Linear Time Series

Christian Francq

CREST-ENSAE

Chapter 2: ARMA Models

Outline

Definition of ARMA models

Definition:

(X_t) is an ARMA(p,q) if (X_t) is 2nd-order stationary and satisfies for all t

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t - \psi_1 \epsilon_{t-1} - \cdots - \psi_q \epsilon_{t-q}$$

where $(\epsilon_t) \sim WN(0,\sigma^2)$. It is said that (X_t) follows an ARMA(p,q) with mean μ if $(X_t - \mu)$ is an ARMA(p,q) (with mean 0).

Problem: given coefficients ψ_i and ϕ_j and a WN, does a solution (X_t) of the ARMA model exist?

Existence depends on the AR and MA polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \psi(z) = 1 - \psi_1 z - \cdots - \psi_q z^q.$$

Backward and Forward operators

Definition:

1. the backward operator B (also noted L for lag operator) transforms X_t into X_{t-1}:

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

2. the forward operator F transforms X_t into X_{t+1}:

$$FX_t = X_{t+1}$$

Properties:

- iteration: $B^k X_t = X_{t-k}, \quad F^k X_t = X_{t+k}$
- identity operator: $I = BF = FB$

Polynomials in B or F

$$\begin{aligned} P(B) &= a_0 + a_1 B + \ldots + a_n B^n \\ \Rightarrow P(B)X_t &= a_0 X_t + a_1 X_{t-1} + \ldots + a_n X_{t-n} \end{aligned}$$

One can add, multiply polynomials in B or F as usual polynomials.

We can also consider infinite sums:
if $\sum_{i=-\infty}^{\infty} |a_i| < \infty$, let the power series in B

$$a(B) = \sum_{i=-\infty}^{\infty} a_i B^i,$$

such that for any stationary process (X_t),

$$a(B)X_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i} \text{ is stationary.}$$

Existence

Multiplicative inverse of a polynomial in B

If $P(B)X_t = Z_t$, can we express X_t as a function of present, past and future values of Z_t?

$$BX_t = Z_t \;\Rightarrow\; X_t = FZ_t = Z_{t+1}$$

Inversion of $1-\phi B$: suppose $(1-\phi B)X_t = Z_t$

1. $|\phi| < 1$

$$(1-\phi B)^{-1} = \sum_{i=0}^{+\infty} \phi^i B^i \quad \Rightarrow \quad X_t = \sum_{i=0}^{+\infty} \phi^i Z_{t-i}$$

2. $|\phi| > 1$

$$(1-\phi B)^{-1} = -\sum_{i=1}^{+\infty} \frac{1}{\phi^i} F^i \quad \Rightarrow \quad X_t = -\sum_{i=1}^{+\infty} \frac{1}{\phi^i} Z_{t+i}$$

3. $|\phi| = 1$ Inversion is not possible.

Inversion of $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$

Suppose that

$$\phi(z) \neq 0 \quad \text{for all complex numbers } z \text{ such that } |z| \leq 1,$$

that is, the roots of $\phi(z)$ are outside the unit disc.

Then, there exists $\delta > 0$ such that

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} c_j z^j, \qquad |z| < 1 + \delta$$

where $\sum_{j=0}^{\infty} |c_j| < \infty$.

We then have

$$\frac{1}{\phi(B)} = \sum_{j=0}^{\infty} c_j B^j.$$

Practical derivation of the inverse

The coefficients c_j of

$$\frac{1}{\phi(B)} = \sum_{j=0}^{\infty} c_j B^j$$

can be obtained by

- Identification
- Partial fraction decomposition

Example: $X_t - 0.6X_{t-1} + 0.08X_{t-2} = \epsilon_t$

$$\to X_t = \epsilon_t + \sum_{i=1}^{+\infty} [2(0.4)^i - 0.2^i]\epsilon_{t-i}$$

$\left(\frac{1}{(1-0.4z)(1-0.2z)} = \frac{2}{1-0.4z} - \frac{1}{1-0.2z}\right)$

Existence of a causal solution

The ARMA(p,q) model

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t - \psi_1 \epsilon_{t-1} - \cdots - \psi_q \epsilon_{t-q}$$

writes $\phi(B)X_t = \psi(B)\epsilon_t$. We assume that $\mathrm{Var}(\epsilon_t) = \sigma^2 > 0$, and that the polynomials $\phi(z)$ and $\psi(z)$ have no common root.

Proposition:

A causal stationary solution (*i.e.* of the form $X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$) exists if and only if

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1.$$

We then have $\frac{\psi(B)}{\phi(B)} = \sum_{j=0}^{\infty} c_j B^j$.

Proof

Invertible solution

Invertibility means that ϵ_t can be expressed as a function of X_t and its past values.

Proposition:

The ARMA model is invertible (*i.e.* we have $\epsilon_t = \sum_{j=0}^{\infty} b_j X_{t-j}$ with $\sum_{j=0}^{\infty} |b_j| < \infty$) if and only if

$$\psi(z) = 1 - \psi_1 z - \cdots - \psi_q z^q \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1.$$

We then have

$$\epsilon_t = X_t + \sum_{j=1}^{\infty} b_j X_{t-j}, \qquad \frac{\phi(z)}{\psi(z)} = \sum_{j=0}^{\infty} b_j z^j.$$

Common roots

If common roots exist in the AR and MA polynomials and **have modulus different from de 1**, we get the same solutions by cancelling those roots.

For instance, the model

$$X_t - \phi X_{t-1} = \epsilon_t - \phi \epsilon_{t-1}, \quad |\phi| \neq 1$$

has the unique solution

$$X_t = \epsilon_t.$$

⚠: coefficient ϕ is not identifiable in this model.

Canonical ARMA

Definition and property:

The series (X_t) admits a canonical ARMA(p,q) representation if

$$\phi(B)X_t = \psi(B)\epsilon_t$$

where ϕ and ψ have no common roots, and

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1,$$

$$\psi(z) = 1 - \psi_1 z - \cdots - \psi_q z^q \neq 0 \quad \text{for all } z \text{ such that } |z| \leq 1.$$

- (ϵ_t) is the linear innovation of (X_t):

$$X_t = \underbrace{EL(X_t \mid X_{t-1}, \ldots)}_{\text{best linear pred.}} + \epsilon_t, \quad \text{with} \quad \mathrm{Cov}(\epsilon_t, X_{t-i}) = 0, \quad \text{for } i > 0$$

- The past of X_t and ϵ_t coincide

ARMA with mean

A stationary ARMA with mean μ can be written as

$$\phi(B)(X_t - \mu) = \psi(B)\epsilon_t$$

or equivalently

$$\phi(B)X_t = c + \psi(B)\epsilon_t$$

since

$$\phi(B)(X_t - \mu) = \phi(B)X_t - \phi(B)\mu = \phi(B)X_t - \phi(1)\mu$$

and $\phi(1) \neq 0$ under the stationarity condition. The intercept c is however more difficult to interpret than the mean μ.

Generality of ARMA models

From the Wold (1938) theorem, any stationary process admits a linear representation after subtraction of a deterministic component:

$$X_t = \epsilon_t + \sum_{j=1}^{\infty} c_j \epsilon_{t-j}, \qquad \sum_{j=1}^{\infty} c_j^2 < \infty.$$

This MA(∞) representation is not easy to use in practice.
Finite order MA models can be seen as an approximation of the Wold representation:

$$X_t = \epsilon_t + \sum_{j=1}^{q} c_j \epsilon_{t-j},$$

But a very large order q may be required.
A more parcimonious approximation is obtained by explicitly involving the past values of X_t.

▸ Wold theorem

Recursive relation between autocorrelations

Let the ARMA be written under the canonical form:

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \epsilon_t - \psi_1 \epsilon_{t-1} - \cdots - \psi_q \epsilon_{t-q}.$$

We have

$$\begin{aligned}\gamma(h) &= Cov(X_t, X_{t-h}) = Cov\left(\sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t - \sum_{i=1}^{q} \psi_i \epsilon_{t-i}, X_{t-h}\right) \\ &= \sum_{i=1}^{p} \phi_i Cov(X_{t-i}, X_{t-h}) + Cov(\epsilon_t, X_{t-h}) - \sum_{i=1}^{q} \psi_i Cov(\epsilon_{t-i}, X_{t-h})\end{aligned}$$

Thus

$$\gamma(h) = \sum_{i=1}^{p} \phi_i \gamma(h-i), \qquad \forall h > q.$$

Characteristic property of ARMA(p,q)

Proposition:

For a stationary and centred process (X_t), the autocorrelations satisfy a recursive equation of order p, starting from rank $q+1$:

$$\rho(h) = \sum_{i=1}^{p} \phi_i \rho(h-i), \quad \forall h > q \ (\text{with } \phi(z) \neq 0 \text{ for } |z| \leq 1)$$

if and only if $X_t \sim$ ARMA(p,q), where the ϕ_i's are the AR coefficients.

Proof

In particular: MA(q) $\iff$ $\rho(h) = 0, \quad h > q$

In an ARMA, the autocorrelations decrease at exponential rate.

Application: identification of a model

Given observations $x_1, \dots, x_n$, one approach to **identify the orders** is to compare the $\hat{\rho}(h)$'s with the $\rho(h)$'s of a given model.

In particular, if $\hat{\rho}(h) \approx 0$, for $h > q$, one can identify a MA(q).

MA(q): the $\hat{\rho}(h)$ are asymptotically Gaussian for $h > q$,

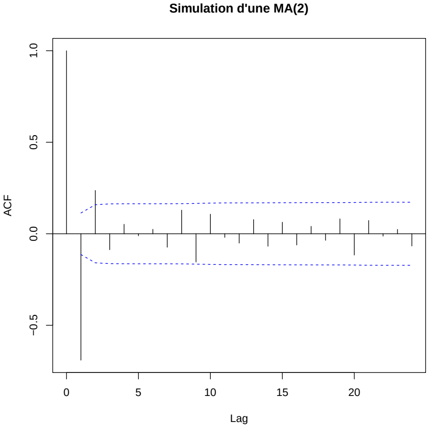
$$\sqrt{n}\hat{\rho}(h) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 + 2\rho^2(1) + \cdots + 2\rho^2(q)) \qquad h > q.$$

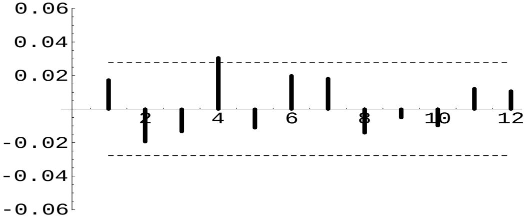
Significance bounds at 95%:

$$\pm 1.96\sqrt{1 + 2\hat{\rho}^2(1) + \cdots + 2\hat{\rho}^2(h-1)}/\sqrt{n}.$$

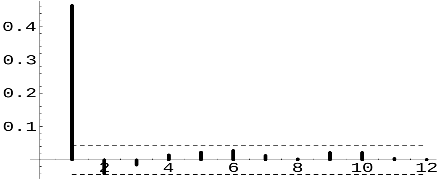
In practice, it is standard to use the smaller bounds $\pm 1.96/\sqrt{n}$.

```
n<-300; epsilon<-rnorm(n)
X<-epsilon[3:n]-1.804*epsilon[2:(n-1)]+0.806*epsilon[1:(n-2)]
acf(X,ci.type="ma",main="Simulation of a MA(2)")
```





Empirical autocorrelations of a strong WN, for n=5000.
In dotted lines, the significance bounds: $\pm 1.96/\sqrt{n}$



Empirical autocorrelations of a MA(1), for $\theta = 0.5$ and n=2000.
In dotted lines, the significance bounds: $\pm 1.96/\sqrt{n}$ (invalid here)

Specific influence of a lagged variable

Like correlations, **partial autocorrelations** convey crucial information on the dependence structure of a stationary process. Both depend only on the second-order properties of the process.

The linear influence of X_{t-h} on X_t can be measured by $\rho(h)$
But between these 2 variables we have: $X_{t-h+1}, \ldots, X_{t-1}$.

$$X_{t-h} \to X_{t-h+1} \to \ldots \to X_{t-1} \to X_t$$

Question: what is the specific influence of X_{t-h} on X_t? (adjusted for the intermediate variables)

Partial autocorrelations of a stationary process

For $h > 1$ let

- $\tilde{X}_t$ = linear regression of X_t on $1, X_{t-1}, \dots, X_{t-h+1}$.
- $\tilde{X}^*_{t-h}$ = linear regression of X_{t-h} on $1, X_{t-1}, \dots, X_{t-h+1}$

To adjust for the influence of the intermediate variables let

$$X_t - \tilde{X}_t \quad \text{and} \quad X_{t-h} - \tilde{X}^*_{t-h}$$

Partial autocorrelation function r

$$\begin{aligned} r(1) &= \rho(1) \\ r(h) &= \mathrm{Corr}(X_t - \tilde{X}_t, X_{t-h} - \tilde{X}^*_{t-h}) := \mathrm{Corr}(X_t, X_{t-h} \mid X_{t-1}, \dots, X_{t-h+1}) \end{aligned}$$

$r(h)$ can be interpreted as a measure of the (linear) dependence between X_t and X_{t-h} that is not conveyed by the intermediate variables

Alternative definition of $r(h)$

Proposition:

$r(h)$ is the coefficient of X_{t-h} in the regression of X_t on $\{1, X_{t-1}, \ldots, X_{t-h}\}$:

$$X_t = \alpha_0 + \sum_{i=1}^{h-1} \alpha_i X_{t-i} + r(h) X_{t-h} + \epsilon_t$$

Proof: see Brockwell Davis (Corr. 5.2.1, 1991)

⚠: the other coefficients α_i cannot be interpreted as $r(i)$'s!

Link with the autocorrelation function

Yule-Walker equations

$$X_t = \alpha_0 + \sum_{i=1}^{h} \alpha_i X_{t-i} + \epsilon_t, \quad \epsilon_t \perp 1, X_{t-1}, \ldots, X_{t-h},$$

entails

$$\begin{bmatrix} \rho(1) \\ \vdots \\ \vdots \\ \vdots \\ \rho(h) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \ldots & & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \ldots & \rho(h-2) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \rho(1) \\ \rho(h-1) & \rho(h-2) & \ldots & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \vdots \\ \alpha_{h-1} \\ r(h) \end{bmatrix}$$

If the matrix R_h is non-singular (generally the case), one can solve the system in $(\alpha_1, \ldots, r(h))$. The solution is unique.

The calculation can be done fastly using Durbin-Levinson's algorithm (see BD, Chapter 5).

Partial autocorrelation of an AR(p)

For a causal AR(p)

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

(ϵ_t) is the innovation:

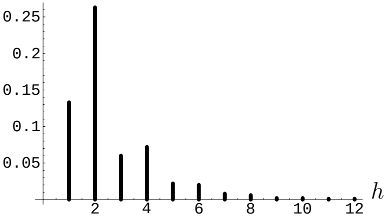
$$\text{Cov}(\epsilon_t, X_{t-i}) = 0, \qquad i > 0.$$

Thus

$$r(p) = \phi_p$$

and

$$r(h) = 0, \qquad \text{for } \; h > p.$$



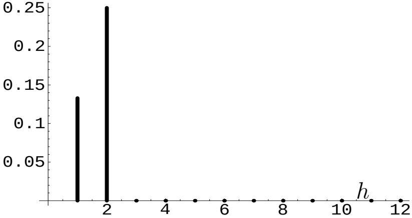


FIG. 1 – Fonction d'autocorrélation (graphe de gauche) et fonction d'autocorrélation partielle (graphe de droite) d'un AR(2) :

$$X_t = 1 + 0.1X_{t-1} + 0.25X_{t-2} + \epsilon_t$$

Empirical partial autocorrelations

The empirical partial autocorrelations, $\hat{r}(h)$, are obtained by replacing the $\rho(k)$ by $\hat{\rho}(k)$ in the previous matrix equation

In a **strong** AR(p) (with strong WN), the asymptotic distribution of the $\hat{r}(h)$, $h > p$, is very simple.

Proposition:

If (X_t) is the causal stationary solution of an AR(p) model with iid WN,

$$\sqrt{n}\hat{r}(h) \xrightarrow{\mathscr{L}} \mathcal{N}(0,1), \quad \forall h > p.$$

Identification of pure MA and AR models

Recall that a MA(q) satisfies

$$\rho(h) = 0 \qquad \text{for all } h > q,$$

and that an AR(p) satisfies

$$r(h) = 0 \qquad \text{for all } h > p.$$

In practice we use $\hat{\rho}(h)$ and $\hat{r}(h)$. If for instance

$$\hat{\rho}(1) \neq 0, \quad \text{and for all } h > 1,\ \hat{\rho}(h) \approx 0,$$

a MA(1) can be fitted.
If now,

$$\hat{r}(3) \neq 0, \quad \text{and for all } h > 4,\ \hat{r}(h) \approx 0,$$

an AR(3) can be fitted.

Identification of mixed models

For (ARMA(p,q) models with $pq \neq 0$), more sophisticated statistical methods can be used:

- the corner method (Béguin, Gouriéroux, Monfort, 1980);
  Corner method
- the spectral density;
  Spectral density

One can also proceed by estimating a lot of models, and by testing
(i) the significance of the estimated parameters, and
(ii) the independence of the residuals.

One can also use information criteria.

Least Squares (LS) estimator

From observations $X_1, X_2, \ldots, X_n$, one can approximate $\epsilon_t(\theta)$, for $0 < t \leq n$, by $e_t(\theta)$ recursively defined by

$$e_t(\theta) = X_t - \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \psi_i e_{t-i}(\theta)$$

where $e_0(\theta) = e_{-1}(\theta) = \ldots = e_{-q+1}(\theta) = X_0 = X_{-1} = \ldots = X_{-p+1} = 0$.

An (approximated) LS estimator $\hat{\theta}_n$ is defined by

$$Q_n(\hat{\theta}_n) = \min_{\theta \in \Theta} Q_n(\theta), \qquad Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} e_t^2(\theta).$$

Case of the AR(1)

Let (X_t) a causal AR(1) with a strong white noise

$$X_t = aX_{t-1} + \epsilon_t, \qquad |a| < 1.$$

We have $\theta = a$, $e_1(a) = X_1$, $e_t(a) = X_t - aX_{t-1}$ for $t = 2, \ldots, n$ and $Q_n(\hat{a}) = 0$ iff

$$\hat{a} = \frac{\sum_{t=2}^{n} X_t X_{t-1}}{\sum_{t=2}^{n} X_{t-1}^2}.^*$$

By the ergodic theorem, $\hat{a} \to a$ a.s. A CLT for non iid sequences shows that

$$\sqrt{n}\{\hat{a} - a\} = \frac{n^{-1/2} \sum_{t=2}^{n} \epsilon_t X_{t-1}}{n^{-1} \sum_{t=2}^{n} X_{t-1}^2} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - a^2).$$

*For other models, for instance the MA(1), the LS estimator is not explictly defined.

Validation

To evaluate the goodness of fit of an ARMA model, 2 types of tests:

- tests on the nullity of the parameters ϕ_j and ψ_j
  [Aim: check if the model cannot be simplified].
- tests on the residuals: if the fit is appropriate, the residuals should resemble a WN sequence.
  [Aim: check if the model is rich enough].

The choice between several appropriate models can be founded on the minimisation of information criteria (to find a balance between the accuracy of the fit and the simplicity of the model).

Tests on the parameters

Goal: compare 2 formulations ARMA(p,q) and ARMA(p',q').

Assume that one model embeds the other one: for instance $p' \leq p$ and $q' \leq q$

i)$p' = p-1, \quad q' = q$: testing the significance of ϕ_p.

Student test: At the level 0.05, the ARMA$(p-1,q)$ is accepted if

$$\frac{|\hat{\phi}_p|}{[\hat{V}(\hat{\phi}_p)]^{1/2}} < 1.96$$

ii)$p' = p, \quad q' = q-1$: idem for testing the significance of ψ_q.

iii)$p' = p-1, \quad q' = q-1$: impossible.
One cannot reduce simultaneously the AR and MA orders (non uniqueness of the ARMA representation).

Tests on the residuals

The residuals $\hat{\epsilon}_t,\ t = 1, \ldots, n$ are obtained from

$$\hat{\epsilon}_t = \frac{\hat{\Phi}(B)}{\hat{\Psi}(B)} X_t, \quad \hat{\Phi}(B) = 1 - \hat{\phi}_1 B - \ldots - \hat{\phi}_p B^p, \hat{\Psi}(B) = 1 - \hat{\psi}_1 B - \ldots - \hat{\psi}_q B^q$$

and $X_t = 0, t \leq 0$.

1. Graph of the residuals: outliers, or deviations of the variance from a constant, are sometimes clearly indicated. Absence of correlations is more difficult to identify.
2. Empirical autocorrelation function of the $\hat{\epsilon}_t$'s: the sample autocorrelations of the ϵ_t's (not the $\hat{\epsilon}_t$'s) are, for n large, approximately iid with distribution $\mathcal{N}(0, 1/n)$.

Because each $\hat{\epsilon}_t$ is a function of the observations, it is not an iid sequence: the asymptotic distribution is not exactly the same as in the iid case (except for large lags).

For large n, the sample autocorrelation of order h of the residuals is (approximately) distributed as

$$\hat{\rho}_{\hat{\epsilon}}(h) \sim \mathcal{N}(0, v)$$

where

- $v < 1/n$ for small h,
- $v \approx 1/n$ for large h.

The variance v can sometimes be explicitly computed.

Portmanteau test (Box-Pierce (1970))

Instead of checking that each $\hat{\rho}_{\hat{\epsilon}}(h)$ is between the bounds $\pm 1.96 v^{1/2}$, one can consider a global statistic depending on all sample autocorrelations.

$$Q = n \sum_{h=1}^{H} \hat{\rho}_{\hat{\epsilon}}^2(h).$$

The asymptotic distribution of Q est approximately a χ^2 with $H - p - q$ degrees of freedom (under the assumption of iid noise). The model is rejected at level $\alpha \in (0,1)$ if

$$Q > \chi^2_{1-\alpha}(H - p - q).$$

Modified Portmanteau test (Ljung-Box (1978))

For finite, even large, n the law of Q is far from the asymptotic distribution. Hence a modified statistic:

$$Q' = n(n+2)\sum_{h=1}^{H}\frac{1}{n-h}\hat{\rho}^2_{\hat{\epsilon}}(h).$$

H must be chosen large enough but not too large (generally between 15 and 30). When the fit is rejected, individual autocorrelations can be inspected to modify the model.

Drawback: lack of power. Even poor fits pass the test. Other tests are aimed at selecting the best model.

Model choice using information criteria

Idea: prevent against over-parametrisation by introducing a cost for the introduction of any additional parameter.

The approach introduced by *Akaike (1969)* is based on the Kullback-Leibler distance between the estimated and true models.

Several estimators of the information were proposed:

1. AIC(p,q) $=\log\hat{\sigma}^2+\frac{2(p+q)}{n}$;
2. BIC(p,q) $=\log\hat{\sigma}^2+(p+q)\frac{\log n}{n}$;
3. ϕ (p,q) $=\log\hat{\sigma}^2+c(p+q)\frac{\log\log n}{n}$, with $c>2$

where

$$\hat{\sigma}^2=\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2.$$

The model will be selected by minimizing the estimated information criterion.

AIC is by far the most widely used. But it often leads to over-parametrisation.

Only the BIC and ϕ criteria lead to consistent estimators of p and q. But for an AR(∞) model, only the AIC criterion leads to an asymptotically efficient estimator.

Remark: sometimes, the criteria do not select the same orders.

Theoretical predictions of an ARMA

Let an ARMA(p,q)

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \epsilon_t - \sum_{i=1}^{q} \psi_i \epsilon_{t-i},$$

assumed to be causal and invertible.

The linear optimal prediction at horizon 1 of X_t is

$$\hat{X}_{t|t-1} = \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{i=1}^{q} \psi_i \epsilon_{t-i}$$

because $\epsilon_t = X_t - \hat{X}_{t|t-1} \perp \mathcal{H}_X(t-1)$ and $\epsilon_{t-i} = \frac{\phi(B)}{\psi(B)} X_{t-i} \in \mathcal{H}_X(t-1)$.†

▸ Complements and predictions at horizon h

†($\mathcal{H}_X(t-1)$: Hilbert space generated by the linear combinations of $1, X_{t-1}, X_{t-2}, \dots$)

Prediction using an estimated ARMA model

Having estimated the ARMA coefficients from $X_1, \dots, X_n$ with $n \leq T$, using initial values (for instance 0) for $X_0, \dots, X_{1-p}, \tilde{\epsilon}_0, \dots, \tilde{\epsilon}_{1-q}$ we compute

$$\tilde{\epsilon}_t = X_t - \sum_{i=1}^{p} \hat{\phi}_i X_{t-i} + \sum_{i=1}^{q} \hat{\psi}_i \tilde{\epsilon}_{t-i}$$

for $t = 1, \dots, T$ which allows to predict X_{T+1} by

$$\hat{X}_{T+1|T} = \sum_{i=1}^{p} \hat{\phi}_i X_{T+1-i} - \sum_{i=1}^{q} \hat{\psi}_i \tilde{\epsilon}_{T+1-i}.$$

End of chapter 2 ☺ **!**

Proof of the necessary and sufficient existence condition for a causal solution

SC: If the condition is satisfied, then there exists $\delta > 0$ such that

$$\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} d_j z^j, \qquad |z| < 1 + \delta$$

where $\sum_{j=0}^{\infty} |d_j| < \infty$. Since $\psi(B)\epsilon_t$ is stationary, multiply by $\phi^{-1}(B)$ the equality $\phi(B)X_t = \psi(B)\epsilon_t$, to get

$$X_t = \phi^{-1}(B)\psi(B)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}.$$

Proof of the necessary and sufficient existence condition for a causal solution

NC: Let X_t a causal solution of the form $X_t = c(B)\epsilon_t$, then

$$\psi(B)\epsilon_t = \phi(B)X_t = \phi(B)c(B)\epsilon_t := \sum_{i=0}^{\infty} e_i\epsilon_{t-i}.$$

Multiply each side by ϵ_{t-h}, $h \geq 0$, and take the expectation, to get $e_0 = 1$, $e_i = -\psi_i$ for $i = 1, \dots, q$ and $e_i = 0$ for $i > q$.
Thus

$$\psi(z) = \phi(z)c(z), \quad |z| \leq 1.$$

Since $\psi(z)$ and $\phi(z)$ have no common root and $|c(z)| < \infty$ for $|z| \leq 1$, we cannot have $\phi(z) = 0$ for $|z| \leq 1$.

Return

Wold decomposition (1938)

$\mathcal{H}_X(t-1)$: linear past of X_t, i.e. the closed subset of L^2 generated by $1, X_{t-1}, X_{t-2}, \dots$.

(X_t) is called deterministic if

$$X_t \in \mathcal{H}_X(-\infty) = \bigcap_{s=-\infty}^{\infty} \mathcal{H}_X(s)$$

Examples: $X_t = m$ (m constant), or $X_t = X$ (X r.r.v. in L^2).

Wold decomposition Theorem (proof in BD p 187)

If (X_t) is a 2nd-order stationary process, then

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} c_i \epsilon_{t-i} + V_t, \quad \sum_{i=1}^{\infty} c_i^2 < \infty,$$

where $(\epsilon_t) \sim WN(0, \sigma^2)$, $\epsilon_t \in \mathcal{H}_X(t)$, and (V_t) is deterministic with $EV_t \epsilon_s = 0 \ \forall t, s$.

Characteristic property of an ARMA(p,q)

It remains to show that

$$\gamma_X(h) - \sum_{i=1}^{p} \phi_i \gamma_X(h-i) = 0, \ \forall h > q \quad \Rightarrow \quad Y_t := X_t - \sum_{i=1}^{p} \phi_i X_{t-i} \sim \text{MA}(q).$$

The linear innovation of Y_t is a WN defined by $\epsilon_t = Y_t - E(Y_t \mid \mathcal{H}_Y(t-1))$. The space $\mathcal{H}_Y(t-1)$ is the direct sum of $\mathcal{H}_Y(t-q-1)$ and of the Hilbert space $\mathcal{H}(\epsilon_{t-1}, \ldots, \epsilon_{t-q})$, generated by $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$. The roots of ϕ being outside the unit circle, we have $\mathcal{H}_Y(t) = \mathcal{H}_X(t)$.
Since $EY_t = 0$ and $EY_t X_{t-h} = 0 \ \forall h > q$, we have

$$Y_t \perp \mathcal{H}_X(t-q-1) = \mathcal{H}_Y(t-q-1).$$

Therefore,

$$Y_t = \epsilon_t + E(Y_t \mid \mathcal{H}(\epsilon_{t-1}, \ldots, \epsilon_{t-q})) = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i},$$

with $E(Y_t - \theta_i \epsilon_{t-i})\epsilon_{t-i} = 0$, that is $\theta_i = EY_t \epsilon_{t-i} / E\epsilon_1^2$.

Return

Proof of the spectral density for a linear tranform

We have

$$\gamma_X(h) = \sum_{j,\ell} a_j a_\ell \gamma_Y(\ell - j + h),$$

thus

$$\begin{aligned}
f_X(\lambda) &= \frac{1}{2\pi} \sum_h \sum_{j,\ell} a_j a_\ell \gamma_Y(\ell - j + h) e^{-\mathrm{i}\lambda(\ell - j + h)} e^{\mathrm{i}\lambda\ell} e^{-\mathrm{i}\lambda j} \\
&= \frac{1}{2\pi} \sum_j a_j e^{\mathrm{i}\lambda j} \sum_\ell a_\ell e^{-\mathrm{i}\lambda\ell} \sum_k \gamma_Y(k) e^{-\mathrm{i}\lambda k} \\
&= \left| \sum_j a_j e^{\mathrm{i}\lambda j} \right|^2 f_Y(\lambda).
\end{aligned}$$

Return

Corner method

The orders p and q of an ARMA(p,q) are characterized by the autocorrelation function:

$$\rho(h) = \sum_{i=1}^{p} \phi_i \rho(h-i), \quad \forall h > q.$$

Let $D(i,j)$ the $j \times j$ Toeplitz matrix

$$D(i,j) = \begin{pmatrix} \rho(i) & \rho(i-1) & \cdots & \rho(i-j+1) \\ \rho(i+1) & \ddots & & \vdots \\ \vdots & & \ddots & \rho(i-1) \\ \rho(i+j-1) & \cdots & \rho(i+1) & \rho(i) \end{pmatrix}$$

and

$$\nabla(i,j) = \det D(i,j).$$

Characterisation of the orders of an ARMA

The $\nabla(i,j)$ can be obtained from the recursion on j

$$\nabla(i,j)^2 = \nabla(i+1,j)\nabla(i-1,j) + \nabla(i,j+1)\nabla(i,j-1),$$

and by setting $\nabla(i,0) = 1, \quad \nabla(i,1) = \rho(|i|)$

Proposition:

The orders p and q are minimal (*i.e.* (X_t) does not admit an ARMA(p',q') representation with $p' < p$ or $q' < q$) if and only if

$$\begin{cases} \nabla(i,j) = 0 & \text{for all } i > q \text{ and all } j > p, \\ \nabla(i,p) \neq 0 & \text{for all } i \geq q, \\ \nabla(q,j) \neq 0 & \text{for all } j \geq p. \end{cases}$$

Corner of zeros in the table of the $\nabla(j, i)$

The minimal orders p and q are characterized by the table

$i \backslash j$	1	2	. . .	q	$q+1$	.	.	.	.
1	ρ_1	ρ_2	. . .	ρ_q	ρ_{q+1}	.	.	.	.
$\vdots$									
p				$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$p+1$				$\times$	0	0	0	0	0
				$\times$	0	0	0	0	0
				$\times$	0	0	0	0	0
				$\times$	0	0	0	0	0

where $\nabla(j, i)$ is at the intersection of row i and column j, and $\times$ denotes a non-zero item.

Table of the studentised statistics

In practice, only a finite number of empirical autocorrelations, $\hat{\rho}(1), \ldots, \hat{\rho}(K)$, are available. This allows to compute the estimates, $\hat{\nabla}(j,i)$, of the $\nabla(j,i)$ for $i \geq 1$, $j \geq 1$ and $i+j \leq K+1$. The table is thus triangular.

The orders p and q are characterized by a corner of "small values". But the determinants concern matrices of different size, and are thus not directly comparable.

We thus consider the studentised statistics defined by

$$t(i,j) = \sqrt{n}\frac{\hat{\nabla}(i,j)}{\hat{\sigma}_{\hat{\nabla}(i,j)}}.$$

Selection of plausible values for the orders

When $\nabla(i,j) = 0$ the statistics $t(i,j)$ follows asymptotically a $\mathcal{N}(0,1)$ (provided EX_t^4 exists).

One can reject the hypothesis $\nabla(i,j) = 0$ at the level $\alpha\%$ if $|t(i,j)|$ is larger than the $(1-\alpha/2)$-quantile, $\Phi^{-1}(1-\alpha/2)$, of a $\mathcal{N}(0,1)$.
(1.96 at level 5%)

One can also detect automatically a corner of small values in the table of the $t(i,j)$ if no value at this corner is larger than $\Phi^{-1}(1-\alpha/2)$ in absolute value.

This approach (not a formal test) allows to select a small number of plausible values for the orders p and q.

Example of use of the corner methode

Simulation of size $n = 1000$ of an ARMA(2,1):

$$X_t - 0.8X_{t-1} + 0.8X_{t-2} = \epsilon_t - 0.8\epsilon_{t-1}, \quad \epsilon_t \sim \mathcal{N}(0,1)$$

.p.|.q..	1	2	3	4	5	6	7	8	9	10	11	12
1	17.6	-31.6	-22.6	-1.9	11.5	8.7	-0.1	-6.1	-4.2	0.5	3.5	2.1
2	36.1	20.3	12.2	8.7	6.5	4.9	4.0	3.3	2.5	2.1	1.8	
3	-7.8	-1.6	-0.2	0.5	0.7	-0.7	0.8	-1.4	1.2	-1.1		
4	5.2	0.1	0.4	0.3	0.6	-0.1	-0.3	0.5	-0.2			
5	-3.7	0.4	-0.1	-0.5	0.4	-0.2	0.2	-0.2				
6	2.8	0.6	0.5	0.4	0.2	0.4	0.2					
7	-2.0	-0.7	0.2	0.0	-0.4	-0.3						
8	1.7	0.8	0.0	0.2	0.2							
9	-0.6	-1.2	-0.5	-0.2								
10	1.4	0.9	-0.2									
11	-0.2	-1.2										
12	1.2											

Example: automatic detection

ARMA(P,Q) MODELS FOUND WITH GIVEN SIGNIFICANCE LEVEL

PROBA	CRIT	MODELS FOUND			
0.200000	1.28	(2, 8)	(3, 1)	(10, 0)	
0.100000	1.64	(2, 1)	(8, 0)		
0.050000	1.96	(1,10)	(2, 1)	(7, 0)	
0.020000	2.33	(0,11)	(1, 9)	(2, 1)	(6, 0)
0.010000	2.58	(0,11)	(1, 8)	(2, 1)	(6, 0)
0.005000	2.81	(0,11)	(1, 8)	(2, 1)	(5, 0)
0.002000	3.09	(0,11)	(1, 8)	(2, 1)	(5, 0)
0.001000	3.29	(0,11)	(1, 8)	(2, 1)	(5, 0)
0.000100	3.72	(0, 9)	(1, 7)	(2, 1)	(5, 0)
0.000010	4.26	(0, 8)	(1, 6)	(2, 1)	(4, 0)

We find the orders $(p,q) = (2,1)$ of the simulated model, but also other plausible values. This is not surprising: the ARMA(2,1) is well approximated by several ARMA models, e.g. AR(6), MA(11) or ARMA(1,8) (but the ARMA(2,1), which is more parcimonious in terms of parameters, can be preferred).

Return

Spectral density

Definition:Fourier transform of γ_X

Let (X_t) a stationary process with absolutely summable autocovariances. The spectral density of (X_t) is

$$\begin{aligned} f_X(\lambda) &= \frac{1}{2\pi}\sum_{h=-\infty}^{+\infty}\gamma_X(h)e^{-i\lambda h} \\ &= \frac{1}{2\pi}\sum_{h=-\infty}^{+\infty}\gamma_X(h)cos(\lambda h) \\ &= \frac{1}{2\pi}\gamma_X(0)+\frac{1}{\pi}\sum_{h=1}^{+\infty}\gamma_X(h)cos(\lambda h), \quad \forall\lambda\in[-\pi;\pi] \end{aligned}$$

Propositions

- f_X is a real, continuous, even, nonnegative function.
- knowledge of f_X is equivalent to the knowledge of the sequence $\{\gamma_X(h)\}$:

$$\begin{aligned}\gamma_X(h) &= \int_{-\pi}^{\pi} f_X(\lambda)\cos(\lambda h)d\lambda \\ &= \int_{-\pi}^{\pi} f_X(\lambda)e^{-\mathrm{i}\lambda h}d\lambda \qquad \forall h \in \mathbb{Z}.\end{aligned}$$

- Intuitively, the spectral density decomposes the autocovariances: if f_X has a peak at frequency λ then $\gamma_X(h)$ is large for h close to $2\pi/\lambda$.
- The spectral density of a WN is constant (no peak, hence no periodicity).

Spectral density of linear transform

Proposition

If (Y_t) has spectral density f_Y and $\sum |a_j| < \infty$, then $X_t := \sum_{j=-\infty}^{\infty} a_j Y_{t-j}$ has spectral density

$$f_X(\lambda) = f_Y(\lambda) \left| \sum_j a_j e^{\mathrm{i}\lambda j} \right|^2 .$$

▸ Proof

Application: if $(X_t) \sim \mathrm{ARMA}(p,q)$ then

$$f_X(\lambda) = \frac{\psi\left(e^{\mathrm{i}\lambda}\right)\psi\left(e^{-\mathrm{i}\lambda}\right)}{\phi\left(e^{\mathrm{i}\lambda}\right)\phi\left(e^{-\mathrm{i}\lambda}\right)} \frac{\sigma^2}{2\pi}.$$

Application to the canonical form

If $(X_t) \sim$ ARMA,

$$\Phi(B)(1-aB)X_t = \Psi(B)(1-bB)\epsilon_t$$

with $\Phi(z)\Psi(z) \neq 0$ for $|z| = 1$ and $ab \neq 0$, then

$$\Phi(B)(1-\frac{1}{a}B)X_t = \Psi(B)(1-\frac{1}{b}B)\epsilon_t^*$$

with ϵ_t^* a WN
Indeed,

$$\epsilon_t^* = \frac{(1-\frac{1}{a}B)(1-bB)}{(1-aB)(1-\frac{1}{b}B)}\epsilon_t$$

and since $(1-e^{i\lambda}/a)(1-e^{-i\lambda}/a) = |1-ae^{i\lambda}|^2/a^2$, we get

$$f_{\epsilon^*}(\lambda) = \frac{b^2}{a^2}\frac{\sigma^2}{2\pi}.$$

Return

Prediction of stationary series

Theoretical setup: (X_t) a stationary time series, with mean $E(X_t) = m$ and autocovariance function γ (supposed to be known).

We look for the best linear combination of $1, X_n, X_{n-1}, \ldots, X_1$ to predict X_{n+h} for $h \geq 1$.

Denote by $\widehat{X}_{n+h|X_n,\ldots,X_1}$ this linear prediction, which has the form:

$$\widehat{X}_{n+h|X_n,\ldots,X_1} = a_0 + a_1 X_n + \cdots + a_n X_1$$

We want to minimize the Mean Square Error, MSE:

$$E(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2$$

Examples of predictions with a fixed number of values:

- $\min_{a_0} E(X_{n+h} - a_0)^2 \quad \Longrightarrow \quad \hat{a}_0 = E(X_t) = m.$

  MSE: $\quad E(X_{n+h} - \hat{a}_0)^2 = \gamma(0).$

- $\min_{a_0,a_1} E(X_{n+h} - a_0 - a_1 X_n)^2$?
  Cancel the derivatives with respect to a_1 and a_2:

$$\begin{cases} E(X_{n+h} - a_0 - a_1 X_n) & = & 0 \\ E\{(X_{n+h} - a_0 - a_1 X_n) X_n\} & = & 0 \end{cases} \Longrightarrow \begin{cases} \hat{a}_1 & = & \rho(h), \\ \hat{a}_0 & = & m(1 - \rho(h)). \end{cases}$$

Linear optimal prediction of X_{n+h} as a function of X_n and MSE:

$$\widehat{X}_{n+h|X_n} = m(1 - \rho(h)) + \rho(h) X_n.$$

$$E(X_{n+h} - \widehat{X}_{n+h|X_n})^2 = \gamma(0)\{1 - \rho(h)^2\}$$

Prediction using all available values:

The solution of

$$\min_{a_0,\dots,a_n} E(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2$$

is obtained for $\hat{a}_0$ and $\hat{\mathbf{a}}_n = (\hat{a}_1, \dots, \hat{a}_n)'$ satisfying

$$\Gamma_n \hat{\mathbf{a}}_n = \gamma_n(h), \quad \hat{a}_0 = m\left(1 - \sum_{i=1}^{n} \hat{a}_i\right)$$

where

$$\Gamma_n = [\gamma(i-j)]_{i,j=1}^{n}, \qquad \gamma_n(h) = (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))'.$$

MSE : $E(X_{n+h} - \widehat{X}_{n+h|X_n,\dots,X_1})^2 = \gamma(0) - \hat{\mathbf{a}}_n' \gamma_n(h).$

Linear prediction using the infinite past

(X_t) stationary time series with mean $E(X_t) = 0$.
The problem is to find the best combination of $X_n, X_{n-1}, \ldots$, to predict X_{n+h}.

Let $\widehat{X}_{n+h|X_n,\ldots}$ this prediction, and suppose it has the forme:

$$\widehat{X}_{n+h|X_n,\ldots} = \sum_{j=1}^{\infty} \alpha_j X_{n+1-j}.$$

The α_j's are characterized by

$$E\left\{\left(X_{n+h} - \sum_{j=1}^{\infty} \alpha_j X_{n+1-j}\right) X_{n+1-i}\right\} = 0, \quad i = 1, 2, \ldots$$

$$\iff \quad \gamma(h+i-1) = \sum_{j=1}^{\infty} \alpha_j \gamma(i-j), \quad i = 1, 2, \ldots$$

Using the MA(∞) representation

$$X_{n+h} = \epsilon_{n+h} + \sum_{j=1}^{\infty} c_j \epsilon_{n+h-j}, \quad \epsilon_t = \text{ innovation of } X_t.$$

Project this relation over the infinite past of X_n:

$$\widehat{X}_{n+h|X_n,\ldots} = \sum_{j=h}^{\infty} c_j \epsilon_{n+h-j}$$

The prediction error is thus: $X_{n+h} - \widehat{X}_{n+h|X_n,\ldots} = \epsilon_{n+h} + \sum_{j=1}^{h-1} c_j \epsilon_{n+h-j}$
In particular: $X_{n+1} - \widehat{X}_{n+1|X_n,\ldots} = \epsilon_{n+1}$

Variance of the prediction error:

$$\mathrm{Var}\left(X_{n+h} - \widehat{X}_{n+h|X_n,\ldots}\right) = \sigma^2(1 + \sum_{j=1}^{h-1} c_j^2)$$

The previous prediction formula is essentially of theoretical interest because the ϵ are not observable, but it allows for simple updating formulas:

$$\widehat{X}_{n+h|X_n,\ldots} - \widehat{X}_{n+h|X_{n-1},\ldots} \quad = \quad c_h\epsilon_n = c_h[X_n - \widehat{X}_{n|X_{n-1},\ldots}]$$

Using the AR(∞) representation

Since

$$X_{n+h} = \sum_{i=1}^{\infty} b_i X_{n+h-i} + \epsilon_{n+h}$$

we get

$$\widehat{X}_{n+h|X_n,\dots} = \sum_{i=1}^{\infty} b_i \widehat{X}_{n+h-i|X_n,\dots}$$

with $\quad \widehat{X}_{n+h-i|X_n,\dots} = X_{n+h-i} \quad$ if $h \le i$.
Neglecting the values anterior to $t = 1$, we deduce

$$\widehat{X}_{n+h|X_n,\dots} \approx \sum_{i=1}^{n+h-1} b_i \widehat{X}_{n+h-i|X_n,\dots}.$$

Using the ARMA form

Suppose $n > p$ and $n > q$

$$X_{n+h} = \sum_{i=1}^{p} \phi_i X_{n+h-i} + \epsilon_{n+h} - \sum_{j=1}^{q} \psi_j \epsilon_{n+h-j}$$

$$\rightarrow \widehat{X}_{n+h|X_n,\ldots} = \sum_{i=1}^{p} \phi_i \widehat{X}_{n+h-i|X_n,\ldots} - \sum_{i=1}^{q} \psi_j \widehat{\epsilon}_{n+h-j|X_n,\ldots}$$

with

$$\begin{aligned} \widehat{\epsilon}_{n+h-j|X_n,\ldots} &= 0, \quad \text{if} \quad h > j \\ &= \epsilon_{n+h-j}, \quad \text{otherwise} \end{aligned}$$

and $\quad \widehat{X}_{n+h-i|X_n,\ldots} = X_{n+h-i} \quad$ if $h \leq i$.

Joint use of these formulas:

We aim at computing predictions at horizon H.

At time n: we have to compute $\widehat{X}_{n+1|X_n,\ldots}, \ldots, \widehat{X}_{n+H|X_n,\ldots}$.

At time $n+1$: we have to
update the previous predictions:

$$\widehat{X}_{n+2|X_{n+1},\ldots}, \ldots, \widehat{X}_{n+H|X_{n+1},\ldots}$$

compute a new prediction: $\widehat{X}_{n+H+1|X_{n+1},\ldots}$.

Several steps

- One can use the AR(∞) representation to compute $\widehat{X}_{n+1|X_n,\dots},\dots,\widehat{X}_{n+H|X_n,\dots}$
  (we will no longer use the observations $X_1,\dots,X_n$)

- At time $n+1$ we observe X_{n+1}: one can use the updating formula deduced from the MA(∞) representation.
  $\Longrightarrow \widehat{X}_{n+2|X_{n+1},\dots},\dots,\widehat{X}_{n+H|X_{n+1},\dots}$

- Remains to compute $\widehat{X}_{n+H+1|X_{n+1},\dots}$: one can use the formula deduced ARMA representation, provided that:
  $H > q \quad (\rightarrow \widehat{\epsilon}_{n+H+1-j|X_{n+1},\dots} = 0, \quad \forall j \leq q)$
  $H > p \quad$ (we do not need to keep the observations)

- The procedure can be iterated when X_{n+2} is observed.

Example: ARMA (1,1)

$$X_t - 0.4X_{t-1} = \epsilon_t - 0.5\epsilon_{t-1}$$

horizon: $H = 3$
8 Observations (obtained by simulation):

$0.480 \quad -0.458 \quad 0.427 \quad -0.159 \quad -0.006 \quad 0.516 \quad -0.499 \quad 0.566$

AR(∞) representation:

$$X_{n+h} = -0.1\sum_{i=1}^{\infty} 0.5^{i-1} X_{n+h-i} + \epsilon_{n+h}$$

$$\rightarrow \quad \hat{X}_{9|8,7,\ldots} = 0.541, \quad \hat{X}_{10|8,\ldots} = 0.592 \quad \hat{X}_{11|8,\ldots} = 0.668$$

Then we observe $X_9 = -0.309$.
Updating formula:

$$\hat{X}_{9+h|9,\ldots} - \hat{X}_{9+h|8,\ldots} = c_h[X_9 - \hat{X}_{9|8,7,\ldots}],$$

with $c_1 = -0.1, \quad c_2 = -0.04$

$$\rightarrow \quad \hat{X}_{10|9,\ldots} = 0.677, \quad \hat{X}_{11|9,\ldots} = 0.702$$

Using the ARMA(1,1) representation :

$\hat{X}_{12|9,\ldots} = 0.4\hat{X}_{11|9,\ldots} = 0.281$ and then repeat ...

◂ Return to the theoretical predictions of an ARMA

Transformation linéaire d'une série stationnaire

Propriété

Soit (X_t) une série stationnaire (au second ordre), $(a_i)_{i \in Z}$ une suite de réels telle que $\sum_{i=-\infty}^{\infty} |a_i| < \infty$. Alors

$$a(B)X_t = \sum_{i=-\infty}^{\infty} a_i B^i X_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i}$$

converge en moyenne quadratique and presque sûrement, ▸ Proof et définit une série stationnaire.

Remarque: En particulier, une combinaison linéaire finie des variables X_{t-i} définit toujours une série stationnaire (donc la stationnarité au second ordre des MA est automatique). ◂ Return

Autocovariance d'une transformation linéaire

Propriété

Soit (X_t) une série stationnaire, $(a_i)_{i\in Z}$ une suite de réels telle que $\sum_{i=-\infty}^{\infty}|a_i| < \infty$. Alors

$$Y_t = a(B)X_t = \sum_{i=-\infty}^{\infty} a_i B^i X_t = \sum_{i=-\infty}^{\infty} a_i X_{t-i}$$

est une série stationnaire de fonction d'autocovariance

$$\gamma_Y(h) = \sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_i a_j \gamma_X(h+i-j).$$

▸ Preuve

Manipulation des séries en B

Les séries en B appliquées à des processus stationnaires se manipulent comme des séries entières. En particulier, if $\sum_{i=-\infty}^{\infty}|a_i|<\infty$, $\sum_{i=-\infty}^{\infty}|b_i|<\infty$, $a(B)=\sum_{i=-\infty}^{\infty}a_iB^i$ and $b(B)=\sum_{i=-\infty}^{\infty}b_iB^i$ alors for all processus stationnaire (X_t)

$$\begin{aligned} a(B)b(B)X_t &= \sum_{i=-\infty}^{\infty} a_i \sum_{j=-\infty}^{\infty} b_j B^{i+j} X_t = b(B)a(B)X_t \\ &= \sum_{k=-\infty}^{\infty}\left(\sum_{j=-\infty}^{\infty} a_{k-j}b_j\right) B^k X_t := c(B)X_t. \end{aligned}$$

De plus $\quad a(B)=b(B) \quad \Longleftrightarrow$

$$a_i=b_i,\ \forall i \quad \Longleftrightarrow \quad a(z)=b(z), \forall 1-\delta<|z|<1+\delta$$

for un $\delta>0$.

Preuve de la convergence de la série

- Par Fubini,
$$E\sum_{i=-\infty}^{\infty}|a_iX_{t-i}|<\infty,$$
donc la série est finie avec probabilité 1
- Pour $0<p<q$,
$$\left\|\sum_{p\leq|i|\leq q}a_iX_{t-i}\right\|_2\leq\sum_{p\leq|i|\leq q}|a_i|\,\|X_{t-i}\|_2\to 0$$
quand $p,q\to\infty$ donc la série converge en moyenne quadratique par le critère de Cauchy.

Retour

Preuve de la stationnarité

En justifiant les interversions de E et $\sum$ par Lebesgue et Fubini, on obtient

$$EY_t = a(B)EX_t = a(1)EX_1 = \sum_{i=-\infty}^{\infty} a_i EX_1$$

puis

$$E(Y_t - EY_1)(Y_{t-h} - EY_1) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i a_j E(X_{t-i} - EX_1)(X_{t-j-h} - EX_1).$$

Retour