

Estimation des modèles ARMA

(1)

On utilise l'estimateur du maximum de vraisemblance, or
verra que dans certains cas cela
revient à un estimateur des
moindres carrés ordinaires.

Vous connaissez déjà les propriétés
de cet estimateur (comportement
asymptotique). Dans le cadre
d'un modèle dynamique les
propriétés en échantillon fini sont
néanmoins différentes des propriétés

connues dans les modèles statiques. ⁽²⁾
On verra par exemple que l'estimateur
des MCO (qui est aussi un estimateur
du maximum de vraisemblance) dans
un modèle AR(1) est biaisé en
échantillon fini. L'estimateur est
quand même convergent asymptotiquement.
C'est une propriété générale des
modèles dynamiques.

La vraisemblance est la fonction
de densité de l'échantillon
conditionnelle au modèle et ses
paramètres.

Supposons de l'échantillon
soit

(5)

$$y_T = \{y_1, y_2, \dots, y_T\}$$

Nous devons donc déterminer
la distribution jointe du vecteur :

$$(Y_1, Y_2, \dots, Y_T)$$

Nous verrons deux définitions de
la fonction de vraisemblance :

⊕ la vraisemblance exacte
 L_0 la densité jointe

⊕ la vraisemblance conditionnelle
 L_0 en conditionne le distribut°
sur des conditions initiales
historiques pour y ou des hypothèses
sur les innovations initiales ε
(non observées).



Nous supposons que le modèle ⁽⁴⁾ est stationnaire au second ordre et gaussien.

• Nous pourrions éventuellement abandonner l'hypothèse de normalité, mais celle-ci simplifie considérablement l'évaluation de la vraisemblance, car :

- une combinaison linéaire de v.a.r. gaussienne est gaussienne
- stabilité par rapport à la marginalisation et au conditionnement.

• L'hypothèse de stationnarité est indispensable pour construire la vraisemblance exacte. On peut éventuellement utiliser la vraisemblance conditionnelle si l'hypothèse de stationnarité n'est pas satisfaite.

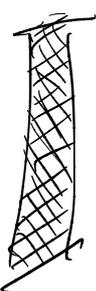
Sous l'hypothèse de stationnarité, les deux définitions de la vraisemblance sont asymptotiquement équivalentes. On verra que le maximum de vraisemblance conditionnelle peut s'interpréter comme un estimateur des MCO.

Il suffit donc de connaître la densité d'une loi normale multivariée pour écrire la fonction de vraisemblance. Si le vecteur X est normalement distribué d'espérance μ_x et de variance $\sum_{n \times n} x$, on note

$$X \sim \mathcal{N}(\mu_x, \Sigma_x)$$

et la fonction de densité est:

(6)


$$f(x) = (2\pi)^{-\frac{n}{2}} |\Sigma_x|^{-\frac{1}{2}} e^{-\frac{1}{2} (x-\mu_x)' \Sigma_x^{-1} (x-\mu_x)}$$

Dans le cas qui nous intéresse, c'est-à-dire pour un modèle dynamique, on retrouvera la fonction d'autocovariance de Y dans la matrice Σ :

$$\Sigma_Y = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(1) & & & \\ \vdots & & & \\ \gamma(T-1) & \gamma(1) & \gamma(0) & \end{pmatrix}$$

C'est pourquoi il est important de ~~calculer~~ calculer les moments d'ordre 2 d'un modèle ARMA.

(1) Vraisemblance des modèles AR

(1.1) Vraisemblance exacte des modèles AR(1)

Le modèle est

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t$$

où $|\varphi| < 1$ et $\varepsilon_t \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$.

L'échantillon utilisé pour l'estimation est constitué de T réalisations successives :

$$Y_T = \{y_1, y_2, \dots, y_T\}$$

Pour écrire la vraisemblance exacte il faut déterminer la distribution jointe du vecteur $(Y_1, Y_2, \dots, Y_T) \equiv \mathbf{y}$

APPROCHE DIRECTE

(8)

Nous savons que ce vecteur est normalement distribué, il nous faut déterminer son espérance et sa variance.

Nous avons :

$$E[Y_t] = \frac{c}{1-\varphi} = \mu \quad \forall t$$

ainsi :

$$E[\mathbf{Y}] = \boldsymbol{\mu}$$

où $\boldsymbol{\mu}$ est un vecteur $T \times 1$ dont chaque élément est égal à μ .

Pour la variance :

$$V[\mathbf{Y}] = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})']$$

$T \times T$

(=) $V[\mathbf{Y}] = \mathbb{I}$

$$\begin{pmatrix} (y_1 - \mu)^2 & (y_1 - \mu)(y_2 - \mu) & \dots & (y_1 - \mu)(y_T - \mu) \\ (y_2 - \mu)(y_1 - \mu) & (y_2 - \mu)^2 & & (y_2 - \mu)(y_T - \mu) \\ \vdots & & \ddots & \vdots \\ (y_T - \mu)(y_1 - \mu) & (y_T - \mu)(y_2 - \mu) & & (y_T - \mu)^2 \end{pmatrix}$$

(=) $V[\mathbf{Y}] =$

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(T-1) \\ \gamma(1) & & & \vdots \\ \vdots & & & \gamma(1) \\ \gamma(T-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}$$

(=) $V[\mathbf{Y}] =$

$$\frac{\sigma^2}{1 - \varphi^2} \begin{pmatrix} 1 & \varphi & \varphi^2 & \dots & \varphi^{T-1} \\ \varphi & & & & \varphi^2 \\ \varphi^2 & & & & \varphi \\ \vdots & & & & \varphi \\ \varphi^{T-1} & \dots & \varphi^2 & \varphi & 1 \end{pmatrix}$$

$$V[\mathbf{y}] = \sigma_\varepsilon^2 \Omega \equiv \Sigma$$

avec $\Omega = \frac{1}{1-\varphi^2} \begin{pmatrix} 1 & \varphi & \dots & \varphi^{T-1} \\ \varphi & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \varphi \\ \varphi^{T-1} & \dots & \varphi & 1 \end{pmatrix}$

La fonction de vraisemblance est donc

$$L(\varphi, \sigma_\varepsilon^2; \mathbf{y}_T) = (2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}-\mu)' \Sigma^{-1} (\mathbf{y}-\mu)}$$

ou encore :

$$L(\varphi, \sigma_\varepsilon^2; \mathbf{y}_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{y}-\mu)' \Omega^{-1} (\mathbf{y}-\mu)}$$

Nous devons donc calculer le déterminant et l'inverse de la matrice Ω pour évaluer la vraisemblance.

On exploite le résultat suivant:

(11)

Il est possible de factoriser la matrice Ω^{-1} sous la forme $\Omega^{-1} = L'L$ avec

$$L = \begin{pmatrix} \sqrt{1-\varphi^2} & 0 & 0 & 0 \\ -\varphi & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

On a alors

• Pour le déterminant:

$$|\Omega|^{-\frac{1}{2}} = |\Omega^{-1}|^{\frac{1}{2}}$$

$$\Rightarrow |\Omega|^{-\frac{1}{2}} = |L'L|^{\frac{1}{2}}$$

$$\Rightarrow |\Omega|^{-\frac{1}{2}} = (|L'| |L|)^{\frac{1}{2}}$$

$$\Rightarrow |\Omega|^{-\frac{1}{2}} = \sqrt{1-\varphi^2}$$

• le déterminant d'un produit est le produit des déterminants
• le déterminant d'une matrice triangulaire est le produit des éléments sur la diagonale.

• Pour la forme quadratique sous l'exponentielle :

$$-\frac{1}{2\sigma_\varepsilon^2} (\gamma - \mu)' \Omega^{-1} (\gamma - \mu) = -\frac{1}{2\sigma_\varepsilon^2} (\gamma - \mu)' L' L (\gamma - \mu)$$

$$\Leftrightarrow -\frac{1}{2\sigma_\varepsilon^2} (\gamma - \mu)' \Omega^{-1} (\gamma - \mu) = -\frac{1}{2\sigma_\varepsilon^2} [L(\gamma - \mu)]' [L(\gamma - \mu)]$$

$$L(\gamma - \mu) = \begin{pmatrix} \sqrt{1-\varphi^2} & 0 & \dots & 0 \\ -\varphi & 1 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & -\varphi & 1 \end{pmatrix} \begin{pmatrix} y_1 - \mu \\ y_2 - \mu \\ \vdots \\ y_T - \mu \end{pmatrix}$$

$$\Leftrightarrow L(\gamma - \mu) = \begin{pmatrix} (y_1 - \mu)\sqrt{1-\varphi^2} \\ (y_2 - \mu) - \varphi(y_1 - \mu) \leftarrow \varepsilon_2 \\ \vdots \\ (y_T - \mu) - \varphi(y_{T-1} - \mu) \leftarrow \varepsilon_T \end{pmatrix}$$

D'où

$$(y_t - \mu) - \varphi(y_{t-1} - \mu) = y_t - (1-\varphi)\mu - \varphi y_{t-1} = y_t - c - \varphi y_{t-1}$$

$$-\frac{1}{2\sigma_\varepsilon^2} (\gamma - \mu)' \Omega^{-1} (\gamma - \mu) = -\frac{1-\varphi^2}{2\sigma_\varepsilon^2} (y_1 - \mu)^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T \underbrace{(y_t - c - \varphi y_{t-1})^2}_{\varepsilon_t}$$

Enfin :

(13)

$$L(\varphi, \sigma_\varepsilon^2; y_T) = \underbrace{\left(2\pi \frac{\sigma_\varepsilon^2}{1-\varphi^2}\right)^{-\frac{1}{2}} e^{-\frac{1-\varphi^2}{2\sigma_\varepsilon^2} (y_1 - \mu)^2}}_{\text{densité de la condition initiale}} \cdot \underbrace{\left(2\pi \sigma_\varepsilon^2\right)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (y_t - \varphi y_{t-1})^2}}_{\text{produit des densités conditionnelles}}$$

APPROCHE ALTERNATIVE (Bayes)

Pour bien comprendre cette écriture de la vraisemblance, on peut arriver au même résultat en utilisant le théorème de Bayes. En effet la densité jointe peut s'écrire comme un produit de densités conditionnelles et une densité marginale pour la condition initiale.

$$\begin{aligned} L(\varphi, \sigma_\varepsilon^2; y_T) &= \int_{y_1, y_2, \dots, y_T} f(y_1, y_2, \dots, y_T) \\ &= \underbrace{f(y_T | y_{T-1}, \dots, y_1) \cdot f(y_{T-1} | y_{T-2}, \dots, y_1) \cdots f(y_2 | y_1)}_{\text{produit des densités conditionnelles}} \cdot \underbrace{f(y_1)}_{\text{densité marginale}} \\ &= f(y_1) \cdot \prod_{t=2}^T f(y_t | y_{t-1}) \end{aligned}$$

• Quelle est la loi de $Y_t | Y_{t-1}$?

On a

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t$$

$$\Rightarrow Y_t | Y_{t-1} \sim \mathcal{N}(c + \varphi Y_{t-1}, \sigma_\varepsilon^2)$$

$$\Rightarrow f_{Y_t | Y_{t-1}}(y_t) = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2}(y_t - c - \varphi y_{t-1})^2}$$

• Quelle est la loi de Y_1 ?

Sous l'hypothèse de stationnarité la distribution de la condition initiale doit être identique à la distribution limite du processus AR(1). On doit donc avoir

$$Y_1 \sim \mathcal{N}\left(\frac{c}{1-\varphi}, \frac{\sigma_\varepsilon^2}{1-\varphi^2}\right)$$

et donc

$$f(y_1) = \left(2\pi \cdot \frac{\sigma_\varepsilon^2}{1-\varphi^2}\right)^{-\frac{1}{2}} \cdot e^{-\frac{1-\varphi^2}{2\sigma_\varepsilon^2} (y_1 - \mu)^2}$$

On obtient donc bien:

$$L(\varphi, \sigma_\varepsilon^2; Y_T) = \underbrace{\left(2\pi \frac{\sigma_\varepsilon^2}{1-\varphi^2}\right)^{-\frac{1}{2}} e^{-\frac{1-\varphi^2}{2\sigma_\varepsilon^2} (y_1 - \mu)^2}}_{\text{densité de la condition initiale}} \cdot \underbrace{\left(2\pi \sigma_\varepsilon^2\right)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (y_t - \varphi y_{t-1})^2}}_{\text{produit des densités conditionnelles}}$$

La factorisation de l'inverse de Ω sous la forme $\Omega^{-1} = L'L$ est, en ce sens, une manifestation du théorème de Bayes -

On peut appliquer ces deux approches pour des modèles autorégressifs plus généraux -

L'estimateur du maximum de vraisemblance est obtenu en maximisant la fonction de vraisemblance par rapport à φ et σ_ε^2 .
 En pratique, on maximise plutôt le logarithme de la vraisemblance :

$$\mathcal{L}(\varphi, \sigma_\varepsilon^2; y_T) = \underbrace{-\frac{T}{2} \log(2\pi) - \frac{1}{2} \log \frac{\sigma_\varepsilon^2}{1-\varphi^2} - \frac{1-\varphi^2}{2\sigma_\varepsilon^2} (y_T - \mu)^2}_{\text{(log) densité de la condition initiale}} - \underbrace{\frac{T-1}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (y_t - c - \varphi y_{t-1})^2}_{\text{(log) densités des distributions conditionnelles}}$$

À cause du terme lié à la condition initiale, la CNO | φ n'est pas linéaire par rapport à φ \rightarrow Pas d'expression analytique pour l'estimateur du maximum de Vraisemblance.

(7)
Il est donc nécessaire d'utiliser
une routine d'optimisation pour obtenir
l'estimateur du Maximum de
Vraisemblance (exacte).

(1.2) Vraisemblance conditionnelle du modèle AR(1)

Dans l'expression de la vraisemblance
exacte la difficulté vient de la
densité marginale de la condition
initiale.

Une façon de simplifier le problème
est d'éliminer ce terme en
considérant la densité jointe de
 (y_1, \dots, y_T) conditionnellement à $y_1 = y_1$
qui est observé (dans l'échantillon).

La vraisemblance conditionnelle
est donc :

(18)

$$L_c(\varphi, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T (y_t - c - \varphi y_{t-1})^2}$$

ou en log :

$$\ln L_c(\varphi, \sigma_\varepsilon^2; Y_T) = -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=2}^T \underbrace{(y_t - c - \varphi y_{t-1})^2}_{\varepsilon_t^2}$$

↳ Maximiser la vraisemblance (cond.) par rapport à c et φ est alors équivalent à minimiser :

$$\sum_{t=2}^T (y_t - c - \varphi y_{t-1})^2$$

la somme
des carrés
des résidus

c'est-à-dire à l'estimateur des
MCO.

On voit donc que l'estimateur du ⁽¹⁸⁾
MV conditionnelle est beaucoup plus
simple (il n'est pas nécessaire d'utiliser
une routine d'optimisation)

Asymptotiquement, quand T tend
vers l'infini, les estimateurs basés
sur la vraisemblance conditionnelle
ou exacte donnent les mêmes
résultats car (sous l'hypothèse
de stationnarité) l'effet de la
condition initiale dans la
vraisemblance s'estompé -

On a donc :

(20)

$$\hat{\varphi}_{mv} = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^T (y_{t-1} - \bar{y})^2}$$

$$\hat{c}_{mv} = \bar{y}(1 - \hat{\varphi}_{mv})$$

$$\hat{\sigma}_{mv}^2 = \frac{1}{T-1} \sum_{t=2}^T (y_t - \hat{c}_{mv} - \hat{\varphi}_{mv} y_{t-1})^2$$

avec la vraisemblance conditionnelle (les MCO).

(1.3) Biais de l'estimateur des MCO

Considérons le cas du modèle AR(1) sans constante, pour simplifier :

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

avec $|\varphi| < 1$ et $\varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$

Nous disposons d'un échantillon

(21)

$$y_T = \{y_1, \dots, y_T\}$$

L'estimateur des MCO de φ est défini par

$$\hat{\varphi}_T = \frac{\sum_{t=2}^T y_t y_{t-1}}{\sum_{t=2}^T y_{t-1}^2}$$

$$\Rightarrow \hat{\varphi}_T = \frac{\varphi \sum_{t=2}^T y_{t-1}^2 + \sum_{t=2}^T y_{t-1} \varepsilon_t}{\sum_{t=2}^T y_{t-1}^2}$$

en substituant le XCP

$$\Rightarrow \hat{\varphi} = \varphi + \frac{\sum_{t=2}^T y_{t-1} \varepsilon_t}{\sum_{t=1}^T y_{t-1}^2}$$

L'estimateur des MCO est sans biais si et seulement si

$$E \left[\frac{\sum_{t=2}^T y_{t-1} \varepsilon_t}{\sum_{t=2}^T y_{t-1}^2} \right] = 0$$

Ce n'est clairement pas le cas car ε_t au numérateur n'est pas indépendant de $y_{t-1}^2, y_{t-2}^2, \dots, y_{T-1}^2$ au dénominateur. (22)

Si $\varphi > 0$, alors $\varepsilon_t > 0$ aura un effet positif sur les valeurs à vendre de y . Cela induit une corrélation négative entre ε_t et $\frac{y_{t-1}}{\sum_{t=2}^T y_{t-1}^2}$ de sorte que

$$\mathbb{E}[\hat{\varphi}] < \varphi$$

le biais est

- d'autant plus important que φ est proche de 1
- d'autant plus faible que T est grand, car ε_t est corrélé avec moins d'éléments dans $\sum_{t=2}^T y_{t-1}^2$.

(1.4) Vraisemblances du modèle AR(p)

On commence par discuter la forme de la vraisemblance exacte. On obtiendra la vraisemblance conditionnelle en retirant les termes liés aux conditions initiales. Comme pour l'AR(1), l'estimateur du maximum de vraisemblance conditionnelle est identique à l'estimateur des MCO.

Le modèle AR(p) est défini par

$$Y_t = c + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

avec $\varepsilon_t \sim \text{BB}(\sigma_\varepsilon^2)$, $(\varphi_i)_{i=1}^p$ tels que le modèle est stable (les racines du polynôme retard sont plus grandes que 1 en

module. On suppose que le 24
modèle est stationnaire au second
ordre (invariance des moments
d'ordre 1 et 2).

On note

$$Y_T = \{y_1, y_2, \dots, y_T\}$$

l'échantillon utilisé pour l'estimation.

Si le modèle a p retards nous
avons besoin de p conditions
initiales :

$$y_1, y_2, \dots, y_p$$

La vraisemblance est la densité de
l'échantillon :

$$L(\phi_1, \dots, \phi_p, \sigma^2; Y_T) = f(y_1, y_2, \dots, y_T)$$

Comme dans le cas de l'AR(1) on peut réécrire (théorème de Bayes) cette densité jointe comme un produit de densités conditionnelles et de la densité jointe des conditions initiales:

$$L(\varphi_1, \dots, \varphi_p, \sigma_\varepsilon^2; Y_T) = f(y_1, \dots, y_p) \cdot \prod_{t=p+1}^T f(y_t | y_{t-1}, \dots, y_{t-p})$$

• Densités conditionnelles

$$Y_t | Y_{t-1}, \dots, Y_{t-p} \sim \mathcal{N}(c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p}, \sigma_\varepsilon^2)$$

$$L_{\mathcal{D}} f(y_t) = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} (y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p})^2}$$

• Densité jointe pour les conditions initiales

On sait que $(Y_1, \dots, Y_p)'$ est normalement distribué. L'espérance de ce

vecteur aléatoire est:

(26)

$$\mathbb{E} \left[\begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \right] = \frac{c}{1 - \varphi_1 - \dots - \varphi_p} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \equiv \mu$$

$p \times 1$

sa variance est:

$$V \left[\begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \right] = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & & & \vdots \\ \vdots & & & \gamma(1) \\ \gamma(p-1) & \dots & \gamma(1) & \gamma(0) \end{pmatrix} \equiv \Sigma$$

où $\gamma(h)$ est la fonction d'autocovariance du processus AR(p), une fonction des paramètres autorégressifs $(\varphi_1, \dots, \varphi_p)$ et de la variance de l'innovation (σ_ε^2) . On a alors

$$f(y_1, \dots, y_p) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(y-\mu)' \Sigma^{-1} (y-\mu)}$$

y_1, \dots, y_p

où $y = (y_1, \dots, y_p)'$.

La vraisemblance (exacte) est donc: (27)

$$L(\varphi_1, \dots, \varphi_p, \sigma_\varepsilon^2; Y_T) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2} (Y-\mu)' \Sigma^{-1} (Y-\mu)} \times \\ (2\pi\sigma_\varepsilon^2)^{-\frac{T-p}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^T (y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p})^2}$$

ou, pour se rapprocher de l'expression obtenue pour l'AR(1), en notant $\Sigma = \sigma_\varepsilon^2 \Omega$:

$$L(\varphi_1, \dots, \varphi_p, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{p}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} (Y-\mu)' \Omega^{-1} (Y-\mu)} \times \\ (2\pi\sigma_\varepsilon^2)^{-\frac{T-p}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^T (y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p})^2}$$

On obtient la vraisemblance conditionnelle en omettant le terme lié aux conditions initiales. La log vraisemblance conditionnelle est donc:

$$\ln_c L(\varphi_1, \dots, \varphi_p, c, \sigma_\varepsilon^2; Y_T) = -\frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log \sigma_\varepsilon^2 \\ - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^T (y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p})^2$$

Maximiser la vraisemblance (conditionnelle) ⁽²⁸⁾
 par rapport à la constante (c) et les
 paramètres autorégressifs ($\varphi_1, \dots, \varphi_p$) revient
 à minimiser la somme des carrés des
 résidus $\varepsilon_t = y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p}$.
 \Leftrightarrow MCO.

(2) Vraisemblance des modèles MA

2.1) Vraisemblance exacte du modèle MA(1)

Le modèle MA(1) est défini par :

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

avec $\varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ et $|\theta| < 1$ (inversibilité).

Rappel La fonction d'autocovariance du MA(1) est :

$$\gamma(h) = \begin{cases} (1 + \theta^2) \sigma_\varepsilon^2 & \text{si } h = 0 \\ \theta & \text{si } h = \pm 1 \\ 0 & \text{sinon.} \end{cases}$$

On dispose d'un échantillon

(29)

$$Y_T = \{y_1, y_2, \dots, y_T\}$$

On note $y = (y_1, y_2, \dots, y_T)'$ vecteur de réalisations

et $Y = (Y_1, Y_2, \dots, Y_T)'$ vecteur aléatoire

des vecteurs $T \times 1$.

L'espérance de Y est nulle et sa variance est :

$$V[Y] = E[YY'] = \sigma^2 \begin{pmatrix} 1+\theta^2 & \theta & 0 & \dots & 0 \\ \theta & 1 & \theta & \dots & 0 \\ 0 & \theta & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \theta & 1+\theta^2 \end{pmatrix} = \Sigma$$

Puisque Y est normalement distribué, la vraisemblance exacte est donc :

$$L(\theta, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} y' \Omega^{-1} y}$$

et donc

$$L(\theta, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} |D|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \tilde{y}' D^{-1} \tilde{y}}$$

où $\tilde{y} = A^{-1} \cdot y$. On peut inverser la matrice A , et donc calculer \tilde{y} , de façon récursive. En effet nous devons avoir

$$A \tilde{y} = y$$

La première ligne de ce système d'équation nous dit que

$$\tilde{y}_1 = y_1$$

La deuxième ligne nous dit que

$$\frac{\rho}{1+\rho^2} \tilde{y}_1 + \tilde{y}_2 = y_2$$

c'est-à-dire que

$$\tilde{y}_2 = y_2 - \frac{\theta}{1+\theta^2} \tilde{y}_1$$

Plus généralement, la t-ième ligne nous dit que :

$$\tilde{y}_t = y_t - \frac{\theta(1+\theta^2+\dots+\theta^{2(t-1)})}{1+\theta^2+\dots+\theta^{2(t-1)}} \tilde{y}_{t-1}$$

On peut donc construire le vecteur \tilde{y} en parcourant le système d'équations du haut vers le bas.

Le déterminant de D est égal au produit des éléments sur sa diagonale :

$$|D| = \prod_{t=1}^T \frac{1 + \sum_{k=1}^t \theta^{2k}}{\sum_{k=1}^t \theta^{2(k-1)}} = \prod_{t=1}^T d_{tt}$$

Finalement, la forme quadratique sous l'exponentielle peut s'écrire :

$$\tilde{y}' D^{-1} \tilde{y} = \sum_{t=1}^T \frac{\tilde{y}_t^2}{1 + \sum_{\tau=1}^t \theta^{2\tau}}$$

$$\sum_{\tau=1}^t \theta^{2(\tau-1)}$$

$$\Rightarrow \tilde{y}' D^{-1} \tilde{y} = \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}$$

Ainsi, la vraisemblance exacte s'écrit :

$$L(\theta, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} \left[\prod_{t=1}^T d_{tt} \right]^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}}$$

La vraisemblance exacte, à nouveau, est une fonction non linéaire des paramètres (surtout θ) que nous cherchons à estimer. Il n'existe pas de solution analytique pour l'estimateur du MV \rightarrow routine d'optimisation.

(2.2) Vraisemblance conditionnelle du MA(1) (35)

Comme pour les processus autorégressifs, on peut définir une vraisemblance conditionnelle (un peu moins simplement).

On note \mathcal{I}_t l'ensemble d'information à la date t . Cet ensemble contient les réalisations de y et ε jusqu'à la date t

$$\mathcal{I}_t = \{y_t, y_{t-1}, \dots, \varepsilon_t, \varepsilon_{t-1}, \dots\}$$

$y_t | \mathcal{I}_{t-1}$ est une variable aléatoire normale (car $\varepsilon_t \notin \mathcal{I}_{t-1}$):

$$y_t | \mathcal{I}_{t-1} \sim \mathcal{N}(\theta \varepsilon_{t-1}, \sigma_\varepsilon^2)$$

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

La densité de $y_t | \mathcal{I}_{t-1}$ (ou $y_t | \varepsilon_{t-1}$) est donc:

$$f_{y_t | \mathcal{I}_{t-1}}(y_t) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \underbrace{(y_t - \theta \varepsilon_{t-1})}_{\varepsilon_t}^2}$$

Supposons que $\epsilon_0 = 0$. On a alors

$$Y_1 | \epsilon_0 \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

Puisque y_1 est observé (dans l'échantillon) nous savons aussi que:

$$\epsilon_1 = y_1$$

En effet, nous savons que

$$\epsilon_t = y_t - \theta \epsilon_{t-1} \quad \forall t$$

Ainsi

$$Y_2 | Y_1 = y_1, \epsilon_0 = 0 \sim \mathcal{N}(\theta \overset{y_1}{\epsilon_1}, \sigma_\epsilon^2)$$

Puisque nous connaissons certainement ϵ_1 (sous l'hypothèse $\epsilon_0 = 0$) nous pouvons déduire ϵ_2 de façon certaine à partir de l'observable y_2 :

$$\epsilon_2 = y_2 - \theta \epsilon_1$$

En poursuivant de cette façon, il est possible d'obtenir la suite $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}$ des innovations à partir de l'échantillon en formulant une hypothèse sur ε_0 (supposé nul). (39)

En notant que les variables aléatoires $Y_t | \mathcal{F}_{t-1}$ et $Y_{t+s} | \mathcal{F}_{t+s-1}$ sont indépendantes pour $s \neq 0$, puisqu'il s'agit de ε_t et ε_{t+s} , on peut écrire la vraisemblance comme un produit de densités conditionnelles (densités des ε)

$$L_c(\theta, \sigma_\varepsilon^2; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \varepsilon_t^2}$$

où ε_t est obtenu récursivement à partir de ε_0 et

$$\varepsilon_t = y_t - \theta \varepsilon_{t-1}$$

en utilisant l'échantillon.

(38)
Il convient de s'interroger sur la pertinence de l'hypothèse $\varepsilon_0 = 0$. En itérant vers le passé, on a :

$$\bullet \quad \varepsilon_t = y_t - \theta (y_{t-1} - \theta \varepsilon_{t-2})$$

$$\Leftrightarrow \varepsilon_t = y_t - \theta y_{t-1} + \theta^2 \varepsilon_{t-2}$$

$$\bullet \quad \varepsilon_t = y_t - \theta y_{t-1} + \theta^2 (y_{t-2} - \theta \varepsilon_{t-3})$$

$$\Leftrightarrow \varepsilon_t = y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 \varepsilon_{t-3}$$

... puis en remontant jusqu'à ε_0 :

$$\varepsilon_t = y_t - \theta y_{t-1} + \theta^2 y_{t-2} + \dots + (-1)^i \theta^i y_{t-i} + \dots + (-1)^t \theta^t \varepsilon_0$$

Si $|\theta| \geq 1$, c'est-à-dire si le processus MA(1) n'est pas inversible, il ne vaut mieux pas utiliser cette approche car au mieux (si $|\theta|=1$) la conséquence sur ε_t du choix pour ε_0 se fait sentir pour tout t ,

ou pire ($|\theta| > 1$) l'effet sur ε_t est de plus important (quand t augmente θ^t augmente). (39)

Si le processus est inversible, $|\theta| < 1$, la valeur choisie pour ε_0 n'a asymptotiquement aucune importance. Évidemment, à la distance finie, ce choix affecte l'inférence.

Remarque 1 En principe nous pourrions aussi estimer ε_0 en traitant cette innovation comme un paramètre supplémentaire.

(2.3) Vraisemblances du MA(q)

(40)

Le modèle est

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

avec $\varepsilon_t \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ et on suppose que les racines de $\Phi(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ sont à l'extérieur du cercle unité.

On fait comme pour le MA(1). Pour la vraisemblance conditionnelle il faut juste postuler plus de valeurs pour les innovations initiales. Le modèle a q retards sur ε , nous avons donc besoin de q conditions initiales sur ε :

$$\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \dots = \varepsilon_{-q+1} = 0$$

Comme dans le modèle $MA(1)$ (41)
on suppose que les innovations initiales
sont nulles... C'est plus simple et
sans conséquence asymptotiquement dans
la mesure où le modèle $MA(q)$ est
invertible (les racines de $\Phi(z)$ sont plus
grandes que 1 en module).

Sous cette hypothèse, en utilisant
l'échantillon $\mathcal{Y}_T = \{y_1, \dots, y_T\}$ on peut
construire récursivement la suite $\{\varepsilon_t\}_{t=1}^T$
car on sait que

$$\varepsilon_t = y_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

La vraisemblance conditionnelle est alors

$$L_c(\theta_1, \dots, \theta_q, \sigma_\varepsilon^2; \mathcal{Y}_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \varepsilon_t^2}$$

Pour la vraisemblance exacte, on doit écrire la densité jointe de (Y_1, \dots, Y_T) . Dans le cas du MA(q) on sait que ce vecteur est normalement distribué d'espérance nulle (puisque'il n'y a pas de constante) et de variance :

$$\Sigma = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(q) & 0 \\ \gamma(1) & \gamma(0) & \dots & \gamma(q-1) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(q) & \gamma(q-1) & \dots & \gamma(0) & 0 \\ 0 & \gamma(q) & \dots & \gamma(1) & \gamma(0) \end{pmatrix} = \sigma_\varepsilon^2 \Omega$$

où

$$\gamma(h) = \sigma_\varepsilon^2 \sum_{i=0}^{1-h} \theta_i \theta_{i-h}$$

avec $\theta_0 = 1$ et $\theta_{-i} = 0 \forall i > 0$.

On notera d_{ii} le i -ème élément sur la diagonale de la matrice D . La vraisemblance peut donc s'écrire

$$L(\theta_1, \dots, \theta_q, \sigma_\varepsilon^2; \mathbf{y}_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} |D|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \tilde{\mathbf{y}}' D^{-1} \tilde{\mathbf{y}}}$$

où les éléments de $\tilde{\mathbf{y}}$ sont obtenus récursivement:

$$\tilde{y}_1 = y_1$$

$$\tilde{y}_2 = y_2 - a_{2,1} \tilde{y}_1$$

$$\tilde{y}_3 = y_3 - a_{3,1} \tilde{y}_1 - a_{3,2} \tilde{y}_2$$

⋮

$$\tilde{y}_t = y_t - a_{t,t-1} \tilde{y}_{t-1} - \dots - a_{t,t-q} \tilde{y}_{t-q}$$

Finalement:

$$L(\theta_1, \dots, \theta_q, \sigma_\varepsilon^2; \mathbf{y}_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} \left(\prod_{t=1}^T d_{tt} \right)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}}$$

(3) Vraisemblance d'un modèle ARMA(p,q) (45)

Soit le processus stochastique

$$Y_t = c + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

avec $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. On note

$$\Lambda = (c, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)'$$

le vecteur des paramètres à estimer. On peut facilement évaluer la vraisemblance conditionnelle. Notons

$$y_0 = (y_0, y_{-1}, \dots, y_{-p+1})$$

un vecteur de conditions initiales pour l'endogène et

$$e_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1})$$

un vecteur de conditions initiales pour les innovations. On peut alors pour tout $t \geq 1$ déduire les innovations

à partir de l'échantillon (et bien sûr des paramètres Λ): (46)

$$E_t = y_t - c - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Conditionnellement à l'ensemble d'information en $t-1$, \mathcal{F}_{t-1} , qui contient y_{t-1}, \dots, y_{t-p} et $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$, y_t est une variable aléatoire gaussienne avec :

$$E[y_t | \mathcal{F}_{t-1}] = c + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$$\text{et } V[y_t | \mathcal{F}_{t-1}] = \sigma_\varepsilon^2$$

De plus $y_t | \mathcal{F}_{t-1}$ est indépendant de $y_s | \mathcal{F}_{s-1}$ $\forall t \neq s$. La vraisemblance conditionnelle s'écrit comme un produit de densités conditionnelles :

$$L(\underline{\Lambda}; Y_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \varepsilon_t^2} \quad (47)$$

où ε_t est obtenu récursivement à partir de e_0 et y_0 .

En général on considère comme condition initiale pour les innovations :

$$\varepsilon_i = 0 \quad \text{pour } i = 0, -1, \dots, -q+1$$

Pour l'endogène on peut utiliser des données (mais on perd alors des observations) ou à l'instar du choix pour ε , on considère l'absence de y comme condition initiale :

$$y_i = \frac{c}{1 - \varphi_1 - \dots - \varphi_p} \quad \text{pour } i = 0, -1, \dots, -p+1$$

(48)
Pour la vraisemblance exacte on fait comme d'habitude... le plus gros du travail est d'écrire la variance du vecteur (Y_1, \dots, Y_T) et éventuellement de factoriser cette matrice (mais l'ordinateur sait le faire pour vous).

(4) Propriétés statistiques de l'estimateur du MV (49)

L'estimateur du maximum de vraisemblance est convergent (même s'il est biaisé à distance finie), il est aussi asymptotiquement normalement distribué :

$$\sqrt{T} (\hat{\Lambda}_T - \Lambda_0) \xrightarrow[T \rightarrow \infty]{\text{loi}} \mathcal{N}(0, I(\Lambda_0)^{-1})$$

↑ vraie valeur des paramètres

où $I(\Lambda_0)$ est la matrice d'information de Fisher, évaluée en Λ_0 (vraie valeur du vecteur des paramètres) :

$$I(\Lambda_0) = -\mathbb{E} \left[\frac{\partial^2 \mathcal{L}(\Lambda; \mathcal{Y}_T)}{\partial \Lambda \partial \Lambda'} \Big|_{\Lambda = \Lambda_0} \right]$$

où $\mathcal{L}(\Lambda; \mathcal{Y}_T) = \log L(\Lambda; \mathcal{Y}_T)$ est le log-vraisemblance. Une définition

alternative de la matrice d'information de Fisher est basée sur le gradient de la log-vraisemblance (le score):

$$I(\lambda_0) = E \left[\left. \frac{\partial \mathcal{L}(\lambda; y_T)}{\partial \lambda} \right|_{\lambda=\lambda_0} \cdot \left. \frac{\partial \mathcal{L}(\lambda; y_T)}{\partial \lambda'} \right|_{\lambda=\lambda_0} \right]$$

En pratique, puisque nous ne connaissons pas la vraie valeur du vecteur de paramètres, on remplace λ_0 par l'estimateur du maximum de vraisemblance. On calcule l'espérance en calculant la moyenne arithmétique des dérivés d'ordre 2 ou « carrés » des gradients de la densité (cond.) des observations.

En exploitant ce résultat asymptotique (51)
On peut définir des tests basés sur la vraisemblance :

- test du ratio de vraisemblance
[estimation des modèles contraint et non contraint]
- test du multiplicateur de Lagrange
[estimation du modèle contraint seulement]

(5) Sélection de modèle

Comment choisir p et q dans un modèle ARMA(p, q)?

- Estimer un ARMA avec p_{\max}, q_{\max} puis tester la nullité des paramètres associés à $y_{t-p_{\max}}$ ou $\varepsilon_{t-q_{\max}}$. Si un des paramètres est nul, spécifier le modèle avec moins de paramètres et reprendre le test...

- Utiliser des tests de ratio de vraisemblance ou du multiplicateur de Lagrange.
- Minimiser un critère d'information.

- Choisir p_{\max} et q_{\max}

- Estimer $(1+p_{\max}) \times (1+q_{\max}) - 1$ modèles

ARMA(p, q) avec $p = 0, 1, \dots, p_{\max}$
 et $q = 0, 1, \dots, q_{\max}$ (en excluant
 le cas $p = q = 0$).

- Pour chaque modèle, calculer
 le critère

$$AIC = 2k - 2 \cdot \log(L)$$

$$\rightarrow AIC_c = AIC + \frac{2k(k+1)}{T-k-1}$$

$$BIC = \log(T) \cdot k - 2 \log(L)$$

$$HQIC = 2k \log(\log(T)) - 2 \log(L)$$

- Choisir le nombre de retard sur la partie MA ou AR en étudiant les moments d'ordre 2 empiriques (autocorrélation, autocorrélation partielle, autocorrélation inverse).

↳ Il existe des procédures automatiques (l'avantage est qu'il n'est pas nécessaire d'estimer des modèles ARMA).