

BIBLIOGRAPHIE :

- Lardic S. et Mignon V. (2002), *Econométrie des Séries Temporelles Macroéconomiques et Financières*, Economica.
- Bourbonnais R. (2000), *Econométrie*, DUNOD.

* * *

I/ ETUDE UNIVARIEE : MODELISATION D'UNE SERIE TEMPORELLE

I.1 / Fonctions d'autocorrélation

Définition 1 : La fonction d'autocorrélation est la fonction notée ρ_k qui mesure la corrélation de la série avec elle-même décalée de k périodes :

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\sigma_{y_t} \sigma_{y_{t-k}}} = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^n (y_t - \bar{y})^2} \sqrt{\sum_{t=k+1}^n (y_{t-k} - \bar{y})^2}}$$

Nous pouvons en déduire que :

$$\rho_0 = 1 \text{ et } -1 \leq \rho_k \leq 1 \quad \forall k$$

Le graphe de la fonction d'autocorrélation est appelé corrélogramme.

Définition 2 : La fonction d'autocorrélation partielle mesure la corrélation entre y_t et y_{t-k} , l'influence des autres variables ($y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$) ayant été retirée.

I.2 / Séries stationnaires

Définition 3 : Une série $\{y_t\}$ pour $t = 1, \dots, T$ est dite stationnaire si :

- (i) $E(y_t) = \mu \quad \forall t$ (constante, ne dépend pas de t) ;
- (ii) $\text{Var}(y_t) = \sigma_y^2 < \infty \quad \forall t$ (constante, ne dépend pas de t) ;
- (iii) $\text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$ (ne dépend pas de t).

La série $\{\varepsilon_t\}$ dont $E(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2$, $\text{Cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ est donc une série stationnaire. Elle est appelée aussi bruit blanc (remarque : un bruit blanc n'est pas nécessairement gaussien).

Une série stationnaire ne doit comporter ni tendance et ni saisonnalité.

Définition 4 : Séries non stationnaires : processus TS et DS

a/ Processus TS :

Le processus TS (Trend Stationary) s'écrit :

$$y_t = \alpha + \beta t + \varepsilon_t$$

où ε_t représente l'erreur du modèle à la date t.

Il présente une non stationnarité de nature déterministe.

Le processus TS est non stationnaire car $E(y_t) = \alpha + \beta t$ dépend du temps t.

Le processus y_t peut être stationnarisé en retranchant à y_t la valeur estimée $\hat{\alpha} + \hat{\beta} t$ par la méthode des Moindres Carrés Ordinaires (voir document économétrie.pdf pour MCO).

b/ Processus DS :

Le processus DS (Differency Stationary) avec dérive ($\beta \neq 0$) s'exprime comme suit :

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

Le processus DS avec dérive est appelé aussi marche au hasard (ou marche aléatoire) avec dérive.

Il présente une non stationnarité de nature stochastique.

Par récurrence, on obtient (dans le cas avec dérive) :

$$y_1 = y_0 + \beta + \varepsilon_1$$

$$y_2 = y_1 + \beta + \varepsilon_2 = y_0 + \beta + \varepsilon_1 + \beta + \varepsilon_2 = y_0 + 2\beta + \varepsilon_1 + \varepsilon_2$$

...

$$y_t = y_0 + \beta t + \sum_{i=1}^t \varepsilon_i \quad \text{où } \varepsilon_i \sim \text{iid}(0, \sigma_\varepsilon^2), \varepsilon_i \text{ est identiquement et indépendamment distribuée.}$$

Le processus DS avec dérive est non stationnaire car on a $E(y_t) = y_0 + \beta t$ qui dépend du temps t. Plus $t \rightarrow \infty$ et plus $E(y_t) \rightarrow \infty$.

Le processus DS (Differency Stationary) sans dérive ($\beta = 0$) s'écrit :

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

Le processus DS sans dérive est appelé aussi marche au hasard (ou marche aléatoire).

Par récurrence, on obtient (dans le cas sans dérive) :

$$y_1 = y_0 + \varepsilon_1$$

$$y_2 = y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2$$

...

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_i \text{ où } \varepsilon_i \sim \text{iid}(0, \sigma_\varepsilon^2)$$

Le processus DS sans dérive est non stationnaire car on a :

$$\text{Var}(y_t) = \text{Var}\left(\sum_{i=1}^t \varepsilon_i\right) = \sum_{i=1}^t \text{Var}(\varepsilon_i) = \sum_{i=1}^t \sigma_\varepsilon^2 = t \sigma_\varepsilon^2.$$

On constate que la variance du processus DS sans dérive dépend du temps t .

Plus $t \rightarrow \infty$ et plus $\text{Var}(y_t) \rightarrow \infty$.

Pour stationnariser le processus DS (avec ou sans dérive), il suffit de le passer en différence première :

$$y_t - y_{t-1} = \beta + \varepsilon_t \text{ (cas avec dérive)}$$

ou

$$y_t - y_{t-1} = \varepsilon_t \text{ (cas sans dérive).}$$

Définition 5 : Une série est dite intégrée d'ordre d (notée $y_t \sim I(d)$) s'il convient de la différencier d fois afin de la stationnariser. La série stationnarisée est alors intégrée d'ordre 0 et est notée $y_t \sim I(0)$.

I.3 / Tests de stationnarité (ou tests de racine unitaire)

Il existe plusieurs tests de racine unitaire : tests de Dickey-Fuller simple et Dickey-Fuller Augmenté, test de Phillips et Perron, test de Kwiatkowski, Phillips, Schmidt et Shin (test de KPSS). Nous n'étudierons ici que les tests de Dickey-Fuller et de Phillips-Perron.

a/ Test de Dickey-Fuller simple :

Le test de Dickey-Fuller permet de savoir si une série est stationnaire ou non et permet aussi de déterminer la bonne manière de stationnariser la série.

Les hypothèses du test sont les suivantes :

$$\left\{ \begin{array}{l} H_0 : \text{processus non stationnaire, il correspond à une de ces formes de non stationnarité :} \\ [1] \quad y_t = \phi_1 y_{t-1} + \varepsilon_t \\ [2] \quad y_t = \phi_1 y_{t-1} + c + \varepsilon_t \\ [3] \quad y_t = \phi_1 y_{t-1} + bt + c + \varepsilon_t \\ \\ \text{où } \phi_1 = 1 \text{ et } \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2) \\ \\ H_1 : |\phi_1| < 1. \end{array} \right.$$

On peut écrire aussi les hypothèses sous la forme suivante :

$$\left\{ \begin{array}{l} H_0 : \text{processus non stationnaire, il correspond à une de ces formes de non stationnarité :} \\ [1] \quad \Delta y_t = (\phi_1 - 1) y_{t-1} + \varepsilon_t \\ [2] \quad \Delta y_t = (\phi_1 - 1) y_{t-1} + c + \varepsilon_t \\ [3] \quad \Delta y_t = (\phi_1 - 1) y_{t-1} + bt + c + \varepsilon_t \\ \\ \text{où } (\phi_1 - 1) = 0 \text{ et } \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2) \\ \\ H_1 : \phi_1 < 1. \end{array} \right.$$

Sous H_0 vraie, la statistique de test pour l'estimateur de ϕ_1 est donnée par :

$$t_{\hat{\phi}_1} = \frac{\hat{\phi}_1 - 1}{\hat{\sigma}_{\hat{\phi}_1}}$$

On commence par étudier le modèle général [3]. On regarde si b est significativement différent de 0 ou non. Si b est significativement non différent de 0, on passe à l'étude du modèle [2] et on cherche à savoir si c est significativement différent de 0 ou pas. Si c est significativement non différent de 0, on étudie le modèle [1].

ATTENTION : Sous H_0 vraie, les t de Student de la constante et de la tendance sont à comparer avec les valeurs de la table de Dickey-Fuller (Pour une taille d'échantillon supérieure à 500 observations, les valeurs critiques sont : 2.78 à 5% pour la tendance du modèle [3], 2.52 pour la constante du modèle [2] et -1.95 pour le paramètre ϕ_1) car sous H_0 vraie le processus étudié est non stationnaire ($y_t \sim I(1)$) et l'estimateur de ϕ_1 ne suit pas la loi normale.

Les règles de décision sont les suivantes :

- Si $t > t_{DF}$ où t_{DF} désigne la valeur critique donnée par table de DF
 \Rightarrow on accepte H_1 : le coefficient de la variable explicative est significativement différent de 0.

Si on a b significativement différent de 0 pour le modèle [3], le test s'arrête ici, on n'étudie pas les autres modèles. De même que si on arrive au modèle [2] et que l'on a la constante qui est significativement différente de 0, le test s'arrête au modèle [2].

- Si $|t_{\theta_1}| > t_{DF} \Rightarrow$ On accepte H_0 : la série est non stationnaire

(ATTENTION : il faut observer ici que pour $|t_{\theta_1}| > t_{DF}$, on n'a pas H_1 ! La règle de décision est ici inversée).

b/ Test de Dickey-Fuller Augmenté :

Dans le test de Dickey-Fuller que nous venons d'étudier, le processus ε_t est par hypothèse un bruit blanc. Or il n'y a aucune raison pour que, a priori, l'erreur soit non corrélée. Le test de Dickey-Fuller Augmenté ne suppose pas que ε_t est un bruit blanc.

Les hypothèses du test de Dickey-Fuller Augmenté se définissent de la façon suivante :

$$\left\{ \begin{array}{l} H_0 : \text{processus non stationnaire, il correspond à une de ces formes de non stationnarité :} \\ [1] \quad \Delta y_t = \rho y_{t-1} - \sum_{k=2}^p \gamma_k \Delta y_{t-k+1} + \eta_t \\ [2] \quad \Delta y_t = \rho y_{t-1} - \sum_{k=2}^p \gamma_k \Delta y_{t-k+1} + c + \eta_t \\ [3] \quad \Delta y_t = \rho y_{t-1} - \sum_{k=2}^p \gamma_k \Delta y_{t-k+1} + bt + c + \eta_t \\ \\ \text{où } \rho = 0, \theta_1 = 1 \text{ et } \eta_t \sim \text{iid}(0, \sigma_\eta^2) \\ H_1 : \theta_1 < 1. \end{array} \right.$$

Démonstration de l'écriture des modèles [1], [2] et [3] :

On a $y_t = \theta_1 y_{t-1} + \varepsilon_t$ où $\varepsilon_t \sim \text{AR}(p-1)$, ε_t n'est pas un bruit blanc :

$$\varepsilon_t = \sum_{i=1}^{p-1} \theta_i \varepsilon_{t-i} + \eta_t \text{ où } \eta_t \sim \text{iid}(0, \sigma_\eta^2)$$

On peut écrire ces équations à l'aide de l'opérateur de décalage B tel que $By_t = y_{t-1}$ et $B^p y_t = y_{t-p}$. On obtient alors :

$$(1 - \theta_1 B)y_t = \varepsilon_t$$

$$\text{et } \varepsilon_t = \theta_1 B \varepsilon_t + \theta_2 B^2 \varepsilon_t + \dots + \theta_{p-1} B^{p-1} \varepsilon_t + \eta_t \Leftrightarrow \left(1 - \sum_{i=1}^{p-1} \theta_i B^i\right) \varepsilon_t = \eta_t$$

$$\text{On peut alors écrire } \left(1 - \sum_{i=1}^{p-1} \theta_i B^i\right)(1 - \theta_1 B)y_t = \eta_t$$

$$\left(1 - \sum_{i=1}^{p-1} \theta_i B^i\right)(1 - \theta_1 B)y_t = \eta_t$$

$$\Leftrightarrow \left(1 - \sum_{i=1}^{p-1} \theta_i B^i\right)(y_t - \theta_1 y_{t-1}) = \eta_t$$

$$\Leftrightarrow (y_t - \theta_1 y_{t-1}) - \sum_{i=1}^{p-1} \theta_i B^i (y_t - \theta_1 y_{t-1}) = \eta_t$$

$$\Leftrightarrow (y_t - \theta_1 y_{t-1}) - \sum_{i=1}^{p-1} \theta_i y_{t-i} + \theta_1 \sum_{i=1}^{p-1} \theta_i y_{t-1-i} = \eta_t$$

$$\Leftrightarrow y_t - \theta_1 y_{t-1} - \theta_1 y_{t-1} - \theta_2 y_{t-2} - \dots - \theta_{p-1} y_{t-(p-1)} + \theta_1 \theta_1 y_{t-2} + \theta_1 \theta_1 y_{t-3} + \dots + \theta_1 \theta_{p-1} y_{t-p} = \eta_t$$

$$\Leftrightarrow y_t = (\theta_1 + \theta_1)y_{t-1} + (\theta_2 - \theta_1 \theta_1)y_{t-2} + \dots + (\theta_{p-1} - \theta_1 \theta_{p-2})y_{t-p+1} - \theta_1 \theta_{p-1} y_{t-p} + \eta_t$$

$$\Leftrightarrow \Delta y_t = [(\theta_1 - 1)(1 - \theta_1 - \theta_2 - \dots - \theta_{p-1})]y_{t-1} - \sum_{k=1}^{p-1} \gamma_k \Delta y_{t-k} + \eta_t$$

car on a :

$$y_t = (\theta_1 + \theta_1)y_{t-1} + (\theta_2 - \theta_1 \theta_1)y_{t-2} + \dots + (\theta_{p-1} - \theta_1 \theta_{p-2})y_{t-p+1} - \theta_1 \theta_{p-1} y_{t-p} + \eta_t$$

$$\Leftrightarrow y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_{p-1} y_{t-p+1} + \alpha_p y_{t-p} + \eta_t.$$

Or on constate que pour un modèle AR(1), on a l'écriture suivante :

$$y_t = \alpha_1 y_{t-1} + \eta_t$$

$$\Leftrightarrow y_t - y_{t-1} = \alpha_1 y_{t-1} - y_{t-1} + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\alpha_1 - 1) y_{t-1} + \eta_t.$$

Pour un modèle AR(2), il vient :

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \eta_t$$

$$\Leftrightarrow y_t - y_{t-1} = -y_{t-1} + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_2 y_{t-1} - \alpha_2 y_{t-1} + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\alpha_1 + \alpha_2 - 1) y_{t-1} + \alpha_2 (y_{t-2} - y_{t-1}) + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\alpha_1 + \alpha_2 - 1) y_{t-1} - \alpha_2 (y_{t-1} - y_{t-2}) + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\alpha_1 + \alpha_2 - 1) y_{t-1} - \alpha_2 \Delta y_{t-1} + \eta_t$$

Pour un modèle AR(3), on obtient :

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\alpha_1 + \alpha_2 + \alpha_3 - 1) y_{t-1} - (\alpha_2 + \alpha_3) \Delta y_{t-1} - \alpha_3 \Delta y_{t-2} + \eta_t$$

On constate donc que pour un modèle AR(p) :

$$\Delta y_t = \left(\sum_{i=1}^p \alpha_i - 1 \right) y_{t-1} - \sum_{k=1}^{p-1} \left(\sum_{i=k+1}^p \alpha_i \right) \Delta y_{t-k} + \eta_t$$

Comme on avait supposé que :

$$\alpha_1 = \theta_1 + \theta_1, \quad \alpha_2 = \theta_2 - \theta_1 \theta_1, \quad \dots, \quad \alpha_p = -\theta_1 \theta_{p-1}$$

On obtient alors :

$$\Delta y_t = (\theta_1 + \theta_1 + \theta_2 - \theta_1 \theta_1 + \dots + \theta_{p-1} - \theta_1 \theta_{p-2} - \theta_1 \theta_{p-1} - 1) y_{t-1} - \sum_{k=1}^{p-1} \left(\sum_{i=k+1}^p \alpha_i \right) \Delta y_{t-k} + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\theta_1 + \theta_1 + \theta_2 - \theta_1 \theta_1 + \dots + \theta_{p-1} - \theta_1 \theta_{p-2} - \theta_1 \theta_{p-1} - 1) y_{t-1} - \sum_{k=1}^{p-1} \gamma_k \Delta y_{t-k} + \eta_t$$

$$\Leftrightarrow \Delta y_t = (\theta_1(1 - \theta_1 - \theta_2 - \dots - \theta_{p-1}) + \theta_1 + \theta_2 + \dots + \theta_{p-1} - 1) y_{t-1} - \sum_{k=1}^{p-1} \gamma_k \Delta y_{t-k} + \eta_t$$

$$\Leftrightarrow \Delta y_t = [(\theta_1 - 1)(1 - \theta_1 - \theta_2 - \dots - \theta_{p-1})] y_{t-1} - \sum_{k=1}^{p-1} \gamma_k \Delta y_{t-k} + \eta_t$$

On trouve alors l'écriture du modèle [1] en posant $\rho = (\theta_1 - 1)(1 - \theta_1 - \theta_2 - \dots - \theta_{p-1})$:

$$\Delta y_t = \rho y_{t-1} - \sum_{k=1}^{p-1} \gamma_k \Delta y_{t-k} + \eta_t$$

ou encore

$$\Delta y_t = \rho y_{t-1} - \sum_{k=2}^p \gamma_k \Delta y_{t-k+1} + \eta_t .$$

Détermination du retard p du test ADF :

La valeur p est déterminée à l'aide du corrélogramme partiel de la série différenciée Δy_t .

Une fois déterminée la valeur p, on procède de la même façon qu'avec le test de Dickey-Fuller simple : on commence par étudier la significativité de b du modèle [3]. La règle de décision est la même que pour le test de DF simple.

La statistique de test pour l'estimateur de ϕ_1 est :

$$t_{\hat{\phi}_1} = \frac{\hat{\phi}_1 - 1}{\hat{\sigma}_{\hat{\phi}_1}}$$

qui est à comparer avec la valeur critique t_{DF} de la table de Dickey-Fuller.

Si $|t_{\phi_1}| > t_{DF} \Rightarrow$ On accepte H_0 : la série est non stationnaire (ATTENTION : il faut observer comme dans le cas du test de DF simple que pour $|t_{\phi_1}| > t_{DF}$, on n'a pas H_1 ! La règle de décision est inversée ici !).

c/ Test de Phillips-Perron :

Le test de Phillips et Perron permet de prendre en compte à la fois l'autocorrélation et l'hétéroscédasticité des erreurs. Il s'appuie sur les mêmes modèles que ceux du test de Dickey et Fuller simple mais propose une correction non-paramétrique de la statistique $t_{\hat{\phi}_1}$.

Le déroulement du test de Phillips-Perron s'effectue en quatre étapes qui sont :

1) Estimation par la méthode des moindres carrés ordinaires les trois modèles du test de Dickey-Fuller simple et calcul des résidus $\hat{\varepsilon}_t$.

2) Détermination de la variance dite de court terme :

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

3) Estimation du facteur correctif s_t^2 appelé variance de long terme :

$$s_t^2 = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2 + 2 \sum_{j=1}^b \left(1 - \frac{j}{b+1}\right) \frac{1}{n} \sum_{t=j+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}$$

où

$$b \approx 4 \left(\frac{n}{100}\right)^{2/9}.$$

4) Calcul de la statistique de Phillips et Perron :

$$t_{\hat{\phi}_1}^{PP} = \sqrt{k} \times \frac{(\hat{\phi}_1 - 1)}{\hat{\sigma}_{\hat{\phi}_1}} + \frac{n(k-1)\hat{\sigma}_{\hat{\phi}_1}}{\sqrt{k}}$$

avec

$$k = \frac{\hat{\sigma}^2}{s_t^2}.$$

Phillips et Perron (1988) montrent que cette correction non-paramétrique apportée à $t_{\hat{\phi}_1}$ ne modifie pas la distribution asymptotique de la statistique qui reste identique à celle qui est observée dans le cas du test de Dickey-Fuller simple. En conséquence, les valeurs critiques tabulées par Dickey et Fuller demeurent également valables pour le test de Phillips-Perron.

I.4 / Processus ARIMA

Lorsque l'on a une série $\{y_t\}$ à non stationnarité stochastique, il convient de la modéliser à l'aide d'un processus ARIMA(p,d,q) où d désigne l'ordre de différenciation (ou d'intégration).

Définition 6 : Un processus ARIMA(p,d,q) ou "Autoregressive Integrated Moving Average" d'ordre p, d, et q pour la série $\{y_t\}$ est un processus de la forme suivante :

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) \nabla^d y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

ou encore

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) (1-B)^d y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

où $\varepsilon_t \sim BB(0, \sigma_\varepsilon^2)$, B est l'opérateur de retard tel que $By_t = y_{t-1}$ et $B^p y_t = y_{t-p}$, ∇^d est l'opérateur de différence de degré d ($d \geq 0$ est un entier positif), (Φ_1, \dots, Φ_p) et $(\theta_1, \dots, \theta_q)$ sont des coefficients à estimer.

La série $\{y_t\}$ est une série non stationnaire alors que la série $w_t = \nabla^d y_t$ est une série stationnaire.

Estimer les paramètres du processus ARIMA(p,d,q) pour la série $\{y_t\}$ non stationnaire revient à estimer les coefficients du processus ARMA(p,q) pour la série $\{w_t\}$ stationnaire.

I.5 / Processus ARMA

Wold (1954) montre que les séries stationnaires peuvent être représentées par les processus ARMA.

Définition 7 : Soit $\{y_t\}$ une série stationnaire. Le modèle AR(p) ou autorégressif d'ordre p est défini par :

$$y_t - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \dots - \Phi_p y_{t-p} = \varepsilon_t$$

ou encore

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) y_t = \varepsilon_t$$

où $\Phi_1, \Phi_2, \dots, \Phi_p$ sont des coefficients (positifs ou négatifs) à estimer et $\varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2)$.

Un modèle AR(p) présente un corrélogramme simple caractérisé par une décroissance géométrique de ses termes et un corrélogramme partiel caractérisé par ses p premiers termes différents de 0.

Définition 8 : Le modèle MA(q) ou "Moving Average" (moyenne mobile) d'ordre q est donné par :

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

ou encore

$$y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

où $\theta_1, \theta_2, \dots, \theta_q$ sont des paramètres à estimer.

Un modèle MA(q) présente un corrélogramme simple défini par ses q premiers termes significativement différents de 0 et un corrélogramme partiel caractérisé par une décroissance géométrique des retards.

Définition 9 : Le modèle ARMA(p,q) est une combinaison des processus AR(p) et MA(q) :

$$y_t - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \dots - \Phi_p y_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

ou encore

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t$$

ou encore

$$\Phi(B) y_t = \theta(B) \varepsilon_t$$

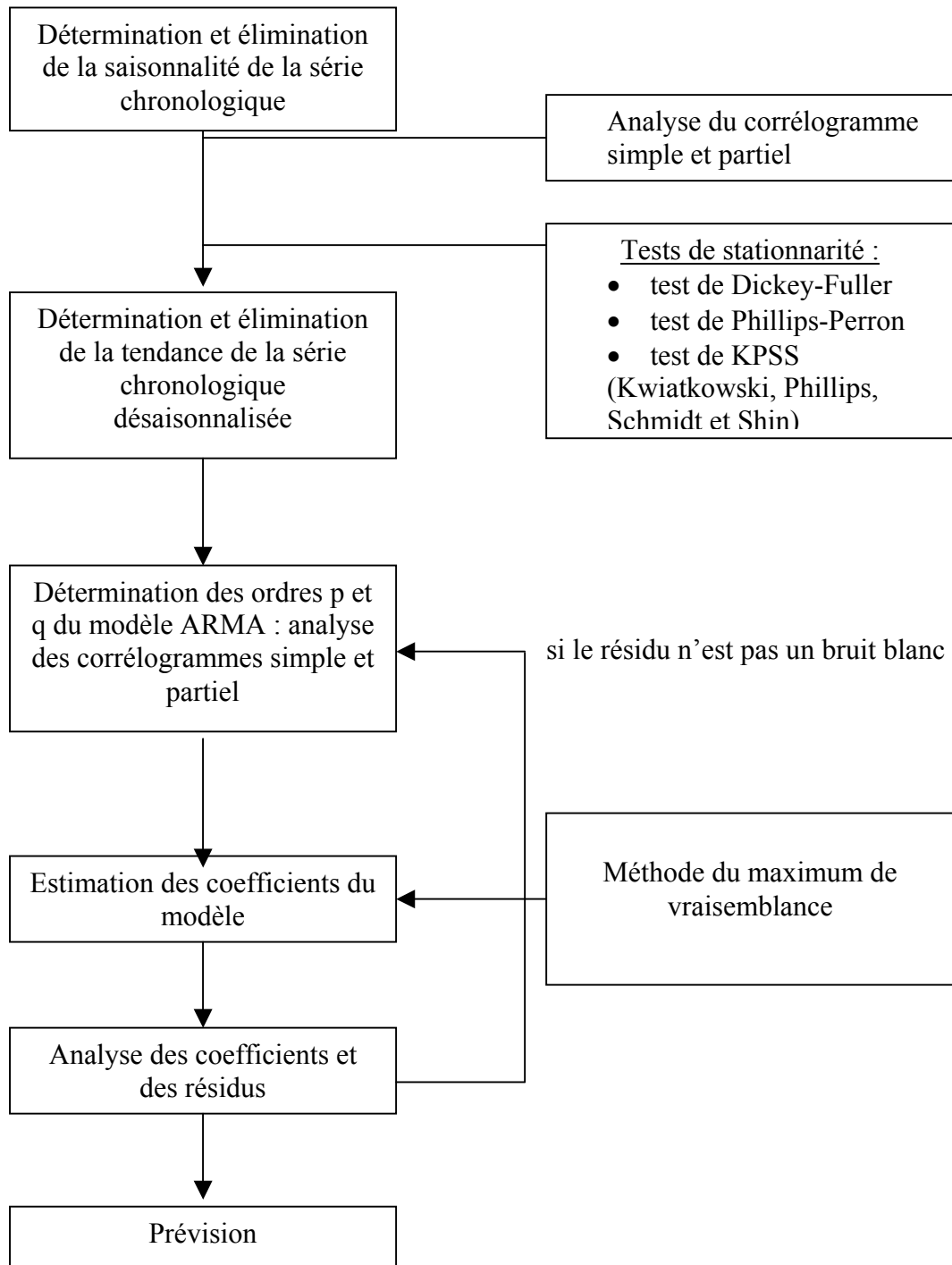
où

$$\varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2).$$

Le modèle ARMA(p,q) présente un corrélogramme simple et partiel qui sont un mélange des deux corrélogrammes des processus AR et MA purs.

I.6 / Méthode de Box et Jenkins

La méthode de Box et Jenkins permet de déterminer le modèle ARIMA pouvant convenir à une série temporelle selon ses caractéristiques. Elle se décompose en plusieurs étapes :



a/ Estimation des paramètres du processus ARMA(p,q) :

L'estimation des coefficients du processus ARMA(p,q) s'effectue principalement à l'aide de la méthode du maximum de vraisemblance. On suppose pour cela que $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

Méthode d'estimation du maximum de vraisemblance :

La méthode du maximum de vraisemblance est couramment utilisée pour estimer les coefficients des modèles des séries temporelles car c'est une méthode simple à mettre en place pour estimer des modèles plus complexes que le modèle linéaire.

Soit le modèle suivant :

$$y_t = a_0 + a_1 x_t + \varepsilon_t.$$

On a alors :

$$E(y_t) = a_0 + a_1 x_t \quad \text{et} \quad \text{Var}(y_t) = \sigma_\varepsilon^2.$$

La fonction de densité de la loi normale de la variable y_t s'écrit :

$$f(y_t) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left[-\frac{(y_t - E(y_t))^2}{2\sigma_\varepsilon^2}\right] = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left[-\frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\varepsilon^2}\right].$$

La fonction de vraisemblance est donnée par :

$$f(y_1, y_2, \dots, y_n; a_0, a_1, \sigma_\varepsilon^2) = \prod_{t=1}^n f(y_t; a_0, a_1, \sigma_\varepsilon^2)$$

où $f(y_t; a_0, a_1, \sigma_\varepsilon^2)$ est la fonction de densité de y_t .

On obtient alors :

$$f(y_1, y_2, \dots, y_n; a_0, a_1, \sigma_\varepsilon^2) = \prod_{t=1}^n \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left[-\frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\varepsilon^2}\right]$$
$$f(y_1, y_2, \dots, y_n; a_0, a_1, \sigma_\varepsilon^2) = \left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}}\right)^n \exp\left[\sum_{t=1}^n -\frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\varepsilon^2}\right].$$

Pour faciliter les calculs, on considère plutôt le logarithme de la fonction de vraisemblance. Il vient alors :

$$\text{Ln } f(y_1, y_2, \dots, y_n; a_0, a_1, \sigma_\varepsilon^2) = n \ln\left(\frac{1}{\sigma_\varepsilon \sqrt{2\pi}}\right) - \left[\sum_{t=1}^n \frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\varepsilon^2}\right]$$

$$\Leftrightarrow \text{Ln } f(y_1, y_2, \dots, y_n; a_0, a_1, \sigma_\varepsilon^2) = -\frac{n}{2} \ln \sigma_\varepsilon^2 - \frac{n}{2} \ln 2\pi - \left[\sum_{t=1}^n \frac{(y_t - a_0 - a_1 x_t)^2}{2\sigma_\varepsilon^2} \right].$$

Cette fonction est à maximiser. Les valeurs des coefficients qui permettent de maximiser la fonction sont issues des conditions du premier ordre suivantes :

$$\frac{\partial \text{Ln } f(\dots)}{\partial a_0} = 0, \quad \frac{\partial \text{Ln } f(\dots)}{\partial a_1} = 0, \quad \frac{\partial \text{Ln } f(\dots)}{\partial \sigma_\varepsilon^2} = 0.$$

b/ Validation du processus ARMA(p,q) :

Lors de la détermination des ordres p et q du processus ARMA(p,q) à l'aide des corrélogrammes simple et partiel, on peut être amené à sélectionner plusieurs ordres possibles p et q pour le processus ARMA(p,q). Après avoir estimé les différents processus ARMA(p,q) possibles, il reste à les valider et à les départager. La validation des processus passe par un examen des coefficients estimés (ils doivent être significativement différent de 0) et par un examen des résidus (les résidus estimés doivent suivre un processus de bruit blanc : $\varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2)$ où ε_t est l'estimateur de l'erreur ε_t puisque l'on a supposé que $\varepsilon_t \sim \text{BB}(0, \sigma_\varepsilon^2)$ lors de la définition du processus ARMA(p,q)).

b.1/ Tests sur les coefficients :

Parmi les processus ARMA estimés, on ne retiendra que ceux dont tous les coefficients ont un t de Student $> 1,96$ (pour un risque de 5% et pour une taille d'échantillon suffisamment grande : $T > 30$).

b.2/ Tests sur les résidus :

❖ *Tests d'autocorrélation :*

Il existe un grand nombre de tests d'autocorrélation, les plus connus sont ceux de Box et Pierce (1970) et Ljung et Box (1978). Nous n'étudierons ici que le test de Box et Pierce. Le test de Ljung et Box est à appliquer lorsque l'échantillon est de petite taille.

Soit une autocorrélation des erreurs d'ordre K ($K > 1$) :

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_K \varepsilon_{t-K} + v_t \quad \text{où } v_t \sim N(0, \sigma_v^2)$$

Les hypothèses du test de Box-Pierce sont les suivantes :

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0 \\ H_1 : \text{il existe au moins un } \rho_i \text{ significativement différent de } 0. \end{cases}$$

Pour effectuer ce test, on a recours à la statistique Q qui est donnée par :

$$Q = n \sum_{k=1}^K \hat{\rho}_k^2$$

où n est le nombre d'observations et $\hat{\rho}_k^2$ est le coefficient d'autocorrélation d'ordre k des résidus estimés e_t .

Sous l'hypothèse H_0 vraie, Q suit la loi du Khi-deux avec K degrés de liberté :

$$Q = n \sum_{k=1}^K \hat{\rho}_k^2 \sim \chi^2(K).$$

La règle de décision est la suivante :

si $Q > k^*$ où k^* est la valeur donnée par la table du Khi-Deux pour un risque fixé et un nombre K de degrés de liberté

\Rightarrow On rejette H_0 et on accepte H_1 (autocorrélation des erreurs).

❖ Tests d'hétéroscédasticité :

Il existe plusieurs tests possibles : test de Goldfeld et Quandt, test de White, test de Breusch et Pagan et test ARCH de Engle. Nous étudierons ici le test ARCH car il est très fréquemment employé en économétrie des séries temporelles financières.

Test ARCH :

Le test ARCH consiste à effectuer une régression autorégressive des résidus carrés sur q retards :

$$e_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2$$

où e_t désigne le résidu à l'instant t issu de l'estimation des paramètres du processus ARMA(p,q).

Pour déterminer le nombre de retards q , on étudie le corrélogramme des résidus au carré.

Les hypothèses du test ARCH sont les suivantes :

$$\begin{cases} H_0 : \text{homoscédasticité et } \alpha_1 = \dots = \alpha_q = 0 \\ H_1 : \text{hétéroscédasticité et il y a au moins un coefficient } \alpha_i \text{ significativement différent de 0.} \end{cases}$$

Pour mener le test, on utilise la statistique de test $n \times R^2$ où n correspond au nombre d'observations de la série e_t et R^2 représente le coefficient de détermination associé à la régression $e_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i e_{t-i}^2$.

Sous l'hypothèse H_0 , la statistique de test $n \times R^2$ suit la loi du Khi-deux à q degrés de liberté.

La règle de décision est alors :

-Si $n \times R^2 \leq \chi^2(q)$ où $\chi^2(q)$ désigne la valeur critique figurant dans la table du Khi-deux, on accepte ici l'hypothèse H_0 d'homoscédasticité.

-Si $n \times R^2 > \chi^2(q)$ où $\chi^2(q)$ désigne la valeur critique valeur figurant dans la table du Khi-deux, on rejette ici l'hypothèse H_0 d'homoscédasticité et on admet qu'il y a de l'hétéroscédasticité.

b.3/ Critères de choix des modèles :

Après examen des coefficients et des résidus, certains modèles sont écartés. Pour départager les modèles restants, on fait appel aux critères standards et aux critères d'information.

❖ *Critères standards :*

- L'erreur absolue moyenne (Mean Absolute Error) :

$$MAE = \frac{1}{n} \sum_t |e_t|$$

où e_t est le résidu du modèle ARMA étudié et n le nombre d'observation.

- Racine de l'erreur quadratique moyenne (Root Mean Squared Error) :

$$RMSE = \sqrt{\frac{1}{n} \sum_t e_t^2}$$

- Ecart absolu moyen en pourcentage (Mean Absolute Percent Error) :

$$MAPE = 100 \frac{1}{n} \sum_t \left| \frac{e_t}{X_t} \right| .$$

Plus la valeur de ces critères est faible, plus le modèle estimé est proche des observations.

❖ *Critères d'information :*

- Le critère d'Akaike :

$$AIC = \ln \sigma_{\varepsilon}^2 + \frac{2(p+q)}{n}$$

où p est l'ordre de la partie AR et q est l'ordre de la partie MA.

- Le critère de Schwarz :

$$SIC = \ln \sigma_{\varepsilon}^2 + (p+q) \frac{\ln(n)}{n}$$

- Le critère d'information de Hannan-Quinn :

$$HQ = \ln \sigma_{\varepsilon}^2 + \alpha(p+q) \ln \left(\frac{\ln(n)}{n} \right)$$

où $\alpha (> 2)$ est une constante.

On choisit le modèle qui minimise les critères standards et les critères d'information. Le modèle sélectionné sera alors utilisé pour la prévision.

I.7 / Processus ARCH

Les processus ARCH (AutoRegressive Conditional Heteroskedasticity) sont utilisés pour modéliser la volatilité d'une série.

On a vu que le modèle ARMA(p,q) s'écrivait

$$\Phi(B)y_t = \theta(B)\varepsilon_t$$

où $\varepsilon_t \sim BB(0, \sigma_{\varepsilon}^2)$.

Or σ_{ε}^2 peut ne pas être constant. On peut avoir $V(\varepsilon_t | \varepsilon_{t-1}) = \sigma_t^2$ qui est une variance conditionnelle.

On utilise alors les processus ARCH pour modéliser cette variance conditionnelle.

En plus du processus ARCH, nous avons aussi les processus GARCH, EGARCH, TGARCH, ARCH-M, ...

Définition 10 : Un processus ARCH(q) s'exprime de la manière suivante :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

avec

$$\alpha_0 > 0 \text{ et } \alpha_i \geq 0 \quad \forall i.$$

Définition 11 : Un processus GARCH(p,q) (Generalized ARCH) est défini comme suit :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

où

$$\alpha_0 > 0, \quad \alpha_i \geq 0 \text{ et } \beta_j \geq 0 \quad \forall i \text{ et } j.$$

Définition 12 : Un processus EGARCH(p,q) (Exponential GARCH) s'écrit de la façon suivante :

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i (\theta z_{t-i} + \gamma(|z_{t-i}| - E|z_{t-i}|)) + \sum_{j=1}^p \beta_j \ln \sigma_{t-j}^2$$

où $z_{t-i} = \frac{\varepsilon_{t-i}}{\sigma_{t-i}}$ représente l'erreur standardisée.

On peut remarquer ici qu'il n'y a pas de contrainte de positivité qui pèse sur les coefficients car l'équation de la variance s'exprime en log. A la différence des processus ARCH et GARCH, le processus EGARCH(p,q) permet à la volatilité de réagir différemment selon le signe des chocs.

Définition 13 : Un processus TGARCH(p,q) (Threshold GARCH) est défini par :

$$\sigma_t = \alpha_0 + \sum_{i=1}^q (\alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^-) + \sum_{j=1}^p \beta_j \sigma_{t-j}$$

où

$$\varepsilon_t^+ = \max(\varepsilon_t, 0) \quad \text{et} \quad \varepsilon_t^- = \min(\varepsilon_t, 0).$$

Le modèle comporte des contraintes de positivité qui sont : $\alpha_0 > 0, \alpha_i^+ \geq 0, \alpha_i^- \geq 0, \beta_j \geq 0 \quad \forall i, j$. Toutefois, le modèle permet ici de considérer les effets asymétriques des chocs sur la volatilité.

Définition 14 : Un processus ARCH-M est donné par :

$$\begin{cases} \Phi(B)y_t = \theta(B)\varepsilon_t + \delta\sigma_t^2 \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad \text{où } \alpha_0 > 0 \text{ et } \alpha_i \geq 0 \forall i \end{cases}$$

La structure ARMA est appelée équation de la moyenne. Pour un processus ARCH-M, on ajoute une variance dans l'équation de la moyenne. On peut avoir aussi un processus GARCH-M, TGARCH-M, ...

Les processus ARCH, GARCH, EGARCH,... sont estimés à l'aide du maximum de vraisemblance (ou plutôt du pseudo maximum de vraisemblance car les erreurs des séries temporelles en finance ne suivent pas pour la plupart une loi normale).

La fonction log-vraisemblance à maximiser s'écrit ici de la façon suivante :

$$\text{Ln } f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n; \theta) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^n \left[\ln(\sigma_t^2) + \left(\frac{\varepsilon_t}{\sigma_t} \right)^2 \right]$$

où θ est un vecteur contenant les paramètres à estimer des processus ARMA et de la volatilité (ARCH, GARCH, ...).

II/ ETUDE MULTIVARIEE : MODELISATION DE LA RELATION ENTRE DEUX SERIES TEMPORELLES :

II.1/ Séries non stationnaires, cointégration et modèle à correction d'erreur (MCE)

Soient $y_t \sim I(1)$, $x_t \sim I(1)$, x_t et y_t sont indépendants, si on estime à l'aide des MCO le modèle suivant :

$$y_t = ax_t + b + \varepsilon_t$$

on obtient :

$$y_t - ax_t - b = \varepsilon_t \sim I(1).$$

$\varepsilon_t \not\sim I(0)$, ε_t n'est donc pas stationnaire (le DW est ici très faible).

De plus, on aboutit à une régression dite fallacieuse ou illusoire ("spurious regression") caractérisée par un R^2 et des t de Student très élevés alors que les deux variables n'ont aucun lien entre elles !

On peut éviter ce problème en passant les variables en différences premières afin de les rendre stationnaires ($\Delta y_t \sim I(0)$ et $\Delta x_t \sim I(0)$ si x_t et y_t sont des processus non stationnaires aléatoires) et en effectuant la régression suivante :

$$\Delta y_t = a \Delta x_t + b + \mu_t.$$

Par ailleurs, on obtient :

$$\Delta y_t - a \Delta x_t - b = \mu_t \rightsquigarrow I(0).$$

Toutefois, il arrive que l'on souhaite travailler avec des variables plutôt en niveau qu'en différences premières (donc plutôt avec des variables non stationnaires). Dans ce cas, comment savoir si la régression effectuée est fallacieuse ou non ? C'est alors que la notion de cointégration prend toute son importance ici. Nous n'avons pas de régression fallacieuse lorsque les variables x_t et y_t sont cointégrées, c'est à dire lorsque l'on a $y_t - ax_t - b = \varepsilon_t \rightsquigarrow I(0)$ alors que $y_t \rightsquigarrow I(1)$ et $x_t \rightsquigarrow I(1)$.

a/ Définition de la cointégration :

Granger a montré que si on avait deux variables non stationnaires ($y_t \rightsquigarrow I(1)$ et $x_t \rightsquigarrow I(1)$), on pouvait avoir :

$$y_t - ax_t - b = \varepsilon_t \rightsquigarrow I(1)$$

ou

$$y_t - ax_t - b = \varepsilon_t \rightsquigarrow I(0) !$$

Définition 15 : Deux séries non stationnaires ($y_t \rightsquigarrow I(1)$ et $x_t \rightsquigarrow I(1)$) sont dites cointégrées si on a :

$$y_t - ax_t - b = \varepsilon_t \rightsquigarrow I(0).$$

Les séries y_t et x_t sont alors notées :

$$x_t, y_t \rightsquigarrow CI(1,1).$$

c/ Test de cointégration entre deux variables :

Etape 1 : tester l'ordre d'intégration des variables :

Une condition nécessaire de cointégration est que les séries doivent être intégrées de même ordre. Si les séries ne sont pas intégrées de même ordre, elles ne peuvent être cointégrées.

Il convient donc de vérifier l'ordre d'intégration des chroniques étudiées à l'aide par exemple du test de Dickey-Fuller (simple ou augmenté).

Si les séries considérées ne sont pas intégrées de même ordre, il n'y a alors pas de risque de cointégration et la procédure s'arrête à cette première étape.

Etape 2 : estimation de la relation de long terme

Si on a :

$$x_t \rightsquigarrow I(1) \quad \text{et} \quad y_t \rightsquigarrow I(1).$$

On estime par les MCO la relation de long terme :

$$y_t = ax_t + b + \varepsilon_t.$$

Pour qu'il y ait cointégration, il faut que le résidu e_t issu de la régression soit stationnaire :

$$e_t = y_t - \hat{ax}_t - \hat{b} \rightsquigarrow I(0).$$

La stationnarité du résidu est testée à l'aide du test DF ou DFA.

On remarque ici que la relation porte sur les résidus estimés et non pas sur les « vrais » résidus de l'équation de cointégration. Par conséquent, nous ne pouvons pas nous référer aux tables de Dickey-Fuller pour mener le test de stationnarité. Il faut regarder ici les tables de MacKinnon.

Si le résidu est stationnaire nous pouvons alors estimer un modèle appelé modèle à correction d'erreur (MCE) qui intègre les variables en variation et en niveau (théorème de la représentation de Granger). L'emploi d'un modèle à correction d'erreur dans le cas de la cointégration permet d'obtenir des prévisions plus fiables que si on avait utilisé la relation de long terme car les résultats de l'estimation de cette relation sont faussés par la non stationnarité des séries.

b/ Modèle à correction d'erreur :

Si on a deux séries cointégrées ($y_t - \hat{ax}_t - \hat{b} \rightsquigarrow I(0)$), on peut estimer le modèle à correction d'erreur (MCE) suivant :

$$\Delta y_t = \gamma \Delta x_t + \delta (y_{t-1} - ax_{t-1} - b) + v_t \text{ avec } \delta < 0.$$

On peut remarquer que le paramètre δ doit être négatif pour qu'il y ait un retour de y_t à sa valeur d'équilibre de long terme qui est $(ax_{t-1} + b)$. En effet, lorsque y_{t-1} est supérieur à $(ax_{t-1} + b)$, il n'y a une force de rappel vers l'équilibre de long terme que si $\delta < 0$.

Le MCE permet de modéliser conjointement les dynamiques de court et long terme.

La dynamique de court terme s'écrit :

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + v_t$$

La dynamique de long terme s'exprime de la manière suivante :

$$y_t = ax_t + b + \varepsilon_t$$

car à long terme, on a $y_{t-1} = y_t$, $x_{t-1} = x_t$ et la dynamique de court terme devient à long terme :

$$y_t = \alpha_0 + \alpha_1 y_t + \alpha_2 x_t + \alpha_3 x_t + v_t$$

$$(1 - \alpha_1)y_t = (\alpha_2 + \alpha_3)x_t + \alpha_0 + v_t$$

$$y_t = ax_t + b + \varepsilon_t$$

$$\text{où } a = \frac{\alpha_2 + \alpha_3}{1 - \alpha_1}, \quad b = \frac{\alpha_0}{1 - \alpha_1}, \quad \varepsilon_t = \frac{v_t}{1 - \alpha_1}.$$

Le MCE s'obtient à partir de la dynamique de court terme :

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + v_t$$

$$y_t - y_{t-1} = \alpha_0 + \alpha_1 y_{t-1} - y_{t-1} + \alpha_2 x_t - \alpha_2 x_{t-1} + \alpha_2 x_{t-1} + \alpha_3 x_{t-1} + v_t$$

$$\Delta y_t = (\alpha_1 - 1)y_{t-1} + \alpha_2 (x_t - x_{t-1}) + \alpha_0 + (\alpha_2 + \alpha_3)x_{t-1} + v_t$$

$$\Delta y_t = -(1 - \alpha_1)y_{t-1} + \alpha_2 (x_t - x_{t-1}) + \alpha_0 + (\alpha_2 + \alpha_3)x_{t-1} + v_t$$

$$\Delta y_t = -(1 - \alpha_1) \left(y_{t-1} - \frac{\alpha_2 + \alpha_3}{1 - \alpha_1} x_{t-1} - \frac{\alpha_0}{1 - \alpha_1} \right) + \alpha_2 \Delta x_t + v_t$$

$$\Delta y_t = \gamma \Delta x_t + \delta (y_{t-1} - ax_{t-1} - b) + v_t$$

$$\text{où } \alpha_2 = \gamma, \quad \delta = -(1 - \alpha_1), \quad a = \frac{\alpha_2 + \alpha_3}{1 - \alpha_1} \quad \text{et} \quad b = \frac{\alpha_0}{1 - \alpha_1}.$$

d/ Estimation du MCE avec une seule variable explicative :

Si les séries Y_t et X_t sont cointégrées :

$$x_t, y_t \sim CI(1,1)$$

nous pouvons estimer le MCE.

Etape 1 : estimation par les MCO de la relation de la relation de long terme :

$$y_t = ax_t + b + \varepsilon_t.$$

Etape 2 : estimation par les MCO de la relation du modèle dynamique de court terme :

$$\Delta y_t = \gamma \Delta x_t + \delta e_{t-1} + v_t \quad \text{avec } \delta < 0$$

$$\text{où } e_t = y_t - \hat{a}x_t - \hat{b}.$$

Le coefficient δ doit être significativement négatif. Dans le cas contraire, la spécification de type MCE n'est pas valable.

II.2/ Modèle VAR

L'absence de cointégration entre deux séries non stationnaires $y_{1,t}$ et $y_{2,t}$ ($y_{1,t} \sim I(1)$ et $y_{2,t} \sim I(1)$), mais l'existence d'une causalité entre les séries stationnaires $\Delta y_{1,t}$ et $\Delta y_{2,t}$ ($\Delta y_{1,t} \sim I(0)$ et $\Delta y_{2,t} \sim I(0)$) nous permet d'estimer un modèle VAR.

a/ Présentation du modèle VAR :

Le modèle VAR("Vector AutoRegressive") à k variables (hors constante) et p retards noté VAR(p) s'écrit :

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + v_t$$

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ \cdot \\ y_{k,t} \end{pmatrix} = \begin{pmatrix} a_{0,1} \\ a_{0,2} \\ \vdots \\ \cdot \\ a_{0,k} \end{pmatrix} + \begin{pmatrix} a_{1,1}^1 & a_{1,2}^1 & \dots & a_{1,k}^1 \\ a_{2,1}^1 & a_{2,2}^1 & \dots & a_{2,k}^1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1}^1 & a_{k,2}^1 & \dots & a_{k,k}^1 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ \cdot \\ y_{k,t-1} \end{pmatrix} + \begin{pmatrix} a_{1,1}^2 & a_{1,2}^2 & \dots & a_{1,k}^2 \\ a_{2,1}^2 & a_{2,2}^2 & \dots & a_{2,k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1}^2 & a_{k,2}^2 & \dots & a_{k,k}^2 \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \\ \vdots \\ \cdot \\ y_{k,t-2} \end{pmatrix} \\ + \dots + \begin{pmatrix} a_{1,1}^p & a_{1,2}^p & \dots & a_{1,k}^p \\ a_{2,1}^p & a_{2,2}^p & \dots & a_{2,k}^p \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1}^p & a_{k,2}^p & \dots & a_{k,k}^p \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ \cdot \\ y_{k,t-p} \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ \cdot \\ v_{k,t} \end{pmatrix}$$

Les variables $y_{1,t}, y_{2,t}, \dots, y_{k,t}$ sont stationnaires. Les perturbations $v_{1,t}, v_{2,t}, \dots, v_{k,t}$ sont des bruits blancs de variances constantes et non autocorrélées.

b/ Test de causalité au sens de Granger :

Soit le modèle VAR(p) pour lequel les variables y_{1t} et y_{2t} sont stationnaires :

$$\begin{cases} y_{1,t} = \gamma_1 + \alpha_{11}y_{1,t-1} + \alpha_{12}y_{1,t-2} + \dots + \alpha_{1p}y_{1,t-p} + \beta_{11}y_{2,t-1} + \beta_{12}y_{2,t-2} + \dots + \beta_{1p}y_{2,t-p} + v_{1,t} \\ y_{2,t} = \gamma_2 + \alpha_{21}y_{1,t-1} + \alpha_{22}y_{1,t-2} + \dots + \alpha_{2p}y_{1,t-p} + \beta_{21}y_{2,t-1} + \beta_{22}y_{2,t-2} + \dots + \beta_{2p}y_{2,t-p} + v_{2,t} \end{cases}$$

Le test consiste à poser ces deux hypothèses :

- $y_{2,t}$ ne cause pas $y_{1,t}$ si l'hypothèse H_0 suivante est acceptée :

$$\beta_{11} = \beta_{12} = \beta_{13} = \dots = \beta_{1p} = 0$$

- $y_{1,t}$ ne cause pas $y_{2,t}$ si l'hypothèse H_0 suivante est acceptée :

$$\alpha_{21} = \alpha_{22} = \alpha_{23} = \dots = \alpha_{2p} = 0$$

On teste ces deux hypothèses à l'aide d'un test de Fisher classique. On peut faire le test équation par équation :

$$\left\{ \begin{array}{l} H_0 : \beta_{11} = \beta_{12} = \beta_{13} = \dots = \beta_{1p} = 0 \text{ et } y_{1,t} = \gamma_1 + \alpha_{11}y_{1,t-1} + \alpha_{12}y_{1,t-2} + \dots + \alpha_{1p}y_{1,t-p} + v_{1,t} \\ H_1 : \text{au moins un des coefficients } \beta \neq 0 \text{ et } y_{2,t} \text{ cause } y_{1,t} \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0 : \alpha_{21} = \alpha_{22} = \alpha_{23} = \dots = \alpha_{2p} = 0 \text{ et } y_{2,t} = \gamma_2 + \beta_{21}y_{2,t-1} + \beta_{22}y_{2,t-2} + \dots + \beta_{2p}y_{2,t-p} + v_{2,t} \\ H_1 : \text{au moins un des coefficients } \alpha \neq 0 \text{ et } y_{1,t} \text{ cause } y_{2,t} \end{array} \right.$$

Si nous sommes amenés à accepter les deux hypothèses que $y_{1,t}$ cause $y_{2,t}$ et que $y_{2,t}$ cause $y_{1,t}$, on parle de boucle rétroactif.

c/ Estimation du modèle VAR(p) :

Dans le cas du modèle VAR, chacune des équations peut être estimée par les MCO, indépendamment les unes des autres (ou par la méthode de vraisemblance).

Comme il y a énormément de coefficients à estimer dans un modèle VAR, il est préférable d'effectuer un test de causalité avant de chercher à estimer le modèle VAR. On pourra ainsi éliminer du modèle à estimer les variables qui n'interviennent pas sur la variable à expliquer.

CONCLUSION :

- Si $x_t, y_t \sim I(0)$ et $y_t - x_t a - b \sim I(0)$: on estime $y_t = x_t a + b + \varepsilon_t$ ou un modèle VAR en niveau pour plusieurs variables.
- Si $x_t, y_t \sim I(1)$ et $y_t - x_t a - b \sim I(0)$: on estime un modèle à correction d'erreur ou un modèle VECM pour plusieurs variables.
- Si $x_t, y_t \sim I(1)$ et $y_t - x_t a - b \not\sim I(0)$ et $\Delta x_t, \Delta y_t \sim I(0)$ avec lien de causalité : on estime la relation $\Delta y_t = \Delta x_t a + b + \eta_t$ ou un modèle VAR en différence première pour plusieurs variables.