Lecture 1: Fundamental concepts in Time Series Analysis (part 2)

Florian Pelgrin

University of Lausanne, École des HEC
Department of mathematics (IMEA-Nice)

Sept. 2011 - Jan. 2012

6. Stationarity, stability, and invertibility

- Consider again a situation where the value of a time series at time t, X_t, is a linear function of a constant term, the last p values of X_t, the contemporaneous and last q values of a white noise process, denoted by ϵ_t :

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

- This process rewrites :

$$\underbrace{\Phi(L) X_t}_{\text{Autoregressive part}} = \mu + \underbrace{\Theta(L)\epsilon_t}_{\text{Moving average part}}$$

with $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$.

Which conditions ?

- Stationarity conditions regard the autoregressive part of the previous (linear) time series model (ARMA(p,q) model).

- Stability conditions also regard the autoregressive part of the previous (linear) time series model (ARMA(p,q) model).

- Stability conditions are generally required to avoid explosive solutions of the stochastic difference equation : $\Phi(L)X_t = \mu + \Theta(L)\epsilon_t$.

- Invertibility conditions regard the moving average part

Implications of the stability (stationarity) conditions

- If stability conditions hold, stationarity conditions are satisfied (the converse is not true) and (X_t) is weakly stationary.

- If the stochastic process (X_t) is stable (and thus weakly stationary), then (X_t) has an infinite moving average representation (MA(∞) representation) :

$$\begin{aligned} X_t &= \Phi^{-1}(L)\left(\mu + \Theta(L)\epsilon_t\right) \\ &= \Phi^{-1}(1)\mu + C(L)\epsilon_t \\ &= \frac{\mu}{1 - \sum_{k=1}^{p}\phi_k} + \sum_{i=0}^{\infty} c_i \epsilon_{t-i} \end{aligned}$$

where the coefficients c_i satisfy $\sum_{i=0}^{\infty} |c_i| < \infty$.

- This representation shows the impact of past shocks, ϵ_{t-i}, on the current value of X_t :

$$\frac{dX_t}{d\epsilon_{t-i}} = c_i$$

- This is an application of the Wold's decomposition

- Omitting the constant term, X_t and ϵ_t are linked by a linear filter in which $C(L) = \Phi^{-1}(L)\Theta(L)$ is called the transfer function.

- The coefficients (c_i) are also referred to as the impulse response function coefficients.

- If $(X_t)_{t\in\mathbb{Z}}$ is weakly stationary but not stable, then (X_t) has the following representation :

$$\begin{aligned} X_t &= \Phi^{-1}(L)\left(\mu + \Theta(L)\epsilon_t\right) \\ &= \Phi^{-1}(1)\mu + C(L)\epsilon_t \\ &= \frac{\mu}{1-\sum_{k=1}^{p}\phi_k} + \sum_{i=-\infty}^{\infty} c_i\epsilon_{t-i}. \end{aligned}$$

where (c_j) is sequence of constants with :

$$\sum_{i=-\infty}^{\infty} |c_i| < \infty$$

- The latter condition insures that $(X_t)_{t\in\mathbb{Z}}$ is weakly stationary.

Determination of the stability and stationarity conditions

Definition

Consider the stochastic process defined by :

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

with $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, and (ϵ_t) is a white noise process. This process is called stable if the modulus of all the roots, λ_i, $i = 1, \cdots, p$, of the (reverse) characteristic equation :

$$\Phi(\lambda) = 0 \Leftrightarrow 1 - \phi_1 \lambda - \phi_2 \lambda^2 - \cdots - \phi_p \lambda^p = 0$$

are greater than one, $|\lambda_i| > 1$ for all $i = 1, \cdots, p$.

Implications of the stability (stationarity) conditions

- Invertibility is the counterpart to stationarity for the moving average part of the process.
- If the stochastic process (X_t) is invertible, then (X_t) has an infinite autoregressive representation ($AR(\infty)$ representation) :

$$\Theta^{-1}(L)\Phi(L)X_t = \Theta^{-1}(L)\mu + \epsilon_t$$

or

$$X_t = \frac{\mu}{1 - \sum_{k=1}^{q} \theta_k} + \sum_{i=1}^{\infty} d_i X_{t-i} + \epsilon_t$$

where the coefficients d_i satisfy $\sum_{i=0}^{\infty} |d_i| < \infty$.

- The AR(∞) representation shows the dependence of the current value X_t on the past values of X_{t-i}.
- The coefficients are referred as the *d*-weights of an ARMA model.
- If the stochastic process (X_t) is not invertible, it may exist a representation of the following form (omitting the constant term) as long as the magnitude of any (characteristic) root of $\Theta(\lambda)$ does not equal unity :

$$X_t = \sum_{i \neq 0} d_i X_{t-i} + \epsilon_t$$

Determination of the invertibility conditions

Definition

Consider the stochastic process defined by :

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

with $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$, and (ϵ_t) is a white noise process. This process is called invertible if the modulus of all the roots, λ_i, $i = 1, \cdots, q$, of the (reverse) characteristic equation :

$$\Theta(\lambda) = 0 \Leftrightarrow 1 + \theta_1\lambda + \theta_2\lambda^2 + \cdots + \theta_q\lambda^q = 0$$

are greater than one, $|\lambda_i| > 1$ for all $i = 1, \cdots, q$.

Three equivalent representations

Definition

A stable (and thus weakly stationary), invertible stochastic process, $\Phi(L)X_t = \mu + \Theta(L)\epsilon_t$ has two other equivalent representations forms :

1. An infinite moving average representation :

$$X_t = \frac{\mu}{1 - \sum_{k=1}^{p} \phi_k} + \sum_{i=0}^{\infty} c_i \epsilon_{t-i}$$

2. An infinite autoregressive representation :

$$X_t = \frac{\mu}{1 - \sum_{k=1}^{q} \theta_k} + \sum_{i=1}^{\infty} d_i X_{t-i} + \epsilon_t$$

- Each of these representations can shed a light on the model from a different perspective.

- For instance, the stationarity properties, the estimation of parameters, and the computing of forecasts can use different representations forms (see further).

Examples

1. An autoregressive process of order 1 :

$$(1 - \rho L)X_t = \epsilon_t$$

where $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ for all t.

- The stability and (weak) stationarity of (X_t) depends on the (reverse) characteristic root of $\Phi(\lambda) = 1 - \rho\lambda$:

$$\Phi(\lambda) = 0 \Leftrightarrow \lambda = \frac{1}{\rho}.$$

- The stability condition writes :

$$|\lambda| > 1 \Leftrightarrow |\rho| < 1.$$

- The stationarity condition writes :

$$|\lambda| \neq 1 \Leftrightarrow |\rho| \neq 1.$$

Therefore,

1. If $|\rho| < 1$, then (X_t) is stable and weakly stationary :

$$(1-\rho L)^{-1} = \sum_{k=0}^{\infty} \rho^k L^k \text{ and } X_t = \sum_{k=0}^{\infty} \rho^k \epsilon_{t-k}$$

2. If $|\rho| > 1$, then a non-causal stationary solution (X_t) exists :

$$(1-\rho L)^{-1} = -\sum_{k=1}^{\infty} \rho^{-k} F^k \text{ and } X_t = -\sum_{k=1}^{\infty} \rho^{-k} \epsilon_{t+k}$$

3. If $|\rho| = 1$, then (X_t) is not stable and stationary : $(1-\rho L)$ cannot be inverted !

2. A moving average process of order 1 :

$$X_t = \epsilon_t - \theta \epsilon_{t-1}$$

where $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ for all t.

- (X_t) is weakly stationary irrespective of the characteristic root of $\Theta(\lambda) = 1 - \theta\lambda$ (why ?).
- The invertibility of (X_t) depends on the (reverse) characteristic root of $\Theta(\lambda) = 1 - \theta\lambda$:

$$\Theta(\lambda) = 0 \Leftrightarrow \lambda = \frac{1}{\theta}.$$

- The invertibility condition writes :

$$|\lambda| > 1 \Leftrightarrow |\theta| < 1.$$

and

$$X_t = -\sum_{k=1}^{\infty} \theta^k X_{t-k} + \epsilon_t.$$

7. Identification tools

- Need some tools to characterize the main properties of a time series.

- Among others :
    - Autocovariance function
    - Autocorrelation function (ACF)
    - Sample autocorrelation function (SACF)
    - Partial autocorrelation function (PACF)
    - Sample partial autocorrelation function (SPACF)
    - Spectral density, etc.

- Keep in mind that one seeks to identify, estimate and forecast the following ARMA(p,q) model :

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{k=0}^{q} \theta_k \epsilon_{t-k}$$

where $\theta_0 = 0$, $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$, p and q are unknown orders to identify, $(\mu, \phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \sigma_\epsilon^2)$ are unknown parameters to estimate (that depends on p and q).

- The ACF and PACF functions are used to identify the appropriate time series model :
  1. The orders p and q can be identified by using the (sample) autocorrelation and (sample) partial autocorrelation function.
  2. The corresponding parameters can be then estimated using several statistical procedures.

The autocovariance function

Definition

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

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

with :

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

Definition

An estimator of $\gamma_X(h)$ (for $h < T$) is defined to be :

$$\hat{\gamma}_X(h) = \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)$$

or

$$\hat{\gamma}_X(h) = \frac{1}{T-h}\sum_{t=h+1}^{T}(X_t - \bar{X}_T)(X_{t-h} - \bar{X}_T).$$

Proposition

If $(X_t)_{t\in\mathbb{Z}}$ *is a second order stationary process with :*

$$X_t = m + \sum_{j\in\mathbb{Z}} a_j \epsilon_{t-j}$$

where the ϵ_t*'s are i.i.d. with mean zero and variance* σ_ϵ^2, $\sum_{j\in\mathbb{Z}} |a_j| < \infty$, *and* $\mathbb{E}(\epsilon_t^4) < \infty$, *then (i)* $\hat{\gamma}_X(h)$ *is an almost surely convergent and asymptotically unbiased estimator, (ii) the asymptotic disbribution is given by :*

$$\sqrt{T}\begin{pmatrix} \hat{\gamma}_X(0) - \gamma_X(0) \\ \hat{\gamma}_X(1) - \gamma_X(1) \\ \vdots \\ \hat{\gamma}_X(h) - \gamma_X(h) \end{pmatrix} \xrightarrow{\ell} \mathcal{N}\left(0, \Omega_\gamma\right)$$

where $\Omega_\gamma = [\Omega_{j,k}]_{0\leq j,k\leq h}$ *is such that :*

$$\Omega_{j,k} = \sum_{\ell=-\infty}^{+\infty} \gamma_X(\ell)\left(\gamma_X(\ell - j + k) + \gamma_X(\ell - j - k)\right).$$

The autocorrelation function

Definition

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

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

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

The autocorrelation function is obtained after re-scaling the autocovariance function by the variance $\gamma_X(0) = \mathbb{V}(X_t)$.

Definition

The autocorrelation function of a (stationary) stochastic process (X_t) satisfies the following 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 (respectively, autocovariance) function of a moving average process of order q, MA(q), is always zero for orders higher than q ($|h| > q$) : MA(q) process has no memory beyond q periods.

- The autocorrelation (respectively, autocovariance) function of a stationary $AR(p)$ process exhibits exponential decay towards zero (but does not vanish for lags greater than p).

- The autocorrelation (respectively, autocovariance) function of a stationary $ARMA(p, q)$ process exhibits exponential decay towards zero : it does not cut off but gradually dies out as h increases.

- The autocorrelation function of a nonstationary process decreases very slowly even at very high lags, long after the autocorrelations from stationary processes have declined to (almost) zero.

The sample autocorrelation function

Definition

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 :

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

Proposition

If $(X_t)_{t\in\mathbb{Z}}$ *is a second order stationary process with :*

$$X_t = m + \sum_{j\in\mathbb{Z}} a_j \epsilon_{t-j}$$

where the ϵ_t*'s are i.i.d. with mean zero and variance* σ_ϵ^2, $\sum_{j\in\mathbb{Z}} |a_j| < \infty$, *and* $\mathbb{E}(\epsilon_t^4) < \infty$, *then*

$$\sqrt{T}\begin{pmatrix} \hat{\rho}_X(1) - \rho_X(1) \\ \vdots \\ \hat{\rho}_X(h) - \rho_X(h) \end{pmatrix} \xrightarrow{\ell} \mathcal{N}\left(0, \Omega_\rho\right).$$

The partial autocorrelation function

- The partial autocorrelation function is another tool for identifying the properties of an ARMA process.
- It is particularly useful to identify pure autoregressive process (AR(p)).
- 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 PACF thus represents the sequence of conditional correlations.
- A correlation coefficient between two random variables at different lags does not adjust for the influence of the intervening lags : the ACF thus represents the sequence of unconditional correlations.

Definition

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

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

$\forall h > 0$.

Definition

The partial autocorrelation function of order h is defined to be :

$$
\begin{aligned}
a_X(h) &= Corr\left[\tilde{X}_t, \tilde{X}_{t-h}\right] \\
&= \frac{\mathbb{C}ov\left(\tilde{X}_t, \tilde{X}_{t-h}\right)}{\left[\mathbb{V}\left(\tilde{X}_t\right)\mathbb{V}\left(\tilde{X}_{t-h}\right)\right]^{1/2}}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{X}_t &= X_t - EL(X_t \mid X_{t-1}, \cdots, X_{t-h+1}) \\
\tilde{X}_{t-h} &= X_{t-h} - EL(X_{t-h} \mid X_{t-1}, \cdots, X_{t-h+1}).
\end{aligned}
$$

□

Definition

The partial autocorrelation function of a second order stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ satisfies :

$$a_X(h) = \frac{|\mathcal{R}^*(h)|}{|\mathcal{R}(h)|}$$

where $\mathcal{R}(k)$ is the autocorrelation matrix of order h définie par :

$$\mathcal{R}(h) \quad = \quad \begin{pmatrix} \rho_X(0) & \rho_X(1) & \cdots & \rho_X(h-1) \\ \rho_X(1) & \rho_X(0) & \cdots & \rho_X(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_X(h-1) & \rho_X(h-2) & \cdots & \rho_X(0) \end{pmatrix}$$

and $\mathcal{R}^*(h)$ is the matrix that is obtained after raplacing the last column of $\mathcal{R}(h)$ by $\rho = (\rho_X(1), \cdots, \rho_X(k))^t$.

Definition

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 $h = 1, 2, \cdots$

Remark : The theoretical partial autocorrelation function of an AR(p) model will be different from zero for the first p terms and exactly zero for higher order terms (why ?).

The sample partial autocorrelation function

- The sample partial autocorrelation function can be obtained by different methods :
  1. Find the ordinary least squares (or maximum likelihood) estimates of a_{hh}
  2. Use the recursive equations of the autocorrelation function (Yule-Walker equations) after replacing the autocorrelation coefficients by their estimates (see further).
  3. Etc.

Proposition

If $(X_t)_{t\in\mathbb{Z}}$ is a second order stationary process with :

$$X_t = m + \sum_{j\in\mathbb{Z}} a_j \epsilon_{t-j}$$

where the ϵ_t's are i.i.d. with mean zero and variance σ^2_ϵ, $\sum_{j\in\mathbb{Z}} |a_j| < \infty$, and $\mathbb{E}(\epsilon_t^4) < \infty$, then

$$\sqrt{T}\begin{pmatrix} \hat{a}_X(1) - r_X(1) \\ \vdots \\ \hat{a}_X(h) - r_X(h) \end{pmatrix} \xrightarrow{\ell} \mathcal{N}\left(0, \Omega_r\right).$$

□

Spectral analysis

Definition (Fourier transform of the autocovariance function)

Let (X_t) be a real-valued stationary process (with absolutely summable autocovariance sequence). The Fourier transform of the autocovariance function $(\gamma_X(h))$ exists and is given by ;

$$
\begin{aligned}
f_X(\omega) &= \frac{1}{2\pi}\sum_{h=-\infty}^{h=+\infty}\gamma_X(h)\exp(-i\omega h)\\
&= \frac{1}{2\pi}\sum_{h=-\infty}^{h=+\infty}\gamma_X(h)cos(\omega h)\\
&= \frac{1}{2\pi}\gamma_X(0)+\frac{1}{\pi}\sum_{h=1}^{h=+\infty}\gamma_X(h)cos(\omega h)
\end{aligned}
$$

$\forall\omega\in[-\pi;\pi]$.

Properties :

1. The spectral density satsifies :
   - $f_X(\omega)$ is continuous, i.e. $| f_X(\omega) |= f_X(\omega)$,
   - $f_X(\omega)$ is real-valued ;
   - $f_X(\omega)$ is a nonnegative function (since the autocovariance function is positive semidefinite).
2. $f_X(\omega) = f_X(\omega + 2\pi)$: f_X is a periodic with period 2π.
3. $f_X(\omega) = f_X(-\omega)\ \forall\omega$: f_X is a symmetric even function.

Properties (cont'd) :

4. The variance satisfies :

$$\mathbb{V}(X_t) = \gamma_X(0) = \int_{-\pi}^{\pi} f_X(\omega) d\omega$$

- The spectrum $f_X(\omega)$ may be interpreted as the decomposition of the variance of a process.
- The term $f_X(\omega)d\omega$ is the contribution to the variance attributable to the component of the process with frequencies in the interval $(\omega, \omega + d\omega)$.
- A pick (in the spectrum) indicates an important contribution to the variance.

Deriving the spectral density function

Definition (Autocovariance generating function)

For a given sequence of autocovariances $\gamma_X(h)$, $h = 0, \pm 1, \pm 2, cdots$, the autocovariance generating function is defined to be :

$$\gamma_X(L) = \sum_{h=-\infty}^{+\infty} \gamma_X(h) L^h$$

where L is the lag operator.

Proposition

Let $(X_t)_{t\in\mathbb{Z}}$ denote a second order stationary process with :

$$X_t = m + \sum_{j\in\mathbb{Z}} \theta_j \epsilon_{t-j}$$

where the ϵ_t's are i.i.d. with mean zero and variance σ_ϵ^2. The autocovariance generating function is defined to be :

$$\gamma_X(L) = \sigma_\epsilon^2 \Theta(L)\Theta(L^{-1}).$$

Definition

Let $\gamma_X(L)$ and $f_X(\omega)$ denote respectively the autocovariance generating function and the spectrum :

$$\gamma_X(L) = \sum_{h=-\infty}^{+\infty} \gamma_X(h) L^h$$

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \exp(-i\omega h) \forall \omega \in [-\pi;\pi].$$

Then

$$f_X(\omega) = \frac{1}{2\pi} \gamma_X(\exp(-i\omega)).$$

Application : Consider a general stationary ARMA(p,q) model

$$\Phi_p(L)X_t = \mu + \Theta_q(L)\epsilon_t$$

with $\Phi_p(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$ and $\Theta_q(L) = 1 + \sum_{j=1}^{q} \theta_j L^j$ sharing no common factor and having their roots outside the unit circle. Then

$$\gamma_X(L) = \sigma_\epsilon^2 \frac{\Theta_q(L)\Theta_q(L^{-1})}{\Phi_q(L)\Phi_q(L^{-1})}$$

and

$$\begin{aligned} f_X(\omega) &= \frac{1}{2\pi}\gamma_X(\exp(-i\omega)) = \frac{\sigma_\epsilon^2}{2\pi}\frac{\Theta_q(\exp(-i\omega))\Theta_q(\exp(i\omega))}{\Phi_q(\exp(-i\omega))\Phi_q(\exp(i\omega))} \\ &= \frac{\sigma_\epsilon^2}{2\pi}\left|\frac{\Theta_q(\exp(-i\omega))}{\Phi_p(\exp(-i\omega))}\right|^2 . \end{aligned}$$

Summary

- Strong and weak stationarity
- Stationarity/nonstationarity
- White noise, Trend stationary processes, Difference stationary processes (random walk)
- Lag operator and writing of time series models
- AR(∞) and MA(∞) representation
- Autocovariance, (sample) autocorrelation, and (sample) partial autocorrelation function
- Spectral density.
- Class of ARMA(p,q) processes.