

1. Processus stochastiques
2. Tests de racine unitaire
3. Problèmes économétriques liés aux variables intégrées
4. Définition de la cointégration
5. Test de cointégration d'Engle et Granger
6. Tests de cointégration de Johansen
7. Modèles à correction d'erreur conditionnel
8. Les modèles à volatilité conditionnelle autorégressive

Source : Econométrie appliquée : Synthèse de cours et exercices corrigés. Eric Dor (2004). Pearson éducation

1. Processus stochastiques

Les séries temporelles observées sont considérées comme des réalisations de processus stochastiques, définis jusqu'à présent comme des successions temporelles de variables aléatoires représentant un même concept. L'objet de cette section est de préciser la définition des processus stochastiques et d'examiner leurs propriétés.

Un processus stochastique X ou $\{X_t\}$ est une variable aléatoire indicée par le temps, multivariée et dont chaque composante univariée toujours le même concept économique quelle que soit la période considérée.

$$X = \{X_t\}_{t=1, \dots, n} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

La réalisation d'un processus stochastique X est une série temporelle observée, dont chaque composante est une réalisation d'une composante différente de X :

Un processus stochastique X ou $\{X_t\}$ est caractérisé par :

- une espérance $E(X_t)$ à chaque période t : $E(X_1), E(X_2) \dots$
- une variance $V(X_t)$ à chaque période t : $V(X_1), V(X_2) \dots$
- des covariances $\text{Cov}(X_t, X_{t-\theta})$ pour toutes les périodes t et tous les retards θ : $\text{Cov}(X_2, X_1), \text{Cov}(X_3, X_1), \text{Cov}(X_3, X_2) \dots$

1.1 PROCESSUS STOCHASTIQUES STATIONNAIRES

Un processus stochastique X ou $\{X_t\}$ est stationnaire si :

- $E(X_t) = \mu_X \forall t$,
- $V(X_t) = \sigma_X^2 \forall t$;
- $\text{Cov}(X_t, X_{t-\theta}) = v(\theta) \forall t, \forall \theta$.

En d'autres termes, X est stationnaire si les conditions suivantes sont réunies :

- L'espérance est constante (elle est inchangée dans le temps).
- La variance est constante dans le temps.
- La covariance entre une composante d'une date t et une composante d'une autre date $t - q$ ne dépend que de l'écart de temps (le « retard ») θ entre les dates, et non de la date t elle-même : par exemple $\text{Cov}(X_{72}, X_{70}) = \text{Cov}(X_{89}, X_{87})$.

Le coefficient de corrélation entre deux composantes de dates différentes, appelé coefficient d'**autocorrélation**, est défini par :

$$\forall t : \rho_{X_t, X_{t-\theta}} = \frac{\text{Cov}(X_t, X_{t-\theta})}{\sqrt{V(X_t) V(X_{t-\theta})}}$$

Lorsque le processus est stationnaire, ce coefficient est dit d'**autocorrélation d'ordre** θ ; il est le même dès lors que l'écart entre les dates est inchangé :

Un cas particulier de processus stochastique stationnaire est le bruit blanc. Un processus stochastique X ou $\{X_t\}$ est un **bruit blanc**.

1.2 TYPOLOGIE DES PROCESSUS STOCHASTIQUES STATIONNAIRES

Représentation de Wold et filtres linéaires

Le théorème de Wold prouve mathématiquement qu'un processus stochastique stationnaire peut toujours être généré par la somme d'une constante et d'une combinaison linéaire, parfois infinie, des composantes courantes et passées d'un bruit blanc. Cela se traduit en termes mathématiques de la façon suivante :

$$\{X_t\} \text{ stationnaire} \Rightarrow$$

\exists un bruit blanc $\{a_t\}$ et des coefficients $\mu_x, \psi_1, \psi_2 \dots$ tels que :

$$\forall t : X_t = \mu_x + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

Cette équation est la **représentation de Wold** du processus stochastique stationnaire $\{X_t\}$. Ici, l'espérance de X_t est représentée par μ_x et la variance de X_t par :

$$\sigma_a^2 \left(\sum_{i=1}^{\infty} \psi_i^2 + 1 \right)$$

On distingue un processus stochastique stationnaire d'une autre par : la spécificité des paramètres qui déterminent la fonction de densité du bruit blanc $\{a_t\}$ qui le génère ;

La spécificité des valeurs des coefficients $\mu, \psi_1, \psi_2 \dots$. Sur cette base, on met en évidence trois types de processus stochastiques stationnaires :

1. Les processus autorégressifs $AR(p)$ ou $ARMA(p,0)$;
2. Les processus en moyenne mobile $MA(q)$ ou $ARMA(0,q)$;
3. Les processus mixtes (autorégressifs et en moyenne mobile) $ARMA(p,q)$.

Processus stochastiques autorégressifs stationnaires

Les processus **autorégressifs** peuvent s'écrire comme la somme d'une constante, de la valeur courante d'un bruit blanc et d'une combinaison linéaire **finie** de leurs valeurs passées.

Un processus stochastique $\{X_t\}$ est dit **autorégressif d'ordre** p , et noté $AR(p)$ ou $ARMA(p,0)$, si :

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + a_t$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Processus $AR(1)$ ou $ARMA(1, 0)$:

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + a_t$$

où $\{a_t\}$ est un bruit blanc.

Exemple

Processus **AR(2)** ou **ARMA(2, 0)** :

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + a_t$$

où $\{a_t\}$ est un bruit blanc.

Soit l'opérateur de retard L , défini par la relation suivante : $L^i X_t = X_{t-i}$. Par exemple,

$$L X_t = X_{t-1}, L^2 X_t = X_{t-2} \dots$$

Si l'on utilise cette notation dans l'équation générique précédente, on obtient :

$$\forall t : X_t = \mu + \varphi_1 L X_t + \varphi_2 L^2 X_t + \dots + \varphi_p L^p X_t + a_t$$

c'est-à-dire :

$$\forall t : X_t - \varphi_1 L X_t - \varphi_2 L^2 X_t - \dots - \varphi_p L^p X_t = \mu + a_t$$

ou :

$$\forall t : (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) X_t = \mu + a_t$$

Le polynôme de degré p en l'opérateur de retard $L(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ est caractéristique du processus **AR(p)**.

Pour qu'un processus stochastique autorégressif soit **stationnaire**, les racines⁽¹⁾ du polynôme caractéristique $(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ doivent être en dehors du cercle unité. Quand ces racines sont réelles, elles doivent être strictement inférieures à -1 ou strictement supérieures à 1 .

Exemple

Soit un processus **AR(1)** ou **ARMA(1,0)**, $\{X_t\}$, défini par la relation :

$$X_t = \mu + \varphi_1 X_{t-1} + a_t$$

où a est un bruit blanc. $\{X_t\}$ se représente aussi de la manière suivante avec l'opérateur de retard :

$$(1 - \varphi_1 L) X_t = \mu + a_t$$

Son polynôme caractéristique est alors $(1 - \varphi_1 L)$ dont la racine⁽¹⁾ est $\frac{1}{\varphi_1}$. Ce processus stochastique $\{X_t\}$ est donc stationnaire si $\frac{1}{\varphi_1} < -1$ ou si $\frac{1}{\varphi_1} > 1$, et donc si $-1 < \varphi_1 < 1$.

$\frac{1}{\varphi_1}$

Un processus stochastique autorégressif d'ordre p a par définition une représentation autorégressive finie (il dépend de ses valeurs passées jusqu'à un retard fini p) :

$$\forall t : X_t = \mu + \varphi_1 L X_t + \varphi_2 L^2 X_t + \dots + \varphi_p L^p X_t + a_t$$

¹. Les racines du polynôme $p(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ sont les valeurs de L telles que $(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) = 0$.

On montre que si ce processus est stationnaire, il a également une représentation en forme de moyenne mobile infinie des valeurs courante et passées du bruit blanc $\{a_t\}$. On dit donc qu'un processus AR stationnaire est **inversible**. Cela s'illustre par l'exemple simple d'un processus AR(1):

$$\begin{aligned}
 X_t &= \mu + \varphi_1 X_{t-1} + a_t \\
 &= \mu + \varphi_1(\mu + \varphi_1 X_{t-2} + a_{t-1}) + a_t \\
 &= \mu + \varphi_1 \mu + \varphi_1^2 X_{t-2} + \varphi_1 a_{t-1} + a_t \\
 &= \mu + \varphi_1 \mu + \varphi_1^2(\mu + \varphi_1 X_{t-3} + a_{t-2}) + \varphi_1 a_{t-1} + a_t \\
 &= \mu + \varphi_1 \mu + \varphi_1^2 \mu + \varphi_1^3 X_{t-3} + \varphi_1^2 a_{t-2} + \varphi_1 a_{t-1} + a_t \\
 &= (1 + \varphi_1 + \varphi_1^2) \mu + \varphi_1^3 X_{t-3} + \varphi_1^2 a_{t-2} + \varphi_1 a_{t-1} + a_t \\
 &= \dots \\
 &= \mu \sum_{i=0}^{\infty} \varphi_1^i + \lim_{i \rightarrow \infty} \varphi_1^i X_{t-i} + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \dots \\
 &= \mu \sum_{i=0}^{\infty} \varphi_1^i + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \quad \text{si } -1 < \varphi_1 < 1 \\
 &= \frac{\mu}{1 - \varphi_1} + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i} \quad \text{si } -1 < \varphi_1 < 1
 \end{aligned}$$

Il s'agit de la forme particulière que prend la représentation de Wold d'un processus stochastique AR(1) stationnaire : $\forall t : X_t = \mu_X + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \frac{\mu_X}{1 - \varphi_1} + \sum_{i=0}^{\infty} \varphi_1^i a_{t-i}$ et $\psi_i = \varphi_1^i \forall i$

Processus stochastiques stationnaires en moyenne mobile

Les processus stochastiques **en moyenne mobile** peuvent s'écrire comme la somme d'une constante, de la valeur courante d'un bruit blanc et d'une combinaison linéaire finie des valeurs passées du bruit blanc.

Un processus stochastique $\{X_t\}$ est dit **en moyenne mobile d'ordre q** , et noté MA(q) ou ARMA(0, q), si :

$$\forall t : X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots + \theta_q a_{t-q}$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Processus MA(1) ou ARMA(0, 1) :

$$\forall t : X_t = \mu + a_t - \theta_1 a_{t-1}$$

où $\{a_t\}$ est un bruit blanc.

Exemple

Processus MA(2) ou ARMA(0, 2) :

$$\forall t : X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

où $\{a_t\}$ blanc.

La stationnarité n'impose aucune condition sur les coefficients, pourvu que tous les θ et la variance de a soient inférieurs à l'infini. Un modèle MA n'a de sens que s'il est inversible, ce qui implique qu'il peut s'écrire comme un modèle autorégressif avec une infinité de retards. Pour cela, les racines⁽²⁾ du polynôme caractéristique $(1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q)$ doivent être en dehors du cercle unité. Quand ces racines sont réelles, elles doivent être strictement inférieures à -1 ou strictement supérieures à 1 .

Processus stochastiques mixtes stationnaires

Un processus stochastique $\{X_t\}$ est dit **mixte** ou **autorégressif en moyenne mobile d'ordres p et q** , et noté $ARMA(p,q)$, si :

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots + \theta_q a_{t-q} \quad (5.6)$$

où le processus stochastique $\{a_t\}$ est un bruit blanc.

Exemple

Un processus $ARMA(1,1)$ évolue de la manière suivante :

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + a_t - \theta_1 a_{t-1}$$

où $\{a_t\}$ est un bruit blanc

Les conditions de stationnarité sont les mêmes que celles des processus autorégressifs.

1.3 TYPOLOGIE DES PROCESSUS STOCHASTIQUES NON STATIONNAIRES

Processus stochastiques non stationnaires à tendance uniquement déterministe

Un processus stochastique non stationnaire à tendance uniquement déterministe évolue de la manière suivante :

$$X_t = g(t) + v_t \quad \text{pour tout } t = 1 \dots n$$

où $g(t)$ est une fonction déterministe (mécanique, non aléatoire) du temps et où v_t est une composante aléatoire stationnaire. v est un processus stochastique stationnaire qui n'est pas nécessairement un bruit blanc : il peut être autocorrélé, lié à ses valeurs passées.

². Les racines du polynôme $p(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ sont les valeurs de L telles que $(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) = 0$.

Le théorème de Wold montre qu'un processus stochastique stationnaire d'espérance nulle peut toujours être écrit comme une combinaison linéaire des valeurs courante et passées d'un bruit blanc :

$$v_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \dots \quad \text{pour tout } t$$

où u_t est un bruit blanc. Si l'on utilise la notation de l'opérateur de retard, on obtient :

$$v_t = (1 + \psi_1 L + \psi_2 L^2 + \dots) u_t = \psi(L) u_t \quad \text{pour tout } t$$

où $\psi(L)$ est un polynôme en l'opérateur de retard L . Un processus stochastique non stationnaire à tendance déterministe uniquement évolue de la manière suivante :

$$X_t = g(t) + \psi(L) u_t \quad \text{pour tout } t = 1 \dots n$$

Vu que la partie aléatoire v_t de X_t est stationnaire, la non-stationnarité de X_t est inhérente uniquement à sa partie déterministe $g(t)$.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un bruit blanc (tous les $\psi_i = 0$ et donc $v_t = u_t$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + u_t \quad \forall t = 1 \dots n$$

Le terme $\mu + \beta t$ est la partie déterministe de X_t , à une tendance déterministe linéaire. Le terme u_t est la partie aléatoire ou stochastique de X_t : elle est stationnaire.

Exemple

Soient $g(t)$ un polynôme du deuxième degré en t et v_t un bruit blanc (tous les $\psi_i = 0$ et donc $v_t = u_t$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta_1 t + \beta_2 t^2 + u_t \quad \forall t = 1 \dots n$$

Le terme $\mu + \beta_1 t + \beta_2 t^2$ est la partie déterministe de X_t , à tendance déterministe quadratique. Le terme u_t est la partie aléatoire ou stochastique de X_t : elle suit un modèle $MA(1)$ stationnaire.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un $MA(1)$ (les $\psi_i = 0$ pour $\psi_i > 1$ et donc $v_t = u_t + \psi_1 u_{t-1}$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + u_t + \psi_1 u_{t-1} \quad \forall t = 1 \dots n$$

Le terme $\mu + \beta t$ est la partie déterministe de X_t , à tendance déterministe linéaire. Le terme $u_t + \psi_1 u_{t-1}$ est la partie aléatoire ou stochastique de X_t : elle est stationnaire.

Exemple

Soient $g(t)$ un polynôme du premier degré en t et v_t un modèle $AR(2)$ stationnaire (tous les ψ_i sont différents de 0, et $v_t = \frac{u_t}{1 - \varphi_1 L - \varphi_2 L^2}$). L'évolution de X_t est décrite par la relation suivante :

$$X_t = \mu + \beta t + v_t \quad \text{où } v_t = \varphi_1 v_{t-1} + \varphi_2 v_{t-2} + u_t \quad \forall t = 1 \dots n$$

La fonction $\mu + \beta t$ est la partie déterministe de X_t , à tendance déterministe linéaire. Le terme $\frac{u_t}{1 - \varphi_1 L - \varphi_2 L^2}$ est la partie aléatoire ou stochastique de X_t : cette partie est stationnaire.

La fonction $\mu + \beta t$ est la partie déterministe de X_t , à tendance déterministe linéaire. Le terme

$u_t 1 - \varphi_1 L - \varphi_2 L^2$ est la partie aléatoire ou stochastique de X_t : cette partie est stationnaire.

Pour simplifier la présentation des caractéristiques d'un processus à tendance déterministe linéaire, on se limite au cas le plus fréquent :

$$X_t = \mu + \beta t + v_t \quad \text{où } v_t = \psi(L)u_t$$

L'espérance de X_t varie dans le temps de manière déterministe :

$$E(X_t) = \mu + \beta t \quad \forall t = 1 \dots n$$

puisque $E(v_t) = 0, \forall t = 1 \dots n$.

On peut expliciter la valeur de X_t et de son espérance à chaque période :

$$t = 1 : X_1 = \mu + \beta + v_1 \quad \text{et } E(X_1) = \mu + \beta$$

$$t = 2 : X_2 = \mu + 2\beta + v_2 \quad \text{et } E(X_2) = \mu + 2\beta$$

$$t = 3 : X_3 = \mu + 3\beta + v_3 \quad \text{et } E(X_3) = \mu + 3\beta$$

...

X_t augmente mécaniquement de β à chaque période. À chaque période, la partie aléatoire de X_t est v_t , qui est stationnaire. On dit que X_t est « stationnaire autour d'une tendance déterministe ». L'augmentation mécanique de l'espérance est la seule cause de la nonstationnarité de X_t , qui est de nature purement déterministe : la partie déterministe de X_t n'est pas stationnaire, contrairement à la partie aléatoire de X_t . La variance de X_t est constante dans le temps parce que v_t est stationnaire :

$$V(X_t) = V(v_t) = \sigma_v^2 \quad \forall t = 1 \dots n$$

Sur les valeurs ultérieures du processus, l'effet d'un choc aléatoire u_t est temporaire, puisque son impact sur les valeurs successives $X_{t+1}, X_{t+2} \dots$ s'estompe progressivement. En effet, un choc u_t en t affecte X_{t+1} uniquement par la manière dont v_t est lié à v_{t+i} . Or v_t est stationnaire : le lien entre v_t et v_{t+i} s'estompe au fur et à mesure que l'écart de temps i s'accroît. Si est un bruit blanc, un choc u_t n'a d'effet que sur X_t , et non sur $X_{t+1}, X_{t+2} \dots$

Processus stochastiques non stationnaires à tendance stochastique

Les processus stochastiques non stationnaires à tendance stochastique sont dits **intégrés** ou à **racines unitaires**.

Un processus stochastique est non stationnaire à tendance stochastique quand une ou plusieurs racines unitaires figurent dans sa partie autorégressive, ce qui implique que la somme de ses coefficients autorégressifs est égale à 1.

Tout processus $\{X_t\}$ peut en effet être représenté approximativement par un modèle $AR(p)$, où p est suffisamment grand pour capter toute la dynamique de $\{X_t\}$:

$$\forall t : X_t = \mu + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + u_t$$

Si l'on utilise la notation de l'opérateur de retard, on obtient :

$$\forall t : (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) X_t = \mu + u_t$$

où le processus stochastique $\{u_t\}$ est un bruit blanc.

X_t est à tendance stochastique si $\varphi_1 + \varphi_2 + \dots + \varphi_p = 1$. Cela implique qu'une racine unitaire figure dans le polynôme caractéristique $1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$ du processus.

Exemple

Soit le processus X_t tel que :

$$X_t = X_{t-1} + \mu + u_t$$

où u_t est un bruit blanc et μ est une dérive. Un tel processus est appelé « marche au hasard avec dérive » ou « marche aléatoire avec dérive ». La précédente équation est équivalente à :

$$X_t - X_{t-1} = \mu + u_t$$

$X_t - X_{t-1}$ est donc stationnaire et il suffit de différencier **une fois** X^t pour obtenir une expression stationnaire. On dit que le processus X_t est intégré d'ordre 1. Il peut être représenté par la relation :

$$(1 - L)X_t = \mu + u_t$$

La racine du polynôme caractéristique $(1 - L)$ est égale à 1. On dit donc que le processus X_t a une racine unitaire. On peut expliciter la valeur du processus et celle de sa variance à chaque période :

$$\begin{aligned} t = 1 : & \quad X_1 = \mu + X_0 + u_1 & \quad \text{et} & \quad E(X_1) = \mu + X_0, \quad V(X_1) = \sigma_u^2 \\ t = 2 : & \quad X_2 = \mu + X_1 + u_2 = 2\mu + X_0 + u_1 + u_2 & \quad \text{et} & \quad E(X_2) = 2\mu + X_0, \quad V(X_2) = 2\sigma_u^2 \\ t = 3 : & \quad X_3 = \mu + X_2 + u_3 = 3\mu + X_0 + u_1 + u_2 + u_3 & \quad \text{et} & \quad E(X_3) = 3\mu + X_0, \quad V(X_3) = 3\sigma_u^2 \\ & \quad \dots & & \end{aligned}$$

Le processus X_t se forme par accumulation des chocs aléatoires u_t , si bien que sa variance augmente avec le temps. Il s'agit d'un cas de non-stationnarité d'origine aléatoire, stochastique : l'accumulation de chocs aléatoires est la tendance stochastique car elle fait augmenter la variance dans le temps. En toute période t , le processus vaut donc :

$$X_t = \mu t + X_0 + (u_1 + u_2 + u_3 + \dots + u_t)$$

Cela implique que :

$$E(X_t) = \mu t + X_0 \quad \text{et} \quad V(X_t) = t\sigma_u^2$$

La somme $(u_1 + u_2 + u_3 + \dots + u_t)$ est la tendance stochastique et la partie déterministe de X_t est $\mu t + X_0$. À la tendance stochastique s'ajoute une tendance déterministe si μ est différent de 0 ; sinon, le processus est une marche aléatoire sans dérive.

Les caractéristiques d'un processus stochastique X_t non stationnaire à tendance stochastique sont :

1. L'espérance de X_t est constante en l'absence de dérive (terme constant) et varie dans le temps en cas de dérive : une tendance déterministe s'ajoute alors à la tendance stochastique.
2. La variance de X_t varie dans le temps.
3. L'effet d'un choc u_t est permanent : il affecte de la même manière toutes les valeurs X_{t+i} ultérieures.

2. Tests de racine unitaire

Lorsqu'on travaille avec une ou plusieurs séries économiques observées, il est important de déterminer, pour chaque série, si le processus stochastique dont elle est la réalisation est :

- stationnaire ;
- non stationnaire à tendance uniquement déterministe (ou stationnaire autour d'une tendance déterministe);
- non stationnaire à tendance stochastique.

L'importance de cette information est évidente dans les cas de figure suivants :

On souhaite spécifier et estimer un modèle où l'évolution d'une série économique n'est « expliquée » que par son passé. Il faut d'abord « stationnariser » la série (établir une transformation stationnaire de cette série) avant de chercher le meilleur modèle *ARMA* qui décrit l'évolution temporelle de cette transformation stationnaire. Or, si une série est non stationnaire, la nature de sa transformation stationnaire diffère en fonction de la tendance (déterministe uniquement ou stochastique) du processus de cette série.

Exemple

Soit un processus stochastique X non stationnaire à tendance stochastique, et tel que $X_t = X_{t-1} + u_t$, où u_t est un bruit blanc. Dans ce cas, la transformation $X_t - X_{t-1}$ est stationnaire. Le modèle *ARMA* appliqué à une telle transformation est alors appelé « modèle *ARIMA* » où le « *I* » rappelle qu'on modélise une transformation stationnaire d'une série « *I* » intégrée.

Soit un processus stochastique X non stationnaire à tendance uniquement déterministe et tel que $X_t = a + bt + u_t$, où u_t est un bruit blanc. Dans ce cas, la transformation $X_t - bt$, ou $X_t - a - bt$, est stationnaire.

-
- On souhaite estimer un modèle linéaire reliant plusieurs séries économiques (par exemple une variable dépendante et une ou plusieurs variables explicatives) et tester des hypothèses sur les paramètres de cette relation. Les techniques d'inférence statistique à utiliser sont différentes en fonction de la nature des processus stochastiques de ces séries :
 - Les techniques d'inférence statistique classique sont valables à condition qu'aucune série de la relation n'ait une tendance stochastique.
 - Les techniques liées à la problématique de la cointégration s'imposent lorsque des séries de la relation ont une tendance stochastique.

Les tests de racine unitaire permettent de tester l'hypothèse H_0 qu'une série économique observée est la réalisation d'un processus stochastique non stationnaire à tendance stochastique, contre l'hypothèse H_1 que ce processus est (i) stationnaire à tendance uniquement déterministe ou (ii) stationnaire.

Ces tests sont basés sur l'estimation préalable d'un modèle autorégressif $AR(p)$ censé approcher le vrai comportement du processus stochastique dont la série observée est une réalisation.

2.2 TESTS DE RACINE UNITAIRE CONTRE STATIONNARITÉ OU TENDANCE DÉTERMINISTE

Cette section décrit les tests de l'hypothèse de racine unitaire contre l'hypothèse de stationnarité ou l'hypothèse de tendance déterministe. Pour détecter si un processus a une tendance stochastique, c'est-à-dire s'il a au moins une racine unitaire, et donc s'il est intégré, on utilise souvent les tests de Dickey et Fuller, basés sur une estimation préalable d'un modèle autorégressif $AR(p)$ avec suffisamment de

retards, donc avec une valeur suffisante de p , pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation.

Parmi les différentes catégories de tests proposés par Dickey et Fuller, on va s'intéresser à la catégorie des tests de la t -stat, les plus utilisés.

Trois types de tests figurent dans cette catégorie selon que le modèle ne contient ni constante ni tendance, qu'il contient une constante mais pas de tendance, ou une constante et une tendance. Ici, on ne présente que deux de ces trois types, ceux dont l'usage est le plus fréquent. On présente ensuite d'autres tests, plus récents, qui ont souvent de meilleures propriétés sur de petits échantillons.

Tests de Dickey et Fuller avec constante et tendance

Ces tests permettent de vérifier l'hypothèse nulle qu'un processus stochastique est à tendance stochastique, avec également une tendance déterministe, contre l'hypothèse que le processus est à tendance uniquement déterministe (donc contre l'hypothèse qu'il est « stationnaire autour d'une tendance déterministe »).

Ces tests sont basés sur l'estimation d'un modèle $AR(p)$, avec constante et tendance, adéquat pour la série à laquelle on s'intéresse, c'est-à-dire avec suffisamment de retards (donc avec une valeur suffisante de p) pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation. Une fois la valeur convenable de p identifiée, on dispose de l'estimation par MCO du modèle $AR(p)$ avec constante et tendance, paramétré de manière à faire apparaître des différences premières :

$$X_t - X_{t-1} = \hat{\mu} + (\hat{\rho} - 1) X_{t-1} + \hat{\lambda}_1 (X_{t-1} - X_{t-2}) + \hat{\lambda}_2 (X_{t-2} - X_{t-3}) + \dots + \hat{\lambda}_{p-1} (X_{t-p+1} - X_{t-p}) + \hat{\beta}t + e_t$$

On teste alors l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\beta = 0$, contre l'hypothèse $H_1 : \rho < 1$ et $\beta \neq 0$, en utilisant la statistique suivante :

$$\tau_\tau = \frac{\hat{\rho} - 1}{S_{\hat{\rho}-1}}$$

Il s'agit du coefficient estimé $\hat{\rho} - 1$ de la variable X_{t-1} , divisé par l'écart type estimé $S_{\hat{\rho}-1}$ de l'estimateur $\hat{\rho} - 1$ de $\rho - 1$. Cette statistique est donc la **t-stat du coefficient de X_{t-1}** dans l'équation estimée précédente.

On prouve en effet que si $\rho = 1$ et $\beta = 0$, τ_τ est distribué selon une table spéciale de Dickey et Fuller⁽³⁾, la même table, quel que soit p . Ce test se réalise du côté gauche de la distribution spéciale de Dickey et Fuller. On prend dans la table la valeur critique telle qu'une réalisation de cette distribution spéciale a 5 chances sur 100 d'être inférieure à la valeur critique, et 95 chances sur 100 d'être supérieure. Si τ_τ est inférieur à la valeur critique (donc plus grand en valeur absolue), on rejette l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\beta = 0$. Si τ_τ est supérieur à la valeur critique (donc plus petit en valeur absolue), on ne la rejette pas. Pour un grand échantillon, la valeur critique est -3.41 .

En fait, des techniques différentes permettent de calculer les valeurs de la table de la distribution de τ_τ , et conduisent à des résultats légèrement distincts. Ainsi Fuller [FUL 1996] et MacKinnon [MAC 1993] proposent des tables différentes de la distribution de τ_τ . Si $\rho = 1$ et donc si tous les λ_i sont nuls, le test de

Dickey et Fuller est appelé « test *DF* avec tendance ». Si $p > 1$, et que des différences retardées figurent donc dans l'équation de base, on parle de « test *ADF*($p - 1$) avec tendance ».

Le test n'est valable que s'il est basé sur l'estimation d'une équation autorégressive avec le p convenable. Il importe donc de bien choisir p . Pour cela, on estime l'équation précédente avec un p assez grand, on la réestime ensuite successivement avec des p plus petits, et l'on s'arrête à la valeur de p qui satisfait l'un des deux critères suivants :

- Il s'agit de la plus petite valeur p pour laquelle le résidu e de l'équation n'est pas autocorrélé.
- Il s'agit de la valeur de p qui optimise un critère d'ajustement statistique du modèle : par exemple celle qui maximise le critère d'information d'Akaike, le critère SBC (Schwarz Bayesian Criterion) ou le critère HQC (Hannan-Quinn Criterion), etc.

Il est indispensable d'utiliser un test *ADF* plutôt qu'un test *DF* dès que le résidu de l'équation de base du test *DF* ($X_t - X_{t-1} = \mu + (\rho - 1)X_{t-1} + \beta t + u_t$) est autocorrélé. En effet, cette autocorrélation induit des distorsions importantes du seuil du test *DF*. Le test *ADF*(p) se comporte alors mieux (moins ou pas de distorsion du seuil du test) si la valeur de p est suffisante, mais perd de la puissance à mesure que p augmente. Si p est grand par rapport à l'effectif n des observations, la puissance du test *ADF*(p) est quasi inexistante.

De manière générale, les tests *DF* ou *ADF* ont peu de puissance sur des petits échantillons (n petit) quand il s'agit de vérifier l'hypothèse $H_0 : \rho = 1$, contre l'hypothèse $H_1 : \rho < 1$, et que le ρ « vrai inconnu » est inférieur à 1 mais proche de 1.

Tests de Dickey et Fuller avec constante sans tendance

Ces tests permettent de tester l'hypothèse nulle qu'un processus stochastique est à tendance purement stochastique, sans avoir également une tendance déterministe, contre l'hypothèse que le processus est stationnaire. Ce test s'utilise donc en pratique pour une série temporelle dont le niveau moyen reste stable dans le temps, qui ne présente donc pas une tendance systématique à augmenter ou à diminuer dans le temps.

De tels tests sont basés sur l'estimation d'un modèle *AR*(p), avec constante et sans tendance, adéquat pour la série à laquelle on s'intéresse, c'est-à-dire avec suffisamment de retards, donc avec une valeur suffisante de p , pour que ce modèle soit une approximation acceptable du vrai processus stochastique dont la série observée est une réalisation. Une fois la valeur convenable de p identifiée, on dispose de l'estimation par *MCO* du modèle *AR*(p) avec constante et tendance, paramétré de manière à faire apparaître des différences premières :

$$X_t - X_{t-1} = \hat{\mu} + (\hat{\rho} - 1) X_{t-1} + \hat{\lambda}_1 (X_{t-1} - X_{t-2}) \\ + \hat{\lambda}_2 (X_{t-2} - X_{t-3}) + \dots + \hat{\lambda}_{p-1} (X_{t-p+1} - X_{t-p}) + e_t$$

On teste alors l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\mu = 0$, contre l'hypothèse $H_1 : \rho < 1$ et $\mu \neq 0$, en utilisant la statistique suivante :

$$\tau_\mu = \frac{\hat{\rho} - 1}{S_{\hat{\rho}-1}}$$

Il s'agit du coefficient estimé $\hat{\rho} - 1$ de la variable X_{t-1} , divisé par l'écart type estimé $S_{\hat{\rho}-1}$ de l'estimateur $\hat{\rho} - 1$ de $\rho - 1$. Cette statistique est donc la **t-stat du coefficient de X_{t-1}** dans l'équation estimée précédente.

On prouve en effet que si $\rho = 1$ et $\mu = 0$, τ_μ est distribué selon une table spéciale de Dickey et Fuller, la même table quel que soit ρ . Ce test se réalise du côté gauche de la distribution spéciale de Dickey et Fuller. On prend dans la table la valeur critique telle qu'une réalisation de cette distribution spéciale a 5 chances sur 100 d'être inférieure à la valeur critique, et 95 chances sur 100 d'être supérieure. Si τ_μ est inférieur à la valeur critique (donc plus grand en valeur absolue), on rejette l'hypothèse de racine unitaire $H_0 : \rho = 1$ et $\mu = 0$. Si τ_μ est supérieur à la valeur critique (donc plus petit en valeur absolue), on ne la rejette pas. Pour un grand échantillon, la valeur critique est -3.41 .

Le test n'est valable que s'il est basé sur l'estimation d'une équation autorégressive avec le p convenable. Il importe donc de bien choisir p . Pour cela, on estime l'équation précédente avec un p assez grand, on la réestime ensuite successivement avec des p plus petits, et l'on s'arrête à la valeur p qui satisfait l'un des deux critères suivants :

- Il s'agit de la plus petite valeur p pour laquelle le résidu e de l'équation n'est pas autocorrélé.
- Il s'agit de la valeur de p qui optimise un critère d'ajustement statistique du modèle : par exemple celle qui minimise le critère d'information AIC d'Akaike.

Si l'on ne rejette pas l'hypothèse de racine unitaire pour X_t avec le test (21), et si ensuite on la rejette pour la variable $X_t - X_{t-1}$ (en lui appliquant par exemple le même test (21)), le résultat du test peut être interprété de la manière suivante : on ne rejette pas l'hypothèse que la première différence de X_t est stationnaire avec une constante nulle (donc que X_t est non stationnaire à tendance stochastique et sans tendance déterministe), contre l'hypothèse que X_t est stationnaire.

Ce test vérifie en effet l'hypothèse que X_t a une espérance constante mais une variance qui augmente, contre l'hypothèse qu'il a une espérance et une variance constantes.

2.3 TESTS DE L'HYPOTHÈSE DE STATIONNARITÉ CONTRE L'HYPOTHÈSE DE RACINE UNITAIRE

Test de Bierens et Guo

Six tests de Bierens et Guo [BIE 1993b] sont disponibles. Les quatre premiers vérifient l'hypothèse H_0 de stationnarité, contre l'hypothèse H_1 de racine unitaire (sans dérive).

$$H_0 : X_t = c + u_t$$

$$H_1 : X_t = X_{t-1} + u_t$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Les deux autres tests de Bierens et Guo vérifient l'hypothèse H_0 de non-stationnarité déterministe, ou de stationnarité autour d'une tendance déterministe, contre l'hypothèse H_1 de racine unitaire avec dérive.

$$H_0 : X_t = c + bt + u_t$$

$$H_1 : X_t = b + X_{t-1} + u_t$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Test de Kwiatkowski, Phillips, Schmidt et Shin

Le premier test de Kwiatkowski, Phillips, Schmidt et Shin [KWI 1992] vérifie l'hypothèse H_0 de stationnarité, contre l'hypothèse H_1 de racine unitaire (sans dérive).

$$H_0 : X_t = c + u_t$$

$$H_1: X_t = X_{t-1} + u_t$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

Le deuxième test [KWI 1992] vérifie l'hypothèse H_0 de non-stationnarité déterministe, ou de stationnarité autour d'une tendance déterministe, contre l'hypothèse H_1 de racine unitaire avec dérive.

$$H_0: X_t = c + dt + u_t$$

$$H_1: X_t = d + X_{t-1} + u_t$$

Dans ces deux équations, u_t est un processus stochastique stationnaire d'espérance nulle.

3. Problèmes économétriques liés aux variables intégrées

Variables intégrées, modèles VAR et cointégration

Cette section définit d'abord la cointégration entre variables intégrées du même ordre. Elle décrit également les tests de cointégration d'Engle et Granger et de Johansen. Grâce à cette structure d'analyse, on peut approfondir la modélisation en définissant les notions de VAR conditionnel à correction d'erreur.

Plusieurs problèmes se posent lorsqu'on veut vérifier l'existence d'une relation linéaire entre plusieurs variables dont certaines ont une tendance stochastique (donc une racine unitaire), et lorsqu'on veut estimer les paramètres de cette relation supposée :

- Même si, dans la réalité, aucune relation linéaire ne lie ces variables, une estimation par MCO peut donner des résultats qui font croire faussement qu'une telle relation existe et qu'elle est importante (R^2 élevé, t -stats significatifs...). C'est le phénomène bien connu de *régression fallacieuse* ou *régression factice*. En fait, l'existence d'une réelle relation à long terme entre des variables intégrées est soumise à certaines conditions, appelées *cointégration entre les variables intégrées*. En d'autres termes, si les variables sont intégrées (ce que l'on vérifie avec les tests de racine unitaire), il faut vérifier leur éventuelle cointégration pour savoir si elles entretiennent réellement une relation à long terme.
- Même en cas de cointégration entre la variable dépendante et les variables explicatives, les estimateurs des moindres carrés ordinaires convergent « trop vite » pour être distribués d'après une loi normale. Par conséquent, les tests d'hypothèse usuels ne suivent pas les lois Student ou asymptotiquement normales : l'inférence statistique classique ne s'applique pas. D'autres techniques d'estimation permettent toutefois de générer des tests d'hypothèse sur les coefficients cointégrants, qui utilisent des distributions connues.

4. Définition de la cointégration

Des processus stochastiques $X_1, X_2 \dots X_w$ **intégrés du même ordre d** sont cointégrés s'il existe une combinaison linéaire de ces processus qui est intégrée d'un ordre inférieur à d .

Il faut donc qu'il existe une valeur $b > 0$ et des valeurs $\beta_1, \beta_2 \dots \beta_w$ vérifiant :

- $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_w X_w$ est $I(d - b)$.
- Chaque variable $X_1, X_2 \dots X_w$ est $I(d)$.

Remarque

En pratique, le cas le plus fréquent est celui où toutes les variables sont intégrées d'ordre 1 (d'ordre 0 ($d = 1$), et où il existe une combinaison linéaire de celles-ci, stationnaire et donc intégrée $b = 1$ et $d - b = 1 - 1 = 0$).

La définition générale permet toutefois de tenir compte d'autres cas possibles, quoique rares. Par exemple, des variables peuvent être chacune $I(2)$ et cointégrées en raison de l'existence d'une combinaison linéaire de celles-ci de type $I(1)$ (dans ce cas, $d = 2$, $b = 1$ et $d - b = 1$) ou $I(0)$ (dans ce cas, $d = 2$, $b = 2$ et $d - b = 0$).

$\beta = (\beta_1, \beta_2 \dots \beta_w)^0$ est le *vecteur cointégrant* (ou de cointégration). $\beta^0 X_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_w X_{wt}$ est l'*expression cointégrée*.

Il peut exister plusieurs vecteurs cointégrants linéairement indépendants pour les mêmes w variables. Le nombre de ces vecteurs est noté r . Dans tous les cas, $r < w$; en d'autres termes on ne peut avoir plus de $w - 1$ vecteurs de cointégration linéairement indépendants.

Toute combinaison linéaire de vecteurs de cointégration pour des variables est aussi un vecteur de cointégration pour ces variables. Tout multiple d'un vecteur de cointégration est également un vecteur de cointégration. Bien entendu, ce nouveau vecteur est linéairement dépendant du premier.

Une même direction vectorielle est représentée par une infinité de vecteurs, tous linéairement dépendants, tous combinaisons linéaires les uns des autres. L'espace de vecteurs de cointégration est donc infini, même si $r = 1$, c'est-à-dire même s'il existe un seul vecteur de cointégration linéairement indépendant. Par « r vecteurs de cointégration linéairement indépendants », on entend donc « r directions vectorielles de cointégration différentes », chacune d'elles étant représentée par une infinité de vecteurs qui sont des multiples les uns des autres. Pour représenter de manière unique chacune de ces r directions vectorielles de cointégration, il faut appliquer une règle de normalisation arbitraire : par exemple, on peut les représenter toutes par leur vecteur dont le premier coefficient (le coefficient de la première variable) vaut 1, ou encore par leur vecteur dont le deuxième coefficient (celui de la deuxième variable) vaut 1, ... On peut aussi représenter la première direction vectorielle (respectivement, la deuxième, la troisième...) par son vecteur dont le premier (respectivement, le deuxième, le troisième...) coefficient vaut 1 (et ainsi de suite). Quand $r > 1$, même si chacune des r directions vectorielles a été normalisée et est donc représentée par un seul vecteur normalisé, toute combinaison linéaire de ces r vecteurs est aussi un vecteur de cointégration. Avoir normalisé les directions initiales ne suffit donc pas à limiter le nombre de vecteurs à r . Pour sélectionner seulement r vecteurs, il faut en plus appliquer une règle arbitraire d'identification, généralement fournie par des restrictions sur les coefficients suggérés par la théorie économique, financière, marketing...

L'exemple le plus fréquent est celui de w variables $X_1, X_2 \dots X_w$, toutes $I(1)$ et cointégrées s'il existe une combinaison linéaire $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_w X_w$ stationnaire, donc une expression $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_w X_w / I(0)$. En pareil cas, chaque variable individuelle est non stationnaire avec une tendance stochastique puisqu'elle est $I(1)$, et a donc tendance à s'éloigner davantage de ses propres conditions initiales, avec de larges fluctuations. Toutefois, étant cointégrées, ces variables divergent ensemble : elles ont tendance à ne pas trop s'éloigner les unes des autres, puisqu'une fonction de ces variables reste stationnaire. Une force économique les relie donc à long terme, de manière qu'elles ne s'écartent pas durablement d'une relation d'équilibre.

La combinaison linéaire $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_w X_w$ représente la déviation par rapport à une relation d'équilibre entre les variables. La cointégration implique que cette déviation est stationnaire: elle fluctue autour

d'une espérance constante, avec une dispersion constante. Cette déviation n'a donc pas tendance à diverger, à devenir trop grande : les variables ne s'écartent pas durablement de leur relation d'équilibre.

Exemple

À long terme, la consommation agrégée réelle doit être proportionnelle au revenu disponible réel des ménages, d'après la théorie économique. À l'équilibre, il faut donc que :

$$C = AY$$

où C est la consommation totale réelle, A une constante et Y le revenu disponible total réel. En effet, la théorie économique montre que, logiquement, le taux de consommation doit être constant à l'équilibre :

$$\frac{C}{Y} = A$$

ou encore que le taux d'épargne doit l'être aussi :

$$\frac{Y - C}{Y} = 1 - A$$

Par conséquent, la relation suivante est vérifiée à l'équilibre :

$$\ln(C) = A' + \ln(Y)$$

où $A' = \ln(A)$. Bien entendu, cette relation prévaut sur le long terme, mais pas à court terme, ni à chaque période t . À court terme, la consommation logarithmique $\ln(C_t)$ peut s'écarter de son équilibre $A' + \ln(Y_t)$ d'un certain montant u_t :

$$\ln(C_t) = A' + \ln(Y_t) + u_t \quad \text{pour tout } t$$

Mais u_t ne peut devenir indéfiniment grand : il doit fluctuer autour de 0, de manière stationnaire, pour permettre à la relation d'équilibre de prévaloir approximativement sur le long terme. Les tests de racine unitaire indiquent que les variables $\ln(C_t)$ et $\ln(Y_t)$ liées par la relation précédente sont intégrées. Cette équation s'écrit encore :

$$\ln(C_t) - \ln(Y_t) = A' + u_t \quad \text{pour tout } t$$

où A' est une constante et u_t un processus stochastique stationnaire. La théorie économique implique ainsi que la combinaison linéaire $\ln(C_t) - \ln(Y_t)$ et $\ln(Y_t)$ sont cointégrées, avec $\beta^1 = 1$ et $\beta^2 = -1$. Les tests de cointégration qui suivent (Y_t est stationnaire, donc que $\ln(C_t)$ permettent de vérifier concrètement, sur les données de la réalité, si cette implication de la théorie économique est vraie.

5. Test de cointégration d'Engle et Granger

5.1 PRINCIPE

Le test d'Engle et Granger est une méthode de vérification de l'existence d'une relation de cointégration entre des variables intégrées et d'estimation de cette relation. Cette méthode est valable sous l'hypothèse arbitraire qu'il existe un seul vecteur de cointégration entre les variables utilisées, et que $b = d$.

Le raisonnement est le suivant : en cas de cointégration entre X_1, X_2, \dots, X_w , il existe des valeurs $\beta_1, \beta_2, \dots, \beta_w$ telles que :

$$\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_w X_w \text{ est } I(0)$$

et donc telles que :

$$X_1 + \frac{\beta_2}{\beta_1} X_2 + \dots + \frac{\beta_w}{\beta_1} X_w \text{ est aussi } I(0)$$

et donc telles que :

$$X_1 + \tilde{\beta}_2 X_2 + \dots + \tilde{\beta}_w X_w \text{ est aussi } I(0)$$

où les coefficients sont normalisés : $\tilde{\beta}_1 = \frac{\beta_1}{\beta_1} = 1, \tilde{\beta}_2 = \frac{\beta_2}{\beta_1} \dots \tilde{\beta}_w = \frac{\beta_w}{\beta_1}$.

Tout processus $I(0)$ est forcément égal à une constante (son espérance) + un processus $I(0)$ d'espérance nulle. La stationnarité de $X_1 + \tilde{\beta}_2 X_2 + \dots + \tilde{\beta}_w X_w$ implique que :

$$X_{1t} + \tilde{\beta}_2 X_{2t} + \dots + \tilde{\beta}_w X_{wt} = \mu + u_t$$

où μ est une constante (non aléatoire) et u un processus stochastique stationnaire $I(0)$ d'espérance nulle. u n'est pas nécessairement un bruit blanc. Il peut être autocorrélé, dès lors qu'il est stationnaire. La cointégration implique donc que :

$$X_{1t} = \mu + (-\tilde{\beta}_2) X_{2t} + \dots + (-\tilde{\beta}_w) X_{wt} + u_t$$

où les coefficients cointégrants normalisés $\tilde{\beta}_2 \dots \tilde{\beta}_w$ sont les seuls pour lesquels u possède les propriétés équivalentes suivantes :

- Il est stationnaire $I(0)$.
- Il n'a pas de tendance stochastique.
- Il n'a pas de racine unitaire.
- Il a une variance constante (voir chapitre 5).

Les coefficients cointégrants de l'équation précédente sont susceptibles d'être estimés par MCO. Cette méthode cherche les valeurs des coefficients qui minimisent la somme des carrés des résidus. Minimiser cette somme revient à chercher les valeurs des coefficients qui minimisent la variance estimée des u_t puisque :

$$\hat{\sigma}_u^2 = \frac{\sum_{t=1}^n e_t^2}{n - w}$$

Or la variance de u est finie uniquement pour les coefficients cointégrants $\tilde{\beta}_2 \dots \tilde{\beta}_w$. Pour d'autres valeurs (d'autres combinaisons linéaires des X_{it}), u est $I(1)$ et a une variance infinie. Par conséquent, en cas de cointégration, et donc s'il existe des coefficients cointégrants $-\tilde{\beta}_2 \dots -\tilde{\beta}_w$ pour lesquels le terme d'erreur u de l'équation précédente a une variance constante, l'estimation de cette équation par MCO doit nécessairement fournir des valeurs estimées de ces coefficients cointégrants.

Tester la cointégration entre les variables revient donc à estimer par MCO une équation linéaire où l'une de ces variables est régressée sur les autres et à tester si le résidu a une racine unitaire. Si l'on ne rejette pas l'hypothèse de racine unitaire pour le résidu estimé, les variables de l'équation ne sont pas cointégrées ; si on la rejette, elles le sont.

Toutefois, les variables étant intégrées, l'inférence statistique classique ne s'applique pas. Les t -ratios des coefficients estimés n'ont pas une distribution Student pour des petits échantillons, ni une distribution gaussienne pour de grands échantillons, sous l'hypothèse de nullité du coefficient vrai correspondant. Les probabilités critiques des t -ratios, indiquées par défaut par les logiciels et calculées à partir d'une

distribution Student, ne sont donc pas fiables. À ce niveau, il est donc impossible de réaliser des tests d'hypothèse et des intervalles de confiance, pour vérifier s'ils admettent des valeurs des coefficients plausibles économiquement.

5.2 COINTÉGRATION ET MÉCANISME À CORRECTION D'ERREUR

Selon le théorème de représentation de Granger, en cas de cointégration entre des variables intégrées d'ordre 1, l'évolution de chacune d'elles est régie nécessairement par un *modèle à correction d'erreur*. Pour chaque variable cointégrée, un modèle linéaire existe donc, où la variation de la variable est une fonction de ses variations passées et des variations passées des autres variables, ainsi que de la valeur passée de l'expression cointégrée :

$X_1, X_2 \dots X_w$ cointégrées $\Leftrightarrow \forall i = 1, \dots, w$:

$$\Delta X_{it} = \mu_i + \sum_{j=1}^w \sum_{\theta=1}^r \varphi_{ij\theta} \Delta X_{jt-\theta} + \lambda_i \left(X_{1t-1} - \tilde{\beta}_2 X_{2t-1} - \dots - \tilde{\beta}_w X_{wt-1} \right) + u_{it}$$

Où u_{it} est un processus stochastique de type bruit blanc. Un tel modèle est dit «à correction d'erreur» (on parle aussi de mécanisme à correction d'erreur).

L'expression cointégrée retardée, notée $\left(X_{1t-1} - \tilde{\beta}_2 X_{2t-1} - \dots - \tilde{\beta}_w X_{wt-1} \right)$, représente l'écart (ou l'« erreur ») par rapport à la relation d'équilibre⁽⁴⁾ à la période $t - 1$. Le paramètre λ_i correspond donc au taux de réaction de la variable X_i à l'écart précédent par rapport à l'équilibre. λ_i mesure ainsi l'intensité avec laquelle la variable X_i varie en t pour corriger l'erreur de la période $t - 1$ par rapport à l'équilibre. C'est pourquoi on parle de modèle à correction d'erreur.

6. Tests de cointégration de Johansen

6.1 PRINCIPE

Les tests de Johansen permettent de vérifier des hypothèses sur le nombre de vecteurs de cointégration dans un système $VAR(p)$ reliant des variables qui sont toutes intégrées du même ordre.

On part d'un vecteur X_t (rassemblant n variables) dont l'évolution est décrite par un modèle $VAR(p)$ du type :

$$X_t = \mu + \sum_{l=1}^p \Phi_l X_{t-l} \left(+ \sum_{i=1}^{s-1} w_i Q_{it} \right) + u_t$$

On peut le reparamétriser ainsi :

$$\Delta X_t = \mu - \Phi(1)X_{t-1} + \sum_{l=1}^{p-1} A_l \Delta X_{t-l} + \left(\sum_{i=1}^{s-1} w_i Q_{it} \right) + u_t$$

avec :

$$A_l = - \sum_{j=l+1}^p \Phi_j \quad \forall l = 1, 2, \dots, p-1$$

$$\Phi(1) = I_n - \sum_{j=1}^p \Phi_j$$

Les variables Q_{it} sont d'éventuelles variables indicatrices (ou dummy), saisonnières (voir chapitre 3), s représentant la fréquence des données (12 pour des données mensuelles par exemple). On peut rarement ajouter une tendance linéaire t . On suppose que $u_t \sim N_n(0, \Sigma)$ et que les u_t de périodes différentes sont indépendants. On suppose aussi que $X_t \sim I(1)$, donc que toutes les variables sont intégrées d'ordre 1. Sous l'hypothèse de normalité de u , la méthode de Johansen permet de tester l'hypothèse de l'existence d'un effectif r quelconque de relations cointégrées, allant de 0 à $n-1$.

Johansen montre que les estimateurs de maximum de vraisemblance des paramètres d'un modèle VAR cointégré sont fonction de quelques matrices assemblées à partir des résultats de l'estimation de régressions auxiliaires par MCO. Pour appliquer la méthode de Johansen, on réalise donc d'abord l'estimation par MCO du système d'équations linéaires suivant :

$$\Delta X_{t-1} = \hat{C}_0 + \hat{C}_{01} \Delta X_{t-1} + \dots + \hat{C}_{0p-1} \Delta X_{t-p-1} + \left(\sum_{i=1}^{s-1} \hat{w}_i Q_{it} \right) + R_{0t}$$

où C_0 est une matrice de type $m \times 1$, chaque C_{0i} est de type $m \times m$ et R_{0t} est un vecteur de m résidus au temps t .

On réalise également l'estimation par MCO du système d'équations linéaires suivant :

$$X_{t-1} = \hat{C}_1 + \hat{C}_{11} \Delta X_{t-1} + \dots + \hat{C}_{1p-1} \Delta X_{t-p-1} + \left(\sum_{i=1}^{s-1} \hat{w}_i Q_{it} \right) + R_{1t}$$

où C_1 est une matrice de type $n \times 1$, chaque C_{1i} est de type $n \times n$ et R_{1t} est un vecteur de m résidus au temps t .

On utilise alors les vecteurs R_{0t} et R_{1t} de résidus calculés pour former les matrices S_{01} , S_{00} , S_{10} et S_{11} , de type $N \times N$, de la manière suivante :

$$S_{ij} = \left(\frac{1}{N} \right) \sum_{t=1}^N R_{it} R'_{jt}$$

où N est l'effectif des observations disponibles (les variables X_i sont observées sur N périodes).

Pourquoi tous ces calculs préliminaires? On montre aisément que la fonction de vraisemblance du système VAR(p) avec $\Phi(1) = -\alpha\beta^0$ est maximisée si l'on choisit les estimateurs :

$$\begin{aligned} \hat{\Sigma}^{MV} &= S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10} \\ \hat{\alpha}^{MV} &= S_{01} \beta (\beta' S_{11} \beta)^{-1} \end{aligned}$$

À ce stade, ces estimateurs sont fonction de β , qui est lui-même une matrice de paramètres inconnus qu'il faut estimer. Si l'on a un estimateur de maximum de vraisemblance pour β , il suffit de le remplacer dans les formules précédentes pour disposer des estimateurs de maximum de vraisemblance de Σ et α .

La fonction de vraisemblance du système VAR(p) avec $\Phi(1) = \alpha\beta$, où α et β ont r colonnes et un rang r , est maximisée si l'on choisit $\beta^{MV} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r)$, où $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r$ sont les vecteurs propres associés aux r valeurs propres $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r$ dont le vecteur $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r)'$ est solution du système d'équations suivant :

$$\left| \hat{\lambda} S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0$$

Elles sont classées par ordre décroissant : $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_r$. Les vecteurs propres $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r$ sont normalisés de manière à ce que leur matrice $V = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_r)$, de type $r \times r$, vérifie :

$$I_r = V' S_{11} V$$

Les r vecteurs cointégrants estimés, donc les r colonnes de β^{MV} , sont ainsi les r premiers vecteurs propres (les vecteurs associés aux r plus grandes valeurs propres parmi les r valeurs propres), parmi les r vecteurs propres existants. (Les r vecteurs de cointégration sont les r vecteurs propres les plus significatifs parmi les m vecteurs propres.) Ces calculs d'estimateurs sont bien sûr conditionnels au choix d'un rang r . Lequel choisir? Pour déterminer le rang r de cointégration, Johansen propose deux tests :

a) Le test de la trace, $LR(r|n) = -N \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i)$ permet de vérifier l'hypothèse H_0 : rang de $\beta = r$ (r vecteurs cointégrants linéairement indépendants) contre l'hypothèse H_1 : rang de $\beta = n$ (n vecteurs cointégrants linéairement indépendants). Sous l'hypothèse H_0 , ce test a une distribution particulière, tabulée par Johansen.

b) Le test de la plus grande valeur propre, $LR(r|n) = -N \ln(1 - \hat{\lambda}_{r+1})$, permet de vérifier l'hypothèse H_0 : rang de $\beta = r$ (r vecteurs cointégrants linéairement indépendants) contre l'hypothèse H_1 : rang de $\beta = r + 1$ ($r + 1$ vecteurs cointégrants linéairement indépendants). Sous l'hypothèse H_0 , ce test a une distribution particulière tabulée par Johansen.

Ces deux outils sont valables lorsque μ est laissé sans restriction, ce qui permet à X_t d'avoir une tendance linéaire. Ils se déclinent en d'autres versions :

- On peut imposer *a priori* que $\alpha \perp \mu = 0$, c'est-à-dire qu'il existe un vecteur β_0 de type $r \times 1$ tel que $\mu = \alpha \beta_0$; sous les mêmes hypothèses H_0 précédentes, les tests ont des distributions particulières tabulées par Johansen.
- On peut également imposer que $\mu = 0$.
- On peut ajouter une tendance linéaire dans le modèle et en contraindre le coefficient pour éviter que les variables de X_t aient une tendance quadratique.

7. Modèles à correction d'erreur conditionnel

Le vecteur u_t des termes d'erreur du système dynamique cointégré est supposé non autocorrélé et distribué d'après une normale, de matrice de variance et de covariance Σ c'est-à-dire : $u_t \sim N(0, \Sigma)$. Il en résulte la distribution conditionnelle suivante de ΔX_t par rapport à « son passé » :

$$\Delta X_t | X_{t-1}, X_{t-2}, \dots \sim N(\mu + \alpha \beta' X_{t-1} + A_1 \Delta X_{t-1}, \Sigma)$$

On partitionne le vecteur X_t en $X_t = \begin{pmatrix} Y_t \\ Z_t \end{pmatrix}$ où Y_t est un vecteur $m \times 1$ et Z_t un vecteur $(n-m) \times 1$:

$$Y_t = \begin{pmatrix} X_{1t} \\ \vdots \\ X_{mt} \end{pmatrix}, \quad Z_t = \begin{pmatrix} X_{(m+1)t} \\ X_{(m+2)t} \\ \vdots \\ X_{nt} \end{pmatrix}$$

On partitionne de même le vecteur μ et la matrice Σ :

$$\mu^1 = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}, \quad \mu^2 = \begin{pmatrix} \mu_{m+1} \\ \mu_{m+2} \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\Sigma_{11} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}, \quad \Sigma_{12} = \begin{pmatrix} \sigma_{1(m+1)} & \sigma_{1(m+2)} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \ddots & \sigma_{2n} \\ \sigma_{m(m+1)} & \sigma_{m(m+2)} & \cdots & \sigma_{mn} \end{pmatrix},$$

$$\Sigma_{21} = \begin{pmatrix} \sigma_{(m+1)1} & \cdots & \sigma_{(m+1)m} \\ \sigma_{(m+1)2} & \cdots & \sigma_{(m+1)m} \\ \vdots & \ddots & \vdots \\ \sigma_{(m+1)2} & \cdots & \sigma_{(m+1)2} \end{pmatrix} \text{ et } \Sigma_{22} = \begin{pmatrix} \sigma_{(m+1)(m+1)} & \sigma_{(m+1)(m+2)} & \cdots & \sigma_{(m+1)n} \\ \sigma_{(m+2)(m+1)} & \sigma_{(m+2)(m+2)} & \cdots & \sigma_{(m+2)n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n(m+1)} & \sigma_{n(m+2)} & \cdots & \sigma_{nn} \end{pmatrix}$$

On partitionne également les matrices α et A_1 :

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

Où α_1 est une matrice $m \times r$ contenant les m premières lignes de α , et α_2 est une matrice $(n-m) \times r$ contenant les $(n-m)$ dernières lignes de α .

$$A_1 = \begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix}$$

Où A_{11} est une matrice $m \times n$ avec les m premières lignes de A , et A_{21} est une matrice $(n-m) \times n$ avec les $n-m$ dernières lignes de A .

L'hypothèse de normalité évoquée précédemment permet de déduire la distribution conditionnelle de ΔY_t conditionnellement à Z_t et au passé de X , et la distribution marginale de ΔZ_t :

$$\begin{aligned} & \Delta Y_t | Z_t, X_{t-1}, X_{t-2} \dots \\ & \sim N \left(\begin{matrix} \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (\Delta Z_t - \mu_2) + (\alpha_1 - \Sigma_{12} \Sigma_{22}^{-1} \alpha_2) \beta' X_{t-1} + (A_{11} - \Sigma_{12} \Sigma_{22}^{-1} A_{21}) \Delta X_{t-1} \\ \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \end{matrix} \right) \end{aligned}$$

$$\Delta Z_t | Z_{t-1}, Z_{t-2} \dots \sim N (\mu_2 + \alpha_2 \beta' X_{t-1} + A_{21} \Delta X_{t-1}, \Sigma_{22})$$

On écrit encore :

$$\begin{aligned} \Delta Y_t &= (\mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2) + (\Sigma_{12} \Sigma_{22}^{-1}) \Delta Z_t + (\alpha_1 - \Sigma_{12} \Sigma_{22}^{-1} \alpha_2) \beta' X_{t-1} \\ &\quad + (A_{11} - \Sigma_{12} \Sigma_{22}^{-1} A_{21}) \Delta X_{t-1} + u_{1,2t} \\ \Delta Z_t &= \mu_2 + \alpha_2 \beta' X_{t-1} + A_{21} \Delta X_{t-1} + u_{2t} \end{aligned}$$

Où :

$$u_{1,2t} = u_t^1 - \Sigma_{12} \Sigma_{22}^{-1} u_t^2.$$

La première équation matricielle est un *modèle à correction d'erreur conditionnel en forme réduite* tandis que la deuxième est un *modèle à correction d'erreur marginal*. L'une et l'autre sont un simple reparamétrage du système VAR cointégré original. Elles contiennent donc la même information, mais présentée différemment.

$(\mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2)$ et $u_{1,2t}$ sont des vecteurs $m \times 1$, $\Sigma_{12} \Sigma_{22}^{-1}$ est une matrice $m \times (n-m)$, $(\alpha_1 - \Sigma_{12} \Sigma_{22}^{-1} \alpha_2)$ est une matrice $m \times m$, $\beta' X_{t-1}$ est un vecteur $m \times 1$, $(\alpha_1 - \Sigma_{12} \Sigma_{22}^{-1} \alpha_2) \beta'$ est une matrice $m \times n$, et $(A_{11} - \Sigma_{12} \Sigma_{22}^{-1} A_{21})$ une matrice $m \times n$.

8. les modèles à volatilité conditionnelle autorégressive

Les modèles à volatilité variable sont particulièrement adaptés à la modélisation des rendements des actifs financiers. Lorsqu'on observe une série temporelle financière de haute fréquence (variation du taux de change, du cours boursier...), on remarque généralement que sa volatilité, représentée statistiquement par la variance, n'est pas constante dans le temps, en d'autres termes, que sa variance change dans le temps. Le processus sous-jacent révèle donc une hétéroscédasticité. En général, on constate que des périodes assez longues où la volatilité est faible sont suivies de périodes assez longues où elle est forte. En fait, des dates où la volatilité est faible ont tendance à se succéder, et des dates où elle est forte ont également tendance à se suivre. On dit également que la volatilité, ou la variance du processus, est autorégressive : elle dépend de ses valeurs passées. Les processus $ARCH(p)$ et $GARCH(p,q)$ sont définis de manière à pouvoir représenter ce type de série observée.

8.1 LES PROCESSUS $ARCH(p)$

Un processus $ARCH(p)$ est un processus à hétéroscédasticité conditionnelle autorégressive d'ordre p . Un processus stochastique $\{X_t\}$ est un processus $ARCH(p)$ quand :

$$X_t = \mu + \varepsilon_t$$

$$\text{où } f_{\varepsilon_t | X_{t-1}, X_{t-2}, \dots} \sim N(0, h_t)$$

avec :

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

sous la condition $\alpha_i \geq 0$ pour tout i

Ce processus $\{X_t\}$ est donc tel que :

$$f_{X_t | X_{t-1}, X_{t-2}, \dots} \sim N(\mu, h_t)$$

avec :

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i (X_{t-i} - \mu)^2$$

sous la condition $\alpha_i \geq 0$ pour tout i

La distribution conditionnelle de X_t , conditionnellement à ses valeurs passées, a ainsi une variance qui dépend des valeurs passées de X_t .

L'espérance conditionnelle de X_t est fournie par :

$$E(X_t | X_{t-1}, X_{t-2}, \dots) = \mu \quad \forall t$$

La variance conditionnelle de X_t est fournie par :

$$V(X_t | X_{t-1}, X_{t-2}, \dots) = h_t \quad \forall t$$

La distribution conditionnelle de X_t à son passé est normale :

$$f_{X_t | X_{t-1}, X_{t-2}, \dots} \sim N(\mu, h_t)$$

Mais la distribution non conditionnelle, ou marginale, de X_t ne l'est pas :

$$f_{X_t} \neq N$$

L'espérance non conditionnelle de X_t est fournie par :

$$E(X_t) = \mu$$

La variance non conditionnelle de X_t est fournie par :

$$V(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i}$$

quand toutes les racines du polynôme en $L, \sum_{i=1}^p \alpha_i L^{p-i}$, sont en dehors du cercle unité.

Remarque : le processus X_t peut s'écrire ainsi :

$$X_t = \mu + W_t U_t \quad \text{où } U_t \sim N(0,1) \quad \forall t$$

et est indépendant d'une période à l'autre :

$$W_t = \left\{ \alpha_0 + \sum_{i=1}^p \alpha_i (X_{t-i} - \mu)^2 \right\}^{\frac{1}{2}}$$

Exemple (Processus ARCH(1))

$$X_t = \mu + \varepsilon_t$$

où $f_{\varepsilon_t} | X_{t-1}, X_{t-2}, \dots \sim N(0, h_t)$ avec $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

et donc :

$$f_{X_t} | X_{t-1}, X_{t-2}, \dots \sim N(\mu, h_t)$$

avec $h_t = \alpha_0 + \alpha_1 (X_{t-1} - \mu)^2$

8.2. LES PROCESSUS GARCH(p,q)

Il s'agit d'une généralisation d'un processus ARCH(p) :

$$X_t = \mu + \varepsilon_t$$

Où $f_{\varepsilon_t} | X_{t-1}, X_{t-2}, \dots \sim N(0, h_t)$. Avec $h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$ sous les conditions $\alpha_i \geq 0$ et $\beta_j \geq 0$ pour tout i

X_t est donc un processus tel que :

$$f_{X_t} | X_{t-1}, X_{t-2}, \dots \sim N(\mu, h_t)$$

avec

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i (X_{t-i} - \mu)^2 + \sum_{j=1}^q \beta_j h_{t-j}$$

On montre que cela implique que ε_t^2 suit un processus ARMA(m,q) avec $m = \max(p,q)$:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + v_t - \sum_{j=1}^q \beta_j v_{t-j} \quad \text{où } v_t = \varepsilon_t^2 - h_t$$

Exemple (Processus GARCH(1,1))

$$X_t = \mu + \varepsilon_t$$

où $f_{\varepsilon_t} | X_{t-1}, X_{t-2}, \dots \sim N(0, h_t)$ avec $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$ et donc :

$$f_{X_t} | X_{t-1}, X_{t-2}, \dots \sim N(\mu, h_t)$$

$$\text{avec } h_t = \alpha_0 + \alpha_1 (X_{t-1} - \mu)^2 + \beta_1 h_{t-1}$$

Exemple (Cas particulier des processus GARCH(1,1) ou processus IGARCH(1,1))

Lorsque $\alpha_0 = 0$, $\alpha_1 + \beta_1 = 1$, on parle de GARCH intégré ou IGARCH(1,1) :

$$h_t = \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1) h_{t-1}$$

ce qui implique :

$$E(h_{t+\theta} | \varepsilon_t, \varepsilon_{t-1}, \dots) = h_{t+1} \forall \theta \geq 1$$

La variance conditionnelle h_t se caractérise par une racine unitaire.