Univariate time series

SARIMA processes

V. Lefieux

Table of contents

Table of contents

Introduction

If $(X_t)_{t\in\mathbb{Z}}$ is a stationary process and if $(a_i)_{i\in\mathbb{Z}}$ is a sequence such that $\sum_{i\in\mathbb{Z}} |a_i| < +\infty$ then the process $(Y_t)_{t\in\mathbb{Z}}$ defined by :

$$Y_t = \sum_{i\in\mathbb{Z}} a_i X_{t-i}$$

is stationary.
One can rewrite :

$$Y_t = \sum_{i\in\mathbb{Z}} a_i X_{t-i} = \sum_{i\in\mathbb{Z}} a_i B^i X_t = P(B) X_t$$

where $P(B) = \sum_{i\in\mathbb{Z}} a_i B^i$.

Lag series

P is said to be a lag series if :

$$P(B) = \sum_{i \in \mathbb{Z}} a_i B^i$$

where $\sum_{i \in \mathbb{Z}} |a_i| < +\infty$.

We have :

- The linear combination of two lag series is a lag series.
- The product of two lag series is a lag series.

Inverse of $I - \lambda B$

We usually need to invert lag poynomials, so we consider the inversion of $I - \lambda B$.
We want to find the STATIONARY process $(Y_t)_{t\in\mathbb{Z}}$ such that :

$$(I - \lambda B)\, Y_t = X_t$$

where $\lambda \in \mathbb{C}$.

Inverse of $I - \lambda B : |\lambda| < 1$ (1/2)

The module root $\frac{1}{\lambda}$ of $1 - \lambda z$ is more than 1.
Consider the sequence $(a_i)_{i \in \mathbb{Z}}$ such that :

$$a_i = \begin{cases} 0 & \text{si } i \in \mathbb{Z} \setminus \mathbb{N} \\ \lambda^i & \text{si } i \in \mathbb{N} \end{cases}$$

The series $\sum_{i=-\infty}^{+\infty} a_i$ is absolutely convergent and :

$$(I - \lambda B) \sum_{i \in \mathbb{N}} \lambda^i B^i = \lambda^0 B^0 = I.$$

Inverse of $I - \lambda B$: $|\lambda| < 1$ (2/2)

The solution $(Y_t)_{t\in\mathbb{Z}}$ of $(I - \lambda B)\, Y_t = X_t$ is the sum of the general case and of the particular case $(I - \lambda B)\, Y_t = 0$.
We have :

$$\begin{aligned}(I - \lambda B)\, Y_t = 0 &\Leftrightarrow Y_t = \lambda Y_{t-1} \\ &\Leftrightarrow Y_t = c\lambda^t\end{aligned}$$

where $c \in \mathbb{R}$.
Thus the solution is :

$$Y_t = \sum_{i\in\mathbb{N}} \lambda^i X_{t-i} + c\lambda^t.$$

However the only stationary solution is obtained with $c = 0$:

$$Y_t = \sum_{i\in\mathbb{N}} \lambda^i X_{t-i}.$$

Inverse of $I - \lambda B$: $|\lambda| > 1$

We have :

$$\begin{aligned} Y_t &= (I - \lambda B)^{-1} X_t \\ &= \left[-\lambda B \left(-\frac{1}{\lambda} F + I \right) \right]^{-1} X_t \\ &= \left(-\frac{1}{\lambda} F \right) \sum_{i \in \mathbb{N}} \left(\frac{1}{\lambda} \right)^i F^i X_t \\ &= -\sum_{i=1}^{+\infty} \frac{1}{\lambda^i} X_{t+i}. \end{aligned}$$

Thus :

$$Y_t = -\sum_{i=1}^{+\infty} \frac{1}{\lambda^i} X_{t+i}.$$

Inverse of $I - \lambda B$: $|\lambda| = 1$

For $|\lambda| = 1$, there is no stationary process $(Y_t)_{t \in \mathbb{Z}}$ such that $(I - \lambda B)\, Y_t = X_t$.

General case 1/3

Consider the lag series :

$$\Phi(B) = I - \varphi_1 B - \ldots - \varphi_p B^p$$

where $(\varphi_1, \ldots, \varphi_p) \in \mathbb{R}^p$.
We want to find the stationary process $(Y_t)_{t \in \mathbb{Z}}$ such that :

$$\Phi(B) Y_t = X_t.$$

Let $z_j = \frac{1}{\lambda_j}$ with $j \in \{1, \ldots, p\}$ be the roots of $\Phi(z)$.
If one of the roots z_j is on the unit circle then there is no stationary solution.

General case 2/3

If there is no roots z_j on the unit circle then there is lag series $\Psi(B) = \sum_{i\in\mathbb{Z}} \psi_i B^i$, where $(\psi_i)_{i\in\mathbb{Z}}$ is a real sequence, such that :

- $\Phi(B)\Psi(B) = I$,
- $Y_t = \Psi(B) X_t$ is stationary.

If all the roots z_j are outside the unit circle then we can write :

$$\Psi(B) = \sum_{i\in\mathbb{N}} \psi_i B^i.$$

If all the roots z_j are inside the unit circle then we can write :

$$\Psi(B) = \sum_{i\in\mathbb{N}^*} \psi_i F^i.$$

General case 3/3

In order to determine Ψ, we can use one of these methods :

- Identification method.
- Partial fraction expansion of $\frac{1}{\Phi(z)}$ and power series expansion.
- Division of 1 by $\Phi(z)$ by increasing power order.

Table of contents

Definition

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise with variance σ^2.
$(X_t)_{t\in\mathbb{Z}}$ is said to be an autoregressive process or a AR process of order p, written $\mathrm{AR}(p)$, if :

- $(X_t)_{t\in\mathbb{Z}}$ is stationary,
- $\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$ where $(\varphi_1, \ldots, \varphi_p) \in \mathbb{R}^p$ and $\varphi_p \neq 0$.

Notation

We generally use the notation $\Phi(B) X_t = \varepsilon_t$ where :

$$\Phi(B) = I - \sum_{i=1}^{p} \varphi_i B^i.$$

Note that :

- Sometimes we find $\Phi(B) = I + \sum_{i=1}^{p} \varphi_i B^i$.
- If $\Phi(B)$ has a root on the unit circle then the process $(X_t)_{t\in\mathbb{Z}}$ isn't stationary, thus it isn't an AR process.

Canonical representation : introduction 1/2

We suppose that the lag polynomial $\Phi(B)$ is invertible (no root on the unit circle).
Let $z_j = \frac{1}{\lambda_j}$, $j \in \{1, \ldots, p\}$, the roots of $\Phi(z)$.
If one root z_j is on the unit circle then there is no stationary solution.
Assume that $(z_j)_{j\in\{1,\ldots,r\}}$ are inside the circle unit and that $(z_j)_{j\in\{r+1,\ldots,p\}}$ are outside the unit circle. We can write :

$$\Phi(B) = \prod_{j=1}^{p} (I - \lambda_j B).$$

Canonical representation : introduction 2/2

This polynomial is invertible but we prefer to have only positive powers of B (the roots must be outside the unit circle).
Thus we prefer to consider the polynomial :

$$\Phi^*(B) = \prod_{j=1}^{r} \left(I - \frac{1}{\lambda_j} B \right) \prod_{j=r+1}^{p} \left(I - \lambda_j B \right).$$

Φ^* is obtained from Φ by inverting the roots which are inside the unit circle.
Result : we obtain a new AR model (the canonical version) :

$$\Phi^*(B) X_t = \eta_t.$$

Canonical representation : definition

Let $(X_t)_{t\in\mathbb{Z}}$ an AR(p) process :

$$\Phi(B)\, X_t = \varepsilon_t.$$

If the roots of Φ are outside the unit circle then we have the canonical representation.
In this case, the associated white noise is the innovation.

From now we consider canonical AR processes.

$MA(\infty)$ representation

Let $(X_t)_{t \in \mathbb{Z}}$ be a canonical AR (p) process :

$$\Phi(B) X_t = \varepsilon_t.$$

It has a $MA(\infty)$ representation :

$$X_t = \Phi^{-1}(B) \varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}$$

where $(\psi_i)_{i \in \mathbb{N}}$ is a real sequence.

Yule-Walker equations

Let $(X_t)_{t\in\mathbb{Z}}$ be a canonical AR(p) process :

$$\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

We can obtain the Yule-Walker equations :

$$\begin{cases} \gamma(0) = \dfrac{\sigma^2}{1 - \sum_{i=1}^{p} \varphi_i \rho(i)} \\ \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix} = R_p \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} \end{cases}.$$

Autocorrelations of an AR process

Autocorrelations are solutions of a simple linear recurrence equation of order p.
If the roots of $\Phi(z)$, $z_i = \frac{1}{\lambda_i}$, $i \in \{1, \ldots, p\}$, are real and unique then we have $\rho(h) = \sum_{i=1}^{p} c_i \lambda_i^h$.
Autocorrelations exponentially decrease to 0.
In the general case, we obtain a damped sine wave.

Partial autocorrelations of an AR process

If $(X_t)_{t\in\mathbb{Z}}$ is a AR(p) process then its partial autocorrelations are zero after p :

$$\begin{cases} r(p) \neq 0 \\ \forall h \in \mathbb{N}, h \geq p+1 : r(h) = 0 \end{cases} .$$

Conversely, it's a necessary and sufficient condition that $(X_t)_{t\in\mathbb{Z}}$ is an AR(p) process.

Proposition

Let $(X_t)_{t\in\mathbb{Z}}$ be a canonical AR$\,(p)$ process :

$$\forall t \in \mathbb{Z} : X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

We have :

$$r\,(p) = \varphi_p.$$

Note that :

- This proposition applies only to canonical process.
- One can't deduce anything for $r\,(h)$, $h \in \{2, \ldots, p\}$.

Table of contents

Definition

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise of variance σ^2.
$(X_t)_{t\in\mathbb{Z}}$ is said to be a MA process of order q, written MA (q), if :

$$\forall t \in \mathbb{Z} : X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where $(\theta_1, \ldots, \theta_q) \in \mathbb{R}^q$ and $\theta_q \neq 0$.

Notation

We generally use the notation $X_t = \Theta(B)\varepsilon_t$ where :

$$\Theta(B) = I + \sum_{i=1}^{q} \theta_i B^i.$$

Note that :

- Sometimes we find $\Theta(B) = I - \sum_{i=1}^{q} \theta_i B^i$.
- A MA process is stationnary.

Canonical representation

Let $(X_t)_{t\in\mathbb{Z}}$ be a MA(q) process :

$$X_t = \Theta(B)\,\varepsilon_t.$$

If the roots of Θ are outside the unit circle then we have the canonical representation.
In this case, the associated white noise is the innovation.

From now we consider canonical MA processes.

AR (∞) representation

Let $(X_t)_{t\in\mathbb{Z}}$ be a canonical MA (q) process :

$$X_t = \Theta(B)\,\varepsilon_t.$$

It has a AR (∞) representation :

$$\varepsilon_t = \Theta^{-1}(B)\,X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i},$$

thus :

$$X_t = -\sum_{i=1}^{+\infty} \pi_i X_{t-i} + \varepsilon_t$$

where $(\pi_i)_{i\in\mathbb{N}}$ is a real sequence.

Autocorrelations of a MA process

Let $(X_t)_{t\in\mathbb{Z}}$ be a canonical MA (q) process :

$$\forall t \in \mathbb{Z} : X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

We have :

$$\gamma(0) = \sigma^2 \left(1 + \sum_{i=1}^{q} \theta_i^2\right)$$

and :

$$\forall h \in \mathbb{N}^* : \gamma(h) = \begin{cases} \left(\theta_h + \sum_{i=h+1}^{q} \theta_i \theta_{i-h}\right) \sigma^2 & \text{if } h \in \{1, \ldots, q\} \\ 0 & \text{otherwise} \end{cases}$$

Autocorrelations of a MA process

If $(X_t)_{t\in\mathbb{Z}}$ is a process MA (q) process then its autocorrelations are zero after q :

$$\begin{cases} \rho(q) \neq 0 \\ \forall h \in \mathbb{N}, h \geq q+1 : \rho(h) = 0 \end{cases} .$$

Conversely, it's a necessary and sufficient condition that $(X_t)_{t\in\mathbb{Z}}$ is a MA (q) process.

Partial autocorrelations of a MA process

Partial autocorrelations are solutions of a simple linear recurrence equation of order q.
They decrease to 0.
In the general case, we obtain an exponential decrease or a damped sine wave.

Table of contents

Definition

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise of variance σ^2.
$(X_t)_{t\in\mathbb{Z}}$ is said to be a ARMA process of order (p,q), written $\mathrm{ARMA}\,(p,q)$, if :

- $(X_t)_{t\in\mathbb{Z}}$ is stationary,
- $\forall t \in \mathbb{Z} : X_t - \sum_{i=1}^{p} \varphi_i X_{t-i} = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$
  where $(\varphi_1, \ldots, \varphi_p) \in \mathbb{R}^p$, $\varphi_p \neq 0$, $(\theta_1, \ldots, \theta_q) \in \mathbb{R}^q$ and $\theta_q \neq 0$.

Notation

We generally use the notation :

$$\Phi(B) X_t = \Theta(B) \varepsilon_t$$

where :

$$\Phi(B) = I - \sum_{i=1}^{p} \varphi_i B^i,$$

$$\Theta(B) = I + \sum_{i=1}^{q} \theta_i B^i.$$

Remarks

- One can consider non centered ARMA processes $(X_t)_{t\in\mathbb{Z}}$. In this case, resultats are the same withe the process $Y_t = X_t - \mu_X$.
  From now we consider centered ARMA processes.
- An AR (p) process is a ARMA $(p, 0)$ process.
- A MA (q) process is a ARMA $(0, q)$ process.

Representations of an ARMA process

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARMA (p, q) process :

$$\Phi\,(B)\,X_t = \Theta\,(B)\,\varepsilon_t.$$

The representation is :

- minimal if Φ and Θ have no common root,
- causal if the roots of Φ are outside the unit circle,
- invertible if the roots of Θ are outside the unit circle,
- canonical if the representation is causal and invertible. In this case, the associated white noise is the innovation.

MA (∞) representation of an ARMA process

Let $(X_t)_{t\in\mathbb{Z}}$ be an minimal canonical ARMA (p,q) process :

$$\Phi(B)X_t = \Theta(B)\varepsilon_t.$$

it has a MA (∞) representation :

$$X_t = \Phi^{-1}(B)\Theta(B)\varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty}\psi_i\varepsilon_{t-i}$$

where $(\psi_i)_{i\in\mathbb{N}}$ is a real sequence.
With $\psi_i = 0$ for $i < 0$, $\theta_0 = 1$ and $\theta_i = 0$ for $i > q$, we have :

$$\forall i \in \mathbb{N} : \psi_i - \sum_{j=1}^{p}\varphi_j\psi_{i-j} = \theta_i.$$

AR (∞) representation of an ARMA process

It has an AR (∞) representation :

$$\varepsilon_t = \Theta^{-1}(B)\,\Phi(B)\,X_t = X_t + \sum_{i=1}^{+\infty} \pi_i X_{t-i},$$

thus :

$$X_t = -\sum_{i=1}^{+\infty} \pi_i X_{t-i} + \varepsilon_t$$

where $(\pi_i)_{i\in\mathbb{N}}$ is a real sequence.
With $\pi_i = 0$ for $i < 0$, $\varphi_0 = -1$ and $\varphi_i = 0$ for $i > p$, we have :

$$\forall i \in \mathbb{N} : \pi_i + \sum_{j=1}^{q} \theta_j \pi_{i-j} = -\varphi_i.$$

Autocorrelations of an ARMA process 1/2

From the $MA(\infty)$ representation :

$$\gamma(h) = \sigma^2 \sum_{i=0}^{+\infty} \psi_i \psi_{i+h}$$

where $\psi_0 = 1$.

Autocorrelations of an ARMA process 2/2

From the AR equation, for $h \in \mathbb{N}$:

$$
\begin{aligned}
&\gamma(h) - \varphi_1 \gamma(h-1) - \ldots - \varphi_p \gamma(h-p) \\
&= \operatorname{Cov}\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, X_{t-h}\right) \\
&= \operatorname{Cov}\left(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}, \sum_{i=0}^{+\infty} \psi_i \varepsilon_{t-h-i}\right) \\
&= \begin{cases} \sigma^2 \sum_{i=0}^{+\infty} \theta_{i+h} \psi_i & \text{if } h \in \{0, \ldots, q\} \\ 0 & \text{otherwise} \end{cases}.
\end{aligned}
$$

Remarks

- Autocorrelations decrease to 0.
  - If $p > q$ then we obtain an exponential decrease or a damped sine wave.
  - If $q \geq p$, the decrease is after the first $q - p$ values.
- There are similar properties for partial autocorrelations.

Corner method 1/3

There is no simple characterization of ARMA processes based on simple and partial autocorrelations. The corner method comes from autocorrelation matrixes properties. Consider, for $(i,j) \in \mathbb{N}^2$:

$$\Omega_{i,j} = \begin{bmatrix} \rho(i) & \rho(i-1) & \dots & \dots & \rho(i-j+1) \\ \rho(i-1) & \rho(i) & \rho(i-1) & \dots & \rho(i-j) \\ \rho(i-2) & \rho(i-1) & \rho(i) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho(i-1) \\ \rho(i-j+1) & \rho(i-j) & \dots & \rho(i-1) & \rho(i) \end{bmatrix}$$

and their determinants :

$$\Delta_{i,j} = \det(\Omega_{i,j}).$$

Corner method 2/3

For a ARMA (p,q) process, we have :

- $\forall (i,j) \in \mathbb{N}^2, i > q, j > p : \Delta_{i,j} = 0,$
- $\forall (i,j) \in \mathbb{N}^2, i \leq q : \Delta_{i,p} \neq 0,$
- $\forall (i,j) \in \mathbb{N}^2, j \leq p : \Delta_{q,j} \neq 0.$

Corner method 3/3

For k enough large, we represent the matrix $M = (\Delta_{i,j})_{(i,j)\in\{1,\dots,k\}^2}$ and a corner appears :

$$M = \left[\begin{array}{ccc|ccc}
\Delta_{1,1} & \cdots & \Delta_{1,p} & \Delta_{1,p+1} & \cdots & \Delta_{1,k} \\
\vdots & & \vdots & \vdots & & \vdots \\
\Delta_{q,1} & \cdots & \Delta_{q,p} & \Delta_{q,p+1} & \cdots & \Delta_{q,k} \\
\hline
\Delta_{q+1,1} & \cdots & \Delta_{q+1,p} & & & \\
\vdots & & \vdots & & 0 & \\
\Delta_{k,1} & \cdots & \Delta_{k,p} & & &
\end{array}\right].$$

Spectral density of an ARMA process

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARMA(p,q) process :

$$\Phi(B)\,X_t = \Theta(B)\,\varepsilon_t.$$

Its spectral density is :

$$f(\omega) = \frac{\sigma^2}{2\pi}\frac{\left|\Theta\left(e^{-i\omega}\right)\right|^2}{\left|\Phi\left(e^{-i\omega}\right)\right|^2}.$$

ARMA process estimation : principle

Let $(X_t)_{t\in\mathbb{Z}}$ un process ARMA (p,q) be a minimal canonical process :

$$\Phi(B)\,X_t = \Theta(B)\,\varepsilon_t.$$

The aim is to estimate Φ and Θ, and σ^2.
Estimations from autocorrelations aren't efficient. We use maximum likelihood estimation after a preliminary estimation.

Preliminary estimation : AR processes 1/2

From Yule-Walker equations :

$$\begin{cases} \begin{pmatrix} \widehat{\varphi}_1 \\ \vdots \\ \widehat{\varphi}_p \end{pmatrix} = \widehat{R}_p^{-1} \begin{pmatrix} \widehat{\rho}(1) \\ \vdots \\ \widehat{\rho}(p) \end{pmatrix} \\ \\ \widehat{\sigma}^2 = \widehat{\gamma}(0) \left(1 - \sum_{i=1}^{p} \widehat{\varphi}_i \widehat{\rho}(i)\right) \end{cases} .$$

Preliminary estimation : AR processes 2/2

With $\varphi = (\varphi_1, \dots, \varphi_p)$ and $\widehat{\varphi} = (\widehat{\varphi}_1, \dots, \widehat{\varphi}_p)$, we have :

$$\sqrt{n}\left(\widehat{\varphi} - \varphi\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2 \Sigma_p^{-1}\right)$$

and :

$$\widehat{\sigma}^2 \xrightarrow{\mathbb{P}} \sigma^2$$

where :

$$\Sigma_p = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma(1) \\ \gamma(p-1) & \dots & \gamma(1) & \gamma(0) \end{bmatrix}.$$

Preliminary estimation : MA and ARMA processes 1/2

Consider a minimal canonical ARMA (p, q) process.
From the MA (∞) representation :

$$X_t = \varepsilon_t + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{t-i}.$$

We use the innovation algorithm in order to estimate coefficients $(\psi_i)_{i \in \{1,\dots,n\}}$.

Preliminary estimation : MA and ARMA processes 2/2

With $\psi_i = 0$ for $i < 0$, $\theta_0 = 1$ and $\theta_i = 0$ for $i > q$, we have :

$$\forall i \in \mathbb{N} : \psi_i - \sum_{j=1}^{p} \varphi_j \psi_{i-j} = \theta_i.$$

We thus obtain a first estimation of $(\varphi_1, \ldots, \varphi_p)$ and $(\theta_1, \ldots, \theta_q)$ from $\left(\widehat{\psi}_1, \ldots, \widehat{\psi}_{p+q}\right)$.

Maximum likelihood estimation

We assume now that the residuals are a gaussian white noise with variance $,\sigma^2$.
Based on $(X_1, \ldots, X_T)$, the likelihood is :

$$\ell\left(x_1, \ldots, x_T; \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2\right) = \frac{1}{(2\pi)^{\frac{T}{2}}} \frac{1}{\sqrt{\det \Sigma_T}} \exp\left(-\frac{1}{2} x^\top \Sigma_T^{-1} x\right)$$

where $x = (x_1, \ldots, x_T)^\top$ and Σ_T is the variance-covariance matrix of $(X_1, \ldots, X_T)$.
It's a nonlinear optimization problem.

Maximum likelihood estimation

Consider :

$$\forall i \in \{1, \ldots, T\} : \widehat{X}_i = \mathbb{E}\left(X_i \Big/ \mathcal{H}_1^{i-1}(X)\right)$$

We use the innovation algorithm for the one step optimal linear forecast and for forecast error $\varepsilon_i = X_i - \widehat{X}_i$, and their variance $v_{i-1} = \mathbb{E}\left[\left(X_i - \widehat{X}_i\right)^2\right]$.

One don't need to calculate Σ_T^{-1} and $\det \Sigma_T$.

Maximum likelihood estimation

We have :

$$\begin{pmatrix} X_1 \\ \vdots \\ X_T \end{pmatrix} = C_T \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}.$$

Maximum likelihood estimation

$(\varepsilon_i)_{i\in\{1,\ldots,T\}} = \left(X_i - \widehat{X}_i\right)_{i\in\{1,\ldots,T\}}$ are uncorrelated, the variance-covariance matrix of $(\varepsilon_i)_{i\in\{1,\ldots,T\}}$ is :

$$V_T = \begin{bmatrix} v_0 & 0 & \ldots & 0 \\ 0 & v_1 & \ddots & \ldots \\ \ldots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & v_{T-1} \end{bmatrix}.$$

We have $\Sigma_T = C_T V_T (C_T)^\top$,thus :

$$\det \Sigma_T = (\det C_T)^2 \det V_T = v_0 \ldots v_{T-1}$$

et :

$$x^\top \Sigma_T^{-1} x = \sum_{i=1}^{T} \frac{(x_i - \widehat{x}_i)^2}{v_{i-1}}.$$

Maximum likelihood estimation

We have :

$$\ell\left(x_1, \ldots, x_T; \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2\right)$$
$$= \frac{1}{(2\pi)^{\frac{T}{2}}} \frac{1}{\sqrt{v_0 \ldots v_{T-1}}} \exp\left(-\frac{1}{2} \sum_{i=1}^{T} \frac{(x_i - \widehat{x}_i)^2}{v_{i-1}}\right)$$

$\widehat{X}_i$ is recursively obtained withe the innovation algorithm.

Maximum likelihood estimation

Consider :

$$v_{i-1} = \sigma^2 r_{i-1}.$$

We can rewrite :

$$\ell\left(x_1, \ldots, x_T; \varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q, \sigma^2\right)$$
$$= \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \frac{1}{\sqrt{r_0 \ldots r_{T-1}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{T} \frac{(x_i - \widehat{x}_i)^2}{r_{i-1}}\right).$$

Maximum likelihood estimation

Finally we have :

$$\left(\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_q\right)$$

$$= \underset{(\varphi_1,\ldots,\varphi_p,\theta_1,\ldots,\theta_q)}{\arg\min} \left\{ \ln\left[\frac{1}{T} S\left(\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_q\right)\right] + \frac{1}{T}\sum_{i=1}^{T} \ln r_i \right\}$$

and :

$$\widehat{\sigma}^2 = \frac{1}{T} S\left(\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_q\right)$$

where $S\left(\widehat{\varphi}_1, \ldots, \widehat{\varphi}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_q\right) = \sum_{i=1}^{T} \dfrac{(x_i - \widehat{x}_i)^2}{r_{i-1}}$.

Estimators are efficient.

Model selection : Kullback criterium

Let f_0 be a probability density which is estimated by an element of the family $\mathcal{F}$.
The Kullback criterium measure the difference between the the trues and the estimated probability density :

$$I\left(f_0, \mathcal{F}\right) = \min_{f \in \mathcal{F}} \int \ln\left(\frac{f_0\left(x\right)}{f\left(x\right)}\right) f_0\left(x\right) \mathrm{d}x.$$

This quantity is positive, 0 if $f_0 \in \mathcal{F}$.

Model selection : Kullback criterium

We have :

$$\int \ln\left(\frac{f_0(x)}{f(x)}\right) f_0(x)\,dx = \mathbb{E}_{f_0}\left[\ln\left(\frac{f_0(X)}{f(X)}\right)\right]$$
$$= \mathbb{E}_{f_0}\left[\ln(f_0(X))\right] - \mathbb{E}_{f_0}\left[\ln(f(X))\right].$$

We need to minimize $-\mathbb{E}_{f_0}\left[\ln(f(X))\right]$.

Model selection : ARMA case

Based on $(X_1, \ldots, X_T)$, estimators of $-\mathbb{E}_{f_0}[\ln(f(X))]$ for ARMA(p,q) processes can be written :

$$\widehat{C}(f_0, \mathcal{F}) = -\frac{1}{T}\ln(f) + \alpha(T)(p+q)$$

where α is a decreasing function.
If f is a gaussian, we obtain :

$$\widehat{C}(f_0, \mathcal{F}) = \ln\left(\widehat{\sigma}^2\right) + \alpha(T)(p+q)$$

where $\widehat{\sigma}^2$ is the estimated variance.

Model selection : ARMA case

We generally consider :

- Akaïke criterium
  With $\alpha\left(T\right)=\dfrac{2}{T}$:

$$AIC\left(p,q\right)=\ln\left(\widehat{\sigma}^{2}\right)+2\frac{p+q}{T}.$$

- Schwarz criterium
  With $\alpha\left(T\right)=\dfrac{\ln\left(T\right)}{T}$:

$$BIC\left(p,q\right)=\ln\left(\widehat{\sigma}^{2}\right)+\left(p+q\right)\frac{\ln\left(T\right)}{T}.$$

Table of contents

Definition

$(X_t)_{t\in\mathbb{N}}$ is said to be an ARIMA process of order (p, d, q), written ARIMA (p, d, q), if :

$$\Phi(B)\,\nabla^d X_t = \Theta(B)\,\varepsilon_t$$

where :

- $\nabla^d = (I - B)^d$,
- $\Phi(B) = I - \varphi_1 B - \ldots - \varphi_p B^p$ where $(\varphi_1, \ldots, \varphi_p) \in \mathbb{R}^p$ and $\varphi_p \neq 0$,
- $\Theta(B) = I + \theta_1 B + \ldots + \theta_q B^q$ where $(\theta_1, \ldots, \theta_q) \in \mathbb{R}^q$ and $\theta_q \neq 0$.

Remarks

- An ARIMA process isn't defined on $\mathbb{Z}$ but on $\mathbb{N}$ by convention. Initial conditions are fixed with :

$$Z = \left(X_{-p}, \ldots, X_{-1}, \varepsilon_{-q}, \ldots, \varepsilon_{-1}\right)^{\top}.$$

- ARIMA models can be applied on times series with a trend.
- $(I - B)^d X_t$ is asympotically equivalent to an ARMA (p, q) process.
- $(X_t)_{t \in \mathbb{N}}$ isn't a stationary process.

Autoregressive and moving average representations

We can't obtain a AR(∞) nor MA(∞) representation but it's possible to obtain :

- Moving average representation :

$$X_t = \varepsilon_t + \sum_{i=1}^{t} \psi_i \varepsilon_{t-i} + \psi^*(t) Z$$

  where $\psi^*(t)$ is a $p+q$ dimensional vector.

- Autoregressive representation :

$$X_t = -\sum_{i=1}^{t} \pi_i X_{t-i} - \pi^*(t) Z + \varepsilon_t$$

  where $\pi^*(t)$ is a $p+q$ dimensional vector.

Definition

$(X_t)_{t\in\mathbb{N}}$ is said to be a SARIMA process of order $(p,d,q)\,(P,D,Q)_s$ if :

$$\Phi(B)\,\Phi'(B^s)\,\nabla^d\nabla_s^D X_t = \Theta(B)\,\Theta'(B^s)\,\varepsilon_t$$

where :

- $\nabla^d = (I-B)^d$,
- $\nabla_s^D = (I-B^s)^D$,
- $\Phi(B) = I - \varphi_1 B - \ldots - \varphi_p B^p$ where $(\varphi_1,\ldots,\varphi_p)\in\mathbb{R}^p$ and $\varphi_p \neq 0$,
- $\Phi'(B) = I - \varphi'_1 B - \ldots - \varphi'_P B^P$ where $(\varphi'_1,\ldots,\varphi'_P)\in\mathbb{R}^P$ and $\varphi'_P \neq 0$,
- $\Theta(B) = I + \theta_1 B + \ldots + \theta_q B^q$ where $(\theta_1,\ldots,\theta_q)\in\mathbb{R}^q$ and $\theta_q \neq 0$,
- $\Theta'(B) = I + \theta'_1 B + \ldots + \theta'_Q B^Q$ where $(\theta'_1,\ldots,\theta'_Q)\in\mathbb{R}^Q$ and $\theta'_Q \neq 0$.

We write : $\mathrm{SARIMA}(p,d,q)\,(P,D,Q)_s$.

Remarks

- SARIMA models can be applied on times series with a trend and a seasonality.
- Estimation of a SARIMA model : estimation of an ARMA model on the differentiated time series.

Table of contents

Lag series
AR process
MA process
ARMA process
ARIMA and SARIMA processes
SARIMA model forecasting

Remark

The method is similar for SARIMA models than for ARIMA models.

To forecast X_{T+h} (with $h \in \mathbb{N}^*$) based on $(X_1, \ldots, X_T)$, we can use the autoregressive or the moving average representation.

Forecasting with the autoregressive representation

$$
\begin{aligned}
&\widehat{X}_T(h) \\
&= \mathbb{E}\left(X_{T+h} \middle/ \mathcal{H}_1^{T*}(X)\right) \\
&= \mathbb{E}\left(-\sum_{i=1}^{T+h} \pi_i X_{T+h-i} - \pi^*(T+h)\, Z + \varepsilon_{T+h} \middle/ \mathcal{H}_1^{T*}(X)\right) \\
&= \mathbb{E}\left(-\sum_{i=1}^{h-1} \pi_i X_{T+h-i} - \sum_{i=h}^{T+h} \pi_i X_{T+h-i} - \pi^*(T+h)\, Z + \varepsilon_{T+h} \middle/ \mathcal{H}_1^{T*}(X)\right) \\
&= -\sum_{i=1}^{h-1} \pi_i \widehat{X}_T(h-i) - \sum_{i=1}^{h-1} \pi_i X_{T+h-i} - \pi^*(T+h)\, Z \\
&\simeq -\sum_{i=1}^{h-1} \pi_i \widehat{X}_T(h-i) - \sum_{i=1}^{h-1} \pi_i X_{T+h-i} \text{ car } \pi^*(T) \overset{T\to+\infty}{\longrightarrow} 0.
\end{aligned}
$$

Forecasting with the moving average representation

$$
\begin{aligned}
\widehat{X}_T(h) &= \mathbb{E}\left(X_{T+h} \middle/ \mathcal{H}_1^{T*}(X)\right) \\
&= \mathbb{E}\left(\varepsilon_{T+h} + \sum_{i=1}^{T+h} \psi_i \varepsilon_{T+h-i} + \psi^*(T+h)\, Z \middle/ \mathcal{H}_1^{T*}(X)\right) \\
&= \sum_{i=h-1}^{T+h} \psi_i \varepsilon_{T+h-i} + \psi^*(T+h)\, Z \\
&\simeq \sum_{i=h-1}^{T+h} \psi_i \varepsilon_{T+h-i} \text{ car } \psi^*(T) \overset{T \to +\infty}{\longrightarrow} 0.
\end{aligned}
$$