

ECO9036: Macroéconométrie

Alain Guay

April 27, 2007

SÉRIES TEMPORELLES

1 Introduction

On observe une dépendance temporelle pour les séries macroéconomique et financières. Cette dépendance peut découler de différents phénomènes économiques.

Exemples:

- Investissement;
- Coûts d'ajustement;
- Habitudes de consommation.

On observe donc des séries qui sont corrélées dans le temps (autocorrélées).

Objectif:

On va chercher une représentation parcimonieuse de cette dépendance temporelle pour des fins de prévisions. On va se situer premièrement dans un contexte univarié. La série sera donc modélisée en fonction de son passé seulement.

2 Présentation des séries temporelles

2.1 Processus stochastique

Un processus stochastique est une séquence de variables aléatoires définies sur un ensemble Ω . Ces variables aléatoires sont soit univariées ou multivariées et sont fonction du temps. On va noter un processus stochastique univarié de la façon suivante:

$$Y = (Y_t, t = t_0, t_{0+1}, t_{0+2}, \dots, 1, 2, \dots, \infty)$$

et t_0 peut être égal à $-\infty$.

Une série temporelle est un échantillon observé à différents points ou intervalles dans le temps noté $Y_t(\omega), t = 1, \dots, T$ où $\omega \in \Omega$.

Exemple:

$$y_1, y_2, \dots, y_T.$$

Les observations y_1, y_2, \dots, y_T sont donc une réalisation d'un processus stochastique Y_t . Voici un exemple d'un processus stochastique:

Exemple

$$\begin{aligned} \varepsilon_t &\sim i.i.d(0, \sigma^2) \quad t = 0, \pm 1, \pm 2, \pm 3, \dots \\ E(\varepsilon_t) &= 0 \quad E(\varepsilon_t \varepsilon_{t-j}) = 0 \quad \forall j. \end{aligned}$$

On peut alors simuler une réalisation de ce processus à l'aide de simulations sur ordinateur. On aura autant de réalisations que de simulations à partir du processus.

L'espérance mathématique de la variable aléatoire est donnée par la formule suivante:

$$\mu_t \equiv E(Y_t) = \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t$$

où $f_{Y_t}(y_t)$ est la fonction de densité marginale de Y_t . Cette espérance correspond à la moyenne de la fonction de répartition. Un estimateur de cette espérance est donnée par la moyenne empirique:

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

La variance de la variable aléatoire Y_t est définie comme étant:

$$\gamma_{0t} \equiv E(Y_t - \mu_t)^2 = \int_{-\infty}^{\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t) dy_t.$$

Un estimateur de cette variance est donnée par:

$$\hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2.$$

2.2 Fonctions d'autocovariance et d'autocorrélation

Fonctions décrivant la dynamique d'un processus stochastique.

Fonction d'autocovariance:

$$\begin{aligned}\gamma_k &= \text{cov}(Y_t, Y_{t-k}) = E[(Y_t - E(Y_t))(Y_{t-k} - E(Y_{t-k}))] \\ \gamma_0 &= \text{var}(Y_t) \\ \gamma_{-k} &= \text{cov}(Y_t, Y_{t+k}) = E[(Y_t - E(Y_t))(Y_{t+k} - E(Y_{t+k}))]\end{aligned}$$

La séquence

$$\gamma_k, \quad k = 0, \pm 1, \pm 2, \dots$$

est appelée fonction d'autocovariance. Cette fonction mesure la dépendance temporelle de la série.

Inconvénient: Elle n'est pas invariante aux unités de mesure.

Exemple

$$z_t = 100y_t$$

On aura que

$$\text{cov}(z_t z_{t+k}) = 10000 \text{cov}(y_t, y_{t+k})$$

Solution:

Si on divise la fonction d'autocovariances par la variance, γ_k , on aura une fonction invariante aux unités de mesure. On obtiendra ainsi la fonction d'autocorrélation.

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad k = 0, \pm 1, \pm 2, \dots$$

On aura que $|\rho_k| \leq 1$.

2.3 Processus stochastique stationnaire

Définition 1 *Un processus stochastique Y_t est stationnaire du second ordre si et seulement si*

1. $E(Y_t) = m$ indépendant de t ,
2. $\text{var}(Y_t) < \infty \quad \forall t$, et
3. $\text{cov}(Y_t, Y_{t-k}) = \gamma_k, \quad \forall k \text{ et } \forall t$.

Donc, les deux premiers moments sont indépendants de t .

Stationnarité au sens stricte

Définition 2 *La loi conjointe de (Y_1, Y_2, \dots, Y_t) est indépendante de t .*

2.4 Opérateurs de retards

On définit un opérateur de retards L de la façon suivante,

$$Ly_t = y_{t-1}.$$

En particulier, on aura

$$L^2 y_t = L(Ly_t) = Ly_{t-1} = y_{t-2}$$

et de façon générale,

$$L^n = y_{t-n}$$

Un polynôme d'ordre p d'opérateurs de retards est défini comme étant:

$$\alpha_p(L) = 1 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$$

Si on a deux polynômes de retards tels que

$$\Phi_p(L)\Psi_q(L) = 1$$

On dira que $\Psi_q(L)$ est l'inverse de $\Phi_p(L)$, c.à.d. que

$$\Psi_q(L) = [\Phi_p(L)]^{-1}$$

Étudions l'inversion de $(1 - \lambda L)$. L'application définie par $(1 - \lambda L)$ est inversible si et seulement si $|\lambda| \neq 1$. Examinons les différents cas. Pour le cas où $|\lambda| < 1$, l'inverse de $(1 - \lambda L)$ notée par $(1 - \lambda L)^{-1}$ est $\sum_{i=0}^{\infty} \lambda^i L^i$ puisque

$$\sum_{i=0}^{\infty} \lambda^i L^i (1 - \lambda L) = L^0 = 1.$$

Ainsi,

$$(1 - \lambda L)^{-1} y_t = \sum_{i=0}^{\infty} \lambda^i L^i y_t = \sum_{i=0}^{\infty} \lambda^i y_{t-i}.$$

Dans le cas où $|\lambda| > 1$. On a la relation suivante:

$$1 - \lambda L = -\lambda L \left[1 - \frac{1}{\lambda} L^{-1} \right].$$

La première partie du terme de droite est inversible et est donnée par $-\frac{1}{\lambda} L^{-1}$. La deuxième partie du terme de droite est également inversible et est donnée par $\sum_{i=0}^{\infty} \frac{1}{\lambda^i} L^{-i}$. On a alors,

$$(1 - \lambda L)^{-1} = \left(-\frac{1}{\lambda L} \right) \left(\sum_{i=0}^{\infty} \frac{1}{\lambda^i} L^{-i} \right) = - \sum_{i=1}^{\infty} \lambda^{-i} L^{-i}.$$

Ainsi,

$$(1 - \lambda L)^{-1} y_t = - \sum_{i=1}^{\infty} \lambda^{-i} L^{-i} y_t = - \sum_{i=1}^{\infty} \lambda^{-i} y_{t+i}.$$

Le terme de droite est donc fonction des valeurs futures de y_t . Dans le cas où $|\lambda| = 1$, l'application $(1 - \lambda L)$ n'est pas inversible.

2.5 Ergodicité du premier et du deuxième moments

L'ergodicité d'un processus s'exprime sous la forme de restrictions sur la mémoire de ce processus. Les valeurs réalisées par le processus à un horizon suffisamment éloigné doivent être pratiquement non corrélées.

On dira qu'un processus stationnaire du second ordre est ergodique en sa moyenne si la moyenne empirique converge en probabilité vers la moyenne théorique. Ainsi,

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t \xrightarrow{p} E(Y_t).$$

Ce processus sera ergodique en sa moyenne si l'autocovariance γ_k tend vers zéro suffisamment rapidement lorsque k devient grand. En particulier, un processus stationnaire du second ordre $\{Y_t\}$ est ergodique si il satisfait la condition suivante:

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty. \quad (1)$$

De la même manière, un processus stationnaire du second ordre sera ergodique si

$$\frac{1}{T-k} \sum_{t=k+1}^T (y_t - \mu)(y_{t-k} - \mu) \xrightarrow{p} \gamma_k$$

pour $\forall k$. Si le processus est gaussien, alors la condition 1 est suffisante pour avoir l'ergodicité de tous les moments. On ne peut cependant tester directement l'hypothèse d'ergodicité.

2.6 Définitions de bruits blancs

Nous allons ici définir différents types de bruit blancs.

Définition 3 $\{\epsilon_t, t \in Z\}$ est un bruit blanc faible, lorsque cette suite de variables aléatoires est de moyenne nulle, homoscédastique et non corrélées. Ainsi,

$$\begin{aligned} EL(\epsilon_t / \epsilon_{t-1}, \epsilon_{t-2}, \dots) &= 0 \\ E(\epsilon_t^2) &= \sigma^2. \end{aligned}$$

Ici, l'opérateur $EL(\cdot)$ signifie l'espérance conditionnelle linéaire. On va montrer que la condition $EL(\epsilon_t / \epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0$ implique que $E(\epsilon_{t-k}\epsilon_t) = 0$ pour tout $k > 0$. Par le loi des projections itérées (voir JH p. 100 et 742):

$$E(\epsilon_{t-k}\epsilon_t) = E[EL(\epsilon_{t-k}\epsilon_t / \epsilon_{t-1}, \epsilon_{t-2}, \dots)].$$

Puisque ϵ_{t-k} fait partie de l'ensemble d'information pour l'espérance conditionnelle, on peut donc faire passer l'opérateur espérance. Ainsi,

$$E[EL(\epsilon_{t-k}\epsilon_t / \epsilon_{t-1}, \epsilon_{t-2}, \dots)] = E\epsilon_{t-k}[EL(\epsilon_t / \epsilon_{t-1}, \epsilon_{t-2}, \dots)].$$

L'espérance conditionnelle linéaire étant égale à zéro par la Définition 3, on a alors directement que $E(\epsilon_{t-k}\epsilon_t) = 0$. Ceci est vrai pour tout $k > 0$.

Définition 4 $\{\epsilon_t, t \in Z\}$ est un bruit blanc conditionnel, lorsque:

$$EL(\epsilon_t/\epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0$$

$$V(\epsilon_t/\epsilon_{t-1}, \epsilon_{t-2}, \dots) = \Sigma$$

Définition 5 $\{\epsilon_t, t \in Z\}$ est un bruit blanc fort si elle est une suite de variables aléatoires indépendantes identiquement distribuées (i.i.d.) centrée à zéro dont les moments du second ordre existent.

Définition 6 $\{\epsilon_t, t \in Z\}$ est un bruit blanc gaussien si cette variable aléatoire suit une loi normale centrée à zéro.

Définition 7 $\{\epsilon_t, t \in Z\}$ est une différence de Martingale, lorsque:

$$E(\epsilon_t/\epsilon_{t-1}, \epsilon_{t-2}, \dots) = 0$$

2.7 Processus autorégressif (AR)

Pour simplifier les notation, on confondra par la suite l'écriture du processus et de sa réalisation.

Définition 8 On appelle processus autorégressif d'ordre p , un processus stationnaire y_t vérifiant une relation du type

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

où ε_t est un bruit blanc faible.

Exemple: processus AR(1) centré à zéro

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t$$

On va étudier la fonction d'autocovariance de ce processus AR(1)

$$\begin{aligned}\gamma_0 &= E(Y_t Y_t) \\ E(Y_t Y_t) &= \phi E(Y_t Y_{t-1}) + E(Y_t \varepsilon_t)\end{aligned}$$

On cherche

$$\begin{aligned}E(Y_t \varepsilon_t) &= \phi \underbrace{E(Y_{t-1} \varepsilon_t)}_0 + \underbrace{E(\varepsilon_t \varepsilon_t)}_{\sigma^2} \\E(Y_t Y_{t-1}) &= \phi E(Y_{t-1} Y_{t-1}) + E(\varepsilon_t Y_{t-1}) \\ \gamma_1 &= \phi \gamma_0\end{aligned}$$

On substitue

$$\gamma_0 = \phi \gamma_0 + \sigma_\varepsilon^2$$

et on obtient l'expression pour la variance

$$\begin{aligned}\gamma_0 &= \phi \gamma_0 + \sigma_\varepsilon^2 \\ \gamma_0 &= \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}\end{aligned}$$

On cherche maintenant à caractériser complètement la fonction d'autocovariance.

Ainsi, pour l'autocovariance d'ordre 1, on a

$$\gamma_1 = E(Y_t Y_{t-1}) = \phi E(Y_{t-1} Y_{t-1}) + \underbrace{E(\varepsilon_t Y_{t-1})}_{=0}$$

$$\phi \gamma(0) = \phi \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}.$$

Pour l'autocovariance d'ordre 2,

$$\gamma_2 = E(Y_t Y_{t-2}) = \phi E(Y_{t-1} Y_{t-2}) + \underbrace{E(\varepsilon_t Y_{t-2})}_{=0}$$

$$\gamma_2 = \phi \gamma(1) = \phi^2 \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}.$$

Et de façon générale:

$$\gamma_k = \phi^k \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}$$

Est-ce un processus stationnaire?

On va vérifier les 3 conditions:

1. On peut le montrer de façon récursive. On a alors que

$$Y_t = \phi^t(Y_0) + \sum_{i=1}^t \phi^{t-i} E(\varepsilon_i).$$

Pour $|\phi| < 1$ et t assez grand, on aura que $E(y_t) = 0$, puisque $E(\varepsilon_i) = 0 \forall i$.

2.

$$\text{var}(Y_t) = \frac{\sigma_\varepsilon^2}{(1-\phi^2)} \Rightarrow |\phi| < 1$$

3.

$$\gamma_k = \phi^k \frac{\sigma_\varepsilon^2}{(1-\phi^2)}$$

Les trois conditions sont vérifiées si $|\phi| < 1$.

Étude d'un processus AR(2)

On a donc le processus AR(2) suivant:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Si Y_t est stationnaire d'ordre deux, la moyenne non conditionnelle de Y_t sera

$$\begin{aligned} E(Y_t) &= \mu + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t) \\ E(Y_t) &= \mu + \phi_1 E(Y_t) + \phi_2 E(Y_t) + E(\varepsilon_t) \\ E(Y_t) &= \frac{\mu}{1 - \phi_1 - \phi_2} \end{aligned}$$

On cherche maintenant la variance du processus centré $\hat{Y}_t = Y_t - \frac{\mu}{1-\phi_1-\phi_2}$

$$\begin{aligned} E(\hat{Y}_t \hat{Y}_t) &= E(\hat{Y}_t \hat{Y}_{t-1} \phi_1) + E(\hat{Y}_t \hat{Y}_{t-1} \phi_2) + E(\varepsilon_t \hat{Y}_t) \\ \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2 \end{aligned}$$

On a

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma_\varepsilon^2, \quad (2)$$

$$\begin{aligned} \gamma_k &= E(\hat{Y}_{t-k} \hat{Y}_t) = \phi_1 E(\hat{Y}_{t-k} \hat{Y}_{t-1}) + \phi_2 E(\hat{Y}_{t-k} \hat{Y}_{t-2}) + E(\hat{Y}_{t-k} \varepsilon_t) \\ &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \end{aligned}$$

On divise par γ_0 , on obtient

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

puisque $\rho_1 = \rho_{-1}$ et $\rho_0 = 1$, on a pour $k = 1$

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ &= \frac{\phi_1}{(1 - \phi_2)} \end{aligned}$$

Pour $k = 2$,

$$\begin{aligned} \rho_2 &= \phi_1 \rho_1 + \phi_2 \\ &= \frac{\phi_1^2}{(1 - \phi_1)} + \phi_2. \end{aligned}$$

Maintenant, l'équation (2) peut être réécrit de la façon suivante:

$$\gamma_0 = \phi_1 \rho_1 \gamma_0 + \phi_2 \rho_2 \gamma_0 + \sigma_\varepsilon^2.$$

En substituant pour les valeurs obtenues de ρ_1 et ρ_2 , on obtient:

$$\gamma_0 = \left[\frac{\phi_1^2}{(1 - \phi_2)} + \frac{\phi_2 \phi_1^2}{(1 - \phi_2)} + \phi_2^2 \right] + \sigma_\varepsilon^2.$$

que l'on peut réécrire comme étant:

$$\gamma_0 = \frac{(1 - \phi_2) \sigma_\varepsilon^2}{(1 + \phi_2) [(1 - \phi_2)^2 - \phi_1^2]}.$$

Étudions maintenant la stationnarité du processus:

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \varepsilon_t$$

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = \varepsilon_t$$

$$(1 - \lambda_1 L)(1 - \lambda_2 L) Y_t = \varepsilon_t$$

De façon similaire au processus AR(1), le processus AR(2) sera stationnaire si $|\lambda_i| < 1$ pour $i = 1$ et 2 (voir Hamilton, chap. 1). On cherche la valeur de λ_1 et λ_2 . On développe l'expression

$$\begin{aligned} (1 - (\lambda_1 + \lambda_2)L + \lambda_1 \lambda_2 L^2) Y_t &= \varepsilon_t \\ \Rightarrow \lambda_1 + \lambda_2 &= \phi_1 \\ \lambda_1 \lambda_2 &= -\phi_2 \end{aligned}$$

On peut réécrire les relations plus haut en substituant L par z . Ainsi,

$$(1 - \phi_1 z - \phi_2 z^2) = (1 - \lambda_1 z)(1 - \lambda_2 z).$$

Pour que l'égalité entre le côté droit et le côté gauche soit respectée, les deux fonctions doivent donner la même valeur peu importe la valeur prise par z . En particulier, en posant soit $z = \lambda_1^{-1}$ ou $z = \lambda_2^{-1}$, le terme de droite est égal à zéro. Donc, si on trouve une valeur de z telle que le terme de droite est égal à zéro, cette valeur devrait donner l'égalité à zéro du terme de gauche. Les valeurs de z donnant un terme de gauche à zéro,

$$(1 - \phi_1 z - \phi_2 z^2) = 0$$

sont données par les formules suivantes:

$$z_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

et

$$z_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}.$$

En fixant $z = z_1$ ou $z = z_2$, le terme de gauche est égale à zéro et, en prenant $z = \lambda_1^{-1}$ ou $z = \lambda_2^{-1}$, le terme de droite est égal à zéro. Ainsi, on obtient les deux égalités suivantes:

$$\begin{aligned} z_1 &= \lambda_1^{-1} \\ z_2 &= \lambda_2^{-1} \end{aligned}$$

On peut obtenir directement les valeurs λ_1 et λ_2 . On cherche λ_1 et λ_2 comme solution de l'équation obtenue à l'aide des contraintes entre les coefficients ϕ et λ à partir de l'égalité:

$$(1 - \phi_1 L - \phi_2 L^2) = (1 - (\lambda_1 + \lambda_2)L + \lambda_1 \lambda_2 L^2).$$

Cette égalité nous donne les relations suivantes:

$$\lambda_1^2 - \lambda_1 \phi_1 - \phi_2 = 0$$

et

$$\lambda_2^2 - \lambda_2\phi_1 - \phi_2 = 0$$

Ces équations sont appelées équations caractéristiques. Les solutions en λ_1 et λ_2 de ces équations sont données par les deux expressions suivantes:

$$\lambda_1 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

et

$$\lambda_2 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}.$$

Les solutions λ_1 et λ_2 sont appelées racines caractéristiques des équations caractéristiques. On a donc la stationnarité d'ordre deux d'un processus AR(2) si les racines sont plus petites que 1. On dira alors que les racines sont à l'intérieur sur cercle unité. Pour l'écriture de l'équation en fonction de z , on aura alors la stationnarité du second ordre si les racines sont à l'extérieur du cercle unité.

On peut montrer que la condition de stationnarité implique que i) $\phi_1 + \phi_2 < 1$, ii) $\phi_2 - \phi_1 < 1$ et $|\phi_2| < 1$.

De façon générale, un processus AR(p):

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t.$$

On peut facilement montrer que la fonction d'autocovariances d'un processus AR(p) est donnée par:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$$

pour $k = 1, 2, \dots$ et

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_{\varepsilon_t}$$

pour $k = 0$. En divisant par γ_0 , on obtient la fonction d'autocorrélations appelée également équations de Yule-Walker. Ainsi,

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}$$

pour $k = 1, 2, \dots$.

Un processus AR(p) peut être écrit comme étant

$$(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L) Y_t = \varepsilon_t.$$

On peut obtenir les équations caractéristiques correspondant à ce processus AR(p). Ces équations sont données par:

$$\lambda_i^p - \phi_1 \lambda_i^{p-1} - \phi_2 \lambda_i^{p-2} - \cdots - \phi_{p-1} \lambda_i + \phi_p = 0.$$

Ces p équations caractéristiques donnent les p solutions λ_i . La condition de stationnarité du second ordre pour un processus AR(p) est la suivante: $|\lambda_i| < 1$ pour $i = 1, \dots, p$.

Dans le cas AR(p) générale, il est difficile d'obtenir les racines caractéristiques. Une condition suffisante pour que $|\lambda_i| < 1$ est que $\sum_{i=1}^p \phi_i < 1$. Donc pour assurer la stationnarité du second ordre, la somme des coefficients autorégressifs devra être inférieure à 1.

Si au moins une racine caractéristique est égale à 1 alors $\sum_{i=1}^p \phi_i = 1$. On dira que ce processus AR(p) possède au moins une racine unité (unit root).

2.7.1 Estimation d'un processus autorégressif

Supposons le processus AR(1) gaussien suivant:

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

et ε_t is *i.i.d.* $N(0, \sigma_\varepsilon^2)$. Le vecteur de paramètres à estimer est donc $\Theta = (c, \phi, \sigma_\varepsilon^2)'$.

Considérons la première observation y_1 . La moyenne et sa variance marginales (non conditionnelles) sont données par les expressions suivantes:

$$E(y_1) = \mu = \frac{c}{(1 - \phi)}$$

et

$$E(y_1 - \mu)^2 = \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}.$$

Puisque ε_t est gaussien, alors y_1 est gaussien. La densité de la première observation est alors donnée par:

$$f(y_1; \Theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_\varepsilon^2/(1-\phi^2)}} \exp\left[\frac{-\{y_1 - c/(1-\phi)\}^2}{2\sigma_\varepsilon^2/(1-\phi^2)}\right].$$

La distribution de la seconde observation conditionnellement à l'observation y_1 découle du processus AR(1). Ainsi,

$$y_2 = c + \phi y_1 + \varepsilon_2.$$

Conditionner sur $Y_1 = y_1$ revient à considérer que la variable aléatoire Y_1 est une constante égale à y_1 . La moyenne conditionnelle de y_2 est alors

$$E(y_2|y_1) = c + \phi y_1$$

et la variance conditionnelle de y_2 est

$$E((y_2 - (c + \phi y_1)|y_1)^2) = \sigma_\varepsilon^2.$$

La densité conditionnelle de y_2 par rapport à y_1 s'écrit:

$$f(y_2/y_1; \Theta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma_\varepsilon^2}\right].$$

La densité conjointe de y_2 et y_1 est alors égale à

$$f(y_2, y_1; \Theta) = f(y_2|y_1; \Theta) * f(y_1; \Theta).$$

De façon générale, la densité conditionnelle de y_t par rapport à y_{t-1} est:

$$f(y_t|y_{t-1}; \Theta) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma_\varepsilon^2}\right].$$

La densité conjoint des observations $t = 1, \dots, T$ est alors donnée par:

$$f(y_t, y_{T-1}, \dots, y_2, y_1; \Theta) = f(y_1; \Theta) * \prod_{t=2}^T f(y_t|y_{t-1}; \Theta).$$

La log-vraisemblance est obtenue en prenant le log de l'expression plus haut:

$$L(\Theta) = \log f(y_1; \Theta) + \sum_{t=2}^T \log f(y_t|y_{t-1}; \Theta).$$

L'estimateur du maximum de vraisemblance de Θ est obtenue en maximisant la log-vraisemblance par rapport à Θ . Cet estimateur sera donc solution des conditions du premier ordre. Le résultat sera obtenu en résolvant un système non linéaire. On doit donc faire appel à des méthodes d'optimisation numérique (voir Hamilton chap. 5).

Une alternative simple au maximum de vraisemblance exact consiste à considérer la valeur observée y_1 comme étant déterministe et de maximiser la vraisemblance conditionnelle par rapport à cette observation. La densité conjointe est alors:

$$f(y_T, y_{T-1}, \dots, y_2 | y_1; \Theta) = \prod_{t=2}^T f(y_t | y_{t-1}; \Theta).$$

L'estimateur de la log-vraisemblance conditionnelle est obtenue en maximisant le log de la densité conjointe par rapport à Θ . On peut facilement montrer que cet estimateur correspond à l'estimateur des moindres carrés ordinaires. Cet estimateur est donc très simple à obtenir comparativement à l'estimateur du maximum de vraisemblance exact. De plus, si la taille de l'échantillon est grande, la première observation joue un rôle négligeable.

Si le terme d'erreur ε n'est pas gaussien, l'estimateur du maximum de vraisemblance conditionnelle sera quand même convergent. Un estimateur convergent obtenu à partir d'une vraisemblance mal spécifiée est appelée estimateur du pseudo-maximum de vraisemblance. Les écart-types ne doivent pas être calculés de la même façon.

Examinons l'estimateur des moindres carrés ordinaires pour un processus AR(p). On a une réalisation d'un processus AR(p)

$$y_1, y_2, y_3, \dots, y_T$$

qu'on modélise comme étant:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

ou en notation matricielle,

$$Y = X\phi + \varepsilon$$

où

$$Y = (y_{p+1}, \dots, y_T), \quad \varepsilon = (\varepsilon_{p+1}, \dots, \varepsilon_T)$$

$$X = \begin{bmatrix} y_p & y_{p-1} & \cdots & y_1 \\ y_{p+1} & y_p & y_{p-1} & \vdots \\ \vdots & & \ddots & \vdots \\ y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{bmatrix}$$

et $\phi = (\phi_1, \phi_2, \dots, \phi_p)'$. On peut estimer ce processus par les moindres carrés linéaires (m.c.o.). L'estimateur des m.c.o. est défini comme étant:

$$\hat{\phi} = (X'X)^{-1}X'Y.$$

On peut montrer que cet estimateur est convergent et asymptotiquement normal pour un processus AR(p).

Cependant, en petit échantillon, $\hat{\theta}$ sera biaisé. On a que

$$\begin{aligned} \hat{\phi} &= (X'X)^{-1}X'Y \\ \hat{\phi} &= (X'X)^{-1}X'X\theta + X'\varepsilon \\ \hat{\phi} - \phi &= (X'X)^{-1}X'\varepsilon. \end{aligned}$$

En présence d'une variable endogène retardée (ce qui est le cas pour les processus autorégressif), la matrice $X'X$ est corrélée avec le vecteur $X'\varepsilon$. De façon asymptotique

$$\text{plim } \hat{\phi} = \text{plim } \phi + \underbrace{\text{plim } (X'X)^{-1} \text{plim } (X'\varepsilon)}_{=0}$$

théorème de Slutsky

où *plim* signifie la probabilité limite, l'estimateur est donc convergent.

Par le théorème central limite, on peut montrer que

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{L} N(0, \Sigma)$$

où

$$\Sigma = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & & \gamma_{p-2} \\ \vdots & \gamma_1 & \gamma_0 & & \vdots \\ \vdots & & & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \cdots & \gamma_0 \end{bmatrix}$$

et

$$\hat{\Sigma} = \hat{\sigma}_\varepsilon^2 \left(\frac{X'X}{T} \right)^{-1}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{(Y - X\hat{\theta})'(Y - X\hat{\theta})}{T - p}$$

Le processus AR(p) peut avoir également une moyenne qui n'est pas centrée à zéro.

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

$$E(y_t) = \mu = \frac{c}{1 - \phi_1 - \dots - \phi_p}$$

2.7.2 Autocorrélation partielle

Le coefficient d'autocorrélation partielle ϕ_{kk} est défini comme étant égal au coefficient de corrélation linéaire entre

$$y_t - E(y_t/y_{t-1}, \dots, y_{t-k+1})$$

et

$$y_{t-k} - E(y_{t-k}/y_{t-1}, \dots, y_{t-k+1}).$$

Le coefficient ϕ_{kk} mesure donc la relation entre y_t et y_{t-k} une fois enlevés les liens transitant par les variables intermédiaires $y_{t-1}, \dots, y_{t-k+1}$.

On obtient le coefficient ϕ_{kk} en effectuant la régression de y_t sur des valeurs passées jusqu'à y_{t-k} . Ainsi

$$y_t = \mu + \phi_{1k} y_{t-1} + \phi_{2k} y_{t-2} + \dots + \phi_{kk} y_{t-k} + \varepsilon_t$$

Pour un processus AR(p), on aura

$$\phi_{kk} \neq 0 \text{ si } k \leq p$$

$$= 0 \text{ si } k > p$$

2.8 Processus moyennes mobiles

Définition 9 On appelle processus moyenne mobile d'ordre q un processus y_t définie par

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où ε_t est un bruit blanc faible.

Exemple: Processus MA(1) centré à zéro,

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Dérivons maintenant la fonction d'autocovariance d'un processus MA(1).

$$\begin{aligned}\gamma_0 = E(y_t y_t) &= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_t + \theta_1 \varepsilon_{t-1})] \\ &= (1 + \theta_1^2) \sigma_\varepsilon^2 \\ \gamma_1 = E(y_t y_{t-1}) &= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2})] \\ &= \theta_1 \sigma_\varepsilon^2 \\ \gamma_2 = E(y_t y_{t-2}) &= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-2} + \theta_1 \varepsilon_{t-3})] \\ &= 0 \\ \gamma_k = E(y_t y_{t-k}) &= 0 \quad k \geq 2.\end{aligned}$$

La dépendance temporelle du processus est donc limitée à l'ordre de ce processus.

Est-ce un processus stationnaire?

On vérifie les 3 conditions:

1. $E(y_t) = E(\varepsilon_t + \theta_1 \varepsilon_{t-1}) = 0 \quad \forall t$
2. $var(y_t) = (1 + \theta_1^2) \sigma_\varepsilon^2 < \infty \quad \forall t$
3. $cov(y_t y_{t-1}) = \theta_1 \sigma_\varepsilon^2$ dépend de k seulement

Donc, peu importe la valeur de θ_1 , ce processus est stationnaire.

Comparaison d'un processus AR avec un processus MA

Prenons un processus AR(1)

$$\begin{aligned}y_t &= \phi y_{t-1} + \varepsilon_t \\ (1 - \phi L)y_t &= \varepsilon_t\end{aligned}$$

Si ce processus est stationnaire du second ordre ($|\phi| < 1$), alors on peut le récrire sous forme d'une moyenne mobile

$$y_t = \sum_{i=0}^{\infty} \phi^i L^i \varepsilon_t = \sum_{i=0}^{\infty} \hat{\phi}^i \varepsilon_{t-i}$$

Implication: Un processus AR stationnaire du second ordre a une représentation moyenne mobile infinie. La dépendance temporelle est plus longue que pour un processus MA(1). De la même façon, on peut inverser le processus MA pour obtenir une représentation autorégressive. Dans le cas où le paramètre MA(1) est plus petit que 1 en valeur absolue, alors le bruit blanc faible à la période T peut être réécrit comme une fonction infinie des retards de y_t . Ainsi,

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \\ \varepsilon_t &= \sum_{i=0}^{\infty} (-\theta_1)^i y_{t-i}. \end{aligned}$$

2.8.1 Détermination de l'ordre q du processus MA

Pour un processus MA(q), la fonction d'autocovariance est donnée

$$\gamma_k = \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k} & \text{pour } k = 0, 1, \dots, q \\ 0 & \text{pour } k > q \end{cases}$$

et la fonction d'autocorrélation par

$$\rho_k = \begin{cases} \frac{\left(\sum_{i=0}^{q-k} \theta_i \theta_{i+k} \right)}{\left(\sum_{i=0}^q \theta_i^2 \right)} & \text{pour } k = 0, 1, \dots, q \\ 0 & \text{pour } k > q \end{cases}$$

En examinant la fonction d'autocovariance, on déduit qu'un processus moyennes mobiles d'ordre fini est un processus stationnaire du second ordre. Sa variance est finie et les autocovariances ne dépendent pas de t mais de k .

On peut estimer ρ_k de la façon suivante:

$$\hat{\rho} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

où

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})$$

et

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

et $\hat{\rho}_k \rightarrow N(0, \frac{1}{T})$.

2.8.2 Estimation d'un processus MA(q)

On a vu que l'estimation d'un processus AR(p) était beaucoup plus simple si on considérait une vraisemblance conditionnelle aux valeurs initiales. Le maximum de vraisemblance conditionnelle correspond alors aux moindres carrés ordinaires. De la même façon, l'estimation de processus moyennes mobiles est plus simple en considérant une vraisemblance conditionnelle aux valeurs initiales ε .

Examinons dans un premier temps un processus MA(1)

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

et ε_t est *i.i.d.* $N(0, \sigma_\varepsilon^2)$. Le vecteur de paramètres à estimer est $\Theta = (\mu, \theta, \sigma_\varepsilon^2)'$. Si la valeur de ε_{t-1} est connu avec certitude alors

$$Y_t | \varepsilon_{t-1} \sim N\left((\mu + \theta\varepsilon_{t-1}), \sigma_\varepsilon^2\right)$$

ce qui donne la densité conditionnelle

$$f(y_t | \varepsilon_{t-1}; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{(y_t - \mu - \theta\varepsilon_{t-1})^2}{2\sigma_\varepsilon^2}\right].$$

On va supposer que $\varepsilon_0 = 0$ est connu. Alors,

$$Y_1 | \varepsilon_0 = 0 \sim N\left(\mu, \sigma_\varepsilon^2\right)$$

Étant donné l'observation y_1 , la valeur de ε_1 est alors connue et est égale à

$$\varepsilon_1 = y_1 - \mu.$$

Connaissant ε_1 , on peut alors écrire la densité conditionnelle de y_2 ,

$$f(y_2|y_1, \varepsilon_0; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left[\frac{-(y_2 - \mu - \theta\varepsilon_1)^2}{2\sigma_\varepsilon^2} \right].$$

Puisque que ε_1 est connu, on peut calculer ε_2 comme étant:

$$\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1.$$

En poursuivant cette procédure, pour une valeur donnée $\varepsilon_0 = 0$, la suite $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}$ peut être obtenue en itérant sur l'expression suivante:

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

pour $t = 1, 2, \dots, T$ et $\varepsilon_0 = 0$.

On peut alors écrire la densité conditionnelle de l'observation t comme étant:

$$f(y_t|\underline{y}_{t-1}, \varepsilon_0 = 0; \Theta) = f(y_t|\varepsilon_{t-1}; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp \left[\frac{-\varepsilon_t^2}{2\sigma_\varepsilon^2} \right],$$

où $\underline{y}_{t-1} = y_{t-1}, y_{t-2}, \dots, y_1$.

La vraisemblance est donnée par la densité conjointe des observations que sera égale au produit des densités conditionnelles:

$$f(y_T, y_{T-1}, \dots, y_1|\varepsilon_0 = 0; \Theta) = f(y_1|\varepsilon_0 = 0; \Theta) \prod_{t=2}^T f(y_t|\underline{y}_{t-1}, \varepsilon_0 = 0; \Theta).$$

La log-vraisemblance conditionnelle est alors donnée par:

$$L(\Theta) = \log f(y_T, y_{T-1}, \dots, y_1|\varepsilon_0 = 0; \Theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_\varepsilon^2) - \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}.$$

Même si les itérations effectuées pour obtenir les termes ε_t sont simples, l'estimateur obtenue est une fonction non linéaire compliquée. On doit donc avoir recours à des routines d'optimisation numérique. Ceci est d'autant plus vrai pour un processus MA(q). On peut remarquer que les itérations pour obtenir ε_t à partir de la valeur initiale ε_0 nous donneront l'expression suivante pour ε_t

$$\varepsilon_t = (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) - \dots (-1)^{t-1} \theta^{t-1} (y_1 - \mu) + (-1)^t \theta^t \varepsilon_0.$$

Si $|\theta|$ est substantiellement plus petit que l'unité, l'effet de la valeur initiale ε_0 devient rapidement négligeable dans le calcul de la log-vraisemblance conditionnelle.

Cette log-vraisemblance conditionnelle sera alors une bonne approximation en grand échantillon de la log-vraisemblance exacte. Si $|\theta| > 1$, ce n'est plus le cas. On peut obtenir un estimateur à partir de la vraisemblance exacte par le filtre de Kalman.

Il est également important de noter que l'estimateur obtenue à l'aide de la log-vraisemblance conditionnelles correspond à l'estimateur des moindres carrés non linéaires. Cet estimateur est convergent même si le terme d'erreur n'est pas gaussien.

2.9 Théorème de Wold

On peut maintenant énoncer un théorème important en série temporelles: le Théorème de Wold.

Théorème 1 *Tout processus $\{Y_t\}$ centré à zéro et stationnaire d'ordre deux peut toujours être décomposé en une composante déterministe (éventuellement fonction du temps t) et une composante moyennne mobile infinie tel que*

$$Y_t = \delta(t) + \Psi(L)\varepsilon_t$$

où $\Psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$, $\psi_0 = 1$ et ε_t est un bruit blanc faible centré à zéro et ayant comme variance σ_ε^2 .

Ce théorème est valide pour tous processus stationnaires linéaires ou non linéaires.

2.10 Fonction génératrice d'autocovariances

Nous avons vu que pour un processus stationnaire d'ordre deux, on peut calculer sans ambiguïté la suite des autocovariances $\{\gamma_k\}_{k=-\infty}^{\infty}$. Cette information peut être résumée à l'aide de la fonction génératrice d'autocovariance définie comme étant:

$$g_y(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k,$$

où z est un scalaire complexe sans signification intrinsèque.

Considérons le calcul de la fonction génératrice d'autocovariances d'un processus MA(1). Pour ce processus, $\gamma_0 = (1 + \theta^2)\sigma_\varepsilon^2$ et $\gamma_1 = \gamma_{-1} = \theta\sigma_\varepsilon^2$. La fonction génératrice est alors donnée par:

$$\begin{aligned} g_y(z) &= [\theta\sigma_\varepsilon^2]z^{-1} + (1 + \theta^2)\sigma_\varepsilon^2 z^0 + [\theta\sigma_\varepsilon^2]z^1 \\ &\quad \sigma_\varepsilon^2[\theta z^{-1} + (1 + \theta^2) + \theta z]. \end{aligned}$$

Cette expression peut être réécrite de la façon suivante:

$$g_y(z) = \sigma_\varepsilon^2(1 + \theta z)(1 + \theta z^{-1}).$$

Pour un processus MA(q)

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

ceci suggère que la fonction génératrice d'autocovariances peut être calculée à partir de l'expression suivante:

$$g_y(z) = \sigma_\varepsilon^2(1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q)(1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \dots + \theta_q z^{-q}).$$

On peut étendre la fonction $g_y(z)$ à un processus MA(∞). Considérons que

$$Y_t = \mu + \psi(L) \varepsilon_t$$

avec

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$$

et $\sum_{k=0}^{\infty} |\psi_k| < \infty$ alors la fonction génératrice d'autocovariances s'écrit de façon compacte:

$$g_y(z) = \sigma_\varepsilon^2 \psi(z) \psi(-z).$$

Par exemple, un processus AR(1) peut être écrit:

$$Y_t - \mu = (1 - \phi L)^{-1} \varepsilon_t,$$

et $\psi(L) = 1/(1 - \phi L)$. La fonction génératrice d'autocovariances de ce processus AR(1) est alors:

$$g_y(z) = \frac{\sigma_\varepsilon^2}{(1 - \phi z)(1 - \phi z^{-1})}.$$

De la même façon, pour un processus AR(p), la fonction génératrice d'autocovariances est:

$$g_y(z) = \frac{\sigma_\varepsilon^2}{(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p)(1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \dots - \phi_p z^{-p})}.$$

2.11 Modèles ARMA (p,q) et ARIMA (p,d,q)

2.11.1 Modèle ARMA (p,q)

Définition 10 *Un processus stationnaire y_t admet une représentation ARMA(p,q) minimale s'il satisfait*

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

où ε_t est un bruit blanc faible.

Exemple: un processus ARMA(1,1) centré à zéro

$$y_t - \phi_1 y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

On peut montrer que la fonction d'autocovariance est donnée par:

$$\begin{aligned} \text{var}(y_t) &= \gamma_0 = \frac{(1 + \theta_1^2 + 2\phi_1\theta_1)}{(1 - \phi_1^2)} \sigma_\varepsilon^2 \\ \text{cov}(y_t y_{t-1}) &= \phi_1 \gamma_0 + \theta_1 \sigma_\varepsilon^2 \\ \text{cov}(y_t y_{t-k}) &= \phi_1 \gamma_{k-1} \quad k \geq 2. \end{aligned}$$

En divisant ces expressions par la variance γ_0 , on obtient la fonction d'autocorrélation ρ_k . La fonction d'autocorrélation et la fonction d'autocorrélation partielle d'un processus ARMA (1,1) décroît de façon géométrique. De plus, on voit bien à partir de l'expression de la variance que ce processus est stationnaire d'ordre deux si $|\phi| < 1$.

2.12 Modèle ARIMA (p,d,q)

Définition 11 *Un processus stationnaire y_t admet une représentation ARIMA(p,q) minimale s'il satisfait*

$$\phi_p(L)(1-L)^d y_t = \theta_q(L)\varepsilon_t$$

où ε_t est un bruit blanc et d est l'ordre de différenciation.

Exemple: un processus ARIMA(1,1,1) centré à zéro.

$$\Delta y_t - \phi_1 \Delta y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

où $\Delta = (1 - L)$.

Les fonctions d'autocovariance et d'autocorrélation partielles de la variable en différence (Δy_t) sont les mêmes que pour un processus ARMA(1,1).

3 Approche de Box-Jenkins

L'approche de Box-Jenkins est une procédure en trois étapes pour obtenir la représentation ARMA pour série observée.

- **Identification:** On identifie à l'aide des fonctions d'autocorrélation et d'autocorrélation partielle un processus ARMA plausible. On différencie si nécessaire. Les guides pour le choix de p et de q sont essentiellement les formes des estimations de fonctions d'autocorrélation $\hat{\rho}_k$ et d'autocorrélation partielle \hat{r}_k de la série stationnaire.

Pour juger si les $\hat{\rho}_k$ et les \hat{r}_k sont significativement différents de zéro, il faut calculer leur écart-type. On peut démontrer que pour un MA(q), on a pour tout $k > q$:

$$V(\rho_K) \approx \frac{1}{T} \left[1 + 2 \sum_{i=1}^q \rho^2(i) \right].$$

On peut donc prendre comme estimation de l'écart-type:

$$\hat{\sigma} [\hat{\rho}_k] = \frac{1}{\sqrt{T}} \left[1 + 2 \sum_{i=1}^q \hat{\rho}^2(i) \right]^{1/2}.$$

Pour identifier l'ordre q d'un processus moyenne mobile, on peut représenter la suite des autocorrélations estimées $\hat{\rho}_k$ et regarder à partir de quelle valeur de k tous les $\hat{\rho}_k$ restent dans l'intervalle

$$\pm \frac{1.96}{\sqrt{T}} \left(1 + 2[\hat{\rho}_1^2 + \hat{\rho}_2^2 + \dots + \hat{\rho}_{k-1}^2] \right)^{1/2}.$$

De même, en utilisant un résultat de Quenouille (1949), on peut prendre $\frac{1}{\sqrt{T}}$ comme écart-type approximatif de \hat{r}_k pour $k > p$, si le processus est un AR(p). Ainsi, pour identifier l'ordre p d'un processus autorégressif, on peut représenter

la suite des autocorrélations partielles estimées \hat{r}_K et examiner à partir de quelle valeur elle reste dans la bande

$$\left[-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right].$$

Il existe également des estimateurs de la variance qui dépendent de la série observée. Nous verrons ces estimateurs plus tard dans le cours.

- **Estimation:** une fois la représentation identifiée, on estime le modèle. Pour un processus AR(p), on estime par les moindres carrés ordinaires. Pour un processus MA ou ARMA, on procède par maximum de vraisemblance ou moindres carrés non linéaires.
- **Validation:** Essentiellement, on vérifie si les résidus sont caractérisés par un bruit blanc. Est-ce que les résidus possèdent encore une certaine dépendance temporelle? Pour ce faire, on examine les fonctions d'autocorrélation et d'autocorrélation partielle des résidus. On effectue également un test appelé test "portmanteau".

Test " Portmanteau"

Ce test a été proposé par Box-Pierce (1970). Son nom, qui en anglais signifie "fourre-tout", fait référence à son aspect global. L'hypothèse nulle considérée est:

$$\begin{aligned} H_0 &: \rho_1 = \rho_2 = \dots = \rho_k = 0 \\ H_1 &: \exists \rho_k \neq 0 \end{aligned}$$

La statistique du test (Q-stat) est définie comme étant:

$$Q^K = T \sum_{k=1}^K \hat{\rho}_k^2 \sim \chi^2(K - p - q)$$

On peut montrer que sous l'hypothèse d'indépendance des ε_t , la statistique Q^K suit asymptotiquement une loi du χ^2 à $K - p - q$ degrés de liberté. Les propriétés à distance finie de Q^K sont assez différentes des propriétés asymptotique, et ce, même pour T relativement grand. Une modification de la statistique a été proposée et elle est donnée par:

$$Q^K = T(T + 2) \sum_{k=1}^K \frac{1}{T - k} \hat{\rho}_k^2 \sim \chi^2(K - p - q).$$

Le nombre K doit être choisi relativement élevé, on le prend habituellement dans la zone de 15-30.

Si les résidus sont toujours autocorrélés, alors on recommence à partir de la première étape jusqu'à obtenir des résidus qui correspondent à un bruit blanc.

3.1 Approches alternatives

Une approche introduite par Akaike (1969) consiste à supposer que les modèles ARMA(p,q) fournissent des approximations de la réalité et que la vraie loi inconnue des observations ne satisfait pas forcément un tel modèle. On peut alors fonder le choix du modèle sur une mesure de proximité entre la vraie loi inconnue et le modèle proposé. La mesure habituellement retenue est la quantité d'information de Kullback. Soit $f_0(Y)$ la densité inconnue des observations, $\{f(Y), f \in F_{p,q}\}$ la famille de densité correspondant au modèle ARMA(p,q), l'écart entre la vraie loi et le modèle est mesurée par:

$$I(f_0, F_{p,q}) = \min_{f \in F_{p,q}} \int \log \frac{f_0(Y)}{f(Y)} f_0(Y) dY.$$

Cette quantité est toujours positive ou ne s'annule que si la vraie loi f_0 appartient à la famille de loi $F_{p,q}$. La valeur de la quantité d'information $I(f_0, F_{p,q})$ est évidemment inconnue, mais si on dispose d'un estimateur $\hat{I}(f_0, F_{p,q})$ avec de bonnes propriétés, il sera possible de l'utiliser comme critère. On retiendra alors le modèle $F_{p,q}$ conduisant à la plus petite valeur de l'estimation $\hat{I}(f_0; F_{p,q})$.

Les estimateurs de la quantité d'information qui ont été proposés sont:

- Akaike (AIC): $\ln \hat{\sigma}^2 + 2 \frac{(p+q)}{T}$,
- Schwartz bayésien criterion (BIC): $\ln \hat{\sigma}^2 + 2(p+q) \frac{\ln T}{T}$,
- HQ: $\ln(\hat{\sigma}^2) + (p+q)c \frac{\ln \ln T}{T}$ où $c > 2$.

Le premier de ces critères introduit par Akaike (1969) est de loin le plus utilisé. Le critère BIC a été proposé en particulier dans Akaike (1977) ou Schwarz (1978) et la troisième méthode a été introduite par Hannan-Quinn (1979). Cependant, les estimations de p et de q déduites des deux derniers critères sont convergentes et

conduisent à une sélection asymptotiquement correcte du modèle. Le premier critère choisira de façon correcte le modèle ou un modèle qui englobe le vrai modèle.

3.2 Prédiction

3.2.1 Principes généraux de prédiction

On va supposer que l'on s'intéresse à la prédiction de la valeur d'une variable aléatoires Y_{t+1} à partir d'un ensemble d'information contenant les variables X_t observées jusqu'à la date t . Par exemple, l'ensemble d'observations X_t peut contenir les retards de la variable Y_t . On va noter la prédiction de Y_{t+1} par rapport à l'ensemble d'information X_t comme étant $Y_{t+1|t}$.

Pour évaluer la qualité d'une prédiction, nous avons besoin d'une métrique, en d'autres mots, d'une fonction de perte. Nous allons utiliser une fonction quadratique. On cherche donc la prédiction qui minimise

$$E[Y_{t+1} - Y_{t+1|t}]^2.$$

Cette expression correspond à l'écart quadratique moyen ("mean squared errors" en anglais), que nous noterons EQM.

On va montrer que la prédiction avec l'écart quadratique moyen minimal est donnée par l'espérance conditionnelle de Y_{t+1} par rapport à l'ensemble d'information X_t :

$$Y_{t+1|t} = E[Y_{t+1}|X_t].$$

Preuve:

Pour vérifier que c'est effectivement le cas, nous allons considérer toute fonction $g(X_t)$ autre que l'espérance conditionnelle. Pour cette fonction quelconque, l'EQM est:

$$\begin{aligned} E[Y_{t+1} - g(X_t)]^2 &= E[Y_{t+1} - E(Y_{t+1}|X_t) + E(Y_{t+1}|X_t) - g(X_t)]^2 \\ &= E[Y_{t+1} - E(Y_{t+1}|X_t)]^2 + 2E([Y_{t+1} - E(Y_{t+1}|X_t)][E(Y_{t+1}|X_t) - g(X_t)]) \\ &\quad + E[E(Y_{t+1}|X_t) - g(X_t)]^2. \end{aligned}$$

Posons pour le terme du centre:

$$E\eta_{t+1} = E([Y_{t+1} - E(Y_{t+1}|X_t)][E(Y_{t+1}|X_t) - g(X_t)]).$$

Considérons l'espérance conditionnelle de η_{t+1} par rapport à l'ensemble d'information X_t . Par rapport à cette ensemble d'information, les termes $E(Y_{t+1}|X_t)$ et $g(X_t)$ sont connus, on peut donc les sortir de l'opérateur et écrire:

$$\begin{aligned} E[\eta_{t+1}|X_t] &= [E(Y_{t+1}|X_t) - g(X_t)]E([Y_{t+1} - E(Y_{t+1}|X_t)]|X_t) \\ &= [E(Y_{t+1}|X_t) - g(X_t)] \times 0 = 0 \end{aligned}$$

Par une application directe de la loi des projections itérées, on obtient donc

$$E[\eta_{t+1}] = E_{X_t}(E[\eta_{t+1}|X_t]) = 0.$$

L'EQM est donc égal à

$$E[Y_{t+1} - g(X_t)]^2 = E[Y_{t+1} - E(Y_{t+1}|X_t)]^2 + E[E(Y_{t+1}|X_t) - g(X_t)]^2.$$

Le deuxième terme à droite est nécessairement supérieur ou égal à zéro impliquant ainsi que l'EQM de l'espérance conditionnelle, donné par le premier terme à droite, est nécessairement plus petit ou égal à l'EQM de $g(X_t)$. La prévision basée sur une fonction de l'ensemble d'information X_t qui minimise l'EQM est donc l'espérance conditionnelle $E(Y_{t+1}|X_t)$.

On va maintenant se restreindre aux prévisions linéaires. Ainsi,

$$E[Y_{t+1}|X_t] = \alpha'X_t.$$

Afin d'exploiter l'ensemble d'information X_t , on cherchera une valeur du vecteur α telle que l'erreur de prévision ne soit pas corrélée avec l'ensemble d'information X_t ,

$$E[(Y_{t+1} - \alpha'X_t)X_t'] = 0. \quad (3)$$

Si les conditions d'orthogonalité (3) sont respectées, alors $\alpha'X_t$ est appelé la projection linéaire de Y_{t+1} sur X_t . On va montrer que cette projection linéaire donne l'écart quadratique minimum parmi les fonctions de prévisions linéaires.

Preuve

Prenons une fonction linéaire arbitraire $g'X_t$. L'écart quadratique moyen pour cette fonction arbitraire est donné par:

$$\begin{aligned} E[Y_{t+1} - g'X_t]^2 &= E[Y_{t+1} - \alpha'X_t + \alpha'X_t - g'X_t]^2 \\ &= E[Y_{t+1} - \alpha'X_t]^2 + 2E([Y_{t+1} - \alpha'X_t][\alpha'X_t - g'X_t]) \\ &\quad + E[\alpha'X_t - g'X_t]^2. \end{aligned}$$

Examinons le terme du centre:

$$E([Y_{t+1} - \alpha'X_t][\alpha'X_t - g'X_t]) = (E[Y_{t+1} - \alpha'X_t]X_t') [\alpha - g] = 0'[\alpha - g],$$

par les conditions d'orthogonalité (3). L'écart quadratique moyen est donc égal à:

$$E[Y_{t+1} - g'X_t]^2 = E[Y_{t+1} - \alpha'X_t]^2 + E[\alpha'X_t - g'X_t]^2.$$

On voit bien que pour tout vecteur g différent de α , l'écart quadratique moyen est supérieur à celui obtenu avec le vecteur α . La projection linéaire $\alpha'X_t$ est donc la fonction linéaire que minimise l'écart quadratique moyen. Il est important de noter que l'EQM de la projection linéaire est toujours supérieur ou égal à l'EQM de l'espérance conditionnelle.

Les représentations ARMA sont souvent utilisées pour effectuer de la prévision basée sur des fonctions linéaires. L'objectif est donc de minimiser l'écart quadratique moyen de la prévision à partir du processus ARMA retenu. Pour un horizon de prévision h , on minimise donc l'expression suivante:

$$E \left[(y_{t+h} - \hat{y}_{t+h})^2 \right]$$

où E_t est l'espérance mathématique au temps t , donc incluant l'information jusqu'en t , et \hat{y}_{t+h} est la prévision linéaire de y_{t+h} conditionnelle à cet ensemble d'information. La prévision optimale selon ce critère sera

$$\hat{y}_{t+h} = E_t [y_{t+h}]$$

Prenons, par exemple, un processus AR(1), on a donc

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t.$$

Pour un horizon $h = 1$, la prévision sera

$$\begin{aligned} E_t y_{t+1} &= E_t [\mu + \phi y_t + \varepsilon_{t+1}] \\ &= \mu + \phi y_t \\ \text{puisque } E_t [\varepsilon_{t+1}] &= 0. \end{aligned}$$

L'erreur quadratique moyenne sera

$$\begin{aligned} E [y_{t+1} - E_t y_{t+1}]^2 &= E_t [y_{t+1} - \mu - \phi y_t]^2 \\ E [\varepsilon_{t+1}]^2 &= \sigma_\varepsilon^2. \end{aligned}$$

Pour un horizon $h = 2$, on aura

$$\begin{aligned} E_t y_{t+2} &= E_t [\mu + \phi y_{t+1} + \varepsilon_{t+2}] \\ &= \mu + \phi E_t (y_{t+1}) + E_t (\varepsilon_{t+2}) \\ &= \mu + \phi [\mu + \phi y_t] \\ &= \mu + \phi \mu + \phi^2 y_t. \end{aligned}$$

L'erreur quadratique moyenne sera:

$$E [y_{t+2} - E_t y_{t+2}]^2 = E [y_{t+2} - \mu - \phi \mu - \phi^2 y_t]^2$$

alors

$$\begin{aligned} E [y_{t+2} - E_t y_{t+2}]^2 &= E [\mu + \phi y_{t+1} + \varepsilon_{t+2} - \mu - \phi \mu - \phi^2 y_t]^2 \\ &= E [\phi (y_{t+1} - \mu - \phi y_t) + \varepsilon_{t+2}]^2 \\ &= E [\phi \varepsilon_{t+1} + \varepsilon_{t+2}]^2 = [1 + \phi^2] \sigma_\varepsilon^2. \end{aligned}$$

De façon générale, la prévision optimale pour un horizon h sera

$$\begin{aligned} E_t y_{t+h} &= E [\mu + \phi y_{t+h-1} + \varepsilon_{t+h}] \\ &= E (1 + \phi + \dots + \phi^{h-1}) \mu + \phi^h y_t \end{aligned}$$

et l'erreur quadratique moyenne sera

$$E [\phi^{h-1} \varepsilon_{t+1} + \phi^{h-2} \varepsilon_{t+2} + \dots + \varepsilon_{t+h}]^2 = [1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)}] \sigma_\varepsilon^2.$$

Si $h \rightarrow \infty$, alors

$$E_t y_{t+h} = \frac{\mu}{(1-\phi)}$$

qui correspond à la moyenne non conditionnelle, et

$$E(y_{t+h} - E_t y_{t+h})^2 = \frac{\sigma^2}{(1-\phi^2)}$$

qui correspond à la variance non conditionnelle de y_t .

Prenons maintenant un processus MA(1). On a donc

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

Pour un horizon $h = 1$, la prévision sera

$$\begin{aligned} E_t y_{t+1} &= E_t [\mu + \varepsilon_{t+1} + \theta\varepsilon_t] \\ &= \mu + \theta\varepsilon_t \text{ puisque } E_t \varepsilon_{t+1} = 0 \end{aligned}$$

L'erreur quadratique moyenne est donnée par

$$\begin{aligned} E[y_{t+1} - E_t y_{t+1}]^2 &= E[\mu + \varepsilon_{t+1} + \theta\varepsilon_t - \mu - \theta\varepsilon_t]^2 \\ &= E(\varepsilon_{t+1})^2 = \sigma_\varepsilon^2 \end{aligned}$$

Pour $h = 2$, la prévision optimale est

$$E_t y_{t+2} = E_t [\mu + \varepsilon_{t+2} + \theta\varepsilon_{t+1}] = \mu$$

qui correspond à la moyenne non conditionnelle et l'erreur quadratique moyenne est

$$\begin{aligned} E[y_{t+2} - E_t y_{t+2}]^2 &= E[\mu + \varepsilon_{t+2} + \theta\varepsilon_{t+1} - \mu]^2 \\ &= [\varepsilon_{t+2} + \theta\varepsilon_{t+1}]^2 = [1 + \theta^2] \sigma_\varepsilon^2 \end{aligned}$$

Pour $h > 2$,

$$E y_{t+h} = \mu \text{ et } E[y_{t+h} - E y_{t+h}]^2 = [1 + \theta^2] \sigma_\varepsilon^2.$$

Examinons maintenant un processus ARMA(1,1). On a donc

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

Pour $h = 1$, la prévision optimale est

$$\begin{aligned} E(y_{t+1}) &= E_t [\mu + \phi y_t + \varepsilon_{t+1} + \theta \varepsilon_t] \\ &= \mu + \phi y_t + \theta \varepsilon_t \text{ puisque} \\ &E_t \varepsilon_{t+1} \end{aligned}$$

et l'erreur quadratique moyenne est

$$\begin{aligned} E [y_{t+1} - E_t y_{t+1}]^2 &= E [y_{t+1} - \mu - \phi y_t - \theta \varepsilon_t]^2 \\ &= E [\mu + \phi y_t + \varepsilon_{t+1} + \theta \varepsilon_t - \phi y_t - \theta \varepsilon_t]^2 \\ &= E [\varepsilon_{t+1}]^2 = \sigma_\varepsilon^2 \end{aligned}$$

Pour un horizon $h = 2$, la prévision optimale est

$$\begin{aligned} E_t(y_{t+2}) &= E_t [\mu + \phi y_{t+1} + \varepsilon_{t+2} + \theta \varepsilon_{t+1}] \\ &= \mu + \phi E_t y_{t+1} = \mu + \phi [\mu + \phi y_t + \theta \varepsilon_t] \\ &= [\mu + \phi \mu + \phi^2 y_t + \phi \theta \varepsilon_t] \end{aligned}$$

L'erreur quadratique moyenne est

$$\begin{aligned} E [y_{t+2} - E_t y_{t+2}]^2 &= E [\mu + \phi y_{t+1} + \varepsilon_{t+2} + \theta \varepsilon_{t+1} - \mu - \phi \mu - \phi^2 y_t - \phi \theta \varepsilon_t]^2 \\ &= E [\mu + \phi [y_{t+1} - \mu - \phi y_t - \theta \varepsilon_t] + \varepsilon_{t+2} + \theta \varepsilon_{t+1} - \mu]^2 \\ &= E [\phi \varepsilon_{t+1} + \varepsilon_{t+2} + \theta \varepsilon_{t+1}]^2 \\ &= E [1 + (\phi + \theta)^2] \sigma_\varepsilon^2 \end{aligned}$$

Pour $h = 3$, la prévision optimale est

$$\begin{aligned} E_t(y_{t+3}) &= E_t [\mu + \phi y_{t+2} + \varepsilon_{t+3} + \theta \varepsilon_{t+2}] \\ &= \mu + E_t \phi y_{t+2} = \mu + \phi \mu + \phi^2 \mu + \phi^3 y_t + \phi^2 \theta \varepsilon_t \end{aligned}$$

et

$$\begin{aligned} E [y_{t+3} - E_t y_{t+3}]^2 &= E [\mu + \phi y_{t+2} + \varepsilon_{t+3} - \mu - \phi \mu - \phi^2 \mu - \phi^3 y_t - \phi^2 \theta \varepsilon_t]^2 \\ &= E [\mu + \phi [y_{t+2} - E_t y_{t+2}] + \varepsilon_{t+3} + \theta \varepsilon_{t+2} - \mu]^2 \\ &= E [\phi [\phi \varepsilon_{t+1} + \varepsilon_{t+2} + \theta \varepsilon_{t+1}] + \varepsilon_{t+3} + \theta \varepsilon_{t+2}]^2 \\ &= E [1 + (\phi + \theta)^2 + (\phi^2 + \phi \theta)^2] \sigma_\varepsilon^2 \end{aligned}$$

On construit l'intervalle de confiance de la prévision à un horizon h comme étant:

$$\hat{y}_{t+h} \pm 1.96 \times (\text{erreur quadratique moyenne})^{1/2}$$

pour un intervalle à 95%.

4 Processus univariés non stationnaires

La plupart des séries macroéconomiques et financières ne sont pas stationnaires. La moyenne et la variance ne sont pas indépendantes de t .

4.1 Processus particuliers

Marche aléatoire

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

Ce processus correspond à un processus AR(1) avec un coefficient égal à 1.

On ne peut prédire les variations de y_t , c.à.d. $\Delta y_t = y_t - y_{t-1}$.

On va étudier ce processus. Réécrivons ce processus:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t \\ &= y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\vdots \\ y_t &= y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \end{aligned}$$

Les chocs ne s'estompent pas dans le temps. On dit que les chocs sont permanents.

Pour un processus AR(1) stationnaire ($|\phi| < 1$), on a que

$$y_t = \phi^t y_0 + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}.$$

Contrairement à la marche aléatoire, l'effet des chocs s'estompent dans le temps puisque ϕ^i tend vers zéro, pour $|\phi| < 1$, lorsque i tend vers l'infini. On dira que les chocs ont un effet transitoire (mean reverting).

Calculons les moments d'ordre 1 et 2 de la marche aléatoire:

$$\begin{aligned} E(y_t) &= y_0 \quad \forall t \\ \text{var}(y_t) &= \text{var} \left[y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i} \right] = t\sigma^2 \end{aligned}$$

Lorsque le nombre d'observations (t) tend vers l'infini, la variance tend vers l'infini.

La variance n'est pas donc bornée. Ce processus est donc non stationnaire.

La fonction d'autocovariance est donnée par:

$$\begin{aligned}
 cov(y_t y_{t-k}) &= E[(y_t - y_0)(y_{t-k} - y_0)] \\
 &= E[(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \dots)(\varepsilon_{t-k} + \varepsilon_{t-k-1} + \varepsilon_{t-k-2} + \dots)] \\
 &= E[(\varepsilon_{t-k}^2 + \varepsilon_{t-k-1}^2 + \varepsilon_{t-k-2}^2 + \dots)] \\
 &= (t-k)\sigma^2.
 \end{aligned}$$

Elle dépend donc de t .

La fonction d'autocorrélation est:

$$\begin{aligned}
 \rho_k &= \frac{cov(y_t y_{t-k})}{[var(y_t)var(y_{t-k})]^{\frac{1}{2}}} = \frac{(t-k)\sigma^2}{[t\sigma^2(t-k)\sigma^2]^{\frac{1}{2}}} \\
 &= \left[\frac{t-k}{t} \right]^{\frac{1}{2}}.
 \end{aligned}$$

La fonction d'autocorrélation décroît donc lentement en k . Si on différencie la marche aléatoire, on obtient un bruit blanc

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t$$

Exemples: Rendement boursier et taux de change.

Marche aléatoire avec dérive (**Random Walk**)

La marche aléatoire avec dérive est définie comme étant:

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

où ε_t est un bruit blanc.

Réécrivons ce processus,

$$\begin{aligned}
 y_t &= \mu + y_{t-1} + \varepsilon_t \\
 &= \mu + \mu + y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\
 &= \mu + \mu + \mu + y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\
 &\vdots \\
 y_t &= y_0 + \mu t + \sum_{i=0}^{t-1} \varepsilon_{t-i}.
 \end{aligned}$$

C'est donc un processus caractérisé par une tendance linéaire et dont les chocs ont un effet permanent sur y_t .

Les fonctions d'autocovariance et d'autocorrélation sont données par les expressions suivantes:

$$\begin{aligned} \text{var}(y_t) &= t\sigma^2 \\ \text{cov}(y_t y_{t-k}) &= (t-k)\sigma^2 \Rightarrow \text{dépend de } t \\ \rho_k &= \left[\frac{t-k}{t} \right]^{\frac{1}{2}}. \end{aligned}$$

Ces expressions sont les mêmes que pour la marche aléatoire sans dérive.

De façon générale, on dira que y_t est intégrée d'ordre d si cette variable doit être différenciée d fois pour être stationnaire. On notera

$$y \sim I(d)$$

En particulier,

$$\begin{aligned} y &\sim I(1) \Rightarrow \Delta y_t \text{ est stationnaire,} \\ y &\sim I(0) \Rightarrow y_t \text{ est stationnaire.} \end{aligned}$$

Exemple: ARIMA(p,d,q)

$$\phi_p(L)(1-L)^d y_t = \theta_q(L)\varepsilon_t$$

On a vu que pour une série $I(1)$, la fonction d'autocorrélation décroît lentement. On doit cependant faire un test formel pour savoir si la série est stationnaire ou pas. On appelle ces tests, des tests de racine unité ou unitaire. Le test consiste à savoir si la plus grande racine est égale à 1.

Exemple: Processus AR(1)

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

$$H_0 : \phi_1 = 1$$

$$H_1 : \phi_1 < 1$$

On peut réécrire le processus AR(1) tel que

$$\Delta y_t = \mu + (\phi - 1)y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \mu + \alpha y_{t-1} + \varepsilon_t$$

$$H_0 : \alpha = 0 \Rightarrow \text{non stationnaire}$$

$$H_1 : \alpha < 0 \Rightarrow \text{stationnaire}$$

On appelle ce test, le test de Dickey-Fuller. La loi asymptotique de la statistique t est non standard. La loi dépend du nombre de termes déterministes dans les équations suivantes:

	5%	10%
$\Delta y_t = \alpha y_{t-1} + \varepsilon_t$	-1.95	-2.6
$\Delta y_t = \mu + \alpha y_{t-1} + \varepsilon_t$	-2.89	-3.51
$\Delta y_t = \mu + \beta t + \alpha y_{t-1} + \varepsilon_t$	-3.45	-4.04

Si on utilise les valeurs critiques standards, on rejette H_0 , donc la racine unité, trop souvent.

Pour un processus AR(p), on aura le test augmenté par les retards de la variable en différence

$$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \varepsilon_t$$

$$H_0 : \alpha = 0$$

$$H_1 : \alpha < 0$$

Les valeurs critiques sont les mêmes que pour le processus AR(1).

Pour effectuer le test de Dickey-Fuller augmenté, on doit choisir le nombre de retards. Ce choix est très important. Si notre équation ne comporte pas assez de retards, alors le test n'aura pas le bon niveau puisque que $\hat{\alpha}$ sera biaisé. Si l'équation comporte trop de retards, alors le test sera moins puissant (perte de degrés de liberté).

Choix du nombre de retards

Campbell et Perron ont proposé une procédure pour fixer le nombre de retards. On fixe un nombre de retards maximal et on effectue un test sur la signification du

dernier retard de l'équation. Si le dernier retard n'est pas significatif, on le retranche. On estime à nouveau et on effectue un test sur la signification du dernier retard. On continue cette procédure jusqu'à trouver un retard significatif. On peut ensuite effectuer le test de racine unité.

4.2 Tests de racine unité

On cherche à savoir si la série a une tendance déterministe ou stochastique. On va utiliser une représentation qui incorpore les deux cas possibles

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

où $v_t = \alpha v_{t-1} + u_t$.

On réécrit

$$\begin{aligned} y_t &= \gamma_0 + \gamma_1 t + \alpha(y_{t-1} - \gamma_0 - \gamma_1(t-1)) + u_t \\ y_t &= \beta_0 + \beta_1 t + \alpha y_{t-1} + u_t \end{aligned}$$

où $\beta_0 = \gamma_0(1 - \alpha) + \gamma_1 \alpha$ et $\beta_1 = \gamma_1(1 - \alpha)$. Si on a une racine unité ($\alpha = 1$) alors $\beta_1 = 0$.

On peut manipuler l'équation pour obtenir directement la statistique t avec un logiciel. On soustrait y_{t-1} de chaque côté ce qui nous donne le test de Dickey-Fuller:

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + u_t$$

où

$$H_0 : \alpha = 1 \text{ ou } (\alpha - 1) = 0$$

$$H_1 : \alpha < 1$$

On peut également considérer les cas où la série varie autour d'une moyenne. On a alors

$$y_t = \gamma_0 + v_t$$

où $v_t = \alpha v_{t-1} + u_t$.

On réécrit

$$y_t = \gamma_0 + \alpha(y_{t-1} - \gamma_0) + u_t$$

$$y_t = \beta_0 + \alpha y_{t-1} + u_t$$

où $\beta_0 = \gamma_0(1 - \alpha)$. Si on a une racine unité, alors $\beta_0 = 0$.

On peut écrire le test de racine unitaire dans ce cas de la façon suivante:

$$\Delta y_t = \beta_0 + (\alpha - 1)y_{t-1} + u_t$$

Enfin, si la série varie autour d'une moyenne égale à zéro, on a alors,

$$y_t = v_t$$

où $v_t = \alpha v_{t-1} + u_t$.

Le test de racine unitaire dans ce cas sera:

$$\Delta y_t = (\alpha - 1)y_{t-1} + u_t$$

Remarque: On peut effectuer le test de racine unité à l'aide de deux statistiques:

1. la statistique t pour $H_0 : (\alpha - 1) = 0$. On aura alors selon le cas t_{nc}, t_c, t_{ct} .
2. la statistique Z où $Z = T(\hat{\alpha} - 1)$. On aura alors les statistiques Z_{nc}, Z_c, Z_{ct} .

Considérons le cas le plus simple et examinons la loi asymptotique de $\hat{\alpha}$. On a donc l'équation suivante:

$$y_t = \alpha y_{t-1} + u_t$$

Alors, pour l'estimateur des M.C.O., on a

$$\begin{aligned} \hat{\alpha} &= \frac{\sum y_{t-1}y_t}{\sum y_{t-1}^2} = \frac{\sum y_{t-1}^2}{\sum y_{t-1}^2} + \frac{\sum u_t y_{t-1}}{\sum y_{t-1}^2} \\ &= 1 + \frac{\sum u_t y_{t-1}}{\sum y_{t-1}^2} \end{aligned}$$

On peut réécrire de façon vectorielle:

$$(\hat{\alpha} - 1) = (Y'_{-1}Y_{-1})^{-1}Y'_{-1}U$$

où Y_{-1} est le vecteur qui contient les valeurs retardées de y_t et U est le vecteur contenant les termes d'erreurs.

Pour le cas où $|\alpha| < 1$, on sait que

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{L} N\left(0, (1 - \alpha^2)\right)$$

Si $\alpha = 1$, alors la variance tend vers zéro. La loi asymptotique est donc dégénérée. On dira que l'estimateur est "super-convergent". Pour mieux comprendre le caractère non standard de la loi asymptotique, nous allons introduire brièvement un processus stochastique appelé, le mouvement Brownien.

Définition 12 *On appelle mouvement Brownien sur $[0, 1]$, un processus $(B(r), r \in [0, 1])$, gaussien, de moyenne nulle, et tel que,*

1. $B(0) = 0$
2. $cov(B(r_1), B(r_2)) = \min(r_1, r_2), \forall r_1, r_2 \in [0, 1]$
3. *pour toute réalisation, $B(r)$ est continu en t avec une probabilité 1.*

En particulier, pour $0 \leq r_1 \leq r_2 \leq 1$, $B(r_2) - B(r_1) \sim N(0, r_2 - r_1)$.

Écrivons maintenant,

$$Y_T(r) = \frac{Y_{[Tr]}}{T^{1/2}} = \frac{1}{\sqrt{T}} \sum_{\tau=1}^{[Tr]} \varepsilon_\tau$$

et $E(\varepsilon_\tau) = 0$, $var(\varepsilon_\tau) = \sigma^2$, où $[x]$ désigne la partie entière de x . On a alors que $Y_T(r)$ est égale à zéro en espérance et sa variance est donnée par $r\sigma^2$. On peut alors énoncer le résultat de convergence fonctionnelle en loi.

Théorème de Donsker

Théorème 2 *Si $(Y_t, t \geq 0)$ est une marche aléatoire fondée sur un bruit blanc indépendant de variance σ^2 , alors:*

$$\frac{1}{\sigma} Y_T(r) \xrightarrow{L} B(r)$$

où $B(r)$ est un mouvement Brownien sur $[0, 1]$.

On a alors les corollaires suivants,

Corollaire 1 Soit $(Y_t, t \geq 0)$ un processus intégré d'ordre un, la moyenne empirique est telle que:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T Y_t \xrightarrow{L} \omega \int_0^1 B(r) dr,$$

où ω^2 est la variance de long terme (à la fréquence zéro) avec $\omega = \sigma$ dans le cas d'une marche aléatoire puisque ε_t est alors un bruit blanc.

Dans le cas d'un processus ARMA pour le processus intégré $\Phi(L)\Delta Y_t = \Theta(L)\varepsilon_t$, on a que $\omega^2 = \sigma^2 \frac{\Theta(1)^2}{\Phi(1)^2}$.

Corollaire 2 Soit $(Y_t, t \geq 0)$ un processus intégré d'ordre un, on a

$$\frac{1}{T^2} \sum_{t=1}^T Y_t^2 \xrightarrow{L} \omega^2 \int_0^1 B^2(r) dr.$$

Et pour terminer,

Corollaire 3 Soit $(Y_t, t \geq 0)$ un processus intégré d'ordre un, alors,

$$\frac{1}{T} \sum_{t=1}^T Y_t \varepsilon_t \xrightarrow{L} \frac{1}{2} \omega^2 B^2(1) - \frac{1}{2} \sigma^2.$$

Démontrons le dernier Corollaire à l'aides des résultats énoncés plus haut.

Preuve:

On a que $Y_T = \sum_{t=1}^T \varepsilon_t$, alors

$$Y_t^2 = (Y_{t-1} + \varepsilon_t)^2 = Y_{t-1}^2 + 2Y_{t-1}\varepsilon_t + \varepsilon_t^2$$

ce qui implique l'égalité suivante:

$$Y_{t-1}\varepsilon_t = \frac{1}{2} (Y_t^2 - Y_{t-1}^2 - \varepsilon_t^2).$$

En sommant, on obtient,

$$\sum_{t=1}^T Y_{t-1}\varepsilon_t = \frac{1}{2} (Y_T^2 - Y_0^2) - \frac{1}{2} \sum_{t=1}^T \varepsilon_t^2,$$

Table 1: Tableau des valeurs critiques

statistiques	1%	5%	10%
t_{nc}	-2.56	-1.94	-1.62
t_c	-3.43	-2.86	-2.57
t_{ct}	-3.96	-3.41	-3.13
t_{ctt}	-4.37	-3.83	-3.55
Z_{nc}	-13.7	-8.0	-5.7
Z_c	-20.6	-14.1	-11.2
Z_{ct}	-29.4	-21.7	-18.2
Z_{ctt}	-36.6	-28.1	-24.2

puisque $\sum_{t=1}^T (Y_t^2 - Y_{t-1}^2) = (Y_T^2 - Y_0^2)$. En posant $Y_0 = 0$, on a alors directement le résultat recherché,

$$\sum_{t=1}^T Y_{t-1} \varepsilon_t \xrightarrow{L} \frac{1}{2} \omega^2 B(1)^2 - \frac{1}{2} \sigma^2.$$

En utilisant les résultats plus haut, on peut alors montrer que

$$T(\hat{\alpha} - 1) \xrightarrow{L} \frac{\frac{1}{2} (B^2(1) - 1)}{\int_0^1 B^2(r) dr}$$

où $B(r)$ est un mouvement Brownien (aussi appelé processus de Wiener standard).

Les statistiques t_{nc} et Z_{nc} auront les lois asymptotiques suivantes:

$$t_{nc} \xrightarrow{L} \frac{\frac{1}{2} (B^2(1) - 1)}{\left(\int_0^1 B^2(r) dr \right)^{\frac{1}{2}}}$$

$$Z_{nc} \xrightarrow{L} \frac{\frac{1}{2} (B^2(1) - 1)}{\int_0^1 B^2(r) dr}$$

Examinons maintenant le cas où u_t n'est pas un bruit blanc. Dans ce cas, y_t ne suit pas un processus AR(1).

On a le cas général

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

où $v_t = \alpha(L)v_t + u_t$ et $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$.

On a alors pour v_t

$$v_t = \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + \dots + \alpha_p v_{t-p} + u_t$$

On peut récrire cette équation comme étant

$$v_t = \underbrace{(\alpha_1 + \alpha_2 + \dots + \alpha_p)}_{\alpha} v_{t-1} + \delta_1 \Delta v_{t-1} + \delta_2 \Delta v_{t-2} + \dots + \delta_{p-1} \Delta v_{t-p+1} + u_t$$

Prenons un exemple simple, v_t suit un processus AR(2), alors

$$v_t = \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + u_t$$

$$v_t = \alpha_1 v_{t-1} + \alpha_2 v_{t-1} - \alpha_2 v_{t-1} + \alpha_2 v_{t-2} + u_t$$

$$v_t = (\alpha_1 + \alpha_2) v_{t-1} - \alpha_2 \Delta v_{t-1} + u_t$$

On a donc comme modèle:

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

$$\text{et } v_t = \alpha v_{t-1} + \delta_1 \Delta v_{t-1} + u_t$$

où $\alpha = \alpha_1 + \alpha_2$ et $\delta_1 = -\alpha_2$.

On peut récrire notre modèle de la façon suivante:

$$y_t = \gamma_0 + \gamma_1 t + \alpha [y_{t-1} - \gamma_0 - \gamma_1(t-1)] + \delta_1 [\Delta y_{t-1} - \gamma_1] + u_t$$

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \delta_1 \Delta y_{t-1} + u_t$$

où $\beta_0 = \gamma_0(1 - \alpha) + \gamma_1 \alpha - \delta_1 \gamma_1$ et $\beta_1 = \gamma_1(1 - \alpha)$. On soustrait de chaque côté y_{t-1} pour obtenir l'équation utilisée pour effectuer le test de Dickey-Fuller augmenté. On obtient

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + \delta_1 \Delta y_{t-1} + u_t$$

et

$$H_0 : (\alpha - 1) = 0 \Rightarrow \alpha = 1$$

$$H_1 : (\alpha - 1) < 0 \Rightarrow \alpha < 1.$$

v_t suivait un processus AR(2), on a corrigé pour la dépendance temporelle dans l'équation du test de Dickey-Fuller par un terme autorégressif d'ordre un pour Δy_t . L'addition de ce terme permet de "blanchir" le résidu.

De façon générale, si v_t suit un processus AR(p), on corrigera par $p - 1$ termes autorégressifs. Ainsi, on a le modèle suivant

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

$$\text{et } v_t = \alpha_1 v_{t-1} + \alpha_2 v_{t-2} + \dots + \alpha_p v_{t-p} + u_t$$

L'équation du test de Dickey-Fuller augmenté est alors

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \dots + \delta_{p-1} \Delta y_{t-p+1} + u_t$$

Lorsque u_t est bien un bruit blanc, on peut alors prendre les mêmes valeurs critiques (tableau 1).

Question: Comment choisir le nombre de termes à ajouter dans l'équation du test?

Réponse: par la procédure de Campbell-Perron.

Si u_t suit plutôt un processus ARMA(p,q), on cherchera à approximer le mieux possible la dynamique par l'ajout de termes autorégressifs (d'où l'importance du choix de retard). En ajoutant des termes autorégressifs, on effectue une correction paramétrique pour tenir compte de la dynamique c.à.d. pour "blanchir" le bruit. On peut également effectuer une correction non paramétrique. C'est la statistique du test proposé par Phillips et Perron (1988).

Prenons le cas suivant:

$$y_t = \gamma_0 + \gamma_1 t + v_t$$

et $v_t = \alpha v_{t-1} + u_t$.

On réécrit

$$y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + u_t$$

Si v_t suit un processus AR(p) au lieu d'un processus AR(1), alors u_t dans l'équation plus haut n'est pas un bruit blanc.

On réécrit l'équation plus haut

$$\Delta y_t = \beta_0 + \beta_1 t + (\alpha - 1)y_{t-1} + u_t$$

La statistique de Phillips-Perron est basée sur une correction non paramétrique.

Pour la statistique Z_{ct} , on aura

$$Z_{ct}^* = T(\hat{\alpha} - 1) - \frac{T^2(\hat{\omega}^2 - \hat{\sigma}^2)}{\underbrace{2 \sum_{t=1}^T y_{t-1}^2}_{\text{terme de cor. non paramétrique}}}$$

où $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2$ est la variance des résidus et $\hat{\omega}$ est un estimateur convergent de

$$\omega^2 = \lim_{T \rightarrow \infty} E \frac{\left[\sum_{t=1}^T u_t \right]^2}{T}$$

Ce terme tient compte de l'autocorrélation des u_t . Ceci correspond à la variance à la fréquence zéro.

L'estimateur de ω^2 est donné par l'expression suivante:

$$\hat{\omega}_T^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 + 2 \frac{1}{T} \sum_{k=1}^l \left(1 - \frac{k}{1+l} \right) \sum_{t=k+1}^T \hat{u}_t \hat{u}_{t-k}.$$

De la même façon, on peut obtenir la statistique t_{ct} pour la correction non paramétrique.

$$t_{ct}^* = \frac{\hat{\sigma} t_{ct}}{\hat{\omega}} - \frac{T(\hat{\omega}^2 - \hat{\sigma}^2)}{2 \left(\hat{\omega} \sum_{t=1}^T y_{t-1}^2 \right)^{1/2}}$$

Cas particulier: Si u_t est effectivement un bruit blanc, alors $\omega^2 = \sigma^2$. On a alors les statistiques habituelles, mais toujours avec les valeurs critiques mentionnées au tableau (1).

Comportement en petit échantillon

- ADF a généralement une meilleure performance à distance finie.
- Si le terme d'erreur v_t a une composante moyenne mobile avec une racine importante, c.a.d. $v_t = u_t - \theta u_{t-1}$ avec θ près de 1, le test de Phillips et Perron a un problème de niveau important.
- Les deux tests sont peu puissants.

Solution pour améliorer la puissance du test de Dickey-Fuller augmenté:

Le test d'Elliot-Rothemberg-Stock (ERS) (1996)

Ce test consiste en une modification du test de Dickey-Fuller (augmenté ou pas).
Le test consiste dans un premier temps à construire des "quasi-différences".

Prenons le cas avec une constante et une tendance. On calcule premièrement

$$\tilde{\alpha} = 1 + \frac{c}{T}$$

où $c = -13.5$, ce qui correspond à une alternative locale. Lorsque $T \rightarrow \infty$, $\alpha = 1$, donc ceci correspond à l'hypothèse nulle.

On construit par la suite les "quasi-différences" suivantes:

$$\tilde{y}_t = y_t - \tilde{\alpha} y_{t-1}$$

et

$$\tilde{z}_t = z_t - \tilde{\alpha} z_{t-1}$$

où $z_t = (1, t)$.

On effectue la régression par les M.C.O. de \tilde{y}_t sur \tilde{z}_t et on obtient l'estimateur $\hat{\beta}$. On construit ensuite une série "détrendée"

$$y_t^d = \tilde{y}_t - \tilde{z}_t \hat{\beta}.$$

On effectue ensuite la régression de Dickey-Fuller

$$\Delta y_t^d = (\alpha - 1)y_{t-1}^d + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i}^d + u_t$$

avec l'hypothèse nulle habituelle,

$$H_0 : (\alpha - 1) = 0$$

Ce test est beaucoup plus puissant que le test de Dickey-Fulley ou de Phillips et Perron. Dans le cas où on a seulement une constante, alors on prend la valeur $c = -7$.

Solution pour améliorer le niveau du test de Phillips et Perron: Tests modifiés (Perron et Ng (1996), Ng et Perron (2001))

Modifications de la statistique et choix de l'estimateur de ω (Perron et Ng (1996)):

Perron et Ng proposent les modifications suivantes aux statistiques de Z^* et t^* :

$$MZ^* = Z^* + \frac{T}{2}(\hat{\alpha} - 1)$$

et

$$Mt^* = t^* + \frac{1}{2} \left(\frac{\sum_{t=1}^T y_{t-1}^2}{\hat{\omega}^2} \right)^{1/2} (\hat{\alpha} - 1).$$

La statistique MZ^* a les mêmes valeurs critiques que la statistique Z^* puisque l'expression $\frac{T}{2}(\hat{\alpha} - 1)$ converge vers zéro sous l'hypothèse nulle. En effet, l'estimateur $\hat{\alpha}$ converge vers 1 à la vitesse de T sous l'hypothèse nulle. Les valeurs critiques de Mt^* sont données dans l'article de Stock (1990).

De plus, les auteurs trouvent que l'estimateur paramétrique de ω basé sur l'équation suivante:

$$\Delta y_t = \beta_0 + \sum_{j=1}^k \delta_j \Delta y_{t-j} + u_t.$$

donne de meilleurs résultats en petit échantillon. Cette estimateur est alors:

$$\hat{\omega}_{AR} = \frac{\hat{\sigma}_u}{(1 - \delta(1))^2}.$$

On doit choisir le nombre de retards k pour effectuer le test de Dickey-Fuller ou pour obtenir l'estimateur de la fréquence zéro défini plus haut. Ng et Perron (2001) montre que le problème de niveau dépend du choix de k .

Les critères d'information BIC et AIC tendent à choisir une valeur de k trop petite en échantillon fini. Ng et Perron proposent les statistiques BIC et AIC modifiées pour le choix de k :

$$MBIC = \ln(\hat{\sigma}_u^2) + \frac{\ln(T - kmax)(\tau_T(k) + k)}{T - kmax}$$

$$MAIC = \ln(\hat{\sigma}_u^2) + \frac{2(\tau_T(k) + k)}{T - kmax}$$

où $\tau_T(k) = (\hat{\sigma}_u^2)^{-1} (\hat{\alpha} - 1) \sum_{t=kmax}^T y_{t-1}^2$. Les deux auteurs proposent également d'utiliser les quasi-différences à la ERS pour effectuer le test.

Discussion sur la fréquence des données:

- La fréquence n'est pas importante mais plutôt la longueur de l'échantillon

Discussion:

- H_0 : la stationnarité (KPSS)
- Changement structurel
- Équivalence observationnelle

Références

- Elliot, G., T.J. Rothenberg et J.H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, **64**, 813-836.
- Ng, S. et P. Perron (2001), "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, **69**, 1519-1554.
- Perron, P. et Serena Ng (1996), "Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties," *Review of Economic Studies*, **63**, 435-463.

4.3 Régression fictive (Spurious Regression)

On a le modèle suivant:

$$y_t = \alpha_0 + \beta x_t + \varepsilon_t$$

Les hypothèses habituelles sont:

- y_t et x_t sont stationnaires
- $E(\varepsilon_t) = 0$ et $var(\varepsilon_t) < \infty$

Si y_t et x_t sont des processus non stationnaires ($I(1)$), et que $\beta = 0$, on aura une régression fictive (Granger et Newbold (1974)).

Une régression fictive est caractérisée par (Phillips (1986)):

- L'estimateur $\hat{\beta}$ ne converge pas en probabilité vers une constante (zéro dans ce cas-ci) mais vers une variable aléatoire. L'incertitude ne s'estompe pas asymptotiquement.
- R^2 tend vers une variable aléatoire lorsque T tend vers l'infini.
- La statistique t_β diverge lorsque t tend vers l'infini
- $DW = \frac{\sum_{t=0}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=0}^T \hat{\varepsilon}_t^2} \rightarrow 0$ en probabilité lorsque t tend vers l'infini.

Ainsi, ces résultats apparaissent intéressants selon les critères du R^2 et de la statistique t . Cependant, ils sont dénués de sens. L'estimateur des M.C.O. est biaisé et la loi asymptotique de $\hat{\beta}$ est non standard. La raison est que le terme d'erreur est non stationnaire. En effet, si $\beta = 0$, on a le modèle suivant:

$$y_t = \alpha_0 + \varepsilon_t.$$

Puisque y_t est $I(1)$, alors ε_t est $I(1)$. On aura également le même problème si y_t est $I(1)$ et x_t est $I(0)$. Le terme d'erreur sera alors non stationnaire.

Solution

On différencie les séries $I(1)$ pour les rendre stationnaires.

5 Cointégration

On a deux variables y_t et x_t qui sont intégrées d'ordre 1. Donc, la variance de ces variables est infinie. Cependant, une combinaison linéaire de ces deux variables peut être stationnaire.

Exemples:

- Revenu disponible et la consommation,
- Taux d'intérêt de long et de court terme
- Parité du pouvoir d'achat.

On a donc que y_t est $I(1)$ et x_t est $I(1)$. La cointégration implique la relation suivante de long terme

$$y_t = \mu + \beta x_t + \varepsilon_t$$

et le terme d'erreur est stationnaire $I(0)$. On peut réécrire de la façon suivante

$$y_t - \beta x_t = \mu + \varepsilon_t$$

On dira que $(1, -\beta)$ est le vecteur de cointégration. On peut également avoir de la cointégration entre plusieurs variables.

Définition 13 Les séries $X_{jt}, j = 1, \dots, m$ où X_t est intégré d'ordre d sont dites cointégrées si et seulement s'il existe une combinaison linéaire non nulle des séries qui est intégrée d'ordre strictement inférieur à d . On dira que

$$X \sim CI(d, b)$$

où b est le degré de cointégration.

Cette combinaison linéaire est appelée vecteur de cointégration. On a le vecteur des séries X_t alors

$$\alpha' X_t = Z_t$$

sera stationnaire si les séries X_t sont $I(1)$ et cointégrées. Z_t peut contenir une constante ou une tendance déterministe. Cependant la variance de Z_t est finie.

5.1 Cointégration et tendances communes

La cointégration implique la présence d'une ou de tendances stochastiques communes entre les séries.

Prenons un exemple avec deux séries y_t et z_t . On suppose que l'on peut décomposer ces deux séries de la façon suivante:

$$\begin{aligned}y_t &= \mu_{yt} + \varepsilon_{yt} \\z_t &= \mu_{zt} + \varepsilon_{zt}\end{aligned}$$

où μ_{it} est une marche aléatoire pour la variable i et ε_{it} est une composante stationnaire du second ordre. Puisque μ_{yt} et μ_{zt} sont une marche aléatoire, on peut les récrire de la façon suivante

$$\begin{aligned}\mu_{yt} &= \mu_{y0} + \sum_{j=1}^t \eta_{yj} \\ \mu_{zt} &= \mu_{z0} + \sum_{j=1}^t \eta_{zj}\end{aligned}$$

On a donc une tendance stochastique pour chaque variable. Si y_t et z_t sont cointégrés $CI(1, 1)$, alors il existe une combinaison linéaire $\alpha'x_t$ qui est stationnaire où $x_t = (y_t, z_t)'$. On aura donc que

$$\alpha'x_t = \alpha_1 y_t + \alpha_2 z_t = \alpha_1 \mu_{y0} + \alpha_1 \sum_{j=1}^t \eta_{yj} + \alpha_2 \mu_{z0} + \alpha_2 \sum_{j=1}^t \eta_{zj} + \alpha_1 \varepsilon_{yt} + \alpha_2 \varepsilon_{zt}$$

est stationnaire. Pour que cette expression soit stationnaire,

$$\alpha_1 \sum_{j=1}^t \eta_{yj} + \alpha_2 \sum_{j=1}^t \eta_{zj},$$

doit nécessairement être égale à zéro pour tout t . Ce qui implique la condition nécessaire et suffisante suivante:

$$\alpha_1 \mu_{yt} + \alpha_2 \mu_{zt} = \alpha_1 \mu_{y0} + \alpha_2 \mu_{z0}$$

Puisque μ_{y0} et μ_{z0} sont les valeurs initiales observées, alors

$$\alpha_1 \mu_{y0} + \alpha_2 \mu_{z0} = C^*$$

où C^* est une constante. On a donc

$$\alpha_1\mu_{yt} + \alpha_2\mu_{zt} = C^*$$

On réécrit

$$\mu_{yt} = \frac{C^*}{\alpha_1} - \frac{\alpha_2}{\alpha_1}\mu_{zt}$$

On voit donc que les variables y_t et z_t ont la même tendance stochastique à un scalaire près $\left(-\frac{\alpha_1}{\alpha_2}\right)$. On a donc l'implication suivante:

*Si deux processus $I(1)$ sont cointégrés $CI(1,1)$, ils ont nécessairement la même tendance stochastique. Toute autre combinaison linéaire α'^*x_t n'éliminera pas la tendance stochastique. Le vecteur de cointégration est donc unique à un scalaire près*

On reprend donc notre lien de cointégration,

$$\alpha'x_t = \alpha_1y_t + \alpha_2z_t = \alpha_1\mu_{y0} + \alpha_2\mu_{z0} + \alpha_1\varepsilon_{yt} + \alpha_2\varepsilon_{zt}$$

et cette combinaison linéaire est bien stationnaire puisque ε_{yt} et ε_{zt} sont deux processus stationnaires. On dira donc que y_t et z_t ont une tendance commune: "common trend" dans la terminologie de Stock-Watson (1988).

Prenons maintenant un exemple avec trois variables y_t , z_t et w_t . On a la décomposition suivante pour ces trois variables:

$$y_t = \mu_{yt} + \varepsilon_{yt}$$

$$z_t = \mu_{zt} + \varepsilon_{zt}$$

$$w_t = \mu_{wt} + \varepsilon_{wt}$$

où μ_{it} est une marche aléatoire pour la variable i et ε_{it} est une composante stationnaire du second ordre.

S'il y a cointégration entre y_t , z_t et w_t , alors il existe une combinaison linéaire $\alpha'x_t$ qui est stationnaire ($I(0)$) où $x_t = (y_t, z_t, w_t)'$. On aura alors

$$\alpha'x_t = \alpha_1y_t + \alpha_2z_t + \alpha_3w_t = \alpha_1\mu_{y0} + \alpha_1 \sum_{j=1}^t \eta_{yj}$$

$$+ \alpha_2 \mu_{z0} + \alpha_2 \sum_{j=1}^t \eta_{zj} + \alpha_3 \mu_{w0} + \alpha_3 \sum_{j=1}^t \eta_{wj} + \alpha_1 \varepsilon_{yt} + \alpha_2 \varepsilon_{zt} + \alpha_3 \varepsilon_{wt}.$$

La condition nécessaire et suffisante pour que $\alpha' x_t$ soit stationnaire est donc que

$$\alpha_1 \sum_{j=1}^t \eta_{yj} + \alpha_2 \sum_{j=1}^t \eta_{zj} + \alpha_3 \sum_{j=1}^t \eta_{wj} = 0$$

ou écrit différemment que

$$\alpha_1 \mu_{yt} + \alpha_2 \mu_{zt} + \alpha_3 \mu_{wt} = \alpha_1 \mu_{y0} + \alpha_2 \mu_{z0} + \alpha_3 \mu_{w0}$$

où le terme de droite est une constante.

Cette égalité implique que la tendance stochastique d'une variable est une combinaison des deux autres tendances stochastiques. On dira alors que y_t , z_t et w_t ont deux tendances stochastiques en commun.

Supposons maintenant que les trois variables partagent la même tendance stochastique, c.à.d.

$$\alpha_1 \sum_{j=1}^t \eta_{yj} = \alpha_2 \sum_{j=1}^t \eta_{zj} = \alpha_3 \sum_{j=1}^t \eta_{wj}$$

alors, il existera deux combinaisons linéaires $\alpha' x_t$ qui sont stationnaires. On aura donc deux vecteurs de cointégration et une tendance en commun. En effet,

$$\alpha_1 \sum_{j=1}^t \eta_{yj} = \alpha_2 \sum_{j=1}^t \eta_{zj}$$

On aura alors comme vecteur de cointégration $(1, -\frac{\alpha_2}{\alpha_1}, 0)$ pour y_t , z_t et w_t .

Ainsi, $\alpha' x_t$ où $\alpha = (1, -\frac{\alpha_2}{\alpha_1}, 0)$ est stationnaire. On a donc

$$\begin{aligned} y_t - \frac{\alpha_2}{\alpha_1} z_t &= \mu_{y0} + \sum_{j=1}^t \eta_{yj} + \varepsilon_{yt} - \frac{\alpha_2}{\alpha_1} \left[\mu_{z0} + \sum_{j=1}^t \eta_{zj} + \varepsilon_{zt} \right] \\ &= \mu_{y0} - \frac{\alpha_2}{\alpha_1} \mu_{z0} + \sum_{j=1}^t \eta_{yj} - \frac{\alpha_2}{\alpha_1} \sum_{j=1}^t \eta_{zj} + \varepsilon_{yt} - \frac{\alpha_2}{\alpha_1} \varepsilon_{zt} \end{aligned}$$

puisque

$$\alpha_1 \sum_{j=1}^t \eta_{yj} = \alpha_2 \sum_{j=1}^t \eta_{zj},$$

alors

$$y_t - \frac{\alpha_2}{\alpha_1} z_t = \mu_{y0} - \frac{\alpha_2}{\alpha_1} \mu_{z0} + \varepsilon_{yt} - \frac{\alpha_2}{\alpha_1} \varepsilon_{zt}$$

est bien un processus stationnaire. De la même façon, on peut montrer qu'un vecteur de cointégration égal à $(1, 0, -\frac{\alpha_3}{\alpha_1})$ donne bien un processus stationnaire. Dans ce cas-ci, on a donc deux vecteurs de cointégration et une tendance en commun pour les trois variables.

De façon générale, supposons la représentation suivante:

$$x_t = \mu_t + \varepsilon_t$$

où x_t est un vecteur de dimension m , $x_t = (x_{1t}, x_{2t}, \dots, x_{mt})$, μ_t est un vecteur de dimension m contenant les tendances stochastiques suivantes $\mu_t = (\mu_{1t}, \mu_{2t}, \dots, \mu_{mt})$. ε_t est un vecteur de m composante stationnaire, $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})$. S'il existe r vecteur de cointégration tel que $\alpha'x_t$ est stationnaire, alors α est de dimension $r \times m$. Il y aura $m - r$ tendances stochastiques en commun entre ces m variables. Il est important de noter que α est de rang r , les vecteurs de cointégration sont donc linéairement indépendants.

5.2 Cointégration et représentation à correction d'erreurs

S'il y a un lien à long terme entre des variables, toutes déviations de ce lien devraient entraîner un ajustement de ces variables vers ce lien de long terme. La dynamique à court terme des variables sera donc influencée par les déviations des liens à long terme. La modélisation de la dynamique de ces variables en tenant compte des liens à long terme est appelée modèle à correction d'erreurs.

Prenons un exemple. On suppose que le taux d'intérêt à court terme et à long terme sont des variables ($I(1)$). La théorie de la structure à terme des taux d'intérêt implique une relation de long terme entre ces deux variables. Il y aura cointégration et le vecteur de cointégration sera $(1, -1)$. Un modèle à correction d'erreurs pour le taux d'intérêt à court terme (r_{st}) et à long terme (r_{Lt}) sera

$$\begin{aligned} \Delta(r_{st}) &= \beta_s(r_{L,t-1} - r_{s,t-1}) + \varepsilon_{st} \\ \Delta(r_{Lt}) &= \beta_L(r_{L,t-1} - r_{s,t-1}) + \varepsilon_{Lt} \end{aligned}$$

où ε_{st} et ε_{Lt} sont des bruits blancs possiblement corrélés.

Ainsi, les variations des taux d'intérêt à court et à long terme sont fonctions des déviations de la relation de long terme à la période $t - 1$,

$$r_{L,t-1} - r_{s,t-1},$$

et β_s et β_L sont appelées paramètres de vitesse d'ajustement. On peut généraliser cette représentation de la façon suivante:

$$\begin{aligned}\Delta(r_{st}) &= b_{10} + b_s(r_{L,t-1} - r_{s,t-1}) + \sum_{i=1}^p b_{11}(i)\Delta r_{st-i} + \sum_{i=1}^p b_{12}(i)\Delta r_{Lt-i} + \varepsilon_{st} \\ \Delta(r_{Lt}) &= b_{20} + b_L(r_{L,t-1} - r_{s,t-1}) + \sum_{i=1}^p b_{21}(i)\Delta r_{st-i} + \sum_{i=1}^p b_{22}(i)\Delta r_{Lt-i} + \varepsilon_{Lt}.\end{aligned}$$

On a donc ajouté des variables retardées en différence pour chaque équation. On remarque que les différentes variables de cette représentation sont stationnaires puisque les variables en différence sont stationnaires et que $r_{L,t-1} - r_{s,t-1}$ est stationnaire. Les paramètres estimés auront donc des lois asymptotiques standards.

Le modèle à correction d'erreurs correspond à une représentation VAR augmentée du lien de cointégration entre les variables. S'il y a cointégration et que l'on modélise les variables en différence sans tenir compte du lien à long terme entre les séries, on aura alors une erreur de spécification correspondant à l'omission de la variable explicative du lien de cointégration.

De façon générale, on aura la représentation suivante:

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \dots + \pi_p x_{t-p} + \varepsilon_t$$

où $x_t = (x_{1t}, x_{2t}, \dots, x_{mt})'$ sont des variables $I(1)$ et $\varepsilon_t = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{mt})'$ est un vecteur de bruits blancs possiblement corrélés entre eux. L'expression πx_{t-1} est stationnaire. La dimension de la matrice π correspond au nombre de vecteur de cointégration.

Il y a trois cas particuliers:

1. $\text{rang}(\pi) = 0$, on a donc une représentation en différence. Chaque variable est $I(1)$ et il n'existe pas de lien de cointégration entre les variable.

2. $0 < \text{rang}(\pi) < m$ et $\text{rang}(\pi) = r$. Chaque variable est $I(1)$ et il existe r vecteurs de cointégration linéairement indépendants entre les m variables.
3. $\text{rang}(\pi) = m$. Chaque variable est stationnaire en niveau.

On peut obtenir la représentation vectorielle à correction d'erreurs (VECM) à partir du VAR pour les variables en niveau. On a donc

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_{p+1} x_{t-p+1} + \varepsilon_t$$

On peut réécrire ce VAR de la façon suivante

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \cdots + \pi_p \Delta x_{t-p} + \varepsilon_t$$

où

- $\pi = \sum_{i=1}^{p+1} A_i - I$,
- $\pi_j = -\sum_{i=j}^{p+1} A_i$.

Prenons l'exemple d'un VAR(1) à deux variables y_t et z_t . On a donc

$$\begin{aligned} y_t &= a_{11} y_{t-1} + a_{12} z_{t-1} + \varepsilon_{yt} \\ z_t &= a_{21} y_{t-1} + a_{22} z_{t-1} + \varepsilon_{zt} \end{aligned}$$

où ε_{yt} et ε_{zt} sont des bruits blancs faibles possiblement corrélés entre eux. On peut réécrire ce VAR de la façon suivante

$$\begin{bmatrix} (1 - a_{11}L) & -a_{12}L \\ -a_{21}L & (1 - a_{22}L) \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Si on inverse pour obtenir la représentation moyenne mobile, on a

$$\begin{aligned} y_t &= \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2} \\ z_t &= \frac{a_{21}L\varepsilon_{yt} + (1 - a_{11}L)\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}. \end{aligned}$$

On remarque que les deux variables ont le même dénominateur

$$(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2.$$

Ce dénominateur détermine la dynamique des deux variables. On peut obtenir les deux racines caractéristiques λ_i de ce dénominateur tel qu'on puisse le réécrire

$$(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2 = (1 - \lambda_1L)(1 - \lambda_2L).$$

Examinons les différents cas possibles

1. $|\lambda_1|, |\lambda_2| < 1$, dans ce cas les deux variables sont stationnaires,
2. $|\lambda_1| = 1, |\lambda_2| = 1$, Les deux variables seront $I(2)$. Il faut donc les "différencier" deux fois pour qu'elles soient stationnaires.
3. Si $a_{12} = a_{21} = 0$, chaque variable est un processus $AR(1)$. Si λ_1 et λ_2 sont égaux à 1, c.à.d. si a_{11} et a_{22} sont égaux à 1, alors les deux variables sont $I(1)$ mais ne sont pas cointégrées.
4. Pour que y_t et z_t soient $CI(1,1)$, il faut qu'une des racines soit égale à 1 et l'autre plus petite que 1 en valeur absolue.

Prenons le cas où $|\lambda_1| = 1$, on a alors

$$\begin{aligned} y_t &= \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - L)(1 - \lambda_2L)} \\ z_t &= \frac{a_{21}L\varepsilon_{yt} + (1 - a_{11}L)\varepsilon_{zt}}{(1 - L)(1 - \lambda_2L)} \end{aligned}$$

ou

$$\begin{aligned} (1 - L)y_t &= \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - \lambda_2L)} \\ (1 - L)z_t &= \frac{a_{21}L\varepsilon_{yt} + (1 - a_{11}L)\varepsilon_{zt}}{(1 - \lambda_2L)}. \end{aligned}$$

Les deux variables sont donc stationnaires en différences puisque $|\lambda_2| < 1$.

Pour examiner la cointégration, récrivons le VAR de la façon suivante:

$$\begin{aligned} \begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} &= \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \\ \Delta y_t &= (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \\ \Delta z_t &= a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt} \end{aligned}$$

On peut montrer que si une des racines est égale à 1, alors

$$a_{11} - 1 = -\frac{a_{12}a_{21}}{(1 - a_{22})}.$$

On obtient donc

$$\begin{aligned}\Delta y_t &= -\frac{a_{12}a_{21}}{(1 - a_{22})}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \\ \Delta z_t &= a_{21}y_{t-1} - (1 - a_{22})z_{t-1} + \varepsilon_{zt}.\end{aligned}$$

En normalisant, on a

$$\begin{aligned}\Delta y_t &= \beta_y(y_{t-1} - \alpha z_{t-1}) + \varepsilon_{yt} \\ \Delta z_t &= \beta_z(y_{t-1} - \alpha z_{t-1}) + \varepsilon_{zt}\end{aligned}$$

où $\beta_y = -\frac{a_{12}a_{21}}{(1-a_{22})}$, $\alpha = \frac{1-a_{22}}{a_{21}}$ et $\beta_z = a_{21}$. Une représentation à correction d'erreurs impose donc des restrictions sur les paramètres du VAR en niveau.

On peut maintenant énoncer le théorème de représentation d'Engel et Granger (1987).

Théorème 3 1. *Un vecteur x_t de séries $I(1)$ ne sont pas cointégrés si et seulement si le modèle s'écrit*

$$\pi(L)\Delta x_t = \varepsilon_t$$

où $\pi(L)$ est un polynôme de degré p et les racines $|\pi(L)|$ sont à l'extérieur du cercle unité.

2. *Si les séries du vecteur x_t sont cointégrées et si α est une matrice de dimension $r \times m$ dont les lignes sont des vecteurs de cointégration indépendants, le modèle admet une représentation à correction d'erreurs du type*

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^p \Delta x_{t-i} + \varepsilon_t$$

Remarques

1. La cointégration implique l'existence d'une représentation à correction d'erreurs. L'existence d'une représentation à correction d'erreurs implique la cointégration entre les variables.

2. Le nombre de relation de cointégration sera égal au rang de la matrice π . On pourra donc effectuer un test de cointégration basé sur le rang de la matrice π pour une représentation vectorielle à correction d'erreurs.

Puisque la matrice π est de rang r , on peut la décomposer de la façon suivante:

$$\pi = \beta\alpha'$$

où β est de dimension $m \times r$ et α est de dimension $m \times r$ également. La matrice α contient les vecteurs de cointégration et la matrice β les paramètres d'ajustement aux relations de long terme.

Remarque

β et α ne sont pas identifiées puisque

$$\pi = \beta\alpha' = \beta FF^{-1}\alpha' = \beta^*\alpha^{*'}$$

où $\beta^* = \beta F$ et $\alpha^{*'} = \alpha(F^{-1})'$.

On aura deux façons d'effectuer un test de cointégration

- Est-ce qu'il existe une combinaison linéaire α tel que $\alpha'x_t$ est stationnaire?
- Un test basé sur le rang de la matrice π dans le VECM.

5.3 Tests de cointégration

5.3.1 Test de Engel et Granger

On suppose que nous avons effectué un test de racine unité sur y_t et x_t et que nous ne pouvons rejeter l'hypothèse de racine unité (non stationnaire). On pense que y_t et x_t sont reliées à long terme. On va donc effectuer un test de cointégration.

On effectue un M.C.O. sur

$$y_t = \mu + \beta x_t + \varepsilon_t$$

et on fait un test de racine unité sur ε_t . Donc,

$$\Delta\varepsilon_t = \rho\varepsilon_{t-1} + \sum_{i=1}^p \delta_i \Delta\varepsilon_{t-i} + \mu_t$$

$$H_0 : \rho = 0$$

$$H_1 : \rho < 0$$

La loi asymptotique est non standard et dépend du nombre de variables dans la première estimation par M.C.O. et de la présence d'une constante et une tendance dans cette régression.

5.3.2 Test de Johansen(1988)

C'est un test basé sur le rang de la matrice π dans la représentation VECM. Le test d'Engel et Granger a le défaut du choix d'une variable de normalisation, ex:

$$y_t = a_{10} + a_{11}z_t + \varepsilon_{1t}$$

ou

$$z_t = a_{20} + a_{21}y_t + \varepsilon_{2t}.$$

En petit échantillon, le résultat peut dépendre de la variable qui est choisie pour la normalisation. Ce qui est peu souhaitable. Bien entendu, ce choix ne fait aucune différence de façon asymptotique.

De plus, si on est en présence de plusieurs variables, il peut exister plus d'un vecteur de cointégration. Il n'y a pas de façon asymptotique de procéder dans l'optique d'Engel et Granger dans ce cas.

Johansen (1988) et Stock et Watson (1988) ont développé simultanément une procédure de test basée sur le rang de la matrice π dans la représentation VECM. Pour comprendre le test de Johansen, examinons un VAR en niveau de 1 retard. On a donc

$$x_t = Ax_{t-1} + \varepsilon_t$$

on récrit ce VAR de la façon suivante

$$\begin{aligned} x_t - x_{t-1} &= Ax_{t-1} - x_{t-1} + \varepsilon_t \\ \Delta x_t &= \underbrace{(A - I)}_{\pi} x_{t-1} + \varepsilon_t. \end{aligned}$$

Le rang de la matrice π est égal au nombre de vecteurs de cointégration. Le test sera donc basé sur le rang de cette matrice. Pour effectuer ce test, on doit calculer les valeurs propres de la matrice π et évaluer si ces valeurs propres sont significativement différentes de zéro. Ce test est comparable au test de Dickey-Fuller mais dans un contexte multivarié.

De façon générale, le test sera basé sur la représentation VECM suivante:

$$\Delta x_t = A_0 + \pi x_{t-1} + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \varepsilon_t.$$

Le test consiste donc à évaluer le nombre de valeurs propres (ou valeurs caractéristiques) différentes de zéro. On calcule les valeurs propres de la matrice π et on les ordonne de façon croissante $\lambda_1 > \lambda_2 > \dots > \lambda_m$. Si le rang de π est égal à zéro, donc il n'y a pas de cointégration, alors toutes les valeurs propres ne seront pas significativement différentes de zéro.

Puisque $\ln(1) = 0$, alors $\ln(1 - \lambda_i)$ sera égal à zéro si λ_i n'est pas significativement différente de zéro. S'il n'y a pas de cointégration, alors

$$\ln(1 - \lambda_1) = \ln(1 - \lambda_2) = \dots = \ln(1 - \lambda_m) = 0.$$

Johansen introduit deux statistiques pour effectuer le test

$$\begin{aligned} \lambda_{\text{trace}}(r) &= -T \sum_{i=r+1}^m \ln(1 - \hat{\lambda}_i) \\ \lambda_{\text{max}}(r, r+1) &= -T \ln(1 - \hat{\lambda}_{r+1}). \end{aligned}$$

La première statistique $\lambda_{\text{trace}}(r)$ cherche à évaluer si le nombre de vecteurs distincts de cointégration est moins que ou égal à r , ce qui constitue l'hypothèse nulle contre une hypothèse alternative générale. La seconde statistique considère l'hypothèse nulle de r vecteurs de cointégration contre l'hypothèse nulle de $r+1$ vecteurs de cointégration.

Les valeurs critiques dépendent des termes déterministes dans le VECM (matrice A_0) et des termes déterministes dans la relation de cointégration.

Si les variables en niveau ont une tendance linéaire, il faudra le terme A_0 dans le VECM pour capter cette tendance. On aura alors

$$\Delta x_t = A_0 + \pi x_{t-1} + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \varepsilon_t$$

où $x_t = (x_{1t}, x_{2t}, \dots, x_{mt})'$. Si les variables en niveau n'ont pas de tendance linéaire mais que la relation de cointégration inclut une constante alors le VECM est

$$\Delta x_t = \pi^* x_{t-1}^* + \sum_{i=1}^p \pi_i \Delta x_{t-i} + \varepsilon_t$$

où $x_t^* = (1, x_{1t}, x_{2t}, \dots, x_{mt})'$. On peut effectuer un test pour la présence d'une constante dans la relation de cointégration.

5.4 Estimation du vecteur de cointégration

Pour fin de simplicité, on suppose le modèle bivarié suivant

$$\begin{aligned} y_{1t} &= \alpha y_{2t} + v_{1t} \\ y_{2t} &= y_{2t-1} + v_{2t} \end{aligned}$$

Les deux termes d'erreurs peuvent être corrélés entre eux de façon contemporaine et dans le temps. Ils peuvent être également autocorrélés. On définit

$$\begin{aligned} v_t &= (v_{1t}, v_{2t})' \\ \text{et } v &= (v_1, v_2, \dots, v_T)'. \end{aligned}$$

La matrice de variance-covariance de v est donnée par

$$E(vv') = \Sigma.$$

On a donc ici un cadre général. On va examiner des cas particuliers pour la matrice Σ .

5.4.1 Premier cas

v_{1t} et v_{2t} ne sont pas corrélés entre eux et ne sont pas autocorrélés. On a donc

$$\begin{aligned} E(v_{1t}v_{2t-j}) &= 0, \quad \forall j \\ E(v_{1t}v_{1t-j}) &= 0, \quad \text{pour } j \neq 0 \\ E(v_{2t}v_{2t-j}) &= 0, \quad \text{pour } j \neq 0 \end{aligned}$$

L'estimateur des moindres carrés ordinaires est donné par l'expression suivante

$$\begin{aligned}\hat{\alpha}_T &= \left[\sum_{t=1}^T y_{2t}y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t}y_{1t} \\ \hat{\alpha}_T &= \left[\sum_{t=1}^T y_{2t}y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t} [y_{2t}\alpha + v_{1t}] \\ \hat{\alpha}_T - \alpha &= \left[\sum_{t=1}^T y_{2t}y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t}v_{1t}\end{aligned}$$

Puisque $E(y_{2t}v_{1t}) = 0$, alors l'estimateur $\hat{\alpha}$ est asymptotiquement sans biais. De plus, cet estimateur converge à la vitesse T . On dira qu'il est "superconvergent". En effet

$$\hat{\alpha}_T - \alpha = \underbrace{\left[\sum_{t=1}^T y_{2t}y_{2t} \right]^{-1}}_{O_p(T^2)} \underbrace{\sum_{t=1}^T y_{2t}v_{1t}}_{O_pT} \Rightarrow \text{à montrer}$$

Il faut donc que $\hat{\alpha} - \alpha$ soit multipliée par T pour que la loi de cet estimateur ne soit pas dégénérée. On peut effectuer de l'inférence standard sur le vecteur de cointégration $(1, -\alpha)$.

5.4.2 Deuxième cas

v_{1t} n'est pas corrélé avec v_{2t} mais il est autocorrélé. On a donc

$$\begin{aligned}E(v_{1t}v_{2t-j}) &= 0, \quad \forall j \\ E(v_{1t}v_{1t-j}) &\neq 0, \quad \text{pour certains } j\end{aligned}$$

L'estimateur des moindres carrés ordinaires est alors biaisé en petit échantillon mais il est toujours "superconvergent". Examinons le biais en petit échantillon. Prenons le cas où le terme v_{1t} suit un processus AR(1). On a donc

$$\begin{aligned}y_{1t} &= \alpha y_{2t} + v_{1t} \quad \text{où} \quad v_{1t} = \rho v_{1t-1} + \varepsilon_{1t} \\ y_{2t} &= y_{2t-1} + v_{2t}.\end{aligned}$$

L'estimateur des M.C.O. est toujours

$$\hat{\alpha}_T - \alpha = \left[\sum_{t=1}^T y_{2t}y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t}v_{1t}$$

On a que

$$v_{1t} = \rho v_{t-1} + \varepsilon_{1t} = \rho [y_{1t-1} - \alpha y_{2t-1}] + \varepsilon_{1t}$$

On substitue pour l'expression de $\hat{\alpha}_T - \alpha$, alors

$$\hat{\alpha}_T - \alpha = \left[\sum_{t=1}^T y_{2t} y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t} [\rho (y_{1t-1} - \alpha y_{2t-1}) + \varepsilon_{1t}].$$

On a que $E(y_{2t} [\rho (y_{1t-1} - \alpha y_{2t-1}) + \varepsilon_{1t}]) \neq 0$ puisque $E(y_{2t} y_{2t-1}) \neq 0$. On a donc un biais en petit échantillon. Cependant, puisque $\sum_{t=1}^T y_{2t} y_{2t}$ augmente à la vitesse T^2 et que le biais augmente à la vitesse T , le biais disparaît de façon asymptotique. L'estimateur est donc "superconvergent".

Comment peut-on corriger ce biais de petit échantillon? Prenons notre exemple où v_{1t} suit un processus AR(1). On a donc

$$\begin{aligned} y_{1t} &= \alpha y_{2t} + v_{1t} \quad \text{où } v_{1t} = \rho v_{1t-1} + \varepsilon_{1t} \\ \Rightarrow y_{1t} &= \alpha y_{2t} + \frac{\varepsilon_{1t}}{(1 - \rho L)} \\ y_{1t} &= \rho y_{1t-1} + \alpha y_{2t} - \alpha \rho y_{2t-1} + \varepsilon_{1t} \\ y_{1t} &= \alpha y_{2t} + \rho (y_{1t-1} - \alpha y_{2t-1}) + \varepsilon_{1t} \end{aligned}$$

On peut donc corriger le biais en introduisant des retards du terme de correction d'erreurs. On effectue alors des M.C. non linéaires. Avec cette correction paramétrique de la dépendance temporelle de v_{1t} , on peut effectuer de l'inférence standard sur le vecteur de cointégration.

5.4.3 Troisième cas

v_{1t} est corrélée de façon contemporaine avec v_{2t} et v_{1t} n'est pas autocorrélée. On a donc,

$$\begin{aligned} E(v_{1t} v_{2t-j}) &= \sigma_{12} \quad \text{pour } j = 0 \\ &= 0 \quad \text{autrement} \end{aligned}$$

et

$$E(v_{1t} v_{1t-j}) = 0 \quad \text{pour } j \neq 0.$$

Notre estimateur des M.C.O. est donné par

$$\hat{\alpha}_T - \alpha = \left[\sum_{t=1}^T y_{2t} y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t} v_{1t}$$

puisque $y_{2T} = \sum_{i=1}^T v_{2t}$, alors $E(y_{2t} v_{1t}) \neq 0$. L'estimateur est donc biaisé en petit échantillon. Cependant, il est toujours superconvergent puisque $\sum y_{2t} y_{2t}$ augmente à la vitesse T^2 et $\sum y_{2t} v_{1t}$ augmente à la vitesse T . Donc,

$$\hat{\alpha} - \alpha \xrightarrow{t \rightarrow \infty} 0.$$

Comment peut-on corriger ce biais de petit échantillon? Puisque v_{1t} est corrélé avec v_{2t} , on peut le projeter dans l'espace engendré par v_{2t} . On aura alors

$$v_{1t} = \delta_{12} v_{2t} + \varepsilon_{1t}$$

où ε_{1t} est orthogonal à v_{2t} . On substitue maintenant dans notre modèle

$$y_{1t} = \alpha y_{2t} + v_{1t}$$

$$y_{1t} = \alpha y_{2t} + \delta_{12} v_{2t} + \varepsilon_{1t}.$$

puisque $\Delta y_{2t} = v_{2t}$, on peut réécrire le modèle comme étant

$$y_{1t} = \alpha y_{2t} + \delta_{12} \Delta y_{2t} + \varepsilon_{1t}.$$

L'estimateur des M.C.O. est alors sans biais en petit échantillon et on peut effectuer de l'inférence standard.

5.4.4 Quatrième cas

v_{1t} est corrélée de façon temporelle avec v_{2t} . On a alors

$$E(v_{1t} v_{2t-j}) = \delta_{12,j} \quad \text{pour } j = -p, \dots, p$$

L'estimateur des M.C.O.

$$\hat{\alpha}_T - \alpha = \left[\sum_{t=1}^T y_{2t} y_{2t} \right]^{-1} \sum_{t=1}^T y_{2t} v_{1t}$$

est biaisé en petit échantillon mais il est toujours “superconvergent”. Examinons le biais de petit échantillon. Prenons l’expression suivante:

$$\sum_{t=1}^T y_{2t} v_{1t} = \sum_{t=1}^T \left(\sum_{j=1}^t v_{2j} \right) v_{1t}.$$

Cette expression est différente de zéro en espérance. On a donc un biais en petit échantillon. Pour la même raison que le cas précédent, l’estimateur est, par contre “superconvergent”.

Comment peut-on corriger ce biais de petit échantillon? On va projeter v_{1t} sur l’espace engendré par v_{2t-j} pour $j = -p, \dots, p$. Ceci nous donne

$$v_{1t} = \sum_{j=-p}^p \delta_{12,j} v_{2t-j} + \varepsilon_{1t}$$

ε_{1t} est alors orthogonale à v_{2t-j} pour $j = -p, \dots, p$. On substitue dans notre modèle. On a alors

$$y_{1t} = \alpha y_{2t} + \sum_{j=-p}^p \delta_{12,j} v_{2t-j} + \varepsilon_{1t}$$

puisque $\Delta y_{2t} = v_{2t}$, le modèle peut être réécrit de la façon suivante:

$$y_{1t} = \alpha y_{2t} + \sum_{j=-p}^p \delta_{12,j} \Delta y_{2t-j} + \varepsilon_{1t}.$$

On doit donc inclure des valeurs retardées et avancées de Δy_{2t} pour corriger le biais de petit échantillon. On peut alors effectuer de l’inférence standard.

5.4.5 Contexte général

Pour une matrice de variance-covariance générale pour v_t c.à.d.

$$E(vv') = \Sigma,$$

on devra tenir compte des biais possibles en petit échantillon. Cette matrice générale de variance-covariance permet les quatre cas étudiés précédemment.

L’estimateur des moindres carrés optimal sera obtenu à partir de l’équation modifiée du modèle. On a donc

$$y_{1t} = \alpha y_{2t} + v_{1t}.$$

L'équation modifiée est donnée par

$$y_{1t} = \alpha y_{2t} + \sum_{i=1}^k \rho_k (y_{1t-i} - \alpha y_{2t-i}) + \sum_{j=-p}^p \delta_{12,j} \Delta y_{2t-j} + \varepsilon_{1t}$$

On devra donc effectuer les moindres carrés non-linéaires pour obtenir nos estimateurs. Cette solution a été proposée par Phillips-Loretan. On pourrait également corriger les biais de petit échantillon de façon non paramétrique (Stock et Watson (1993)). Les tests usuels pour le vecteur de cointégration ont alors une loi asymptotique standard.

Si notre système comporte plus qu'une relation de cointégration, on procède alors avec le VECM de Johansen. Les estimateurs ainsi obtenus sont optimaux.

5.4.6 Inférence pour la relation de cointégration

On veut effectuer des tests pour certaines restrictions sur le vecteur de cointégration. Prenons par exemple la parité des pouvoirs d'achat. On a

$$e = \frac{p}{p^*}$$

Le taux de change, les prix domestiques et étrangers sont I(1). Cependant, selon la parité des pouvoirs d'achat, il devrait y avoir cointégration entre ces trois variables. Récrivons cette relation sous forme logarithmique

$$\ln e_t = \ln p_t - \ln p_t^*$$

On peut donc effectuer un test de cointégration sur l'équation suivante:

$$\ln e_t = \alpha + \beta_1 \ln p_t + \beta_2 \ln p_t^* + v_t$$

Si v_t est I(0), alors il y a cointégration.

On peut maintenant faire un test correspondant à la parité des pouvoirs d'achats.

On a donc comme restrictions

$$H_0 = \beta_1 = 1 \text{ et } \beta_2 = -1$$

Si on utilise comme méthode d'estimation les estimateurs optimaux de Phillips-Loretan ou Johansen (VECM), alors les tests ont une loi asymptotique standard.

On peut donc effectuer les tests de façon habituelle.

6 Contexte multivarié

6.1 Introduction

On cherche à estimer les liens dynamiques entre plusieurs séries temporelles.

Exemple: la croissance de la monnaie et la croissance de la production.

6.2 Forme structurelle

On a par exemple la forme structurelle suivante:

$$\begin{aligned}\Delta y_t &= b_0 + b_1 \Delta m_{t-1} + b_2 \Delta y_{t-1} + \eta_{1t} \\ \Delta m_t &= a_0 + a_1 \Delta y_t + a_2 \Delta m_{t-1} + \eta_{2t}\end{aligned}$$

où la matrice de variance-covariance de $\eta_t = (\eta_{1t}, \eta_{2t})$ donné par $E(\eta_t \eta_t') = \Omega$ est diagonale. Les chocs η_{1t} et η_{2t} sont donc orthogonaux.

6.3 Forme réduite

On peut réécrire la forme structurelle sous la forme réduite suivante:

$$\begin{aligned}\Delta y_t &= b_0 + b_1 \Delta m_{t-1} + b_2 \Delta y_{t-1} + \eta_{1t} \\ \Delta m_t &= (a_0 + a_1 b_0) + (a_2 + a_1 b_1) \Delta m_{t-1} + a_1 b_2 \Delta y_{t-1} + \eta_{2t} + a_1 \eta_{1t}\end{aligned}$$

Une forme réduite se caractérise par une écriture des variables endogènes en fonction des variables exogènes ou prédéterminées.

On peut considérer une représentation incluant plusieurs retards.

$$\begin{aligned}\Delta y_t &= \mu_y + \delta_{21} \Delta m_{t-1} + \dots + \delta_{2p} \Delta m_{t-p} + \psi_{21} \Delta y_{t-1} + \dots + \psi_{2p} \Delta y_{t-p} + \epsilon_{1t} \\ \Delta m_t &= \mu_m + \delta_{11} \Delta m_{t-1} + \dots + \delta_{1p} \Delta m_{t-p} + \psi_{11} \Delta y_{t-1} + \dots + \psi_{1p} \Delta y_{t-p} + \epsilon_{2t}\end{aligned}$$

et $E(\epsilon_t \epsilon_t') = \Sigma$ où $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t})$.

De façon générale, on peut considérer la représentation dynamique sous forme réduite de m variables contenues dans le vecteur X_t

$$X_t = \mu + \Theta_1 X_{t-1} + \dots + \Theta_p X_{t-p} + \epsilon_t.$$

On appelle cette représentation vectorielle dynamique un V.A.R. (Vector Autoregressif Process).

Pour les m équations, on aura alors le VAR(p) suivant:

$$X_t = \mu + \Theta_1 X_{t-1} + \dots + \Theta_p X_{t-p} + \epsilon_t$$

où

$$\begin{aligned} X_t &= (X_{1t}, X_{2t}, \dots, X_{mt})', \\ \epsilon_t &= (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{mt})' \\ E(\epsilon_t) &= 0 \text{ et} \\ E(\epsilon_t \epsilon_t') &= \Sigma. \end{aligned}$$

ϵ_t est un vecteur de bruits blancs non corrélés entre eux de façon contemporaine et la matrice de variance-covariance Σ est de dimension $m \times m$.

6.4 Stationnarité

Un processus stochastique vectorielle est stationnaire si

1. $E(X_t) = m \quad \forall t$
2. $var(X_t) < \infty, \quad \forall m \text{ et } \forall t$
3. $cov(X_t, X_{t+k}) = E(X_t - m)(X_{t+k} - m)' = \Gamma_k$ dépend de k et non de t

On note que $\Gamma_k = \Gamma'_{-k}$

6.5 Estimation

Si on examine la forme réduite, c.à.d. la représentation VAR, on a un système d'équations de type S.U.R.E.. Étant donné que les variables explicatives sont les mêmes pour chaque équation, l'estimateur M.C.O. est équivalent à l'estimateur M.C.G.. Donc, l'estimateur des M.C.O. est optimal.

6.6 Choix du nombre de retards

$$AIC(p) = \ln \det(\Sigma_p) + \frac{2m^2p}{T}$$
$$SIC(p) = \ln \det(\Sigma_p) + \frac{m^2p \ln T}{T}$$

On minimise AIC ou SIC.

6.7 Causalité

Granger (1969) a proposé le concept de “causalité à la Granger” qui fait intervenir les prévisions des variables à partir de leurs passés. Prenons la représentation incluant les variables de la croissance de la monnaie et la croissance de la production.

Définition 14 *On dira que Δm_t cause Δy_t si et seulement si*

$$E(\Delta y_t / \underline{\Delta y_{t-1}}, \underline{\Delta m_{t-1}}) \neq E(\Delta y_t / \underline{\Delta y_{t-1}})$$

avec la notation $\underline{x_{t-1}} = (x_{t-1}, x_{t-2}, x_{t-3}, \dots)$.

Ceci implique que les retards de la variables Δm_t aide à prédire la variable Δy_t en tenant compte des retards de cette variable. On peut effectuer un test de causalité en testant si les retards de la variables Δm_t sont significatifs dans l'équation de Δy_t . Malheureusement ce concept de “causalité à la Granger” n'a pas d'interprétation économique.

6.8 Fonction de réponses (“Impulse responses”)

On peut obtenir la représentation moyenne mobile de la forme réduite en inversant la représentation VAR. On peut réécrire la représentation VAR comme étant:

$$\Theta_p(L)X_t = \epsilon_t.$$

où $\Theta_p(L) = I + \Theta_1L + \Theta_2L^2 + \dots + \Theta_pL^p$. Alors, la représentation moyenne mobile est

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

ou

$$X_t = C(L)\epsilon_t$$

Cette représentation nous donne l'impact des chocs ϵ_t sur les variables X_t .

Reprenons l'exemple avec la monnaie et la production. On peut inverser la représentation VAR de la forme réduite pour obtenir

$$\begin{bmatrix} \Delta m_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \mu_m^* \\ \mu_y^* \end{bmatrix} + \begin{bmatrix} C_{11}(L) & C_{12}(L) \\ C_{21}(L) & C_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

où μ_m^* et μ_y^* représentent la moyenne non conditionnelle de Δm_t et Δy_t respectivement.

Examinons maintenant l'impact du choc ϵ_{1t} sur Δm_t à l'aide de la représentation moyenne mobile. Cet impact est mesuré par $C_{11}(L)$.

$$\begin{aligned} C_{11}(L)\epsilon_{1t} &= C_{11}^0\epsilon_{1t} + C_{11}^1L\epsilon_{1t} + C_{11}^2L^2\epsilon_{1t} + \dots \\ &= +C_{11}^0\epsilon_{1t} + C_{11}^1\epsilon_{1,t-1} + C_{11}^2\epsilon_{1,t-2} + C_{11}^3\epsilon_{1,t-3} + \dots \end{aligned}$$

Le terme C_{11}^0 mesure l'impact instantané du choc ϵ_{1t} sur la variable Δm_t .

L'effet cumulatif du choc ϵ_{1t} est donné par la somme des coefficients moyenne mobile correspondants. Ainsi, cet effet est égal à

$$C_{11}(1) = C_{11}^0 + C_{11}^1 + C_{11}^2 + C_{11}^3 + \dots$$

On appelle $C_{11}(1)$ le multiplicateur de long terme puisqu'il mesure l'impact à long terme d'un choc ϵ_{1t} pour une série qui est stationnaire en différence.

On peut donc mesurer l'impact de chaque choc sur chaque variable. Cependant, ces chocs sont corrélés. Il est donc difficile d'interpréter la fonction de réponse obtenue. On devra donc chercher à retrouver la forme structurelle à l'aide de la forme réduite.

6.9 VAR structurels

On a donc la forme réduite

$$X_t = C(L)\epsilon_t.$$

On peut obtenir cette forme en estimant le VAR et en inversant la partie autorégressive. Pour simplifier, on suppose ici que la moyenne non conditionnelle du vecteur X_t est égale à zéro. Comme on le mentionnait plus tôt, les erreurs sont corrélées. Ainsi

$$E(\epsilon_t \epsilon_t') = \Sigma,$$

n'est pas une matrice diagonale.

On cherche à obtenir la forme structurelle

$$X_t = A(L)\eta_t$$

où $E(\eta_t \eta_t') = \Omega$ est une matrice diagonale. Ainsi, pour la forme structurelle, les erreurs ne sont pas corrélées.

Examinons les liens entre la forme réduite et la forme structurelle. L'effet contemporain des chocs pour la forme réduite est donné par

$$C(0)\epsilon_t = \epsilon_t$$

puisque $C(0) = I$ et l'effet contemporain des chocs pour la forme structurelle est donné par $A(0)\eta_t$. On a donc le lien suivant entre les erreurs de la forme réduite et ceux de la forme structurelle

$$\epsilon_t = A(0)\eta_t.$$

Si on peut identifier $A(0)$, alors on aura les chocs structurels η_t puisque

$$\eta_t = A(0)^{-1}\epsilon_t.$$

On obtient par la suite les coefficients moyennes mobiles structurelles par les relations suivantes:

$$A(L) = C(L)A(0).$$

Comment obtenir la forme structurelle à partir de la forme réduite?

On fait face à un problème d'identification. On veut obtenir à partir de chocs corrélés des chocs structurels donc non corrélés.

Première solution possible: Sims (1980)

On va procéder à l'aide de restrictions de court terme. Pour ce faire, on va utiliser la matrice de variance-covariance de l'impact contemporain. Cette matrice de variance-covariance pour la forme réduite et la forme structurelle sera

$$\Sigma = A(0)\Omega A(0)'$$

On peut réécrire ceci comme étant,

$$\Sigma = A^*(0)A^*(0)'$$

où $A^*(0) = A(0)\Omega^{\frac{1}{2}}$. On notera par la suite $A^*(0)$ par $A(0)$.

Le problème d'indentification est le suivant: la matrice Σ est une matrice symétrique, elle a donc $\frac{m(m+1)}{2}$ éléments différents. Pour sa part, la matrice $A(0)$ contient m^2 éléments différents. On cherche donc à identifier m^2 éléments à l'aide de $\frac{m(m+1)}{2}$ éléments différents. On a donc un problème d'identification. On devra imposer $\frac{m(m-1)}{2}$ restrictions.

La stratégie de Sims (1980) consiste à appliquer une décomposition de Choleski pour la matrice de variance-covariance Σ . La décomposition de Choleski permet de décomposer une matrice F symétrique définie positive de la façon suivante:

$$F = GG'$$

où G est une matrice triangulaire par le bas. Les éléments au-dessus de la diagonale sont donc égaux à zéro.

En appliquant la décomposition de Choleski à la matrice de variance-covariance de l'effet contemporain de la forme réduite, on obtient une matrice $A(0)$ triangulaire

par le bas, ainsi,

$$A(0) = \begin{bmatrix} a_{11} & 0 & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & 0 & \cdots & \vdots \\ a_{31} & a_{32} & a_{33} & 0 & \cdots & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m,m-2} & a_{m,m-1} & a_{mm} \end{bmatrix}.$$

Par la relation $\epsilon_t = A(0)\eta_t$, cette représentation particulière de $A(0)$ impose donc que seulement le premier choc structurel a un effet contemporain sur la première variable de la représentation VAR. Le premier et le deuxième chocs structurels ont un effet contemporain sur la deuxième variable et ainsi de suite. L'ordonnement des variables dans le VAR est donc cruciale.

De façon générale, on peut imposer $\frac{m(m-1)}{2}$ valeurs de la matrice $A(0)$ et résoudre le système d'équations résultants de l'égalité

$$\Sigma = A(0)A(0).$$

Ces restrictions peuvent provenir de raisonnements économiques (Bernanke et Mishov, QJE (1998)).

Autre solution possible: restrictions de long terme (Blanchard Quah, 1989)

On va vu que l'impact à long terme d'un choc est donné par la somme des coefficients moyennnes mobiles de ce choc. On a donc les relations suivantes pour l'impact à long terme des chocs de la forme réduite et de la forme structurelle:

$$C(1)\epsilon_t = A(1)\eta_t.$$

À partir de cette égalité, on obtient l'égalité des matrices de variance-covariance de long terme donnée par:

$$C(1)\Sigma C(1)' = A(1)A(1)' \tag{4}$$

ce qui implique que

$$C(1)A(0)A(0)C(1)' = A(1)A(1)'$$

et donc

$$C(1)A(0) = A(1).$$

En imposant un nombre suffisant de restrictions sur la matrice $A(1)$ mesurant l'impact à long terme des chocs structurels, on peut identifier la matrice $A(0)$. Par exemple, Blanchard et Quah (1989) applique une décomposition de Choleski sur la matrice de variance-covariance de long terme de la forme réduite donnée par (4). Ce type de restrictions implique, par exemple, que seulement le premier choc à un impact à long terme sur la première variable de la représentation VAR. Il est important de comprendre que l'on peut parler d'impact à long terme seulement pour des séries non stationnaires en niveau et stationnaires en différence. Ces séries seront donc incluent en différence dans la représentation VAR.

On peut maintenant généraliser le problème d'identification des chocs structurels. De façon générale, on a l'égalité suivante entre les matrices de variance-covariance de la forme réduite et de la forme structurelle:

$$C(z)\Sigma C(z)' = A(z)A(z)'$$

pour un opérateur z quelconque. Les restrictions de court terme à la Sims correspondent au cas où $z = 0$. En effet, pour ce cas, on a

$$C(0)\Sigma C(0)' = A(0)A(0)'$$

et $C(0) = I$. Les restrictions de long terme correspondent au cas où $z = 1$,

$$C(1)\Sigma C(1)' = A(1)A(1)'.$$

De la même façon, on pourrait envisager des restrictions dans le domaine des fréquences. Ainsi,

$$C(\exp(i\omega))\Sigma \overline{C(\exp(i\omega))} = A(\exp(i\omega))\overline{A(\exp(i\omega))}$$

et donc $z = \exp(i\omega)$ (Guay et Pelgrin (2005)).

Une fois que l'identification de la matrice $A(0)$ est réalisée, on peut obtenir les fonctions de réponse et la décomposition de variance des chocs structurels à partir de la représentation moyenne mobile structurelle.

6.10 Fonctions de réponses structurelles ("impulse response functions")

Maintenant que nous avons récupéré la représentation moyenne mobile structurelle, on peut analyser la réponse dynamique des variables aux chocs structurels.

On a donc la représentation moyenne mobile structurelle suivante:

$$X_t = A(L)\eta_t.$$

Cette représentation nous donne l'impact des chocs structurels η_t sur les variables X_t .

Reprenons l'exemple avec la monnaie et la production. On a la représentation moyenne mobile structurelle suivante:

$$\begin{bmatrix} \Delta m_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \mu_m^* \\ \mu_y^* \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}$$

où μ_m^* et μ_y^* représentent la moyenne non conditionnelle de Δm_t et Δy_t respectivement.

Examinons maintenant l'impact du choc η_{1t} sur Δm_t à l'aide de la représentation moyenne mobile structurelle. Cet impact est mesuré par $A_{11}(L)$.

$$\begin{aligned} A_{11}(L)\epsilon_{1t} &= A_{11}^0\eta_{1t} + A_{11}^1L\eta_{1t} + A_{11}^2L^2\eta_{1t} + \dots \\ &= A_{11}^0\eta_{1t} + A_{11}^1\eta_{1,t-1} + A_{11}^2\eta_{1,t-2} + A_{11}^3\eta_{1,t-3} + \dots \end{aligned}$$

Le terme A_{11}^0 mesure l'impact instantané du choc η_{1t} sur la variable Δm_t .

L'effet cumulatif du choc η_{1t} est donné par la somme des coefficients moyenne mobile correspondants. Ainsi, cet effet est égal à

$$A_{11}(1) = A_{11}^0 + A_{11}^1 + A_{11}^2 + A_{11}^3 + \dots$$

Lorsque la variable X_t est en différence, comme ici avec Δy_t et Δm_t , on appelle $A_{11}(1)$ le multiplicateur de long terme puisqu'il mesure l'impact à long terme d'un choc η_{1t} .

Les fonctions de réponses ("impulse response functions") mesurent donc l'impact de chaque choc structurel sur chaque variable.

6.11 Décomposition de la variance

On cherche ici à mesurer la proportion de la variance de la série qui est produit par un choc structurel. En particulier, on cherche à mesurer la contribution de chaque structurel à la variance pour différent horizon.

Reprenons la représentation moyenne mobile structurelle: On a donc la représentation moyenne mobile structurelle suivante:

$$X_t = A(L)\eta_t.$$

À partir de cette représentation, on peut calculer la prévision de X_{t+1} conditionnelle à l'ensemble d'information au temps t . Cette prévision est donnée par:

$$E_t X_{t+1} = A^1 \eta_t + A^2 \eta_{t-1} + A^3 \eta_{t-2} + \dots$$

L'erreur de la prévision est donc

$$X_{t+1} - E_t X_{t+1} = A^0 \eta_{t+1}$$

La prévision de X_{t+2} conditionnelle à l'ensemble d'information au temps t est donnée par

$$E_t X_{t+2} = A^2 \eta_t + A^3 \eta_{t-1} + A^4 \eta_{t-2} + \dots$$

L'erreur de prévision est alors

$$X_{t+2} - E_t X_{t+2} = A^0 \eta_{t+2} + A^1 \eta_{t+1}.$$

De façon générale, la prévision de X_{t+h} conditionnelle à l'ensemble d'information au temps t est

$$E_t X_{t+h} = A^h \eta_t + A^{h-1} \eta_{t-1} + A^{h-2} \eta_{t-2} + \dots$$

et l'erreur de prévision correspondante est

$$X_{t+h} - E_t X_{t+h} = A^0 \eta_{t+h} + \dots + A^1 \eta_{t+h-1} + A^2 \eta_{t+h-2} + \dots + A^{h-1} \eta_{t+1}.$$

À partir de ce résultat, on peut calculer la variance de la série provenant de chaque choc structurel pour différents horizons h de prévision.

Reprenons maintenant notre exemple. A partir de la représentation moyenne mobile structurelle de Δm_t , on peut obtenir la variance de l'erreur de prévision à différents horizons pour les deux chocs. Ainsi, de façon générale pour un horizon "h":

$$\begin{aligned} \Delta m_{t+h} - E_t \Delta m_{t+h} = & A_{11}^0 \eta_{1,t+h} + A_{11}^1 \eta_{1,t+h-1} + \dots + A_{11}^{h-1} \eta_{1,t+1} \\ & + A_{12}^0 \eta_{2,t+h} + A_{12}^1 \eta_{2,t+h-1} + \dots + A_{12}^{h-1} \eta_{2,t+1}. \end{aligned}$$

On peut donc obtenir la variance de l'erreur de prévision due aux deux chocs structurels comme étant

$$\begin{aligned} \sigma_{\Delta m}^2(h) = & \left[(A_{11}^0)^2 + (A_{11}^1)^2 + \dots + (A_{11}^{h-1})^2 \right] \sigma_{\eta_1}^2 \\ & + \left[(A_{12}^0)^2 + (A_{12}^1)^2 + \dots + (A_{12}^{h-1})^2 \right] \sigma_{\eta_2}^2. \end{aligned}$$

On peut ainsi avoir la proportion qui provient de chaque choc structurel. Pour le choc structurel 1, on a

$$\frac{\left[(A_{11}^0)^2 + (A_{11}^1)^2 + \dots + (A_{11}^{h-1})^2 \right] \sigma_{\eta_1}^2}{\sigma_{\Delta m}^2(h)}$$

et pour le choc structurel 2,

$$\frac{\left[(A_{12}^0)^2 + (A_{12}^1)^2 + \dots + (A_{12}^{h-1})^2 \right] \sigma_{\eta_2}^2}{\sigma_{\Delta m}^2(h)}.$$

Cette mesure nous indique la proportion des mouvements de la série qui sont produits par chaque choc structurel pour différents horizons. Plus cette proportion est grande pour un choc structurels, plus celui-ci est important dans les fluctuations de la série.

7 Non-linéarité des données financières

Dans les chapitres précédents, on a considéré une structure linéaire pour caractériser les données. Cependant on retrouve très souvent des non-linéarités dans les données.

Le comportement des agents économiques face au risque et aux rendements anticipés peut être non-linéaire. De plus, plusieurs contracts financiers sont non-linéaires, ex: produits dérivés.

On a également vu que les séries financières étaient caractérisées par une distribution asymétrique et leptokurtique. Pour obtenir une telle distribution, on doit considérer une modélisation non-linéaire.

Les modèles non-linéaires sont beaucoup plus difficiles à étudier. On a habituellement pas de forme analytique pour les estimateurs. On doit utiliser des algorithmes d'optimisation. Ces modèles sont donc beaucoup plus lourds à manipuler.

8 Modèles non-linéaires univariés

Par le théorème de Wold, tout processus stationnaire faible admet une représentation moyenne mobile infinie:

$$Y_t = \delta(t) + A(L)\epsilon_t$$

o ϵ_t est une suite de variables aléatoires de moyenne nulle $E(\epsilon_t) = 0$, homoscedastique $V(\epsilon_t) = \Omega$ (non conditionnelle).

Si la série a un comportement non-linéaire, les non-linéarités se retrouveront dans le terme d'erreur ϵ_t .

Lorsqu'on cherche à modéliser la série de façon non linéaire, on expliquera cette série en fonction de terme d'erreur i.i.d.. Ainsi, on aura la représentation générale suivante:

$$x_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$$

o ϵ_t est i.i.d.(0,1) et $f()$ est une fonction quelconque.

Cette fonction est trop générale pour être utilisable. On peut considérer le cas particulier suivant:

$$x_t = g(\epsilon_{t-1}, \epsilon_{t-2}, \dots) + \epsilon_t h(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$$

La fonction $g()$ représente la moyenne conditionnelle:

$$E_{t-1}x_t = g(\epsilon_{t-1}, \epsilon_{t-2}, \dots)$$

et la fonction $h()$, la variance conditionnelle:

$$E_{t-1}[(x_t - E_{t-1}x_t)^2] = h(\epsilon_{t-1}, \epsilon_{t-2}, \dots)^2$$

Le modèle sera non-linéaire dans sa moyenne si $g()$ est non-linéaire et non-linéaire dans sa variance si $h()^2$ est non-linéaire.

Exemple de modèle non-linéaire dans sa moyenne: un processus moyenne mobile d'ordre 1 non-linéaire:

$$x_t = \epsilon_t + \alpha \epsilon_{t-1}^2$$

Dans cette exemple, $g() = \alpha \epsilon_{t-1}^2$ et $h() = 1$.

Les modèles ARCH tel que le ARCH(1) suivant:

$$x_t = \epsilon_t \sqrt{\alpha \epsilon_{t-1}^2}$$

est non-linéaire dans sa variance. Ainsi, $g() = 0$ et $h() = \sqrt{\alpha \epsilon_{t-1}^2}$.

8.1 Quelques modèles paramétriques non-linéaires

On veut représenter la fonction $g()$ par l'expansion de Volterra. Ainsi, on peut représenter tout processus stationnaire strict par une approximation polynomiale dans un bruit blanc gaussien, c.à.d.:

$$x_t = \mu + \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \theta_{ij} \epsilon_{t-i} \epsilon_{t-j} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=j}^{\infty} \theta_{ijk} \epsilon_{t-i} \epsilon_{t-j} \epsilon_{t-k} + \dots$$

On peut également avoir une représentation de type autorégressive. On aura alors:

$$x_t = \mu + \sum_{i=1}^{\infty} a_i x_{t-i} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} b_{ij} x_{t-i} x_{t-j} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \sum_{k=j}^{\infty} c_{ijk} x_{t-i} x_{t-j} x_{t-k} + \dots$$

On peut également avoir une représentation de type ARMA. Par exemple, le modèle bilinéaire a la forme suivante:

$$x_t = \mu + \sum_{i=1}^{\infty} \alpha_i \epsilon_{t-i} + \sum_{i=1}^{\infty} \beta_i x_{t-i} + \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \gamma_{ij} x_{t-i} \epsilon_{t-j}$$

Une généralisation particulière de ce modèle va être caractérisée par de l'hétéroscédasticité conditionnelle (voir plus bas):

$$x_t = \mu + \sum_{i=1}^p \alpha_i \epsilon_{t-i} + \sum_{i=1}^r \beta_i \epsilon_{t-i} \epsilon_t + \epsilon_t$$

dans lequel:

$$E(x_t|I_{t-1}) = \mu + \sum_{i=1}^p \alpha_i \epsilon_{t-i}$$

$$Var(x_t|I_{t-1}) = \left(1 + \sum_{i=1}^r \beta_i \epsilon_{t-i}\right)^2 \sigma_\epsilon^2$$

8.1.1 Modèles linéaires par morceaux

Modèle TAR (Threshold Autoregressive Representation)

Le modèle TAR(1) sera du type:

$$x_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + \epsilon_t & \text{si } x_{t-1} < k \\ \alpha_2 + \beta_2 x_{t-1} + \epsilon_t & \text{si } x_{t-1} \geq k \end{cases}$$

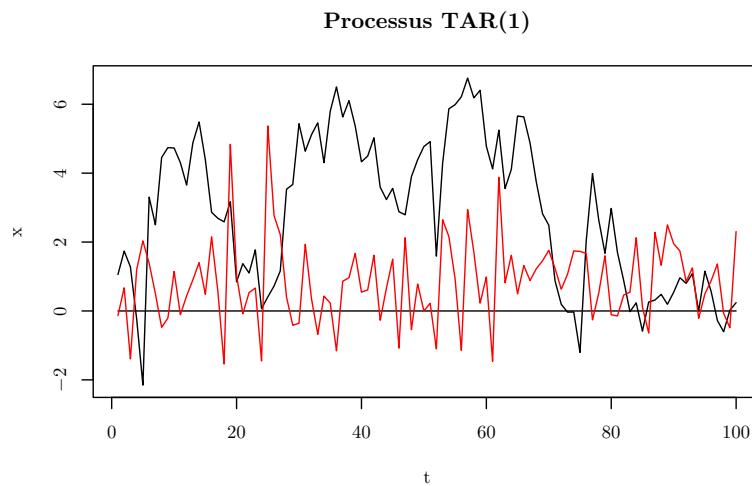
Dans le graphique qui suit, la ligne noire est une série générée à l'aide du processus TAR suivant:

$$x_t = \begin{cases} 0,9x_{t-1} + \epsilon_t & \text{si } x_{t-1} \geq 0 \\ -0,9x_{t-1} + \epsilon_t & \text{si } x_{t-1} < 0 \end{cases}$$

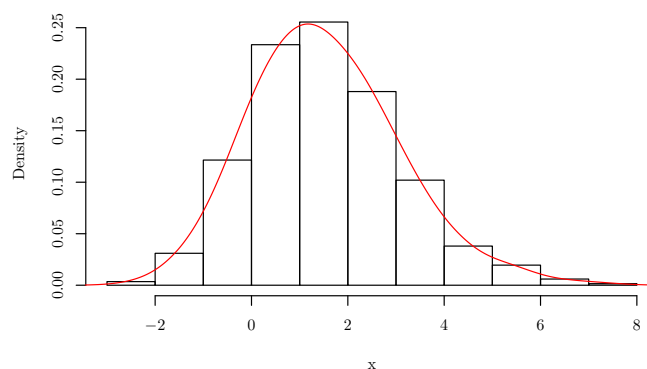
et la ligne rouge représente le processus suivant:

$$x_t = \begin{cases} 0,5x_{t-1} + \epsilon_t & \text{si } x_{t-1} \geq 0 \\ -1,9x_{t-1} + \epsilon_t & \text{si } x_{t-1} < 0 \end{cases}$$

avec $\epsilon_t \sim N(0,1)$



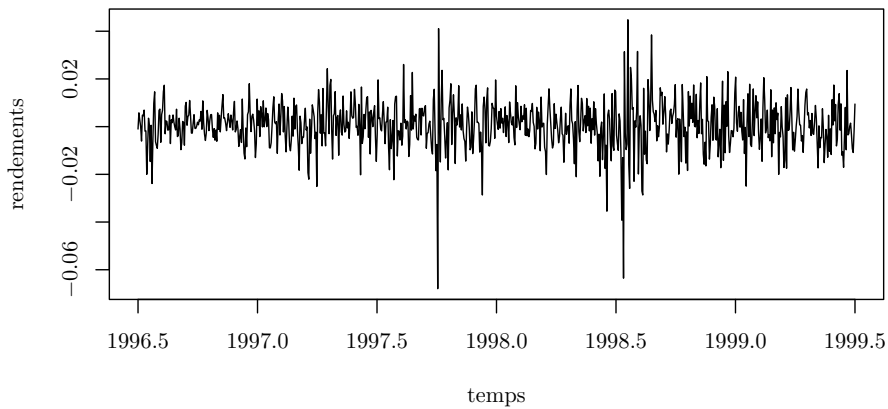
Dans ce type de processus, il y a beaucoup de persistance lorsque x_t est positif alors que l'autocorrélation devient inférieure à zéro lorsque que la série est négative. La conséquence est que le passage sous zéro est de très courte durée. De plus, comme on peut voir sur le graphique qui suit, la nonlinéarité fait en sorte que la densité de x_t est de moyenne non nulle et est asymétrique (j'ai utilisé le premier processus). Selon cette simulation, la skewness tend approximativement vers 1,9 et la kurtosis vers 3,3.¹



Prenons l'exemple des rendements journaliers de l'indice de la bourse de New-York entre juin 1996 et septembre 1999:

¹Il est à noter qu'on pourrait calculer formellement ces deux valeurs mais l'objectif est uniquement d'illustrer l'effet de la nonlinéarité.

Bourse de New-York



Le processus qui minimise le critère d'Akaike est le AR(1). Le résultat de l'estimation est:

	Estimation	É-type	test-t	valeur-p
ar1	0.0299688	0.0349428	0.858	0.3911
constante	0.0006485	0.0003526	1.839	0.0659

Regardons maintenant le résultat de l'estimation du modèle TAR(1) suivant:

$$r_t = \begin{cases} \delta + \phi^+ r_{t-1} + \epsilon_t & \text{si } r_{t-1} \geq 0 \\ \delta + \phi^- r_{t-1} + \epsilon_t & \text{si } r_{t-1} < 0 \end{cases}$$

qu'on a obtenu en estimant le modèle suivant par les MCO:

$$r_t = \delta + \phi^+ r_{t-1} 1_{(r_{t-1} \geq 0)} + \phi^- r_{t-1} 1_{(r_{t-1} < 0)} + \epsilon_t$$

o 1_{exp} est une fonction indicatrice qui est égale à 1 si l'expression "exp" est vraie et est égale à 0 sinon.

	Estimation	É-type	test-t	Valeur-p
δ	-0.0003375	0.0005138	-0.657	0.51140
ϕ^-	-0.0983034	0.0599229	-1.640	0.10129
ϕ^+	0.1693816	0.0633457	2.674	0.00765 **

Le processus est donc similaire aux deux processus qu'on a simulés plus haut.

Pour voir si ce processus reproduit les caractéristiques du rendement de l'indice, j'ai généré un processus en supposant que les estimations sont les vraies valeurs des coefficients et que $\epsilon_t \sim iid N(0, (0, 1))$. On remarque que la moyenne échantillonnale est positive malgré le fait que la constante soit négative. De plus, le processus est légèrement asymétrique comme pour le rendement du portefeuille. Par contre, la simulation ne reproduit pas la kurtosis du rendement. En effet, la kurtosis de ce dernier est égale à 8,4 alors que celle de la série simulée est de 2,9. On va donc devoir considérer d'autres types de nonlinéarité afin de reproduire le comportement leptokurtique des rendements.

Modèle TMA (Threshold Moving Average)

Le modèle sans effet contemporain (Wicker 1981) a la représentation suivante:

$$x_t = \begin{cases} \mu + \epsilon_t + \alpha_1 \epsilon_{t-1} & \text{si } \epsilon_{t-1} < 0 \\ \mu + \epsilon_t + \alpha_2 \epsilon_{t-1} & \text{si } \epsilon_{t-1} \geq 0 \end{cases}$$

Avec effet contemporain (Guay-Scaillet 2003)

$$X_t = \mu + \alpha_0^- \epsilon_t 1_{\epsilon_t < 0} + \alpha_0^+ \epsilon_t 1_{\epsilon_t \geq 0} + \alpha_1^- \epsilon_{t-1} 1_{\epsilon_{t-1} < 0} + \alpha_1^+ \epsilon_{t-1} 1_{\epsilon_{t-1} \geq 0}$$

On a également des modèles avec changements structurels. Par exemple:

$$x_t = \begin{cases} \mu_1 + \beta_1 x_{t-1} + \epsilon_t & \text{pour } t = 1, \dots, T_1 \\ \mu_2 + \beta_2 x_{t-1} + \epsilon_t & \text{pour } t = T_1 + 1, \dots, T \end{cases}$$

Modèles à changement de régime (Hamilton 1989)

On a une modélisation du type suivant:

$$x_t = \begin{cases} \alpha_1 + \beta_1 x_{t-1} + \epsilon_{1t} & \text{si } S_t = 1 \\ \alpha_2 + \beta_2 x_{t-1} + \epsilon_{2t} & \text{si } S_t = 2 \end{cases}$$

où S_t est une variable d'état non observable ayant une matrice de transition P modélisée comme étant une chaîne de Markov. Ainsi:

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}$$

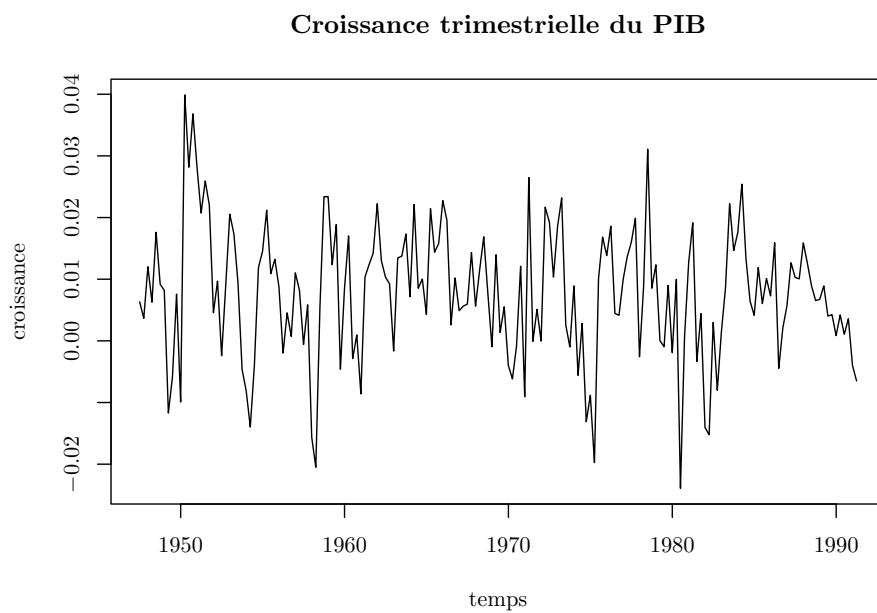
o

$$p_{11} = P\{S_t = 1 | S_{t-1} = 1\}$$

$$p_{22} = P\{S_t = 2 | S_{t-1} = 2\}$$

Remarque: la variance change selon le régime. (référence: Hamilton, *Time Series Analysis* chapitre 22)

Exemple: Prenons l'exemple de la croissance trimestrielle du PIB aux États-Unis de 1947 à 1991. Les régimes sont: $S_t = 1$: croissance positive et $S_t = 2$: croissance négative. ²



Cette série représente bien les processus à changement de régime car il y a de longues périodes consécutives de croissance positive suivies de périodes consécutives

²L'exemple est tiré de "Analysis of financial time series" par Ruey S. Tsay, chapitre 4.

durant lesquelles la croissance est négative. Les résultats de l'estimation du AR(4) à changement de régime sont:

régime 1							
Coefficients	c_1	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ_1	p_{21}
Estimation	0.909	0.265	0.029	-0.126	-0.110	0.816	0.118
É-type	0.202	0.113	0.126	0.103	0.109	0.125	0.053
régime 2							
Coefficients	c_1	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ_1	p_{12}
Estimation	-0.420	0.216	0.628	-0.073	-0.097	1.017	0.286
É-type	0.324	0.347	0.377	0.364	0.404	0.293	0.064

o $p_{21} = 1 - p_{11}$ est la probabilité de tomber en récession alors que $p_{12} = 1 - p_{22}$ est la probabilité de sortir d'une récession. On peut démontrer que l'inverse de ces probabilités représente la durée moyenne de chaque régime. Les récessions ont donc une durée moyenne de 3,69 trimestres alors que les périodes d'expansion durent en moyenne 11,31 trimestres.

9 Estimation non-paramétrique

(ref. Tsay, chap 4)

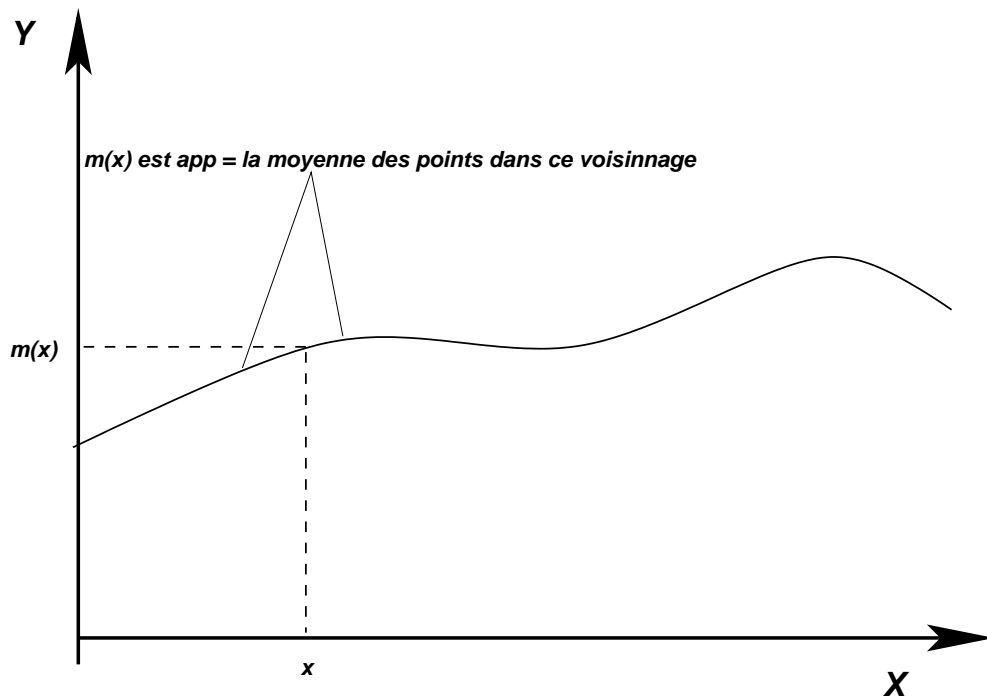
Si on ne connaît le type de nonlinéarité, on peut de façon générale supposer que Y_t est une fonction quelconque de X_t :

$$Y_t = m(X_t) + u_t$$

Si on avait n valeurs Y_i pour un X_t donnée, on pourrait estimer $m(X_t)$ par :

$$\hat{m}(X_t) = \frac{1}{n} \sum_{i=1}^n Y_i$$

En pratique on a qu'un seul Y_t pour chaque X_t . Mais si on suppose que la fonction $m(x)$ est suffisamment lisse, on peut estimer $m(x)$ en calculant la moyenne des Y_t dans un voisinage de x comme le démontre le graphique suivant:



On va donc estimer $m(x)$ de la façon suivante:

$$\hat{m}(x) = \frac{1}{T} \sum_{i=1}^T \omega_t(x) Y_t$$

o $\omega_t(x)$ représente une pondération qui accorde plus d'importance aux Y_t associés aux X_t proches de x . On peut représenter $\omega_t(x)$ de la façon suivante:

$$\omega_t(x) = \frac{K_h(x - x_t)}{\frac{1}{T} \sum_{t=1}^T K_h(x - x_t)}$$

o $K_h(z)$ est un noyau (ou kernel) qui possède la propriété d'une fonction de densité centrée à zéro et h est un paramètre (bandwidth) qui détermine l'étendu du voisinage. On a donc:

$$\int K_h(z) dz = 1$$

On a donc:

$$\hat{m}(x) = \frac{\sum_{t=1}^T K_h(x - x_t) Y_t}{\sum_{t=1}^T K_h(x - x_t)}$$

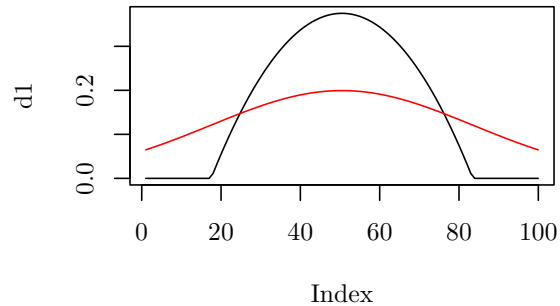
Les deux noyaux proposés par Tsay sont le Gaussien:

$$K_h(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(\frac{-x^2}{2h^2}\right)$$

et celui de Epanechnikov:

$$K_h(x) = \frac{0.75}{h} \left(1 - \frac{x^2}{h^2}\right) \left\{ \left| \frac{x}{h} \right| \leq 1 \right\}$$

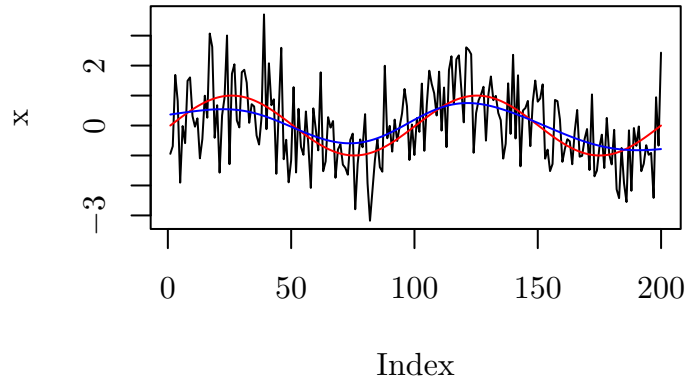
o $\{ \}$ représente une fonction indicatrice. Le graphique qui suit illustre l'allure des deux noyaux pour $h = 2$:



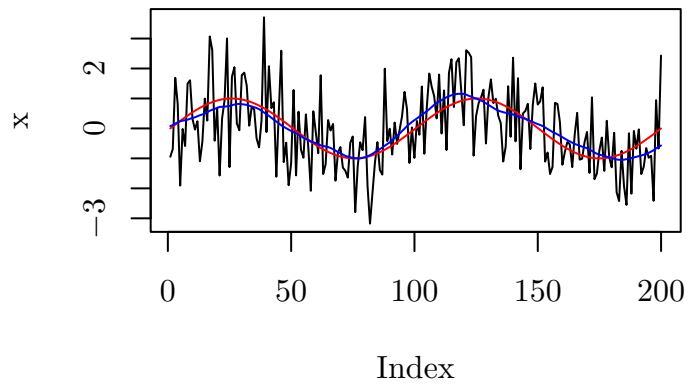
Pour examiner la performance de cet estimateur, considérons le processus suivant:

$$Y_t = \sin(X_t) + u_t$$

avec $X_t \in [-2\pi, 2\pi]$ et $u_t \sim N(0, 1)$. Le graphique suivant illustre le processus Y_t , l'estimation de $m(x)$ à l'aide du noyau gaussien et $h = 1$ et la vraie fonction, $\sin(x)$ (la ligne bleu est $\hat{m}(x)$ alors que la ligne rouge est $m(x)$):



Voici l'estimation de la même fonction avec le noyau de Epanechnikov et $h = 1$:



Il reste à présenter comment choisir h . Ce sera dans une version future du document.

10 Tests univariés de non-linéarité

10.1 Test de Hsieh

On considère des tests basés sur des moments d'ordre supérieurs à deux. Comme on a vu dans les exemples de la section précédente, la non-linéarité a un impact sur les moments d'ordre 3 et 4. Par exemple, Hsieh a introduit un test basé sur une statistique considérant le moment d'ordre 3. La statistique du test est la suivante:

$$Q(i, j) = \frac{E[x_t x_{t-i} x_{t-j}]}{E[x_t^2]^{3/2}}$$

Pour des données qui sont i.i.d. ou pour un modèle de martingale (non-linéaire seulement dans la variance) alors $Q(i, j) = 0$.

L'estimateur sera l'équivalent empirique de cette statistique:

$$\hat{Q}(i, j) = \frac{[\frac{1}{T} \sum_t x_t x_{t-i} x_{t-j}]}{[\frac{1}{T} \sum_t x_t^2]^{3/2}}$$

et

$$\sqrt{T} \hat{Q}(i, j) \xrightarrow{Loi} N(0, \hat{V})$$

o

$$\hat{V} = \frac{[\frac{1}{T} \sum_t x_t^2 x_{t-i}^2 x_{t-j}^2]}{[\frac{1}{T} \sum_t x_t^2]^3}$$

On veut tester simultanément $Q_{i,j} = 0 \quad \forall \quad i, j = 1, \dots, k$ pour un k donné. Ceci peut se faire à l'aide du test χ^2 suivant:

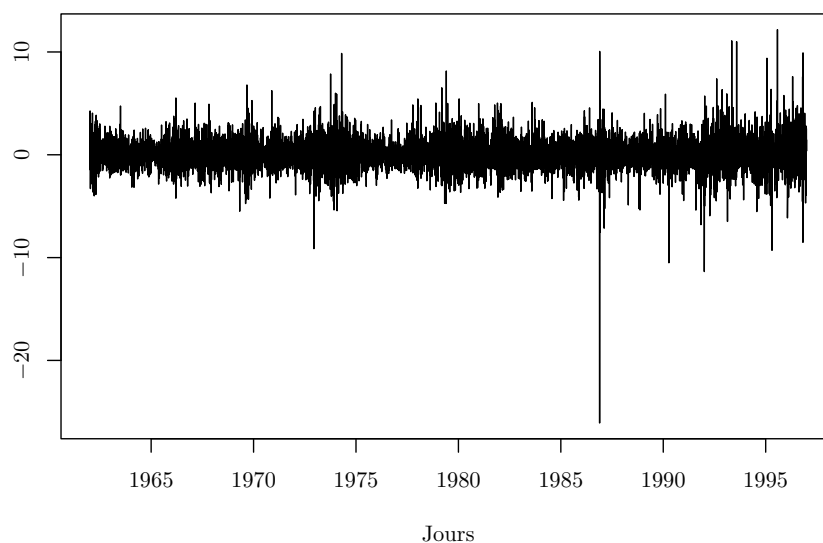
$$\hat{Q} = \sum_{i=1}^k \sum_{j=1}^k T \frac{\hat{Q}_{i,j}^2}{\hat{V}} \sim \chi_{k^2}^2$$

Le choix du paramètre k se fait de la façon habituelle en tenant compte du fait que la puissance diminue lorsque k augmente alors que le niveau est biaisé si k est trop faible.

Exemple: Considérons la série des rendements journaliers du titre de IBM de 1962 à 1997 ³.

³Source: <http://www.gsb.uchicago.edu/fac/ruey.tsay/teaching/fts/>

Rendement IBM



La statistique pour $k=4$ est égale à 15,03 pour une valeur-p de 0,57. Donc on ne rejette pas H_0 :linéarité. Si on augmente k , on obtient des résultats similaires. Le résultat ne signifie pas que le rendement du titre est linéaire car ce test ne détecte pas la nonlinéarité dans la variance.

10.2 Test BDS (Brock, Dechert, Scheinkman (1987))

Ceci est une autre test non-paramétrique mais qui test l'hypothèse iid d'une série temporelle. Il devrait donc pouvoir détecter la nonlinéarité dans la variance et dans la moyenne. Sa construction est quelque peu complexe car elle fait appel à la "correlation integral" dont l'équivalent empirique est:

$$C_l(\delta, T) = \frac{2}{(T-l+1)(T-l)} \sum_{i < j} I_\delta(X_i^l, X_j^l) \quad l = 1, k$$

o $X_i^l = (x_i, x_{i+1}, \dots, x_{i+l-1})'$ et $I_\delta(X_i, X_j)$ est une fonction indicatrice égale à 1 si $\|X_i^l, X_j^l\| < \delta$ et zéro sinon o $\|\cdot\|$ signifie supnorm (la plus grand distance entre les éléments des deux vecteurs)). Les auteurs ont démontré que

$$\sqrt{T} [C_k(\delta, T) - C_1(\delta, T)] \implies N(0, v)$$

avec

$$v = 4 \left[N^k + 2 \sum_{j=1}^{k-1} N^{k-j} C^{2j} + (k-1)^2 C^{2k} - k^2 N C^{2k-2} \right]$$

o C peut être estimé par $C_1(\delta, T)$ et N par:

$$N(\delta, T) = \frac{6}{(T-k+1)(T-k)(T-k-1)} \sum_{t < s < u} I_\delta(x_t, x_s) I_\delta(x_s, x_u)$$

On doit retirer la dépendance linéaire avant d'effectuer le test. On estime donc le ARMA(p,q) approprié et on teste les résidus. Le vecteur $(x_1, \dots, x_T)'$ utilisé plus haut représente donc les résidus et non la série originale.

Les propriétés du test dépendent du choix de k et δ . Par exemple, pour $T = 800$ le niveau du test est de 5% (pour une valeur critique de 1,96) si δ tourne autour de 1 à 1,5 fois l'écart type de x et ce pour des valeurs de k comprises entre 4 et 6. Ce niveau est maintenu pour différentes densités de x . Il n'y a que pour une variable iid chi-carré que la taille de l'échantillon doit dépasser 1000. De plus, la puissance semble excellente lorsque $T = 1000$ pour différente alternative: (i) ARCH, (ii) TAR, (iii) TMA, etc.

Exemple: Si on applique le test au rendement du titre de IBM, on obtient:

k	δ	
	σ_x	$1.5\sigma_x$
2	12,7	13,4
3	16,1	16,4
4	18,8	18,6
5	21,7	20,9
6	24,5	22,8

La linéarité est donc fortement rejetée. Étant donné que le test précédent n'a pas rejeté l'hypothèse nulle, il est donc fort probable que la nonlinéarité soit dans la variance.

10.3 Test Reset: Ramsey(1969)

Ceci est un test paramétrique. La procédure est la suivante: On estime le AR(P) suivant:

$$X_t = \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$$

et on obtient la somme des carrés des résidus SCR_1 , la moyenne conditionnelle de X_t :

$$\hat{X}_t = \hat{\delta} + \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p}$$

et les résidus $\hat{\epsilon}_t$. Ensuite on estime:

$$\begin{aligned} \hat{\epsilon}_t = & \delta + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \\ & + \gamma_2 \hat{X}_t^2 + \dots + \gamma_s \hat{X}_t^s + \eta_t \end{aligned}$$

et on obtient la somme des carrés des résidus SCR_2 . Le test est:

$$F = \frac{(SCR_1 - SCR_2)/(s + p)}{SCR_2/(T - s)}$$

On peut également tester directement les modèles paramétrique qu'on a vus plus haut (TAR, TMA, etc.)

11 Modèles à hétéroscédasticité conditionnelle

Pourquoi l'hétéroscédasticité conditionnelle peut-être intéressante en finance?

On va supposer un modèle AR(1):

$$y_t = u + \phi y_{t-1} + \epsilon_t \quad \forall t, \quad |\phi| < 1$$

où ϵ_t est un bruit blanc faible satisfaisant la condition de différence de martingale:

$$E(\epsilon_t | \epsilon_{t-1}) = 0$$

On va introduire une liaison temporelle de la variance conditionnelle par l'intermédiaire d'un processus autorégressif:

$$\epsilon_t^2 = c + a\epsilon_{t-1}^2 + u_t \quad \forall t$$

où u_t est un bruit blanc fort.

On a besoin de certaines conditions:

1. $|a| < 1$
2. Il faut assurer la positivité de ϵ_t^2

Les conditions suffisantes pour le point 2 est $a > 0$ et $c + u_t \geq 0$ pour toute valeur admissible de u_t . On a donc une contrainte pour le support de la loi de u_t .

Propriétés du processus d'innovation ϵ_t

On a fait l'hypothèse que $E(\epsilon_t | \epsilon_{t-1}) = 0$. Ceci implique que $E(\epsilon_t | \epsilon_{t-h}) = 0$ par le théorème de projection itérée:

$$E(\epsilon_t | \epsilon_{t-h}) = E[E(\epsilon_t | \epsilon_{t-1}) | \epsilon_{t-h}] = E(0 | \epsilon_{t-h}) = 0$$

Par remplacement récursif, on obtient de l'équation de ϵ_t^2 que:

$$\epsilon_t^2 = c[1 + a + \dots + a^{h-1}] + a^h \epsilon_{t-h}^2 + u_t + au_{t-1} + \dots + a^{h-1} u_{t-h+1}$$

Alors

$$E(\epsilon_t^2 | \epsilon_{t-h}) = c \frac{1 - a^h}{1 - a} + a^h \epsilon_{t-h}^2$$

donc

$$V(\epsilon_t | \epsilon_{t-h}) = c \frac{1 - a^h}{1 - a} + a^h \epsilon_{t-h}^2$$

Si $h \rightarrow \infty$, on obtient la variance non conditionnelle:

$$V(\epsilon_t) = \frac{c}{1 - a} \rightarrow \text{homoscédasticité}$$

Propriétés du processus y_t

$$E(y_t | y_{t-h}) = \mu \frac{1 - \phi^h}{1 - \phi} + \phi^h y_{t-h}$$

$$V(y_t | y_{t-1}) = V(\epsilon_t | y_{t-1}) = c + a \epsilon_{t-1}^2$$

Distribution des erreurs

On peut obtenir d'autres résultats si on précise la loi des termes d'erreurs ϵ_t . Engel(1982) suppose que ϵ_t est conditionnellement normal, c.à.d.:

$$\epsilon_t | \epsilon_{t-1} \sim N(0, c + a \epsilon_{t-1}^2)$$

Ce processus conditionnellement normal n'est pas marginalement gaussien.

On a que

$$E(\epsilon_t^2) = \frac{c}{1 - a}$$

$$E(\epsilon_t^4) = \frac{3c^2(1 - a^2)}{(1 - a)^2(1 - 3a^2)}$$

On calcule le Kurtosis:

$$K = \frac{E(\epsilon_t^4)}{3E(\epsilon_t^2)^2} = \frac{1 - a^2}{1 - 3a^2} > 1$$

La distribution est dite leptokurtique.

11.1 Extension des modèles ARCH

11.1.1 Modèle ARCH(q) Engel(1982)

Généralisation du modèle ARCH(1). On aura que:

$$\epsilon_t^2 = c + \sum_{i=1}^q a_i \epsilon_{t-i}^2 + u_t$$

et

$$E(\epsilon_t | \epsilon_{t-1}) = 0$$

où u_t est une différence de martingale. La variance conditionnelle est donc:

$$V(\epsilon_t|\epsilon_{t-1}) = c + \sum_{i=1}^q a_i \epsilon_{t-i}^2$$

Différence de martingale

$$E(\epsilon_t|\epsilon_{t-1}) = 0$$

$$V(\epsilon_t) = \Omega \quad \text{indépendant de } t$$

11.1.2 Modèle GARCH(p,q) Bollerslev(1986)

On peut introduire une partie moyenne mobile. Le modèle est défini par:

$$E(\epsilon_t|\epsilon_{t-1}) = 0$$

$$V(\epsilon_t|\epsilon_{t-1}) = h_t = c + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$$

Ce modèle peut être réécrit sous une forme plus facile à comparer avec l'écriture autorégressive. Introduisons l'innovation suivante:

$$u_t = \epsilon_t^2 - h_t$$

On remplace dans la formulation GARCH:

$$\epsilon_t^2 - u_t = c + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j (\epsilon_{t-j}^2 - u_{t-j})$$

et donc

$$\epsilon_t^2 = c + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + u_t - \sum_{j=1}^p \beta_j u_{t-j}$$

C'est une sorte de représentation ARMA pour le processus ϵ_t^2 .

11.1.3 Modèles ARCH-M Engel-Lilien-Robins(1987)

L'idée qu'un actif risqué produit un rendement supérieur suggère que la variance conditionnelle devrait être une variable explicative pour un processus de rendement. Le ARCH-M (ARCH-in-mean) nous permet de modéliser cette caractéristique. Exemple:

$$y_t = x_t' \beta + \delta h_t + \epsilon_t$$

où ϵ_t suit un processus GARCH.

11.2 Stationnarité d'un processus GARCH(p,q)

On a que:

$$E(\epsilon_t | \epsilon_{t-1}) = 0$$
$$V(\epsilon_t | \epsilon_{t-1}) = h_t = c + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$$

Propriété Un processus ϵ_t satisfaisant un modèle GARCH(p,q) avec des coefficients positifs: $c \geq 0$, $\alpha_i \geq 0$ et $\beta_j \geq 0$, $\forall i, \forall j$, est asymptotiquement stationnaire au second ordre si:

$$\alpha(1) + \beta(1) = \sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$$

11.3 Kurtosis

On suppose un processus GARCH conditionnellement gaussien. Dans un tel cas, les moments d'ordre 2 et 4 du processus sont liés par:

$$E(\epsilon_t^4 | \epsilon_{t-1}) = 3[E(\epsilon_t^2 | \epsilon_{t-1})]^2$$

On prend l'espérance non conditionnelle de chacun des membres:

$$\begin{aligned} E(\epsilon_t^4) &= E[E(\epsilon_t^4 | \epsilon_{t-1})] \\ &= 3E[E(\epsilon_t^2 | \epsilon_{t-1})]^2 \\ &\geq 3\{E[E(\epsilon_t^2 | \epsilon_{t-1})]\}^2 \rightarrow \text{Inégalité de Jensens} \\ &= 3[E(\epsilon_t^2)]^2 \end{aligned}$$

On peut en déduire que

$$K = \frac{E(\epsilon_t^4)}{3[E(\epsilon_t^2)]^2} = 1 + \frac{V[E(\epsilon_t^2 | \epsilon_{t-1})]}{[E(\epsilon_t^2)]^2}$$

Le kurtosis est donc lié à l'hétéroscédasticité conditionnelle.

Inégalité de Jensens Si $h(\cdot)$ est une fonction convexe, alors:

$$E[h(x)] \geq h[E(x)]$$

De la même façon que pour le processus ARCH(1), on peut montrer pour le processus GARCH(1,1) que si $1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 > 0$ alors le Kurtosis est donnée par:

$$K = \frac{E(\epsilon_t^4)}{3E(\epsilon_t^2)^2} = \frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 1$$

La distribution est donc leptokurtique.

12 Estimation par le pseudo-maximum de vraisemblance

On détermine l'estimateur du (pseudo) maximum de vraisemblance en se basant sur l'hypothèse de la loi conditionnelle normale. La vraisemblance est:

$$L(Y; \theta) = \prod_{t=1}^T l_t(y; \theta)$$

où l_t est la densité conditionnellement au passé. L'estimateur est obtenu comme étant:

$$\hat{\theta}_T = \arg \max_{\theta} [\ln L(Y; \theta)]$$

Cette estimateur du pseudo-maximum de vraisemblance dépend à la fois de la vraie distribution sous jacente et de celle utilisée pour calculer la vraisemblance (ici la loi normale). Sous diverses conditions de régularité, cet estimateur est convergent, même si la véritable loi sous jacente n'est pas conditionnellement normale. De plus, cet estimateur est asymptotiquement normal et a comme variance asymptotique:

$$V_{as}[\sqrt{T}(\hat{\theta}_T - \theta)] = J^{-1} I J^{-1}$$

où

$$J = E_0 \left[-\frac{\partial^2 \ln l_t(Y; \theta)}{\partial \theta \partial \theta'} \right]$$

et

$$I = E_0 \left[\frac{\partial \ln l_t(Y; \theta)}{\partial \theta} \frac{\partial \ln l_t(Y; \theta)}{\partial \theta'} \right]$$

E_0 désignant l'espérance par rapport à la vraie loi.

Généralement, J et I sont différentes. Ces deux matrices sont semblables si la vraie loi est la loi utilisée pour obtenir l'estimateur du PMV. Alors, on a:

$$V_{as}[\sqrt{T}(\hat{\theta}_T - \theta)] = J^{-1} = I^{-1}$$

Exemple de l'utilisation du P.M.V.:

On suppose une suite de variables aléatoire indépendantes de même loi Y_t , $t = 1, \dots, T$, de moyenne m et de variance σ^2 inconnus. La log-vraisemblance basée sur l'hypothèse de normalité est:

$$\ln L = \frac{-T}{2} \ln \sigma^2 - \frac{T}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{t=1}^T (Y_t - m)^2$$

les C.P.O. sont:

$$\frac{\partial \ln L}{\partial m} = \frac{1}{\sigma^2} \sum_{t=1}^T (Y_t - m)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^T (Y_t - m)^2$$

On obtient à partir de ces C.P.O., les estimateurs du P.M.V.:

$$\hat{m}_T = \frac{1}{T} \sum_{t=1}^T Y_t \quad , \quad \hat{\sigma}_T^2 = \frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y}_T)^2$$

Ces estimateurs sont convergents même si la vraie loi n'est pas normale.

On obtient pour I et J:

$$I = \begin{bmatrix} \frac{1}{\sigma^2} E \left(\frac{Y-m}{\sigma} \right)^2 & E \left\{ \frac{Y-m}{\sigma^2} \left[-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (Y-m)^2 \right] \right\} \\ E \left\{ \frac{Y-m}{\sigma^2} \left[-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (Y-m)^2 \right] \right\} & E \left[-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} (Y-m)^2 \right]^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & \frac{1}{2\sigma^3} E(u^3) \\ \frac{1}{2\sigma^3} E(u^3) & \frac{3k-1}{4\sigma^4} \end{bmatrix}$$

où $u = \frac{Y-m}{\sigma}$ et $k = \frac{Eu^4}{3(Eu^2)^2} = \frac{Eu^4}{3}$.

J a la forme usuelle:

$$J = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

La matrice de variance-covariance asymptotique est:

$$\begin{aligned} V_{as} \left[\sqrt{T} \begin{pmatrix} \hat{m}_T - m \\ \hat{\sigma}_T^2 - \sigma^2 \end{pmatrix} \right] &= \begin{bmatrix} \sigma^2 & \sigma^3 E u^3 \\ \sigma^3 E u^3 & (3k-1)\sigma^4 \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & m_3 \\ m_3 & (3k-1)\sigma^4 \end{bmatrix} \end{aligned}$$

où $m_3 = E(Y - m)^3$.

Si la vraie loi est normale, on a $m_3 = 0$ et $k = 1$; les deux estimateurs \hat{m}_T et $\hat{\sigma}_T^2$ sont asymptotiquement non corrélés. Ce n'est pas le cas en général. Il y aura corrélation entre les deux estimateurs si la loi présente de l'asymétrie.

12.1 P.M.V. avec de l'hétéroscédasticité conditionnelle

On a le modèle suivant:

$$E(Y_t | Y_{t-1}, X) = m_t(\theta) = m(Y_{t-1}, X_t; \theta)$$

et

$$V(Y_t | Y_{t-1}, X) = h_t(\theta) = h(Y_{t-1}, X_t; \theta)$$

Ces deux moments conditionnels peuvent dépendre à la fois des valeurs passées du processus et des valeurs présentes et passées de certaines caractéristiques exogènes X . Ce modèle inclut les différentes spécifications présentées avec de l'hétéroscédasticité conditionnelle.

On va écrire la vraisemblance comme si la distribution conditionnelle de Y_t sachant Y_{t-1} et X était normale. On a donc:

$$\ln L = \frac{-T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln h_t(\theta) - \frac{1}{2} \sum_{t=1}^T \frac{[Y_t - m_t(\theta)]^2}{h_t(\theta)}$$

Les C.P.O. sont:

$$\frac{\partial \ln L}{\partial \theta} = \frac{-1}{2} \sum_{t=1}^T \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} + \frac{1}{2} \sum_{t=1}^T \frac{[Y_t - m_t(\theta)]^2}{h_t(\theta)^2} \frac{\partial h_t(\theta)}{\partial \theta} + \sum_{t=1}^T \frac{Y_t - m_t(\theta)}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta}$$

Les C.P.O. peuvent être facilement écrites en fonction des résidus "standardisés":

$$\hat{u}_t = \frac{Y_t - m_t(\hat{\theta})}{h_t(\hat{\theta})^{1/2}}$$

On obtient

$$\frac{1}{2} \sum_{t=1}^T \frac{1}{h_t(\hat{\theta})} \frac{\partial h_t(\hat{\theta})}{\partial \theta} [\hat{u}_t^2 - 1] + \sum_{t=1}^T \frac{1}{h_t(\hat{\theta})^{1/2}} \frac{\partial m_t(\hat{\theta})}{\partial \theta} \hat{u}_t = 0$$

⇒ Conditions d'orthogonalité avec les résidus et les résidus aux carrés centrés.

Si on veut diviser θ tel que $\theta = (\alpha', \beta)'$, où α n'apparaît que dans la moyenne conditionnelle et β dans la variance conditionnelle, on a:

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{t=1}^T \frac{Y_t - m_t(\alpha)}{h_t(\beta)} \frac{\partial m_t(\alpha)}{\partial \alpha} = \sum_{t=1}^T \frac{1}{h_t(\beta)^{1/2}} \frac{\partial m_t(\alpha)}{\partial \alpha} u_t = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{2} \sum_{t=1}^T \frac{1}{h_t(\beta)} \frac{\partial h_t(\beta)}{\partial \beta} (u_t^2 - 1) = 0$$

Matrice de covariance asymptotique

On a que

$$J = E_0 \left[\frac{1}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} + \frac{1}{2h_t(\theta)^2} \frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} \right]$$

La matrice I a des expressions qui font intervenir le moment conditionnel d'ordre 3 de Y_t :

$$M_{3t}(\theta) = E[(Y_t - m_t(\theta))^3 | Y_{t-1}, X_t]$$

et sa kurtosis conditionnelle:

$$K_t(\theta) = \frac{1}{3h_t(\theta)^2} E[(Y_t - m_t(\theta))^4 | Y_{t-1}, X_t]$$

Cette matrice est donnée par:

$$I = E_0 \left\{ \frac{1}{4} \frac{1}{h_t(\theta)^2} \frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} [3K_t(\theta) - 1] + \frac{1}{h_t(\theta)} \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} + \frac{1}{2h_t(\theta)^3} \left[\frac{\partial h_t(\theta)}{\partial \theta} \frac{\partial m_t(\theta)}{\partial \theta'} + \frac{\partial m_t(\theta)}{\partial \theta} \frac{\partial h_t(\theta)}{\partial \theta'} \right] M_{3t}(\theta) \right\}$$

Ces formules sont une généralisation de celles obtenues dans le cas d'échantillonnage.

Si $K_t(\theta) = 1$ et $M_{3t}(\theta) = 0$, alors $J = I$.

Si le vecteur θ peut être écrit $\theta = (\alpha', \beta)'$ et que α apparaît seulement dans la moyenne conditionnelle et β dans la variance conditionnelle, alors:

$$J = \begin{bmatrix} E_0 \left[\frac{1}{h_t(\beta)} \frac{\partial m_t(\alpha)}{\partial \alpha} \frac{\partial m_t(\alpha)}{\partial \alpha'} \right] & 0 \\ 0 & E_0 \left[\frac{1}{2h_t(\beta)^2} \frac{\partial h_t(\beta)}{\partial \beta} \frac{\partial h_t(\beta)}{\partial \beta'} \right] \end{bmatrix}$$

$$I = \begin{bmatrix} E_0 \left[\frac{1}{h_t(\beta)} \frac{\partial m_t(\alpha)}{\partial \alpha} \frac{\partial m_t(\alpha)}{\partial \alpha'} \right] & E_0 \left[\frac{1}{2h_t(\beta)^2} \frac{\partial m_t(\alpha)}{\partial \alpha} \frac{\partial h_t(\beta)}{\partial \beta'} M_{3t}(\theta) \right] \\ E_0 \left[\frac{1}{2h_t(\beta)^2} \frac{\partial m_t(\alpha)}{\partial \alpha} \frac{\partial h_t(\beta)}{\partial \beta'} M_{3t}(\theta) \right] & E_0 \left[\frac{1}{4h_t(\beta)^2} \frac{\partial h_t(\beta)}{\partial \beta} \frac{\partial h_t(\beta)}{\partial \beta'} (3K_t(\theta) - 1) \right] \end{bmatrix}$$

La variance asymptotique de l'estimateur $\hat{\alpha}_t$ est

$$V_{as}(\sqrt{T}(\hat{\alpha}_T - \alpha)) = \left\{ E_0 \left[\frac{1}{h_t(\beta)} \frac{\partial m_t(\alpha)}{\partial \alpha} \frac{\partial m_t(\alpha)}{\partial \alpha'} \right] \right\}^{-1}$$

et celle de $\hat{\beta}_T$

$$\begin{aligned} V_{as}(\sqrt{T}(\hat{\beta}_T - \beta)) &= \left\{ E_0 \left[\frac{1}{2h_t(\beta)^2} \frac{\partial h_t(\beta)}{\partial \beta} \frac{\partial h_t(\beta)}{\partial \beta'} \right] \right\}^{-1} \\ &X \left\{ E_0 \left[\frac{1}{4h_t(\beta)^2} \frac{\partial h_t(\beta)}{\partial \beta} \frac{\partial h_t(\beta)}{\partial \beta'} (3K_t(\theta) - 1) \right] \right\} \\ &X \left\{ E_0 \left[\frac{1}{2h_t(\beta)^2} \frac{\partial h_t(\beta)}{\partial \beta} \frac{\partial h_t(\beta)}{\partial \beta'} \right] \right\}^{-1} \end{aligned}$$

On estime I et J par la moyenne empirique:

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial \ln l_t(Y, \hat{\theta}_T)}{\partial \theta} \frac{\partial \ln l_t(Y, \hat{\theta}_T)}{\partial \theta'}$$

$$\hat{J}_T = \frac{-1}{T} \sum_{t=1}^T \frac{\partial^2 \ln l_t(Y, \hat{\theta}_T)}{\partial \theta \partial \theta'}$$

et

$$V_{as}(\sqrt{T}(\hat{\theta}_T - \theta)) = \hat{J}_T^{-1} \hat{I}_T \hat{J}_T^{-1}$$

Considérons un modèle de régression avec des erreurs ARCH(p). On a que:

$$Y_t = X_t b + \epsilon_t$$

où

$$E(\epsilon_t | \epsilon_{t-1}) = 0$$

$$V(\epsilon_t | \epsilon_{t-1}) = c + a_1 \epsilon_{t-1}^2 + \dots + a_p \epsilon_{t-p}^2$$

ainsi

$$m_t(\theta) = X_t b$$

$$h_t(\theta) = c + a_1(Y_{t-1} - X_{t-1}b)^2 + \dots + a_p(Y_{t-p} - X_{t-p}b)^2$$

On a deux groupes de paramètres $\alpha = b$ et $\beta = (c, a')'$. Le premier groupe apparaît à la fois dans la moyenne et la variance, alors que le second est spécifique à la variance.

Les C.P.O. sont:

$$\frac{\partial \ln L(\hat{\theta})}{\partial \alpha} = - \sum_{t=1}^T \frac{1}{\hat{h}_t^2} (\hat{\epsilon}_t^2 - \hat{h}_t) \left(\sum_{j=1}^p \hat{a}_j x'_{t-j} \hat{\epsilon}_{t-j} \right) + \sum_{t=1}^T \frac{\hat{\epsilon}_t}{\hat{h}_t} x'_t = 0$$

$$\frac{\partial \ln L(\hat{\theta})}{\partial \beta} = \sum_{t=1}^T \frac{1}{2\hat{h}_t^2} (\hat{\epsilon}_t^2 - \hat{h}_t) \begin{bmatrix} 1 \\ \hat{\epsilon}_{t-1}^2 \\ \vdots \\ \hat{\epsilon}_{t-p}^2 \end{bmatrix} = 0$$

Le premier sous-système correspond aux C.P.O. usuelles avec hétéroscédasticité exogène. Le second sous-système a l'interprétation suivante: après remplacement du paramètre b par son estimateur, les équation autorégressives définissant le moment d'ordre deux peuvent être approchées par:

$$\hat{\epsilon}_t^2 = c + a_1 \hat{\epsilon}_{t-1}^2 + \dots + a_p \hat{\epsilon}_{t-p}^2 + u_t$$

où

$$\hat{\epsilon}_t = Y_t - X_t \hat{b}_t$$

$$V(u_t) = V(\epsilon_t^2 | \epsilon_{t-1}) = 2h_t^2$$

sous l'hypothèse de normalité conditionnelle

On peut donc effectuer des moindres carrés pondérés (généralisés) avec comme poids $1/2\hat{h}_t^2$.

On peut également montrer que les paramètres figurant dans la moyenne et ceux figurant dans la variance sont non corrélés asymptotiquement comme c'est le cas pour la loi gaussienne, et ce, même si le paramètre b apparaît dans la moyenne

conditionnelle et la variance conditionnelle dans le cas où ϵ_t est conditionnellement gaussien. Donc on aura que:

$$J_{\alpha\beta} = 0$$

où $J_{\alpha\beta}$ est la dérivée de la vraisemblance par rapport à α et β .

12.2 Application à un modèle GARCH

Considérons un modèle GARCH(p,q) conditionnellement gaussien:

$$\epsilon_t | \epsilon_{t-1} \sim N(0, h_t)$$

avec

$$h_t = c + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$$

On réécrit à l'aide de l'opérateur retard:

$$B(L)h_t = c + \alpha(L)\epsilon_t^2$$

ou

$$h_t = c^* + \delta(L)\epsilon_t^2$$

où $\delta(L)$ est un polynôme d'ordre infini. Cette expression donne la variance conditionnelle en terme de paramètres et de variables observables. Ainsi $h_t(\theta)$ dépend de toutes les valeurs passées. Il faudra donc remplacer $h_t(\theta)$ par une approximation tronquée dans laquelle les valeurs ϵ_t^2 correspondant aux dates négatives sont prises égales à zéro. Appelons cette approximation $\tilde{h}_t(\theta)$.

La log vraisemblance sera:

$$\ln \tilde{L} = \frac{-T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^T \ln \tilde{h}_t(\theta) - \frac{1}{2} \sum_{t=1}^T \frac{\epsilon_t^2}{\tilde{h}_t(\theta)}$$

12.3 Test d'hétéroscédasticité pour $h(\theta)$

LM-test (Engle (1982))

On suppose le modèle ARCH(p) suivant:

$$y_t = m_t(\alpha) + \epsilon_t(\alpha)$$

$$h_t(\alpha) = \beta_0 + \sum_{j=1}^p \beta_j \epsilon_{t-j}^2(\alpha).$$

L'hypothèse nulle homoscedasticité conditionnelle est:

$$H_0 : \beta_j = 0 \quad \forall j = 1, \dots, p.$$

On peut facilement effectuer le test de type LM de la façon suivante: on effectue premièrement les moindres carrés ordinaires sur la première équation pour obtenir les résidus estimés $\hat{\epsilon}_t$. Ensuite, on effectue un test sur les retards des résidus aux carrés dans l'équation des résidus aux carrés. On a alors,

$$\hat{\epsilon}_t^2 = \beta_0 + \beta_1 \hat{\epsilon}_{t-1}^2 + \beta_2 \hat{\epsilon}_{t-2}^2 + \dots + \beta_p \hat{\epsilon}_{t-p}^2 + u_t.$$

On peut montrer que la statistique LM du test est égale à

$$LM_T = TR^2 \xrightarrow{L} \chi^2(p)$$

où R^2 est le coefficient de détermination de la régression des résidus aux carrés sur ces retards.

Q-Statistique des résidus aux carrés

Mcleod et Li (1983) ont proposé la statistique de test de Ljung-Box pour les carrés des résidus comme test d'hétéroscedasticité conditionnelle. Dans ce contexte, la statistique est donc:

$$Q(m) = T(T+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2}{T-i}$$

où T est le nombre d'observations et $\hat{\rho}_i^2$ est l'estimateur de l'autocorrélation d'ordre i des résidus $\hat{\epsilon}_t$ aux carrés. Cette statistique suit de façon asymptotique une loi du chi-deux à $m - p - q$ degrés de liberté. p et q étant l'ordre du processus ARMA caractérisant la moyenne conditionnelle. Ce test est asymptotiquement équivalent au test de type LM proposé par Engle (1982).

12.4 Validation de la représentation à hétéroscedasticité conditionnelle

De façon générale, la représentation retenue a la forme GARCH suivante:

$$y_t = X_t' \beta + \epsilon$$

et

$$h_t = c + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}.$$

On appelle alors les résidus "standardisés":

$$\nu_t = \frac{\epsilon_t}{\sqrt{h_t}}.$$

Si la spécification ARCH ou GARCH est adéquate ces résidus standardisés devraient suivre une loi normale ou une student selon le cas. On peut donc tester la spécification à hétéroscédasticité conditionnelle retenues à l'aide des résidus standardisés. Par exemple, on veut vérifier si ces résidus se comportent selon une loi normale.

13 Modèles à hétéroscédasticité conditionnelle multivariés

Jusqu'à maintenant, on a considéré la volatilité d'un seul actif. De façon générale, on s'intéressera à l'évolution de la volatilité dans le temps pour plusieurs actifs financiers.

13.1 Modèle non contraint: GARCH multivarié

On a le modèle suivant:

$$E(\epsilon_t | \epsilon_{t-1}) = 0$$

o ϵ_t est de dimension n et

$$V(\epsilon_t | \epsilon_{t-1}) = H_t$$

et la matrice H_t contient les éléments suivants:

$$h_{k,l,t} = c_{k,l} + \sum_{i=1}^q \left[\sum_{k',l'} \alpha_{klk'l',i} \epsilon_{k',t-i} \epsilon_{l',t-i} \right] + \sum_{i=1}^p \left[\sum_{k',l'} \beta_{klk'l',i} h_{k'l',t-i} \right]$$

Exemple: H_t est de dimension 2x2 et on considère un processus ARCH(1)

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix}$$

$$h_{11t} = c_{11} + \alpha_{1111}\epsilon_{1,t-1}\epsilon_{1,t-1} + \alpha_{1112}\epsilon_{1,t-1}\epsilon_{2,t-1} + \alpha_{1122}\epsilon_{2,t-1}\epsilon_{2,t-1} + \alpha_{1121}\epsilon_{2,t-1}\epsilon_{1,t-1}$$

$$h_{21t} = c_{21} + \alpha_{2111}\epsilon_{1,t-1}\epsilon_{1,t-1} + \alpha_{2112}\epsilon_{1,t-1}\epsilon_{2,t-1} + \alpha_{2122}\epsilon_{2,t-1}\epsilon_{2,t-1} + \alpha_{2121}\epsilon_{2,t-1}\epsilon_{1,t-1}$$

$$h_{12t} = c_{12} + \alpha_{1211}\epsilon_{1,t-1}\epsilon_{1,t-1} + \alpha_{1212}\epsilon_{1,t-1}\epsilon_{2,t-1} + \alpha_{1222}\epsilon_{2,t-1}\epsilon_{2,t-1} + \alpha_{1221}\epsilon_{2,t-1}\epsilon_{1,t-1}$$

$$h_{22t} = c_{22} + \alpha_{2211}\epsilon_{1,t-1}\epsilon_{1,t-1} + \alpha_{2212}\epsilon_{1,t-1}\epsilon_{2,t-1} + \alpha_{2222}\epsilon_{2,t-1}\epsilon_{2,t-1} + \alpha_{2221}\epsilon_{2,t-1}\epsilon_{1,t-1}$$

La matrice H_t est symétrique. Ceci implique des contraintes entre les coefficients:

$$\begin{cases} c_{kl} = c_{lk} \\ \alpha_{klk'l',i} = \alpha_{lkk'l',i} \quad , \quad \alpha_{klk'l',i} = \alpha_{kll'k',i} \\ \beta_{klk'l',i} = \beta_{lkk'l',i} \quad , \quad \beta_{klk'l',i} = \beta_{kll'k',i} \end{cases}$$

Sous ces restrictions, on a

$$\frac{n(n+1)}{2} + (p+q) \left[\frac{n(n+1)}{2} \right]^2$$

paramètres différents. Pour un ARCH(1), on aura $1 + [n(n+1)]/2$ paramètres pour chaque équation. Pour un ARCH(1) ceci donne:

n	1	2	3	4
$1 + \frac{n(n+1)}{2}$	2	4	7	11
Nombre total de paramètres	2	12	42	110

On peut réécrire la formulation GARCH générale de façon plus compact:

$$vech(H_t) = vech(C) + \sum_{i=1}^q A_i vech(\epsilon_{t-i}\epsilon'_{t-i}) + \sum_{i=1}^p B_i vech(H_{t-i})$$

o $vech()$ indique la "vectorisation" de la partie triangulaire par le bas. A_i et B_i sont des matrices contenant les paramètres correspondants.

On peut également représenter sous la forme matricielle:

$$H_t = C + \sum_{i=1}^q (I_n \otimes \epsilon'_{t-i}) A_i (I_n \otimes \epsilon_{t-i}) + \sum_{i=1}^p E \left[(I_n \otimes \epsilon'_{t-i}) B_i (I_n \otimes \epsilon_{t-i}) | \epsilon_{t-i-1} \right]$$

o \otimes représente le produit Kronecker.

Exemple: ARCH(1), n=2

On a

$$\begin{pmatrix} h_{11t} \\ h_{12t} \\ h_{21t} \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \end{pmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1t-1}^2 \\ \epsilon_{1t-1}\epsilon_{2t-1} \\ \epsilon_{2t-1}^2 \end{bmatrix}$$

On a 4 paramètres par équation et 12 paramètres à estimer. On doit donc contraindre le modèle dans la plupart des cas pour l'estimer.

13.2 Modèles contraints

13.2.1 Modèle diagonal:

Les matrices A_i et B_i sont diagonales de telle sorte que h_{ki} dépend seulement de lui-même.

Exemple: ARCH(1)

$$h_{11,t} = c_{11} + b_{11}h_{11,t-1}$$

$$h_{12,t} = c_{12} + b_{12}h_{12,t-1}$$

$$h_{22,t} = c_{12} + b_{22}h_{22,t-1}$$

Problème:

- Un portefeuille d'un certain poids α_1 et α_2 des deux actifs ne respecte pas cette représentation. La contrainte n'est pas invariante à la composition du portefeuille.
- Ne tient pas compte de la substitution des risques entre les actifs
- Ne respecte pas les conditions de positivité.

Exemple: On construit un portefeuille d'un poids α_1 du premier actif et d'un poids α_2 du deuxième actif. La variance conditionnelle du portefeuille est donc:

$$\begin{aligned} h_{\alpha,t} &= \alpha_1^2 h_{11,t} + \alpha_2^2 h_{22,t} + 2\alpha_1\alpha_2 h_{12,t} \\ &= \alpha_1^2 c_{11} + \alpha_2^2 c_{22} + 2\alpha_1\alpha_2 c_{12} \\ &\quad + \alpha_1^2 b_{11} h_{11,t-1} + \alpha_2^2 b_{22} h_{22,t-1} + 2\alpha_1\alpha_2 b_{12} h_{12,t-1} \end{aligned}$$

qui ne correspond pas à une forme diagonale:

$$h_{\alpha,t} = c_{\alpha} + b_{\alpha} h_{\alpha,t-1}$$

13.2.2 Modèle avec des corrélations conditionnelles constantes (Bollerslev 1987)

Les variances et les covariances conditionnelles varient dans le temps. Par contre, on impose des corrélations conditionnelles invariantes dans le temps. On a alors:

$$\epsilon_t | \Omega_{t-1} \sim N(0, H_t)$$

Les termes d'erreurs suivent une normale conditionnelle par hypothèse. La matrice de variance-covariance est définie comme étant:

$$H_t = D_t R D_t$$

où D_t est une matrice diagonale contenant le processus GARCH du rendement i à la position (i, i) et R est la matrice des coefficients de corrélation.

Exemple: On suppose un vecteur contenant les rendements de deux titres financiers. Ainsi:

$$r_t = \begin{pmatrix} r_{1t} \\ r_{2t} \end{pmatrix}$$

La matrice de variance-covariance conditionnelle est alors:

$$H_t = \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} h_{11,t}^{1/2} & 0 \\ 0 & h_{22,t}^{1/2} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} h_{12,t}^{1/2} & 0 \\ 0 & h_{22,t}^{1/2} \end{bmatrix}$$

Si l'hétéroscédasticité conditionnelle du rendement de chaque titre est modélisée comme un GARCH(1,1), alors:

$$h_{11,t} = c_{11} + a_{11} r_{1,t-1}^2 + b_{11} h_{11,t-1}$$

$$h_{22,t} = c_{22} + a_{22} r_{2,t-1}^2 + b_{22} h_{22,t-1}$$

$$h_{12,t} = \rho_{12} h_{11,t}^{1/2} h_{22,t}^{1/2}$$

Pour cette spécification, on a $[n(n+5)]/2$ paramètres à estimer.

Problème: Les corrélations constantes sont souvent rejetées.

13.2.3 Modèle avec des corrélations conditionnelles dynamiques (Engle-Sheppard 2001)

On a donc le vecteur

$$r_t | \Omega_{t-1} \sim N(0, H_t)$$

Dans la modélisation introduite par Engle-Sheppard, les corrélations changent dans le temps. Ainsi:

$$H_t = D_t R_t D_t$$

où D_t est la matrice diagonale contenant le processus GARCH $h_{ii,t}^{1/2}$ du rendement i à la position (1, 1) et R_t est la matrice des coefficients de corrélation.

La stratégie des deux auteurs consiste à spécifier une structure de type GARCH pour modéliser les corrélations dynamiques. On pose

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}$$

et

$$Q_t = \mu + \sum_{m=1}^M \alpha_m (r_{t-m} r'_{t-m}) + \sum_{n=1}^N \beta_n Q_{t-n}$$

et

$$Q_t^{*-1} = \begin{bmatrix} \sqrt{q_{11,t}} & & & 0 \\ & \sqrt{q_{22,t}} & & \\ & & \ddots & \\ 0 & & & \sqrt{q_{NN,t}} \end{bmatrix}$$

La matrice Q_t^* contient les éléments diagonaux de la matrice Q_t élevés à la puissance 1/2.

Ainsi, on aura la matrice des corrélations dynamiques R_t où:

$$R_t = \begin{bmatrix} \rho_{11,t} & \rho_{12,t} & \cdots & \rho_{1N,t} \\ \rho_{12,t} & \rho_{22,t} & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_{1N,t} & \rho_{2N,t} & \cdots & \rho_{NN,t} \end{bmatrix}$$

avec

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}$$

13.2.4 Modèle à facteurs

Le vecteur ϵ_t dépend d'un facteur F_t (ou variables d'état) et d'un bruit tel que:

$$\begin{array}{ccccccc} \epsilon_t & = & \lambda & F_t & + & e_t & \\ (n \times 1) & & (n \times 1) & (1 \times 1) & & (n \times 1) & \end{array}$$

et λ mesure la sensibilité au facteur F_t (ou β), $E(F_t e_t) = 0$, et

$$F_t | F_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i F_{t-i}^2$$

$$E(e_t | e_{t-1}) = 0 \quad , \quad V(e_t | e_{t-1}) = \Omega$$

Ainsi

$$V(\epsilon_t | \epsilon_{t-1}) = H_t = \sigma_t^2 \lambda \lambda' + \Omega$$

On peut généraliser à plusieurs facteurs.

Ce type de représentation est difficile à estimer étant donné que le ou les facteurs sont inobservables. On va utiliser le filtre de Kalman. La distribution de ϵ_t contient un grand nombre d'intégrales à résoudre. On peut facilement écrire la densité de ϵ_t et F_t . La densité des observations ϵ_t cependant comporte des intégrales multiples:

$$l(\epsilon_T, \epsilon_{T-1}, \dots, \epsilon_1 | F_0) = \int \dots \int l(\epsilon_T, F_T, \epsilon_{T-1}, F_{T-1}, \dots, \epsilon_1, F_1 | F_0) dF_T \dots dF_1$$

On a un nombre d'intégrales égal au nombre d'observations.

14 La méthode des moments généralisés

La méthode des moments généralisés (GMM) consiste à estimer des paramètres d'intérêt à l'aide de conditions de moments appelées également conditions d'orthogonalité. L'estimateur est obtenu à l'aide des conditions de moments empiriques correspondantes aux conditions de moments théoriques.

Prenons, par exemple, une variable y_t où $t = 1, \dots, T$. Le premier moment est donnée par:

$$E(y_t) = \mu.$$

Le scalaire μ est donc la moyenne non conditionnelle de la loi des y_t . On obtient un estimateur de μ par l'équivalent empirique de la condition de moment théorique. Ainsi, on peut réécrire le moment théorique sous la forme:

$$E(y_t - \mu) = 0$$

et l'estimateur de la méthode des moments est:

$$\frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu}) = 0 \Rightarrow \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t.$$

L'estimateur $\hat{\mu}$ satisfait la condition de moment empirique.

De la même façon, la variance théorique de y_t est donnée par

$$E(y_t - \mu)^2 = \sigma^2.$$

Cette condition de moment peut naturellement être réécrite comme étant:

$$E((y_t - \mu)^2 - \sigma^2) = 0.$$

L'estimateur de la méthode des moments est obtenu en égalisant à zéro la condition de moment empirique correspondante,

$$\frac{1}{T} \sum_{t=1}^T [(y_t - \hat{\mu})^2 - \hat{\sigma}^2] = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{\mu})^2$$

Examinons maintenant le modèle linéaire suivant:

$$y_t = \beta' x_t + \varepsilon_t$$

où β est un vecteur de paramètres de dimension $K \times 1$, x_t est un vecteur de variables explicatives aléatoires de dimension $K \times 1$, $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_{t-j}) = \sigma^2$ pour $j = 0$ et 0 autrement (bruit blanc faible).

Pour que l'estimateur des moindres carrés ordinaires soit sans biais, il doit respecter la condition suivante:

$$E(\varepsilon_t/x_t) = 0.$$

On peut obtenir des conditions de moments non conditionnelles en prémultipliant par le vecteur x_t et en prenant l'espérance. Ainsi, qui correspond aux conditions de moments suivantes:

$$E x_t E(\varepsilon_t/x_t) = E(x_t \varepsilon_t) = 0.$$

On a donc ici K conditions de moments.

L'estimateur de la méthode des moments généralisés est celui qui égalise les conditions de moments empiriques à zéro,

$$\frac{1}{T} \sum_{t=1}^T x_t \hat{\varepsilon}_t = \frac{1}{T} \sum_{t=1}^T x_t (y_t - \hat{\beta}' x_t) = 0.$$

On obtient ainsi l'estimateur de la méthodes des moments généralisés

$$\hat{\beta} = (X'X)^{-1} X'Y$$

écrit de façon matricielle où X est une matrice $T \times K$ contenant les variables explicatives et Y est le vecteur contenant les observations de la variables dépendantes. Cet estimateur correspond à l'estimateur des moindres carrés ordinaires.

Si $E(x_t \varepsilon_t) \neq 0$, on peut utiliser un vecteur de variables instrumentales z_t de dimension égal ou plus grande que K , tel que $E(x_t z_t') \neq 0$ et $E(z_t \varepsilon_t) = 0$.

L'estimateur à variables instrumentales sera sans biais s'il respecte les conditions d'orthogonalité suivantes:

$$E(z_t \varepsilon_t) = 0.$$

Un vecteur de variables instrumentales doit donc être orthogonale au terme d'erreur ε_t . De plus, ce vecteur doit contenir le plus possible la même information que le vecteur de variables explicatives x_t . Le vecteur de variables instruments doit donc être le plus fortement corrélé possible avec le vecteur de variables explicatives.

L'estimateur est donné par le vecteur de paramètres qui égalise les conditions de moments empiriques à zéro. Ainsi,

$$\frac{1}{T} \sum_{t=1}^T z_t \hat{\varepsilon}_t = \frac{1}{T} \sum_{t=1}^T z_t (y_t - \hat{\beta}' x_t) = 0.$$

On peut réécrire les conditions de moments de façon matricielle:

$$\frac{1}{T} Z' \hat{\varepsilon} = \frac{1}{T} Z' (Y - X \hat{\beta}) = 0$$

où Z est la matrice contenant les variables instrumentales de dimension $T \times Q$.

Si le vecteur de variables instrumentales z_t a la même dimension que le vecteur x_t , alors l'estimateur de la méthode des moments généralisés (estimateur à variables instrumentales) est donné par

$$\hat{\beta}_{V.I} = (Z' X)^{-1} Z' Y.$$

Cet estimateur est également connu comme étant l'estimateur des moindres carrés indirects.

Lorsque le nombre d'instruments est plus grand que le nombre de paramètres d'intérêt, on peut utiliser une combinaison des instruments telle que

$$E(AZ' \varepsilon) = 0$$

où

$$E(Z^{*\prime} \varepsilon)$$

et $Z^* = ZA'$, A est de dimension $K \times Q$ et de rang égal à K .

L'estimateur est alors donnée par

$$\frac{1}{T} AZ' \hat{\varepsilon} = \frac{1}{T} AZ'(Y - X\hat{\beta}) = 0,$$

alors

$$\hat{\beta}_{V.I} = (AZ'X)^{-1} AZ'Y.$$

On a autant d'estimateur qu'il existe de matrice A . Il y a donc une infinité d'estimateur. On peut montrer que l'estimateur optimal est obtenu avec

$$A = X'Z(Z'Z)^{-1},$$

alors l'estimateur de la méthode des moments généralisés est:

$$\hat{\beta}_{V.I} = [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'Y,$$

qui correspond à l'estimateur des doubles moindres carrés (Two stage least squares).

Soit un modèle général non linéaire:

$$Y = h(X, \beta) + \varepsilon$$

où $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)'$ et $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon^1) = \Omega$ où Ω est une matrice définie positive.

On peut donc avoir de l'autocorrélation et/ou de l'hétéroscédasticité. Il est de plus possible que $E(X'\varepsilon) \neq 0$. Cependant, il existe un vecteur de variables instrumentales z_t de dimension q tel que $E(Z'\varepsilon) = 0$.

On peut donc utiliser les conditions de moments empiriques pour obtenir un estimateur de β , ainsi

$$\frac{1}{T} \sum_{t=1}^T z_t \hat{\varepsilon}_t = \frac{1}{T} \sum_{t=1}^T z_t (y_t - h(x_t \hat{\beta})) = 0.$$

Si $Q = K$, c.a.d., si la dimension du vecteur de variables instrumentales est égale à la dimension de β , alors on aura l'égalité à zéro et l'estimateur est unique. Si $Q > K$, on aura besoin d'une mesure de distance par rapport à zéro. On prendra une mesure quadratique. L'estimateur de la méthode des moments généralisés sera donné comme la solution du problème suivant:

$$\hat{\beta} = \arg \min \left(\frac{1}{T} \sum_{t=1}^T z_t \varepsilon_t \right)' W_T \left(\frac{1}{T} \sum_{t=1}^T z_t \varepsilon_t \right)$$

où W_T est une matrice définie positive qui peut dépendre des observations.

Il existe autant d'estimateur qu'il existe de matrice de pondération W . Hansen (1982) a montré que l'estimateur optimal est obtenu pour $W = S^{-1}$ où

$$S = \lim_{T \rightarrow \infty} Var \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t \varepsilon_t \right).$$

Pour le cas avec hétéroscédasticité seulement, on a que

$$S = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{i=1}^T \sigma_{ii} z_i z_i'$$

White (1982) a proposé l'estimateur convergent suivant:

$$S_T = \frac{1}{T} \sum_{i=1}^T \hat{\varepsilon}_i^2 z_i z_i'$$

et il a montré que $S_T \xrightarrow{P} S$.

Lorsqu'il y a autocorrélation des erreurs, Newey-West (1987) ont démontré que l'estimateur suivant était convergent

$$S_T = \frac{1}{T} \sum_{l=0}^r \omega(l) \sum_{i=l}^T \hat{\varepsilon}_i \hat{\varepsilon}_{i-l} (z_i z_{i-l}' + z_{i-l} z_i')$$

où $\omega(l) = 1 - \frac{l}{r+1}$ (fenêtre de Bartlett). On peut choisir r de façon endogène (Newey-West (1994)).

On peut maintenant examiner la méthode des moments généralisés dans un contexte général. De façon générale, on a les conditions de moments théoriques suivantes:

$$E[f(x_t, \theta_0)] = 0$$

où θ_0 est un vecteur de paramètres d'intérêt de dimension p . x_t est un vecteur de séries observées stationnaires et f est une fonction continue de dimension q où $q \geq p$.

$\hat{\theta}_T$ sera choisi tel que les conditions de moments empiriques sont le plus proche de zéro, c.a.d.

$$\frac{1}{T} \sum_{t=1}^T f(x_t, \theta).$$

Si $q = p$, alors on aura l'égalité à zéro, l'estimateur sera unique. Si $q > p$, on minimise une certaine mesure de distance par rapport à zéro.

Définition 15 *Étant donné une matrice symétrique définie positive W_T de dimension $q \times q$ dépendant éventuellement des observations, on appelle l'estimateur de la méthode des moments généralisés associé à W_T , une solution $\hat{\theta}_T(W_T)$ du problème*

$$\min_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \theta) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \theta) \right).$$

La matrice W_T mesure l'importance relative donnée aux conditions de moments.

On peut montrer que:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{loi} N \left(0, (D'WD)^{-1} D'WSD(D'WD)^{-1} \right)$$

où

$$\frac{1}{T} \sum_{t=1}^T \frac{\partial f(x_t, \hat{\theta}_T)}{\partial \theta'} \xrightarrow{p} D = E \left[\frac{\partial f(x_t, \theta_0)}{\partial \theta'} \right]$$

et $W_T \xrightarrow{p} W$, et

$$S = \lim_{T \rightarrow \infty} Var \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T f(x_t, \theta) \right).$$

Il existe autant d'estimateur qu'il existe de matrice W_T . Un estimateur optimal est obtenu avec la matrice $W_T = S_T^{-1}$. On a alors

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{loi} N \left(0, (D'S^{-1}D)^{-1} \right).$$

Pour ce choix de W_T , la matrice de variance-covariance est la plus petite possible.

Lorsque $q > p$, on obtient un test de spécification basé sur les conditions de suridentification. Ce test est donné par la statistique suivante:

$$J_T = T \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \hat{\theta}_T) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \hat{\theta}) \right)$$

et suit une loi du $\chi^2(q - p)$.

La méthode des moments généralisés est une méthode en deux étapes.

1. À la première étape, on estime θ pour une matrice W quelconque. Habituellement, on utilise la matrice identité. Puisque l'estimateur obtenu est convergent, on peut l'utiliser pour construire un estimateur convergent de S .
2. Ayant obtenu un estimateur convergent de S , on réestime θ avec $W_T = S_T^{-1}$. Ainsi, on obtient un estimateur optimal.

Exemple:

Supposons une économie composée d'un agent représentatif qui maximise une fonction d'utilité intertemporelle sous une contrainte budgétaire.

Le problème de maximisation de l'agent est le suivant:

$$\max_{C_{t+j}, A_{t+j+1}} E_t \left[\sum_{j=0}^{\infty} \beta^j U(C_{t+j}) \right]$$

où β est le taux d'escompte, C_{t+j} la consommation au temps $t + j$ et $U(C_{t+j})$ est la fonction d'utilité dépendant de la consommation en $t + j$. La contrainte budgétaire est donnée par l'expression suivante:

$$C_{t+j} + A_{t+j+1} = (1 + R_{t+j-1,t+j})A_{t+j} + Y_{t+j}$$

où A_{t+j} est la quantité détenue d'un actif financier sans risque venant à échéance en $t + j$, $R_{t+j-1,t+j}$ est le rendement de l'actif financier sans risque détenu de la période $t + j - 1$ à la période $t + j$ et Y_{t+j} représente le revenu exogène à la période $t + j$.

On considère la forme fonctionnelle suivante pour la fonction d'utilité:

$$U(C_t) = \frac{C_t^{1-\sigma} - 1}{1-\sigma}$$

où σ est le paramètre d'aversion au risque.

Par les conditions du premier ordre, on peut obtenir l'équation d'Euler suivante:

$$E_t \left[\beta (1 + R_{t,t+1}) C_{t+1}^{-\sigma} \right] = C_t^{-\sigma}.$$

On peut réécrire cette équation sous la forme suivante:

$$E_t \left[\beta (1 + R_{t,t+1}) \left(\frac{C_t}{C_{t+1}} \right)^\sigma - 1 \right] = 0.$$

On a donc une espérance conditionnelle par rapport à l'ensemble d'information en t . On cherche à estimer les paramètres structurelles $\theta = (\beta, \sigma)'$. On peut obtenir une espérance non conditionnelle en projetant l'équation d'Euler sur des éléments de l'ensemble d'information en t . Prenons par exemple les instruments suivants $z_t = (C_t/C_{t-1}, C_{t-1}/C_{t-2}, R_{t-1,t}, R_{t-2,t-1})'$. On aura donc comme moments théoriques:

$$E z_t E_t (\varepsilon_{t+1}/z_t) = E (z_t \varepsilon_{t+1}) = 0$$

où

$$\varepsilon_{t+1} = \beta (1 + R_{t,t+1}) \left(\frac{C_t}{C_{t+1}} \right)^\sigma - 1.$$

On peut donc estimer les paramètres structurelles par GMM avec l'équivalent empirique des moments théoriques. On pourra également effectuer un test de spécification puisque nous avons quatre moments et deux paramètres à estimer.

Problèmes:

- Instruments faibles: le choix des instruments est très important pour les propriétés de l'estimateur GMM en petit et en grand échantillons. Les variables instrumentales doivent être corrélées avec les variables dépendantes sinon le comportement en petit et grand échantillons est très mauvais (voir Staiger et Stock (1997)). Ces deux auteurs proposent d'effectuer une régression des variables dépendantes sur les instruments et de calculer une statistique F pour la significativité des instruments dans cette régression.

- Biais: le biais de l'estimateur GMM augmente avec le nombre de moments. Newey et Smith (2004) donnent une justification asymptotique à ce comportement.
- L'estimateur n'est pas invariant à la normalisation choisie.

Hansen, Heaton et Ogaki (1996) ont proposé un estimateur GMM invariant à la normalisation choisie. Cet estimateur est appelé "Continuous Updated Estimator" (CUE). Il prend la forme suivante:

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \theta) \right)' W_T(\theta) \left(\frac{1}{T} \sum_{t=1}^T f(x_t, \theta) \right).$$

La matrice de poids est donc estimée conjointement. La procédure d'estimation s'effectue en une étape. Newey et Smith (2004) montrent également que le biais de cet estimateur n'augmente pas avec le nombre de conditions de moments. Cet estimateur est cependant plus fragile numériquement.