynamics of the initial series might be described with an ARIMA(p,1,q) model.[16]

Definition 15 (ARIMA). *A stochastic process $(X_t)_{t \geq -p-1}$ is said to be an ARIMA(p,1,q)—an autoregressive integrated moving average model—if it satisfies the following equation (with some further conditions):*

$$\Delta X_t = \mu + \sum_{k=1}^{p} \phi_k \Delta X_{t-k} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \ \forall t \geq 0$$

where ϵ_t is a (weak) white noise process with variance σ_ϵ^2, $\theta_0 = 1$, $\theta_q \neq 0$ and $\phi_p \neq 0$.

Bibliography

Brooks, C., 2014, Introductory Econometrics for Finance (chapters 6-8), Cambridge University Press.

Hamilton, J., 2005, Time Series Analysis, Princeton University Press.

Tsay, R.S., 2002, Analysis of Financial Time Series, Wiley Series in Probability and Statistics, Wiley.

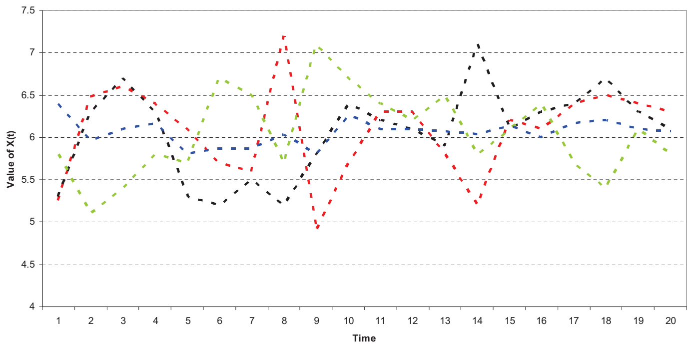
[15] In theory, a time series can be integrated of order d, i.e. one might use subsequently d first-difference transformations to get a weakly stationary series, i.e. $\Delta^d X_t = (1 - L)^d X_t$ where L is the lag operator such that $LX_t = X_{t-1}$. In practise, financial time series are either integrated of order one or order zero.

[16] This can be generalized to an ARIMA(p,d,q) process.

Appendix 1: Stationarity

Suppose that one can observe multiple time-series data of the same process over the same time period (Figure 11).

Figure 11: Multiple time series data for the same stochastic process



In this respect, one could obtain the mean (variance or covariance) values for each t (with $t = 1, \cdots, T$). For instance, Table 2 reports the first twenty realizations of each time series, $X_{1t}, \cdots, X_{4t}$, of the same stochastic process. The last column displays the mean for each time period.

Table 2: Multiple time-series of the same stochastic process

Time	X_{1t}	X_{2t}	X_{3t}	X_{4t}	mean
1	5.25	5.3	5.8	6.4	5.69
2	6.48	6.3	5.1	6.0	5.96
3	6.6	6.7	5.4	6.1	6.20
4	6.4	6.3	5.8	6.2	6.17
5	6.1	5.3	5.7	5.8	5.73
⋮	⋮	⋮	⋮	⋮	⋮
10	5.7	6.4	6.7	6.3	6.27
⋮	⋮	⋮	⋮	⋮	⋮
15	6.2	6.1	6.1	6.1	6.13
16	6.1	6.3	6.4	6.0	6.20
17	6.4	6.4	5.7	6.2	6.17
18	6.5	6.7	5.4	6.2	6.20
19	6.4	6.3	6.1	6.1	6.23
20	6.3	6.1	5.8	6.1	6.07
mean	6.07	6.06	6.05	6.06	

One can observe that the mean (expectation) is roughly the same for $t = 1, \cdots, 20$. This could be done also for the variance and the autocorrelation function. This can be interpreted as $\mathbb{E}[X_t] = m$ for all t, etc. In practise, one only observes one set of realizations for any particular series and multiple series of the same stochastic process cannot be observed. At the same time, if (X_t) is a stationary series, the mean of each column (i.e. the mean based on one series) should be close to the constant value reported in the last column.

Appendix 2: Lag, Lead and first-difference operators

The lag, lead and first-difference (d-difference) operators are quite convenient in time series analysis. Notably, these operators can be used for notation. More importantly, lag polynomial are generally used to characterize the main properties of linear time series models, especially the class of ARMA or ARIMA models and their multivariate generalization (VAR or VECM models).

Lag operator The lag operator can be defined as follows.

Definition 16. *The lag operator or backward (shift) operator, denoted by L, is an operator that shifts the time index backward by one unit.*

Examples:

1. Applying the backward operator to a variable at time t, say X_t, yields the value of that variable at time $t-1$:

$$LX_t = X_{t-1}$$

2. Applying the backward operator twice (L^2) amounts to lagging the variable twice:

$$L^2X_t = L(LX_t) = LX_{t-1} = X_{t-2}.$$

3. More generally,

$$L^kX_t = X_{t-k}$$

More formally, the lag operator transforms one time series, say

$$(Y_t)_{t\in\mathbb{Z}} = (Y_{-\infty}, \cdots, Y_{-1}, Y_0, Y_1, \cdots, Y_{+\infty}),$$

into another time series, say $(X_t)_{t\in\mathbb{Z}}$ where:

$$X_t = Y_{t-1}$$

Note that a constant can be viewed as a special series, namely:

$$(Y_t)_{t\in\mathbb{Z}}$$

where $Y_t = c$ for all t. Therefore,

$$Lc = c.$$

Applications:

- Consider an AR(1) process

$$X_t = \rho X_{t-1} + \epsilon_t$$

where ϵ_t is a weak white noise. Then it can be written as

$$\begin{aligned} X_t = \rho L X_t + \epsilon_t &\Leftrightarrow X_t - \rho L X_t = \epsilon_t \\ &\Leftrightarrow (1 - \rho L) X_t = \epsilon_t \\ &\Leftrightarrow \phi(L) X_t = \epsilon_t \end{aligned}$$

where $\phi(L)$ is a lag polynomial of order one.

- Consider an ARMA(p,q) process

$$\begin{aligned} X_t &= \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ &\quad + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\ &= \mu + \sum_{k=1}^{p} \phi_k X_{t-k} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}. \end{aligned}$$

This can be written as:

$$\Phi(L) X_t = \mu + \Theta(L) \epsilon_t$$

with Φ and Θ are some lag polynomials of order p and q, respectively:

$$\begin{aligned} \Phi(L) &= 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \\ \Theta(L) &= 1 + \theta_1 L + \cdots + \theta_q L^q \end{aligned}$$

The lead operator The lead operator can be defined as follows.

Definition 17. *The lead operator or forward (shift) operator, denoted by F, is an operator that shifts the time index forward by one unit.*

Examples:

1. Applying the forward operator to a variable at time t, say X_t, yields the value of that variable at time $t+1$:

$$F X_t = X_{t+1}$$

2. Applying the forward operator twice (F^2) amounts to leading the variable twice:

$$F^2 X_t = F(F X_t) = F X_{t+1} = X_{t+2}.$$

3. More generally,

$$F^k X_t = X_{t+k}$$

It is worth noting that (i) $F = L^{-1}$ and (ii) $F \circ L X_t = X_t$.

First-difference operator

Definition 18. *The difference operator, denoted by* Δ*, is used to express the difference between two consecutive realizations of a time series:*

$$\Delta X_t = X_t - X_{t-1}$$

More generally, the differentiation of order k *is defined by:*

$$\Delta^k = (1-L)^k$$

Appendix 3: Stationarity, invertibility, infinite order moving average or autoregressive representation and ARMA models

A non exhaustive list of properties of ARIMA(p,d,q) is provided below.

AR(p) models

Definition

An AR(p) model is defined to be

$$\Phi(L)X_t = \mu + \epsilon_t$$

where ϵ_t is a weak white noise, $\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$, μ is a constant term and $\phi_p \neq 0$.

Stability and stationarity Stability and stationarity conditions always apply on the autoregressive part of the stochastic process. In linear time series models, the "noise component" is a linear combination of some weak white noise, and thus is always weakly stationarity. Stability conditions are essential to avoid explosive solutions (e.g., bubbles) or non-bounded solutions. On the other hand, stationary conditions are fundamental since the order of integration matters for estimation, testing procedures, etc.[17] As explained below, all of these conditions require to look at the autoregressive lag polynomial, $\Phi(L)$ and to characterize its roots with either the characteristic equation or the inverse characteristic equation.[18]

Stability requires that all roots of the characteristic equation (inverse characteristic equation) associated to Φ are of modulus less than one (larger than one):

- Characteristic equation

$$\begin{aligned} z^p\Phi(z^{-1}) = 0 \quad &\Leftrightarrow \quad z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1}z - \phi_p = 0 \\ &\Leftrightarrow \quad |z_i| < 1 \qquad \text{for } i = 1, \cdots, p. \end{aligned}$$

[17] It is often tedious to make use of Definition 4.

[18] There is a nice theorem, which states that there is a correspondence between lag polynomials and polynomials, meaning that studying the roots of a lag polynomial is more or less the same as determining the roots of a polynomial of order p.

- Inverse characteristic equation

$$\begin{aligned}\Phi(z) = 0 \quad &\Leftrightarrow \quad 1 - \sum_{j=1}^{p} \phi_j z^j = 0 \\ &\Leftrightarrow \quad |z_i^*| > 1 \qquad \text{for } i = 1, \cdots, p.\end{aligned}$$

Stability implies stationarity.

There exists a (weakly) stationary solution if and only if all roots of the characteristic equation (inverse characteristic equation) are of modulus different from one

- Characteristic equation

$$\begin{aligned}z^p \Phi(z^{-1}) = 0 \quad &\Leftrightarrow \quad z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1} z - \phi_p = 0 \\ &\Leftrightarrow \quad |z_i| \neq 1 \qquad \text{for } i = 1, \cdots, p.\end{aligned}$$

- Inverse characteristic equation

$$\begin{aligned}\Phi(z) = 0 \quad &\Leftrightarrow \quad 1 - \sum_{j=1}^{p} \phi_j z^j = 0 \\ &\Leftrightarrow \quad |z_i^*| \neq 1 \qquad \text{for } i = 1, \cdots, p.\end{aligned}$$

Fundamental representation The fundamental representation is also quite fundamental in time series analysis. Indeed, it insures that the error term can be interpreted as the innovation at time t (kind of "exogeneity" assumption): this is helpful for estimation, forecasting exercises, etc.

The representation of (X_t) is said to be fundamental if and only if all roots of the characteristic equation (inverse characteristic equation) are of modulus less than one (larger than one)

$$\begin{aligned}z^p \Phi(z^{-1}) = 0 \quad &\Leftrightarrow \quad z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1} z - \phi_p = 0 \\ &\Leftrightarrow \quad |z_i| < 1 \qquad \text{for } i = 1, \cdots, p.\end{aligned}$$

In other words, the representation is fundamental if and only if (X_t) is stable. In this case, (ϵ_t) is the innovation process of (X_t) and one obtains the infinite order moving average representation

$$X_t = \frac{\mu}{\Phi(1)} + \sum_{k=0}^{\infty} a_k \epsilon_{t-k}$$

where $a_0 = 1$, $\sum_{k=0}^{\infty} |a_k| < \infty$, $\Phi(1) = 1 - \sum_{j=1}^{p} \phi_j$.

<u>Remark</u> : If the representation is not fundamental but there exists a (weakly) stationary solution, the infinite moving average representation writes

$$X_t = \frac{\mu}{\Phi(1)} + \sum_{k=-\infty}^{\infty} a_k \epsilon_{t-k}$$

In this case, ϵ_t is not the innovation of X_t since (among others) $\mathbb{E}\left[\epsilon_t X_{t-k}\right] \neq 0$ for $k > 0$.

Moments Suppose that the representation is fundamental.

The **unconditional mean** is defined to be

$$\mathbb{E}\left[X_t\right] = \frac{\mu}{\Phi(1)}.$$

In doing so, one solves the equation

$$\mathbb{E}\left[X_t\right] = \mu + \sum_{j=1}^{p} \phi_j \mathbb{E}\left[X_{t-j}\right] + \mathbb{E}\left[\epsilon_t\right]$$

or (using $\mathbb{E}\left[\epsilon_t\right] = 0$ and the notation $\mathbb{E}\left[X_{t-j}\right] \equiv m_x$ for all j)

$$m_x \left(1 - \sum_{j=1}^{p} \phi_j\right) = \mu.$$

The **autocovariance (autocorrelation) function** satisfies a difference equation of order p (Yule-Walker equation) for $|h| \geq p$

$$\gamma_X(h) - \sum_{j=1}^{p} \phi_j \gamma_X(h-j) = 0.$$

The roots of the characteristic equation are exactly the same as the ones characterizing stability! If the roots are real and distinct, then the general solution is

$$\gamma_X(h) = \sum_{j=1}^{p} C_j z_j^h.$$

This requires p initial conditions (to obtain the C_j's terms, $j = 1, \cdots, p$), which are provided by $\gamma_X(0), \cdots, \gamma_X(p-1)$. Then the fundamentalness of the representation implies that the autocovariance function[19] dies out as $h \to \infty$

$$\gamma_X(h) \underset{h \to \infty}{\longrightarrow} 0.$$

Remark: The derivation of the Yule-Walker equation proceeds as follows

- In the presence of a constant term ($\mu \neq 0$)
  - Define the process in mean-deviation;
  - Multiply by $X_t - m_X$ or $\tilde{X}_t$ and take the expectation on both sides;
  - Repeat the previous step for $h = 1, \cdots, p-1$ (to get the initial conditions)
  - Derive the general expression after multiplying by $X_{t-h} - m_X$ or $\tilde{X}_{t-h}$.
- Without a constant term ($\mu = 0$)

[19] This property can be understood using the infinite moving average representation

- Multiply by X_t and take the expectation on both sides;
- Repeat the previous step for $h = 1, \cdots, p-1$ (to get the initial conditions)
- Derive the general expression after multiplying by X_{t-h}.

The **partial autocorrelation function** is defined by

$$a_X(h) \begin{cases} \neq 0 & \text{for } 1 \leq h \leq p \\ 0 & \text{for } h > p. \end{cases}$$

An AR(p) is thus fully identified by its partial autocorrelation function. Moreover, one has

$$\begin{aligned} a_X(1) &= \rho_X(1) \\ a_X(p) &= \phi_p. \end{aligned}$$

MA(q) models

Definition A MA(q) model is defined to be

$$X_t = \mu + \Theta(L)\epsilon_t$$

where ϵ_t is a weak white noise, $\Theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j$, μ is a constant term and $\theta_q \neq 0$. This can also be written as

$$X_t = \mu + \Theta^*(L)\epsilon_t$$

where ϵ_t is a weak white noise, $\Theta^*(L) = 1 - \sum_{j=1}^{q} \theta_j^* L^j$, μ is a constant term, $\theta_q^* \neq 0$ and $\theta_j^* = -\theta_j$ for $j = 1, \cdots, q$.

Stationarity and invertibility **Stationary**: A MA(q) is always (weakly) stationary (as a finite linear combination of a weak white noise and its past).

Invertibility conditions always concern the moving average lag polynomial. It requires that all roots of the characteristic equation (inverse characteristic equation) associate to Θ are of modulus different from one[20]

$$\begin{aligned} z^q\Theta(z^{-1}) = 0 &\Leftrightarrow z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1}z + \theta_q = 0 \\ &\Leftrightarrow |z_i| \neq 1 \quad \text{for } i = 1, \cdots, q. \end{aligned}$$

In practise, one often impose the stronger requirement (as in the course) that the roots of the characteristic equation (inverse characteristic equation) associated to Θ are of modulus less than one (larger than one) in order to get the fundamental representation

$$\begin{aligned} z^q\Theta(z^{-1}) = 0 &\Leftrightarrow z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1}z + \theta_q = 0 \\ &\Leftrightarrow |z_i| < 1 \quad \text{for } i = 1, \cdots, q. \end{aligned}$$

[20]The inverse characteristic equation writes

$$\Theta(z) = 0 \Leftrightarrow 1 + \sum_{j=1}^{q} \theta_j z^j = 0.$$

Fundamental representation The representation of (X_t) is said to be fundamental if and only if all roots of the characteristic equation (inverse characteristic equation) are of modulus less than one (larger than one)

$$\begin{aligned} z^q \Theta(z^{-1}) = 0 &\Leftrightarrow z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1} z + \theta_q = 0 \\ &\Leftrightarrow |z_i| < 1 \qquad \text{for } i = 1, \cdots, q. \end{aligned}$$

In this case, (ϵ_t) is the innovation process of (X_t) and one obtains the infinite autoregressive representation

$$X_t = \frac{\mu}{\Theta(1)} + \sum_{k=1}^{\infty} b_k X_{t-k} + \epsilon_t$$

where $\Theta(1) = 1 + \sum_{j=1}^{q} \theta_j$.

Remark : If the representation is not fundamental but there exists an invertible solution, the infinite order autoregressive representation writes

$$X_t = \frac{\mu}{\Theta(1)} + \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} a_k X_{t-k} + \epsilon_t$$

In this case, ϵ_t is not the innovation of X_t.

Moments Suppose that the representation is fundamental.

The **unconditional mean** is defined to be

$$\mathbb{E}[X_t] = \mu.$$

The **autocovariance (autocorrelation) function** satisfies

$$\gamma_X(h) = \begin{cases} \sigma_\epsilon^2 \left(1 + \sum_{i=1}^{q} \theta_i^2\right) \text{ if } h = 0 \\ \sigma_\epsilon^2 \left(\theta_h + \sum_{i=h+1}^{q} \theta_i \theta_{i-h}\right) \text{ if } 1 \leq |h| < q \\ \theta_q \sigma_\epsilon^2 \text{ if } |h| = q \\ 0 \text{ if } |h| > q \end{cases}$$

The **partial autocorrelation function** satisfies[21]

$$a_X(h) \underset{h \to \infty}{\longrightarrow} 0.$$

A MA(q)) is thus fully identified by its autocorrelation function.

[21] This can be understood using the infinite autoregressive representation.

ARMA(p,q) models

Definition

An ARMA(p,q) model is defined to be

$$\Phi(L)X_t = \mu + \Theta(L)\epsilon_t$$

where ϵ_t is a weak white noise, $\Phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$, $\Theta(L) = \sum_{k=0}^{q} \theta_k L^k$ (with $\theta_0 = 1$), μ is a constant term, $\phi_p \neq 0$ and $\theta_q \neq 0$.

Stability, stationarity, and invertibility The stability and stationary conditions are the same as the ones of an AR(p). The invertibility conditions are the same as the ones of a MA(q).

Stability requires that all roots of the characteristic equation (inverse characteristic equation) associated to Φ are of modulus less than one (larger than one):

$$\begin{aligned} z^p\Phi(z^{-1}) = 0 &\Leftrightarrow z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1}z - \phi_p = 0 \\ &\Leftrightarrow |z_i| < 1 \quad \text{for } i = 1, \cdots, p. \end{aligned}$$

Stability implies stationarity.

There exists a (weakly) stationary solution if and only if all roots of the characteristic equation (inverse characteristic equation) are of modulus different from one

$$\begin{aligned} z^p\Phi(z^{-1}) = 0 &\Leftrightarrow z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1}z - \phi_p = 0 \\ &\Leftrightarrow |z_i| \neq 1 \quad \text{for } i = 1, \cdots, p. \end{aligned}$$

Invertibility requires that all roots of the characteristic equation (inverse characteristic equation) associate to Θ are of modulus different from one:

$$\begin{aligned} z^q\Theta(z^{-1}) = 0 &\Leftrightarrow z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1}z + \theta_q = 0 \\ &\Leftrightarrow |\tilde{z}_i| \neq 1 \quad \text{for } i = 1, \cdots, q. \end{aligned}$$

In practise, one often impose the stronger requirement (as in the course) that all roots of the characteristic equation (inverse characteristic equation) associated to Θ are of modulus less than one (larger than one) in order to get the (minimal) fundamental representation

$$\begin{aligned} z^q\Theta(z^{-1}) = 0 &\Leftrightarrow z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1}z + \theta_q = 0 \\ &\Leftrightarrow |\tilde{z}_i| < 1 \quad \text{for } i = 1, \cdots, q. \end{aligned}$$

Minimal and fundamental representation The representation of (X_t) is said to be **minimal and fundamental** if and only if

1. All roots of the characteristic equation (inverse characteristic equation) associated to Φ are of modulus less than one (larger than one)

$$\begin{aligned} z^p\Phi(z^{-1}) = 0 &\Leftrightarrow z^p - \phi_1 z^{p-1} - \cdots - \phi_{p-1}z - \phi_p = 0 \\ &\Leftrightarrow |z_i| < 1 \quad \text{for } i = 1, \cdots, p. \end{aligned}$$

2. All roots of the characteristic equation (inverse characteristic equation) associated to Θ are of modulus less than one (larger than one) in order to get the (minimal) fundamental representation

$$z^q\Theta(z^{-1} = 0 \;\Leftrightarrow\; z^q + \theta_1 z^{q-1} + \cdots + \theta_{q-1}z + \theta_q = 0$$
$$\Leftrightarrow\; |\tilde{z}_i| < 1 \qquad \text{for } i = 1, \cdots, q.$$

3. The two characteristic equations have no common roots.

In particular, (ϵ_t) is the innovation process of (X_t). Using the minimal and fundamental representation, one obtains two equivalent representations

1. The infinite moving average representation

$$X_t = \frac{\mu}{\Phi(1)} + \sum_{k=0}^{\infty} c_k \epsilon_{t-k}$$

where $c_0 = 1$, $\sum_{k=0}^{\infty} |c_k| < \infty$, $\Phi(1) = 1 - \sum_{j=1}^{p} \phi_j$.

2. The infinite autoregressive representation

$$X_t = \frac{\mu}{\Theta(1)} + \sum_{k=1}^{\infty} d_k X_{t-k} + \epsilon_t$$

where $\Theta(1) = 1 + \sum_{j=1}^{q} \theta_j$.

Moments Suppose that the representation is fundamental.

The **unconditional mean** is defined to be (as in the case of an AR(p))

$$\mathbb{E}[X_t] = \frac{\mu}{\Phi(1)}.$$

The **autocovariance (autocorrelation) function** satisfies a difference equation of order p (Yule-Walker equation) for $|h| \geq \max(q+1, p)$

$$\gamma_X(h) - \sum_{j=1}^{p} \phi_j \gamma_X(h-j) = 0.$$

The roots of the characteristic equation are exactly the same as the ones characterizing stability! If the roots are real and distinct, then the general solution is

$$\gamma_X(h) = \sum_{j=1}^{p} D_j z_j^h.$$

This requires p initial conditions (to obtain the D_j's terms), which are provided by $\gamma_X(0), \cdots, \gamma_X(p-1)$. Then the fundamentalness of the representation implies that the autocovariance function[22] dies out as $h \to \infty$

$$\gamma_X(h) \underset{h\to\infty}{\longrightarrow} 0.$$

[22] This property can be understood using the infinite order moving average representation of an ARMA(p,q).

Remark: The initial conditions are different from the ones of an AR(p) (even if the difference equation is the same)! The **partial autocorrelation function** satisfies[23]

$$a_X(h) \underset{h\to\infty}{\longrightarrow} 0.$$

An ARMA(p,q) cannot be fully identified by its partial autocorrelation function or its autocorrelation function.

ARIMA(p,d,q)

This section is not for the midterm!

Definition A stochastic process $(X_t)_{t\geq -p-d}$ is an autoregressive integrated moving average model if it satisfies the following equation:

$$\Phi(L)(1-L)^d X_t = \mu + \Theta(L)\epsilon_t \ \forall t \geq 0$$

where ϵ_t is a (weak) white noise process with variance σ_ϵ^2, the lag polynomials are given by:

$$\begin{aligned} \Phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p \text{ with } \phi_p \neq 0 \\ \Theta(L) &= 1 + \theta_1 L + \cdots + \theta_q L^q \text{ with } \theta_q \neq 0, \end{aligned}$$

and the initial conditions:

$$Z_{-1} = \{X_{-1}, \cdots, X_{-p-d}, \epsilon_{-1}, \cdots, \epsilon_{-q}\}$$

are such that:

$$\mathbb{C}ov(\epsilon_t, Z_{-1}) = 0 \ \forall t \geq 0.$$

Remark: Initial conditions matter!

Properties Broadly speaking, the properties (stability, stationarity and invertibility of the d-differenced process) are the same as the ones of an ARMA(p,q) (see below). This is also the case for the minimal and fundamental representation.

Equivalence with ARMA(p,q) Let $(X_t)_{t\geq -p-d}$ denote a minimal causal $ARIMA(p,d,q)$ stochastic process:

$$\Phi(L)(1-L)^d X_t = \mu + \Theta(L)\epsilon_t.$$

The stochastic process defined by:

$$Y_t = \Delta^d X_t = (1-L)^d X_t$$

is asymptotically equivalent to an $ARMA(p,q)$ process.

[23] This can be understood using the infinite order autoregressive representation of an ARMA(p,q).

Contents