

Économétrie des séries temporelles non-stationnaires

Chapitre 2: Modélisation des séries non-stationnaires

Gilles de Truchis

Master 2 EIPMC

Les chapitres du cours

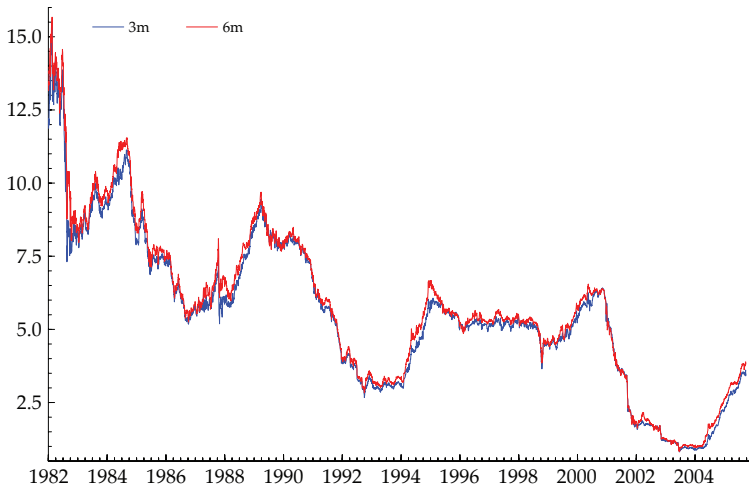
- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

- | | | | |
|---|-----------------------------|---|-------------------|
| 1 | Théorie de la cointégration | 4 | Analyse canonique |
| 2 | Engle et Granger | 5 | Johansen |
| 3 | Endogénéité | 6 | Sum up |

Intuition graphique

M.Ø. Nielsen / Journal of Econometrics 155 (2010) 170–187



Time series plots of U S interest rates covering all business days from 1/4/82 to 9/23/05

Intuition de la cointégration

- Dans le modèle (1) une combinaison de variables $I(1)$ est $I(0)$
- En renversant le raisonnement on obtient l'idée de la cointégration

Definition (Cointégration bivariée et paradigme $I(1)/I(0)$)

Soit y_t et x_t deux variables $I(1)$. S'il existe un coefficient $\beta \neq 0$ produisant une combinaison linéaire

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

tel que $\varepsilon_t \sim I(0)$ alors y_t et x_t possèdent un équilibre de long terme commun et sont cointégrées

- Par son existence, cette équilibre de long terme commun assure que la régression n'est pas factice

Cointégration et vecteur cointégrant

- Dans un papier fondateur, Engle et Granger (1987) définissent ainsi la cointégration

Definition (Cointégration selon Engle et Granger, 1987)

Les éléments du vecteur de variables aléatoires $X_t = (x_{1,t}, \dots, x_{k,t})$ sont dits cointégrés d'ordres (δ, b) , noté $CI \sim (\delta, b)$ si

- a) *tous les éléments du vecteur X_t sont $I(\delta)$*
- b) *il existe un vecteur $\beta \neq 0$ tel que $\varepsilon_t = \beta' X_t \sim I(\delta - b)$ avec $b > 0$*

Le vecteur β est de dimension k et se nomme vecteur cointégrant.

- Dans ce chapitre, on supposera $\delta = 1$ et $b = 1$

Composition du vecteur cointégrant

- Pour mieux comprendre le modèle $\varepsilon_t = \beta' X_t$, détaillons β
- Le plus souvent on normalisera β_1 à 1 est on s'attend à avoir

$$\beta' = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ \vdots \\ -\beta_k \end{pmatrix} \text{ et } X_t = (x_{1,t}, \dots, x_{k,t})$$

de sorte que $\varepsilon_t = \beta' X_t$ s'écrit

$$\varepsilon_t = x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_k x_{k,t} \text{ ou } x_{1,t} = \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t$$

- Dans le cas le plus simple, on a $x_{1,t} = \beta_2 x_{2,t} + \varepsilon_t$

Représentation VAR

- Considérons un modèle VAR(1) bivarié

$$X_t = \Phi X_{t-1} + \varepsilon_t \quad (2)$$

avec $(x_{1,t}, x_{2,t})$, $\varepsilon_t \sim$ i.i.d. $(0, \sigma_\varepsilon^2 < \infty)$ et

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$$

- A partir du polynôme caractéristique $\det(\Phi - zI) = 0$ on sait que

- La relation est factice si les deux racines sont sur le cercle unité

- X_t est stationnaire si les racines ne sont pas sur le cercle unité

\Rightarrow Étudions le cas intermédiaire où $z_1^{(\Phi)} = 1$ et $z_2^{(\Phi)} = \lambda$ avec $|\lambda| < 1$

VAR et diagonalisation

■ Puisque $z_1^{(\Phi)} = 1$, $x_{1,t} \sim I(1)$ et $x_{2,t} \sim I(1)$

■ Puisque $z \neq 0$, Φ est de rang plein (valeurs propres non-nulles)

$\Rightarrow \Phi$ admet une représentation de la forme

$\Phi = P\Lambda P^{-1}$ avec $\Lambda = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ et $P = \begin{pmatrix} \beta & \alpha \\ \gamma & \delta \end{pmatrix}$, des vecteurs propres

■ Sans perte de généralité on suppose $\det P = \beta\delta - \gamma\alpha = 1$ et donc

$$P^{-1} = \begin{pmatrix} \delta & -\alpha \\ -\gamma & \beta \end{pmatrix}$$

\Rightarrow On en déduit aisément une expression pour Φ

$$\Phi = \begin{pmatrix} \beta & \alpha \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \delta & -\alpha \\ -\gamma\lambda & \beta\lambda \end{pmatrix} = \begin{pmatrix} \beta\delta - \lambda\alpha\gamma & -\beta\alpha(1 - \lambda) \\ \gamma\delta(1 - \lambda) & -\gamma\alpha + \lambda\beta\delta \end{pmatrix} \quad (3)$$

- Soit une matrice carrée \mathcal{M} avec $\dim(\mathcal{M}) = k \times k$
- Les solutions de la représentation

$$\mathcal{M}\mathbf{v} = \lambda\mathbf{v} \text{ ou } (\mathcal{M} - \lambda I)\mathbf{v} = 0$$

donne λ un scalaire nommé valeur propre et \mathbf{v} un vecteur propre

- Rappelons que cette représentation admet une solution *ssi*

$$\det(\mathcal{M} - \lambda I) = 0$$

- Il s'agit d'un polynôme caractéristique qui se factorise comme

$$\det(\mathcal{M} - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_k - \lambda)$$

avec $\lambda_1, \dots, \lambda_k$ les k valeurs propres associées à \mathcal{M}

VAR cointégré

- En multipliant tous les termes de X_t par P^{-1} on obtient

$$P^{-1}X_t = P^{-1}\Phi X_{t-1} + P^{-1}\varepsilon_t$$

- On observe alors que $P^{-1}\Phi = P^{-1}P\Lambda P^{-1} = \Lambda P^{-1}$

- En notant $W_t = (w_{1t}, -w_{2t})' = P^{-1}X_t$ on obtient

$$W_t = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} W_{t-1} + \eta_t, \text{ avec } \eta_t = P^{-1}\varepsilon_t$$

- A partir de Λ , on voit que pour ce nouveau VAR,

$$\underbrace{w_{1t} = \delta x_{1t} - \alpha x_{2t}}_{\text{non-stationnaire}} \text{ et } \underbrace{-w_{2t} = \gamma x_{1t} - \beta x_{2t}}_{\text{stationnaire}}$$

- On sait que $(x_{1t}, x_{2t})' \sim I(1)$ mais $w_{2t} \sim I(0)$ car $|\lambda| < 1$

$\Rightarrow X_t \sim CI(1, 1)$ de vecteur cointégrant $(\gamma, -\beta)'$

Représentation en tendance stochastique commune

- A partir de $W_t = P^{-1}X_t$ on peut exprimer X_t comme

$$X_t = PW_t$$

- On a alors,

$$x_{1t} = \beta \underbrace{w_{1t}}_{I(1)} - \alpha \underbrace{w_{2t}}_{I(0)} \text{ et } x_{2t} = \gamma \underbrace{w_{1t}}_{I(1)} - \delta \underbrace{w_{2t}}_{I(0)}$$

$\Rightarrow X_t = (x_{1t}, x_{2t})'$ contient donc deux composantes

- une composante stationnaire $I(0) : w_{2t}$

- une composante non-stationnaire $I(1) : w_{1t}$

\Rightarrow cointégration = existence d'une tendance stochastique commune

Représentation ECM

- Pour rappel, $\det P = \beta\delta - \gamma\alpha = 1$ qu'on peut réécrire comme

$$\beta\delta - \gamma\alpha + \gamma\alpha - \lambda\gamma\alpha = 1 + \gamma\alpha - \lambda\gamma\alpha \Rightarrow \beta\delta - \lambda\gamma\alpha = 1 + \gamma\alpha(1 - \lambda)$$

ou encore

$$\beta\delta - \gamma\alpha - \beta\delta + \lambda\beta\delta = 1 - \beta\delta + \lambda\beta\delta \Rightarrow -\gamma\alpha + \lambda\beta\delta = 1 - \beta\delta(1 - \lambda)$$

- Dans l'expression de Φ (équation 3), cette réécriture donne

$$\Phi = \begin{pmatrix} \beta\delta - \lambda\alpha\gamma & -\beta\alpha(1 - \lambda) \\ \gamma\delta(1 - \lambda) & -\gamma\alpha + \lambda\beta\delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - \lambda) \begin{pmatrix} \alpha \\ \delta \end{pmatrix} \begin{pmatrix} \gamma & -\beta \end{pmatrix}$$

$\Rightarrow X_t = \Phi X_{t-1} + \varepsilon_t$ possède une forme à correction d'erreur (ECM)

$$X_t - X_{t-1} = (1 - \lambda) \begin{pmatrix} \alpha \\ \delta \end{pmatrix} \underbrace{(\gamma x_{1,t-1} - \beta x_{2,t-1})}_{\text{EC : correction d'erreur}} + \varepsilon_t \quad (4)$$

\Rightarrow EC = correction du déséquilibre de long terme passé

Interprétation de l'ECM

- Afin d'interpréter les coefficients de l'ECM, réécrivons ΔX_t

$$\Delta X_t = (1 - \lambda) \begin{pmatrix} \alpha \\ \delta \end{pmatrix} (\gamma \quad -\beta) X_{t-1} + \varepsilon_t = \alpha \beta' X_{t-1} + \varepsilon_t \quad (5)$$

- avec α la force de rappel vers l'équilibre de long terme

$$\alpha = \begin{pmatrix} (1 - \lambda)\alpha \\ (1 - \lambda)\delta \end{pmatrix} \quad (6)$$

- et β le vecteur cointégrant décrivant l'équilibre de long terme

$$\beta = \begin{pmatrix} \gamma \\ -\beta \end{pmatrix} \quad (7)$$

- Par convention, on note $\Pi = \alpha \beta'$ et on parle de VECM(0)

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t \quad (8)$$

Cointégration et rang réduit

- Remarquons tout d'abord que

$$\Pi = \alpha\beta' = \Phi - I \quad (9)$$

- D'après les racines de $|\Phi - zI| = 0$, on déduit que

$$z_1^{(\Pi)} = 0 \text{ et } z_2^{(\Pi)} = (\beta\delta - \alpha\gamma)(\lambda - 1) = \lambda - 1$$

- La valeur propre $z_1^{(\Pi)}$ étant nulle, Π est de rang réduit non nul

\Rightarrow le rang de Π est révélateur de la présence de cointégration ou non

\Rightarrow Si $z_1^{(\Phi)} = 1$ et $|z_2^{(\Phi)}| < 1$, $\text{rang}(\Pi) = 1$ et il y a cointégration

\Rightarrow Si $z_1^{(\Phi)} = z_2^{(\Phi)} = 1$, $\text{rang}(\Pi) = 0$ et il n'y a pas de cointégration

\Rightarrow Si $|z_1^{(\Phi)}| < 1$ et $|z_2^{(\Phi)}| < 1$, $\text{rang}(\Pi) = 2$ et le VAR est stationnaire

Engle et Granger (1987), Johansen (1991), Hansen (2005)

- Soit un VAR k -varié

$$X_t = \Phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma_\varepsilon^2 < \infty) \quad (10)$$

Theorem (Théorème de représentation de Granger)

Soit un VAR de dimension k décrit par l'équation (10) et admettant une forme à correction d'erreur décrite par l'équation (8) avec $\Pi = \alpha\beta'$, le polynôme caractéristique de X_t ayant $k - r$ racines égales à l'unité et toutes les autres en dehors du cercle unité. Alors,

- ΔX_t est $I(0)$ de représentation de Wold $\Delta X_t = C(L)\varepsilon_t$
- $\beta'X_t$ est $I(0)$ et le rang de cointégration de X_t est r , $0 \leq r < k$
- X_t est $I(1)$ de forme $X_t = X_0 + C(1) \sum_{j=1}^t \varepsilon_j$
- $\alpha_\perp \sum_{j=1}^t \varepsilon_j$ représente les tendances stochastiques communes (α_\perp étant le complément orthogonal de α : $\alpha'_\perp \alpha = 0$)

Le théorème énoncé ici est très simplifié. La preuve dans sa version originale comme dans sa version simplifiée dépasse le niveau M2.

VECM et dynamique de court terme

- Le modèle (8) est simple et sans dynamique de court terme

⇒ Supposons un système cointégré bivarié de la forme

$$\begin{aligned}\Delta x_{1t} &= \alpha_1(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{11} \Delta x_{1,t-1} + \gamma_{12} \Delta x_{2,t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= \alpha_2(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{21} \Delta x_{1,t-1} + \gamma_{22} \Delta x_{2,t-1} + \varepsilon_{2t}\end{aligned}\quad (11)$$

- Sous forme vectoriel, il s'agit d'un VECM($p = 1$)

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t$$

avec les coefficients de la dynamique de court terme

$$\Gamma_1 = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

et $\beta = (\gamma, -\beta)' = (1, -\beta_2)'$

Coefficients attendus de l'ECM

$$\Delta x_{1t} = \alpha_1(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{11}\Delta x_{1,t-1} + \gamma_{12}\Delta x_{2,t-1} + \varepsilon_{1t}$$

$$\Delta x_{2t} = \alpha_2(x_{1,t-1} - \beta_2 x_{2,t-1}) + \gamma_{21}\Delta x_{1,t-1} + \gamma_{22}\Delta x_{2,t-1} + \varepsilon_{2t}$$

■ On s'attend à trouver $\alpha_1 < 0$ car

■ on suppose qu'on était à l'équilibre de long terme et que $x_{2t} \nearrow$

$\Rightarrow z_{t-1} = x_{1,t-1} - \beta_2 x_{2,t-1} < 0 \Rightarrow$ déséquilibre de long terme négatif

■ si $\alpha_1 < 0$, $\alpha_1 z_{t-1} > 0$ on observera en t une augmentation de x_{1t}

\Rightarrow cela viendra compenser le déséquilibre

■ On s'attend à trouver $\alpha_2 > 0$ car

■ on suppose qu'on était à l'équilibre de long terme et que $x_{1t} \nearrow$

$\Rightarrow z_{t-1} = x_{1,t-1} - \beta_2 x_{2,t-1} > 0 \Rightarrow$ déséquilibre de long terme positif

■ si $\alpha_2 > 0$, $\alpha_2 z_{t-1} > 0$ on observera en t une augmentation de x_{2t}

\Rightarrow cela viendra compenser le déséquilibre

Représentation VAR d'un VECM

- Précédemment, en partant d'un VAR(1) cointégré (modèle 2)

... nous avons obtenu un VECM(0) (modèle 8)

⇒ Il existe une représentation VAR(2) d'un VECM(1)

$$X_t = (I + \Gamma_1 + \alpha\beta')X_{t-1} - \Gamma_1X_{t-2} + \varepsilon_t = A_1X_{t-1} + A_2X_{t-2} + \varepsilon_t$$

- Comme précédemment la remarque de l'équation 9 s'applique

$$\Pi = -(I - A_1 - A_2) = -(I - I - \Gamma_1 - \alpha\beta' + \Gamma_1) = \alpha\beta'$$

- Cela tient pour le passage d'un VECM(p) vers un VAR($p+1$)

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \Rightarrow X_t = \sum_{j=1}^{p+1} A_j X_{t-j} + \varepsilon_t \quad (12)$$

avec $A_1 = I + \Pi + \Gamma_1$, $A_j = \Gamma_j - \Gamma_{j-1}$ et $A_{p+1} = -\Gamma_{p+1}$

Exogénéité faible

- La notion **exogénéité faible** est définie par Engle et al. (1983)

⇒ Une variable X_t sera faiblement exogène

- pour θ un ensemble de paramètre d'intérêt
- si elle peut être prise comme **prédéterminée** dans le modèle

⇒ ne pas décrire le modèle marginale de X_t n'engendre aucune perte d'information

- A titre d'illustration considérons l'équation ECM suivante

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1}) + \gamma_1 \Delta x_{t-1} + \varepsilon_t$$

- Si le fait de ne pas modéliser x_t conjointement à y_t n'engendre aucune perte d'information, x_t est **faiblement exogène**

Exogénéité faible dans les VECM

- Partons du VECM bivarié (11) avec $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{2 \times 1}, \Sigma)$
- On peut réécrire $x_t = (x_{1,t}, x_{2,t})$ sous une forme orthogonalisée

$$\Delta x_{1t} = \alpha_1^+ (x_{1,t-1} - \beta_1 x_{2,t-1}) + \gamma_{11}^+ \Delta x_{1,t-1} + \gamma_{12}^+ \Delta x_{2,t-1} + \eta_{1t} \quad (13)$$

$$\Delta x_{2t} = \alpha_2 (x_{1,t-1} - \beta_1 x_{2,t-1}) + \gamma_{21} \Delta x_{1,t-1} + \gamma_{22} \Delta x_{2,t-1} + \varepsilon_{2t} \quad (14)$$

■ avec $\eta_{1t} = \varepsilon_{1t} - \Sigma_{12} \Sigma_{22}^{-1} \varepsilon_{2t}$ et $\Sigma_{11}^+ = \Sigma_{11} - \Sigma_{12} \Sigma_{22}$

... $\gamma_{11}^+ = \gamma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \gamma_{21}$ et $\gamma_{12}^+ = \gamma_{12} - \Sigma_{12} \Sigma_{22}^{-1} \gamma_{22}$

... et $\alpha_1^+ = \alpha_1 - \Sigma_{12} \Sigma_{22}^{-1} \alpha_2$ de sorte que $(\eta_{1t} \quad \varepsilon_{2t})' \sim \mathcal{N}(\mathbf{0}_{2 \times 1}, \Sigma^+)$

- On voit que (13) est une équation **conditionnelle** : $x_{1t} | x_{2t}, x_{t-1}$
- Par ailleurs, (14) est une équation **marginale** pour x_{2t} si $\alpha_2 = 0$

$\Rightarrow x_{2t}$ ne réagit pas au déséquilibre de long terme car **prédéterminé**
et donc **faiblement exogène**

Si on utilise la représentation VECM standard de l'équation (11) , l'exogénéité faible de x_{2t} est vérifiée pour $\alpha_2 = 0$ et $\Sigma_{12} = 0$

Notons que

$$\Sigma^+ = \begin{pmatrix} \Sigma_{11}^+ & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

Causalité dans les VECM

Definition (Causalité de court terme)

En considérant le VECM bivarié (13,14), x_{1t} (x_{2t}) ne cause pas x_{2t} (x_{1t}) à court terme au sens de Granger si et seulement si $\gamma_{21} = 0$ ($\gamma_{12} = 0$)

Definition (Causalité de long terme)

En considérant le VECM bivarié (13,14), x_{1t} (x_{2t}) ne cause pas x_{2t} (x_{1t}) à long terme au sens de Granger si et seulement si $\alpha_2\beta_{x_1} = 0$ ($\alpha_1\beta_{x_2} = 0$)

- La causalité à court terme implique une contrainte linéaire, elle est donc simple à tester
- La causalité à long terme implique une contrainte non-linéaire, elle est donc difficile à tester

Système triangulaire à la Phillips (1991)

- Phillips (1991) propose une représentation triangulaire
- Dans un cadre à 1 dimension le système est bivarié de la forme

$$x_{1t} = \gamma' x_{2,t} + \varepsilon_{1t}$$

$$\Delta x_{2t} = \varepsilon_{2t}$$

avec $\gamma = (\beta_2, \dots, \beta_k)$ et $x_{2t} = (x_{2t}, \dots, x_{kt})$

- Sous l'hypothèse que ε_{1t} et ε_{2t} sont faiblement dépendants
 - cette représentation est **semi-paramétrique**
 - cette représentation permet une représentation ECM simple
- ⇒ plus générale que Engle et Granger (1987)

Note Un à priori est fait sur la variable **faiblement exogène**

Système triangulaire de dimension k

- La représentation triangulaire existe pour toute dimension

$$X_{1t} = \Upsilon X_{2t} + \varepsilon_{1t}$$

$$\Delta X_{2t} = \varepsilon_{2t}$$

- Υ une matrice de coefficient $k \times (k - r)$
- X_{1t} est un ensemble de r régresseurs cointégrés avec X_{2t}
- X_{2t} est un ensemble de $k - r$ régresseurs non cointégrés entre eux
- $X_t = (X_{1t}, X_{2t})'$ est un ensemble de k variables aléatoire $I(1)$
- $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ des erreurs faiblement dépendantes

Note Un à priori est fait sur les $k - r$ variables **faiblement exogènes**

Système triangulaire et forme ECM

- La forme ECM des systèmes à la Phillips (1991) s'obtient par différenciation de X_t

$$\Delta X_t = -\Xi \aleph X_{t-1} + v_t$$

avec

$$\Xi = \begin{pmatrix} I_r \\ 0 \end{pmatrix}, \quad \aleph = (I \quad -\Upsilon) \text{ et } v_t = \begin{pmatrix} I & \Upsilon \\ 0 & I \end{pmatrix} \varepsilon_t$$

Plan

- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

Approche de Engle et Granger (1987)

- Comment estimer les modèles de cointégration ?
 - commençons par les systèmes cointégrés de dimension 1
- Soit $x_t \sim I(1)$ un processus stochastique d'innovations v_t et

$$y_t = \zeta + \beta x_t + \varepsilon_t \text{ avec } \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2 < \infty)$$

et $\beta \neq 0$ de sorte que $(y_t, x_t)' \sim CI(1, 1)$

- Il s'agit d'un équilibre de long terme de vecteur cointégrant

$$\beta = \begin{pmatrix} 1 \\ -\beta \end{pmatrix}$$

- L'estimation de β par OLS paraît alors envisageable car
 - $\hat{\beta}$ devrait être consistant si $\beta \neq 0$
 - on sait que $\varepsilon_t \sim I(0)$ si $\beta \neq 0$

Normalisation

- L'exemple précédent repose sur une hypothèse importante

- $y_t = \beta x_t + \varepsilon_t$ et $\beta = (1, -\beta)'$

⇒ il s'agit d'une normalisation

- On aurait pu considérer $x_t = \theta y_t + v_t$ et $\beta = (1, -\theta)'$

⇒ mais attention, $\theta \neq 1/\beta$

- La normalisation a un impact chez Engle et Granger (1987)

⇒ car les valeurs de $\beta = (\beta_1, \beta_2)'$ tel que $\beta' X_t$ est stationnaire ne sont pas uniques

- L'approche d'Engle et Granger (1987) n'est interprétable facilement que dans un système de dimension 1

Procédure en deux étapes

- Engle et Granger (1987) proposent une estimation en 2 étapes

1 on estime β par OLS

... on récupère les résidus estimés $\hat{\varepsilon}_t = y_t - \hat{\zeta} - \hat{\beta}x_t$

2 on teste si $\hat{\varepsilon}_t \sim I(0)$

- On a vu que si $(y_t, x_t) \sim I(1)$ et $\beta \neq 0$ les OLS sont super-consistants

- $T(\hat{\beta} - \beta) \xrightarrow{p} 0$

- On montre par la suite que

- $\hat{\beta}$ est biaisé si ε_t et v_t ne sont pas indépendants

- la distribution de $\hat{\beta}$ étant non-standard, tester $\hat{\varepsilon}_t \sim I(0)$ implique des valeurs critiques non-standards

Dans une 3ème étape on peut également estimer la forme ECM par OLS :

$$\Delta y_t = \hat{\alpha}_1 \hat{\varepsilon}_{t-1} + \hat{\gamma}_{11} \Delta y_{t-1} + \hat{\gamma}_{12} \Delta x_{t-1} + \zeta_t$$

$$\Delta x_t = \hat{\alpha}_2 \hat{\varepsilon}_{t-1} + \hat{\gamma}_{21} \Delta y_{t-1} + \hat{\gamma}_{22} \Delta x_{t-1} + \xi_t$$

Étape 1 de Engle et Granger (1987) : intuition

- Si $\beta \neq 0$ on sait qu'il y a cointégration

\Rightarrow au sein du paradigme $I(1)/I(0)$, $\varepsilon_t \sim I(0)$

- Les OLS sont consistants mais quelle est la distribution de $\hat{\beta}$?

$$x_{1t} = \zeta + \beta x_{2t} + \varepsilon_t, \quad (15)$$

$$\Delta x_{2t} = v_t \quad (16)$$

- Avant de répondre à cette question, précisons que

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} = \Psi(L)\epsilon_t, \text{ avec } \mathbb{E}(\epsilon_t \epsilon_t') = \Sigma = PP' < \infty$$

et $\epsilon_t \sim$ i.i.d. , $\Psi(L)$ absolument sommable et $\Psi(1)$ de rang plein

- On permet ici que les éléments non-diagonaux de Σ soient $\neq 0$

\Rightarrow cela introduit 2 problèmes : endogénéité + autocorrélation via $\Psi(L)$

Étape 1 de Engle et Granger (1987) : théorie

- Supposons qu'on peut partitionner $\Psi(1) \cdot P$ ainsi

$$\Psi(1) \cdot P = \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

avec $\dim \lambda_1 = 1 \times k$ et $\dim \lambda_2 = k - 1 \times k$ pour $k = 2$ ici

- Stock (1987) montre alors que

$$\begin{pmatrix} \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \int_0^1 \mathbf{W}(r)' dr \cdot \lambda_2' \\ \lambda_2 \int_0^1 \mathbf{W}(r)' dr & \lambda_2 \int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \lambda_2' \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

avec $\mathbf{W}(r)'$ un Wiener k -varié, $h_1 := \lambda_1' \mathbf{W}(1)$ et

$$h_2 := \lambda_2 \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \lambda_1 + \sum_{i=0}^{\infty} \mathbb{E}(v_t \varepsilon_{t+i}) \quad (17)$$

- La distribution n'est pas centrée sur 0 : **biais de non-centralité**

\Rightarrow donc sauf si $\varepsilon_t \perp v_t$, $\hat{\beta}$ fait face à un biais $O_p(n^{-1})$

On trouve également la preuve de ce résultat dans Hamilton (1994), p. 618-619

Étape 1 de Engle et Granger (1987) : intuition si $\varepsilon_t \perp v_t$

- Les résultats de Stock (1987) décrivent le cas général
- Pour voir le lien avec le **Théorème 12** du Chapitre 1 repartons de

$$\begin{aligned}x_{1t} &= \zeta + \beta x_{2t} + \varepsilon_t, \\ \Delta x_{2t} &= v_t\end{aligned}$$

- 1 Supposons ε_t et v_t i.i.d. avec

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix} \right)$$

- 2 Supposons également que $\Psi(L) = I$ et

$$\Sigma = PP' \text{ avec } P = \begin{pmatrix} \sigma_\varepsilon & 0 \\ 0 & \sigma_v \end{pmatrix} \Rightarrow \Psi(1) \cdot P = P$$

- On en déduit que $\lambda_1 = (\sigma_\varepsilon \ 0)$ et $\lambda_2 = (0 \ \sigma_v)$

Étape 1 de Engle et Granger (1987) : théorie si $\varepsilon_t \perp v_t$

- Ces 2 hypothèses éliminent le problème d'endogénéité
- La distribution limite devient alors

$$\begin{pmatrix} \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \sigma_v \int_0^1 W_2(r) dr \\ \sigma_v \int_0^1 W_2(r) dr & \sigma_v^2 \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \sigma_\varepsilon W_1(1) \\ \sigma_v \sigma_\varepsilon \int_0^1 W_2(r) dr \int_0^1 W_1(r) dr \end{pmatrix}$$

- En se focalisant sur $\hat{\beta}$ on retrouve le résultat du **Théorème 12**

$$n(\hat{\beta} - \beta) \xrightarrow{d} \left(\sigma_v \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\sigma_\varepsilon \int_0^1 W_2(r) W_1(r) dr \right) \quad (18)$$

mais sans le problème lié à la normalisation par n

- Dans la pratique, on rencontre rarement une situation si simple
 \Rightarrow nous verrons par la suite comment traiter l'endogénéité

Étape 2 de Engle et Granger (1987) : intuition

- Si $\beta = 0$ comment se comporte ε_t dans le système

$$x_{1t} = \beta x_{2t} + \varepsilon_t,$$

$$\Delta x_{2t} = v_t$$

- Afin de répondre à cette question, supposons la régression OLS

$$x_{1t} = \hat{\beta} x_{2t} + \hat{\varepsilon}_t \text{ ou } \hat{\beta}' X_t = \hat{\varepsilon}_t$$

- $\hat{\varepsilon}_t$ devrait être $I(1)$ en l'absence de cointégration donc

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \epsilon_{1t} = \hat{\varepsilon}_{t-1} + \epsilon_{1t}$$

- Dans la pratique on peut estimer ρ par OLS et tester si $\hat{\rho} = 1$

\Rightarrow les résidus étant estimés, quelle sera la distribution de $\hat{\rho}$?

$$\hat{\rho} = \left(\sum_{t=1}^n \hat{\varepsilon}_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} \right)$$

Étape 2 de Engle et Granger (1987) : théorie

- Phillips et Ouliaris (1990) montrent alors que

$$(n-1)(\hat{\rho}-1) \xrightarrow{d} \left[\frac{1}{2} \left((1 \quad -h_2') \cdot \mathbf{W}(1) \mathbf{W}(1)' \cdot \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right) \right. \\ \left. - h_1 \mathbf{W}(1)' \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right. \\ \left. - \frac{1}{2} (1 \quad -h_2') \mathbf{L}' \left(\mathbb{E}(\Delta X_t \Delta X_t') \right) \mathbf{L} \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right] H_n^{-1} \quad (19)$$

avec $(\mathbf{\Lambda} \mathbf{\Lambda}')^{-1} = \mathbf{L} \mathbf{L}'$, $\mathbf{W}(r) = (W_1(r) \quad W_2(r))'$ et

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} := \begin{pmatrix} 1 & \int_0^1 W_2(r) dr \\ \int_0^1 W_2(r) dr & \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 W_1(r) dr \\ \int_0^1 W_2(r) W_1(r) dr \end{pmatrix}$$

et finalement

$$H_n = \int_0^1 W_1(r)^2 dr - \begin{pmatrix} \int_0^1 W_1(r) dr & \int_0^1 W_1(r) W_2(r) dr \end{pmatrix} \begin{pmatrix} 1 \\ h_2 \end{pmatrix}$$

On trouve également la preuve de ce résultat dans Hamilton (1994), p. 621-623

Étape 2 de Engle et Granger (1987) : théorie si $\varepsilon_t \perp v_t$

- Si $\varepsilon_t \perp v_t$, $\Psi(L) = I$ et $\Lambda\Lambda' = \mathbb{E}(\Delta X_t \Delta X_t')$ ce qui implique

$$L' \left(\mathbb{E}(\Delta X_t \Delta X_t') \right) L = L' \Lambda \Lambda' L = I \text{ car } (\Lambda \Lambda')^{-1} = L L'$$

- La distribution asymptotique de $(n-1)(\hat{\rho} - 1)$ devient alors

$$(n-1)(\hat{\rho} - 1) \xrightarrow{d} \left[\frac{1}{2} \left((1 \quad -h_2') \cdot \mathbf{W}(1) \mathbf{W}(1)' \cdot \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} \right) \right. \\ \left. - h_1 \mathbf{W}(1)' \begin{pmatrix} 1 \\ -h_2 \end{pmatrix} - \frac{1}{2} (1 + h_2' h_2) \right] H_n^{-1}$$

avec $\mathbf{W}(r) = (W_1(r) \quad W_2(r))'$

On trouve également la preuve de ce résultat dans Hamilton (1994), p. 621-623

Plan

- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

Correction de Saikkonen (1991) pour l'endogénéité

- L'équation (17) montre qu'en présence d'endogénéité $\hat{\beta}$ est biaisé
- Saikkonen (1991) propose d'ajouter des lags et des leads de Δx_{2t}

$$x_{1t} = \zeta + \beta x_{2t} + \sum_{j=-l}^l \gamma_j \Delta x_{2,t-j} + \tilde{\varepsilon}_t$$

\Rightarrow cela va rendre $\tilde{\varepsilon}_t$ et v_{t-j} orthogonaux pour $j \in \{-l, l\}$ car

$$\varepsilon_t = \sum_{j=-l}^l \varphi_j v_{t-j} + \tilde{\varepsilon}_t$$

- La sélection de l peut se faire à l'aide d'un critère d'information
- Grâce à cette approche **paramétrique**, à présent on a

$$\begin{pmatrix} \tilde{\varepsilon}_t \\ v_t \end{pmatrix} = \tilde{\Psi}(L) \tilde{\varepsilon}_t, \text{ avec } \mathbb{E}(\tilde{\varepsilon}_t \tilde{\varepsilon}_t') = \tilde{\Sigma} = \tilde{P} \tilde{P}' < \infty$$

et $\tilde{\varepsilon}_t \sim \text{i.i.d.}$, $\tilde{\Psi}(L)$ absolument sommable et $\tilde{\Psi}(1)$ de rang plein

Théorie limite des dynamic OLS (DOLS)

- L'approche dite **dynamic OLS** de Saikkonen (1991) assure que

$$\tilde{P} = \begin{pmatrix} \tilde{\sigma}_{11} & 0 \\ 0 & \tilde{\sigma}_{22} \end{pmatrix} \text{ et } \tilde{\Psi}(L) = \begin{pmatrix} \tilde{\psi}_{11}(L) & 0 \\ 0 & \tilde{\psi}_{22}(L) \end{pmatrix}$$

- Pour $\varphi = (\varphi_{-l}, \dots, \varphi_l)'$, Saikkonen (1991) montre que

$$\begin{pmatrix} \sqrt{n}(\hat{\varphi} - \varphi) \\ \sqrt{n}(\hat{\zeta} - \zeta) \\ n(\hat{\beta} - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Q^{-1}h_1 \\ \tilde{\lambda}_{11}v_1 \\ \tilde{\lambda}_{11}v_2 \end{pmatrix}, \text{ avec } Q^{-1} = \mathbb{E}(w_t w_t')$$

et $n^{-1/2} \sum w_t \tilde{\epsilon}_t \rightarrow h_1$ pour $w_t = (v_{t-l}, \dots, v_{t+l})'$

- $\tilde{\lambda}_{11} = \tilde{\psi}_{11}(1)\tilde{\sigma}_{11}$, $\tilde{\lambda}_{22} = \tilde{\psi}_{22}(1)\tilde{\sigma}_{22}$ et $(v_1 \ v_2)'$ donné par

$$\begin{pmatrix} 1 & \tilde{\lambda}_{22} \int_0^1 W_2(r)dr \\ \tilde{\lambda}_{22} \int_0^1 W_2(r)dr & \tilde{\lambda}_{22}^2 \int_0^1 W_2(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1(1) \\ \tilde{\lambda}_{22} \int_0^1 W_2(r)W_1(r)dr \end{pmatrix}$$

Théorie limite des DOLS en absence d'autocorrélation

- Supposons une corrélation contemporaine entre ε_t et v_t

$$\Sigma = PP' \text{ avec } P = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

mais qu'à présent $\tilde{\Psi}(1) = I$ et donc $\tilde{\lambda}_{11} = \tilde{\sigma}_{11}$ et $\tilde{\lambda}_{22} = \tilde{\sigma}_{22}$

- Les **DOLS** nous ramènent ici aux résultats de l'équation (18)
- En effet, en se focalisant sur β on obtient pour $\tilde{\lambda}_{11}v_2$

$$\begin{aligned} \tilde{\lambda}_{11}v_2 &= \left(\tilde{\lambda}_{22}^2 \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\lambda}_{11} \tilde{\lambda}_{22} \int_0^1 W_2(r) W_1(r) dr \right) \\ &= \left(\tilde{\sigma}_{22}^2 \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\sigma}_{11} \tilde{\sigma}_{22} \int_0^1 W_2(r) W_1(r) dr \right) \\ &= \left(\tilde{\sigma}_{22} \int_0^1 W_2(r)^2 dr \right)^{-1} \left(\tilde{\sigma}_{11} \int_0^1 W_2(r) W_1(r) dr \right) \end{aligned}$$

Correction de Phillips et Hansen (1990) pour l'endogénéité

- L'équation (17) montre qu'en présence d'endogénéité, $\hat{\beta}$ fait face
 - à un biais de **second-ordre** (endogénéité) : Ξ_0
 - et un biais de **non-centralité** si autocorrélation sérielle : Ξ_1
- Phillips et Hansen (1990) proposent une double correction car
 - le biais de second-ordre vient des éléments non-diagonaux non-nuls

dans $\Sigma = \mathbb{E}(\epsilon_t \epsilon_t')$ venant de $(\epsilon_t, v_t)' = \Psi(L)\epsilon_t$
 - la non-centralité vient de $\Xi_1 = \sum_{i=0}^{\infty} \mathbb{E}(v_t \epsilon_{t+i}) \neq 0$ dans l'Eq. (17)
- Concernant Σ , la correction est simple
 - Dans l'Eq. (15) on soustrait $\Xi_0 = \sigma_{21}^2 (\sigma_{22}^2)^{-1} \Delta x_{2t} = \sigma_{21}^2 (\sigma_{22}^2)^{-1} v_t$
 - On obtient alors l'équation **modifiée**

$$x_{1t}^* = \zeta + \beta^* x_{2t} + \varepsilon_t^* \text{ où } x_{1t}^* = x_{1t} - \Xi_0 \text{ et } \varepsilon_t^* = \varepsilon_t - \Xi_0 \quad (20)$$

Correction de l'endogénéité via Ξ_0

- Dans le système triangulaire (15,16), la correction Ξ_0 donne

$$\begin{pmatrix} \varepsilon_t^* \\ v_t \end{pmatrix} = \boldsymbol{\xi} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \text{ avec } \boldsymbol{\xi} = \begin{pmatrix} 1 & -\sigma_{21}^2(\sigma_{22}^2)^{-1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

- Phillips et Hansen (1990) montrent alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^* - \zeta) \\ n(\hat{\beta}^* - \beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 & \left(\int_0^1 \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_2 \\ \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \int_0^1 \mathbf{W}(r) dr & \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_2 \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \boldsymbol{\xi}_1 \boldsymbol{\Lambda} \mathbf{W}(1) \\ \boldsymbol{\xi}_2' \boldsymbol{\Lambda} \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right) \boldsymbol{\Lambda}' \boldsymbol{\xi}_1 + \Xi_1^* \end{pmatrix} \quad (21)$$

avec $\Psi(1) \cdot P = \boldsymbol{\Lambda}$ sachant $\Sigma = PP'$ et Ξ_1^* donné par

$$\Xi_1^* = \sum_{i=0}^{\infty} \mathbb{E}(v_t \varepsilon_{t+i}^*) = \sum_{i=0}^{\infty} \mathbb{E}\left(v_t (\varepsilon_t - \sigma_{21}^2(\sigma_{22}^2)^{-1} v_{t+i})\right)$$

Correction de l'endogénéité via Ξ_1

- L'Eq. (21) fait toujours apparaître le biais de non-centralité, Ξ_1^*
- Phillips et Hansen (1990) proposent un estimateur de Ξ_1^*

1 estimer le système triangulaire (15,16)

2 estimer $\hat{\Sigma}$ via $\hat{\Sigma} = \hat{\Omega}_0 + \sum_{i=1}^q (1 - i(q+1)^{-1})(\hat{\Omega}_i + \hat{\Omega}_i')$ avec

$$\hat{\Omega}_i = n^{-1} \sum_{t=i+1}^n \begin{pmatrix} \hat{\varepsilon}_t \hat{\varepsilon}_{t-i} & \hat{\varepsilon}_t \hat{v}_{t-i} \\ \hat{v}_{t-i} \hat{\varepsilon}_t & \hat{v}_{t-i} \hat{v}_{t-i} \end{pmatrix} = \begin{pmatrix} \hat{\omega}_{11}^{(i)} & \hat{\omega}_{12}^{(i)} \\ \hat{\omega}_{21}^{(i)} & \hat{\omega}_{22}^{(i)} \end{pmatrix} \quad (22)$$

3 on note alors que $\Xi_1^* = \sum_{i=0}^{\infty} \begin{pmatrix} \omega_{12}^{(i)} & \omega_{22}^{(i)} \end{pmatrix} (1 \quad -\sigma_{21}^2(\sigma_{22}^2)^{-1})'$ car

$$\Xi_1^* = \sum_{i=0}^{\infty} \mathbb{E} \left(v_t \begin{pmatrix} \varepsilon_t & v_{t+i} \end{pmatrix} \right) (1 \quad -\sigma_{21}^2(\sigma_{22}^2)^{-1})'$$

ce qui donne un estimateur pour Ξ_1^*

$$\hat{\Xi}_1^* = \sum_{i=0}^q (1 - i(q+1)^{-1}) \begin{pmatrix} \hat{\omega}_{12}^{(i)} & \hat{\omega}_{22}^{(i)} \end{pmatrix} (1 \quad -\hat{\sigma}_{21}^2(\hat{\sigma}_{22}^2)^{-1})' \quad (23)$$

Fully modified OLS (FMOLS)

- Via $\hat{\Xi}_0$ et $\hat{\Xi}_1^*$ obtenu par l'Eq. (23) on corrige les 2 biais

⇒ contrairement aux **DOLS** cette correction est non-paramétrique

⇒ on parle de **fully modified OLS** de $\theta^* = (\zeta, \beta)'$

$$\hat{\theta}^* = \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t} \mathbf{x}_{2t}' \right)^{-1} \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t} \underbrace{(x_{1t} - \hat{\Xi}_0)}_{x_{1t}^*} - n \hat{\Xi}_1^* \right), \quad \mathbf{x}_{2t} = (1, x_{2t})'$$

- Phillips et Hansen (1990) montrent alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^* - \zeta) \\ n(\hat{\beta}^* - \beta) \end{pmatrix} \xrightarrow{d} \sigma_{11}^* \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad (24)$$

avec $(v_1 \quad v_2)'$ donné par

$$\begin{pmatrix} 1 & \sigma_{22} \int_0^1 W_2^*(r) dr \\ \sigma_{22} \int_0^1 W_2^*(r) dr & \sigma_{22}^2 \int_0^1 W_2^*(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1^*(1) \\ \sigma_{22} \int_0^1 W_2^*(r) W_1^*(r) dr \end{pmatrix}$$

- Les **FMOLS** nous ramènent également aux résultats de l'Eq. (18)

Correction de Park (1992) pour l'endogénéité

- Park (1992) propose une approche très similaire aux FMOLS
- Le biais de **second-ordre** est corrigé similairement
- Le biais de **non-centralité** est corrigé différemment
 - A partir de l'équation (22) de $\hat{\Omega}$ et $\hat{\Sigma}$ on définit

$$\hat{\Lambda} = \hat{\Omega}_0 + \sum_{i=1}^q (1 - i(q+1)^{-1}) \hat{\Omega}_i = \begin{pmatrix} \hat{\lambda}_{11} & \hat{\lambda}_{12} \\ \hat{\lambda}_{21} & \hat{\lambda}_{22} \end{pmatrix} = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}$$

- On construit les variables **modifiées**

- $x_{1t}^+ = x_{1t} - \left(\hat{\beta}_* \hat{\lambda}_2 \hat{\Omega}_0^{-1} + \begin{pmatrix} 0 & \hat{\sigma}_{12}^2 (\hat{\sigma}_{22}^2)^{-1} \end{pmatrix} \right) \begin{pmatrix} \hat{\varepsilon}_{*t} \\ v_t \end{pmatrix}$

- $x_{2t}^+ = x_{2t} - \hat{\lambda}_2 \hat{\Omega}_0^{-1} \begin{pmatrix} \hat{\varepsilon}_{*t} \\ v_t \end{pmatrix}$ avec $v_t = \Delta x_{2t}$

- $\mathbf{x}_{2t}^+ = \begin{pmatrix} 1 & x_{2t}^+ \end{pmatrix}'$

- $\hat{\beta}_*$ et $\hat{\varepsilon}_{*t}$ proviennent d'une estimation OLS naive de l'Eq. (15)

Régression canonique cointégrante (CCR)

- Chez Park (1992) on voit que $\hat{\Xi}_0^+ = \hat{\Xi}_0$ mais $\hat{\Xi}_1^+ = \hat{\lambda}_2 \hat{\Omega}_0^{-1}$

⇒ on parle de **Canonical Cointegration Regression (CCR)**

$$\hat{\theta}^+ = \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t}^+ \mathbf{x}_{2t}^{+'} \right)^{-1} \left(n^{-1} \sum_{t=1}^n \mathbf{x}_{2t}^+ \mathbf{x}_{1t}^+ \right) \text{ avec } \theta^+ = (\zeta, \beta)'$$

- Park (1992) montre alors que

$$\begin{pmatrix} n^{1/2}(\hat{\zeta}^+ - \zeta) \\ n(\hat{\beta}^+ - \beta) \end{pmatrix} \xrightarrow{d} \sigma_{11}^+ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad (25)$$

avec $(v_1 \ v_2)'$ donné par

$$\begin{pmatrix} 1 & \sigma_{22}^+ \int_0^1 W_2^+(r) dr \\ \sigma_{22}^+ \int_0^1 W_2^+(r) dr & (\sigma_{22}^+)^2 \int_0^1 W_2^+(r)^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W_1^+(1) \\ \sigma_{22}^+ \int_0^1 W_2^+(r) W_1^+(r) dr \end{pmatrix}$$

- Les **CCR** nous ramènent également aux résultats de l'Eq. (18)

Bilan des estimateurs dans l'approche d'Engle et Granger

- Chez Engle et Granger (1987), l'estimation de β est cruciale
 - Nous avons analysé 4 estimateurs
 - **OLS** : biaisé en présence d'endogénéité
 - **DOLS** : corrige paramétriquement le biais d'endogénéité
 - **FMOLS** : corrige non-paramétriquement le biais d'endogénéité
 - **CCR** : corrige non-paramétriquement le biais d'endogénéité
 - Kurozumi et Hayakawa (2009) rappellent et démontrent que
 - **DOLS**, **FMOLS** et **CCR** sont asymptotiquement équivalents
 - **DOLS**, **FMOLS** et **CCR** subissent un biais
 - si l'autocorrélation sérielle des erreurs est très forte
 - pouvant ramener leurs distributions limite à celle des **OLS**
- ⇒ ce biais apparaît donc en échantillon fini

Plan

- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

Approche de Johansen (1991)

- Soit un $VECM(p)$ pour X_t un vecteur de k variables $I(1)$

$$\Delta X_t = \Xi + \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \quad (26)$$

- Johansen (1991) propose une procédure d'estimation de

$$\theta = (\Xi, \Gamma_1, \dots, \Gamma_p, \Pi, \Sigma)$$

- le paramètre Ξ (constante)
 - les paramètres de court terme $\Gamma_1, \dots, \Gamma_p$
 - les paramètres de long terme α et β pour $\Pi = \alpha\beta$
 - les paramètres de Σ sous l'hypothèse que $\varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$
- Cette procédure est basée sur une **analyse canonique**
- Revient à implémenter un **maximum de vraisemblance**

L'analyse canonique : introduction

- Soit \mathbf{y} et \mathbf{x} des variables dont on ignore le degré de corrélation
- L'**analyse canonique** permet de rechercher les combinaisons linéaires entre \mathbf{y} et \mathbf{x} les plus corrélées possibles
- Ces combinaisons forment des **variables canoniques**
- Formellement, soit \mathbf{y} et \mathbf{x} deux vecteurs de variables tels que

$$\dim(\mathbf{y}) = k_y \times 1 \text{ et } \dim(\mathbf{x}) = k_x \times 1$$

⇒ La matrice de covariance jointe résume l'ensemble des corrélations

$$\Sigma = \begin{pmatrix} \mathbb{E}(\mathbf{y}\mathbf{y}') & \mathbb{E}(\mathbf{y}\mathbf{x}') \\ \mathbb{E}(\mathbf{x}\mathbf{y}') & \mathbb{E}(\mathbf{x}\mathbf{x}') \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}$$

- Σ est bien sûr complexe à étudier : solution ?

L'analyse canonique : intuition

- L'**analyse canonique** permet une approche parcimonieuse
- Invoquons deux nouveaux vecteurs η_t et ζ_t tels que

$$\eta_t = \mathcal{V}'y_t \text{ et } \zeta_t = \mathcal{A}'x_t$$

avec $\dim(\mathcal{V}) = k_\wedge \times k_y$ et $\dim(\mathcal{A}) = k_\wedge \times k_x$ et $k_\wedge = k_y \wedge k_x$

- η_t et ζ_t constituent les **variables canoniques**
- \mathcal{V} et \mathcal{A} sont choisis telles que η_t et ζ_t sont de variances unitaires et leurs élément respectifs sont non-corrélés entre eux :

$$\mathbb{E}(\eta_t \eta_t') = \mathcal{V}' \Sigma_{yy} \mathcal{V} = I \text{ et } \mathbb{E}(\zeta_t \zeta_t') = \mathcal{A}' \Sigma_{xx} \mathcal{A} = I$$

$\Rightarrow \mathcal{R}_{k_\wedge \times k_\wedge} = \mathcal{A}' \Sigma_{xy} \mathcal{V} = \mathbb{E}(\zeta_t \eta_t')$ résume les *corr*(η_i, ζ_j) pour $i = j$

- On peut ordonner η_t et ζ_t tel que $(1 \geq r_1 \geq \dots \geq r_{k_\wedge} \geq 0)$

\Rightarrow l'élément r_i de $\text{diag}(\mathcal{R})$ est la i ème **corrélation canonique**

$$k_y \wedge k_x \equiv \min(k_y, k_x)$$

\mathcal{R} est une matrice diagonale de la forme

$$\mathcal{R} = \begin{pmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{k_{\wedge}} \end{pmatrix}$$

Rappel sur le problème des valeurs propres

- Soit une matrice carrée \mathcal{M} avec $\dim(\mathcal{M}) = k \times k$
- Les solutions de la représentation

$$\mathcal{M}\mathbf{v} = \lambda\mathbf{v} \text{ ou } (\mathcal{M} - \lambda I)\mathbf{v} = 0$$

donne λ un scalaire nommé valeur propre et \mathbf{v} un vecteur propre

- Rappelons que cette représentation admet une solution *ssi*

$$\det(\mathcal{M} - \lambda I) = 0$$

- Il s'agit d'un polynôme caractéristique qui se factorise comme

$$\det(\mathcal{M} - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_k - \lambda)$$

avec $\lambda_1, \dots, \lambda_k$ les k valeurs propres associées à \mathcal{M}

- Dans le cadre de la régression canonique, déterminer \mathcal{V} , \mathcal{A} et donc \mathcal{R} , peut se faire à partir des valeurs propres de Σ_{yy} , Σ_{xx} et Σ_{xy}

L'analyse canonique : fonctionnement

- On note $\lambda_1, \dots, \lambda_{k_y}$ et μ_1, \dots, μ_{k_x} les valeurs propres et v_1, \dots, v_{k_y} et $\alpha_1, \dots, \alpha_{k_x}$ les vecteurs propres de

$$(\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}) v_i = \lambda_i v_i \text{ et } (\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}) \alpha_i = \mu_i \alpha_i$$

- Notons que si v_i est un vecteur propre, $c v_i$ l'est aussi $\forall c$
- Dans l'analyse canonique on choisit c tel que

$$\tilde{v}_i' \Sigma_{yy} \tilde{v}_i = 1 \text{ avec } \tilde{v}_i = v_i / \left(\sqrt{v_i' \Sigma_{yy} v_i} \right), \quad \forall i$$

- \mathcal{V} et \mathcal{A} sont alors obtenus grâce aux vecteurs propres normalisés

$$\mathcal{V}' = (\tilde{v}_1, \dots, \tilde{v}_{k_y}) \text{ et } \mathcal{A}' = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_{k_x})$$

- Les corrélations canoniques sont simplement déduites de

$$\mathcal{R}^2 = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_{k_{\wedge}} \end{pmatrix} \Rightarrow r_i = \sqrt{\lambda_i}$$

Plan

- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

Approche de Johansen (1991)

- On part d'un VECM(p) pour X_t un vecteur de k variables $I(1)$

$$\Delta X_t = \Xi + \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \text{ avec } \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$$

et $\theta = (\Xi, \Gamma_1, \dots, \Gamma_p, \Pi, \Sigma)$

- L'hypothèse de normalité mène à la **log-vraisemblance** suivante

$$\mathcal{L}(\theta; X_{2-p}, X_{3-p}, \dots, X_0) = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^n \left(\varepsilon_t' \Sigma^{-1} \varepsilon_t \right)$$

avec $\varepsilon_t = \Delta X_t - \Xi - \Pi X_{t-1} - \sum_{j=1}^p \Gamma_j \Delta X_{t-j}$

- A l'aide de l'analyse canonique, Johansen (1991) développe une procédure en 3 étapes pour obtenir

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta; X_{2-p}, X_{3-p}, \dots, X_0)$$

Johansen (1991) : étape 1 (régression auxiliaire)

- On construit une première régression auxiliaire

$$\Delta X_t = \Omega_0 + \sum_{j=1}^p \Omega_j \Delta X_{t-j} + \omega_t \quad (27)$$

où $\hat{\theta}_\omega = (\hat{\Omega}_0, \hat{\Omega}_1, \dots, \hat{\Omega}_p)$ et $\hat{\omega}_t$ sont obtenus par OLS

- On construit une seconde régression auxiliaire

$$X_{t-1} = \Psi_0 + \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \psi_t \quad (28)$$

où $\hat{\theta}_\psi = (\hat{\Psi}_0, \hat{\Psi}_1, \dots, \hat{\Psi}_p)$ et $\hat{\psi}_t$ sont obtenus par OLS

Johansen (1991) : étape 2 (corrélation canonique)

- A partir de $\hat{\omega}_t$ et $\hat{\psi}_t$, on peut construire

$$\hat{\Sigma}_{\omega\omega} = \frac{1}{n} \sum_{t=1}^n \hat{\omega}_t \hat{\omega}_t', \quad \hat{\Sigma}_{\psi\psi} = \frac{1}{n} \sum_{t=1}^n \hat{\psi}_t \hat{\psi}_t', \quad \hat{\Sigma}_{\psi\omega} = \hat{\Sigma}_{\omega\psi} = \frac{1}{n} \sum_{t=1}^n \hat{\omega}_t \hat{\psi}_t'$$

- A l'aide de la régression canonique, déterminer les valeurs propres ordonnées $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$ de

$$\hat{\Sigma}_{\psi\psi}^{-1} \hat{\Sigma}_{\psi\omega} \hat{\Sigma}_{\omega\omega}^{-1} \hat{\Sigma}_{\omega\psi}$$

- Johansen (1991) montre que $\mathcal{L}(\theta; \cdot)$ atteint un maximum pour

$$\mathcal{L}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i)$$

avec r le **nombre de relations de cointégration**

⇒ Possibilité d'un test \mathcal{LR} sur le nombre de relation de cointégration

Johansen (1991) : étape 3 (MLE)

- L'analyse canonique a révélé les r premiers vecteurs propres normalisés et ordonnés $\hat{\mathcal{V}}'_r = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_r)$ de

$$(\hat{\Sigma}_{\psi\psi}^{-1} \hat{\Sigma}_{\psi\omega} \hat{\Sigma}_{\omega\omega}^{-1} \hat{\Sigma}_{\omega\psi}) \hat{v}_i = \hat{\lambda}_i \hat{v}_i$$

- Johansen (1991) montre alors que

- 1 Le MLE de Π est donné par

$$\hat{\Pi} = \hat{\Sigma}_{\omega\omega} \hat{\mathcal{V}} \hat{\mathcal{V}}' = \hat{\alpha} \hat{\beta}' \text{ avec } \hat{\alpha} = \hat{\Sigma}_{\psi\omega} \hat{\mathcal{V}}_r \text{ et } \hat{\beta} = \hat{\mathcal{V}}_r$$

- 2 Le MLE de Γ_i pour $i = 1, 2, \dots, p$ est donné par

$$\hat{\Gamma}_i = \hat{\Omega}_i - \hat{\Pi} \hat{\Psi}_i$$

- 3 Le MLE de Ξ est donné par

$$\hat{\Xi} = \hat{\Omega}_0 - \hat{\Pi} \hat{\Psi}_0$$

- 4 Le MLE de Σ est donné par

$$\hat{\Sigma} = n^{-1} \sum_{t=1}^n (\hat{\omega}_t - \hat{\Pi} \hat{\psi}_t)(\hat{\omega}_t - \hat{\Pi} \hat{\psi}_t)'$$

Test de trace de Johansen (1988)

- On a vu que $\mathcal{L}(\theta; \cdot)$ atteint un maximum pour

$$\mathcal{L}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i)$$

avec r le nombre de relations de cointégration

- Définissons que sous H_0 il y a r relations de cointégration
- Définissons que sous H_1 il y a $r = k$ relations de cointégration

$$\mathcal{L}_{H_1}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^k \log(1 - \hat{\lambda}_i)$$

- Un test du ratio de vraisemblance est alors donné par

$$\mathcal{LR}_g^{(\tau)} = 2(\mathcal{L}_{H_1}^* - \mathcal{L}_{H_0}^*) = -n \sum_{i=r+1}^k \log(1 - \hat{\lambda}_i)$$

avec $g = k - r$, le nombre de racines unitaires dans le VECM

Distribution limite du test de trace

- Considérons le modèle $\text{VECM}(p)$ avec $\Xi = 0$

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^p \Gamma_j \Delta X_{t-j} + \varepsilon_t \text{ avec } \varepsilon_t \sim \mathcal{N}(\mathbf{0}_{k \times 1}, \Sigma)$$

- La distribution limite de $\mathcal{LR}_g^{(\tau)}$ est donnée par la **trace** de

$$\mathbf{Q} = \left(\int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right)' \left(\int_0^1 \mathbf{W}(r) \mathbf{W}(r)' dr \right)^{-1} \left(\int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right)$$

avec $\mathbf{W}(r)$ un Wiener de dimension g

- Dans le cas $g = 1$, la statistique \mathbf{Q} devient un scalaire

$$Q = \frac{\left(\int_0^1 W(r) dW(r) \right)^2}{\int_0^1 W(r)^2} = \frac{(1/2)^2 \left((W(1)^2 - 1) \right)^2}{\left(\int_0^1 W(r)^2 dr \right)}$$

\Rightarrow Il s'agit de $(\hat{t}_\rho)^2$ dans la théorie limite des OLS non-standard!!

Test $\lambda - \max$ de Johansen (1988)

- Dans le même esprit que le test de trace on peut considérer

$$\mathcal{L}_{H_1}^* = -\frac{nk}{2} \log(2\pi) - \frac{n}{2} \log |\hat{\Sigma}_{\omega\omega}| - \frac{1}{2} \sum_{i=1}^{r+1} (1 - \hat{\lambda}_i)$$

- Un test du ratio de vraisemblance est alors donné par

$$\mathcal{LR}_g^{(\lambda)} = 2(\mathcal{L}_{H_1}^* - \mathcal{L}_{H_0}^*) = -n \log(1 - \hat{\lambda}_{r+1})$$

- La distribution limite de $\mathcal{LR}_g^{(\lambda)}$ est donnée par la **valeur propre la plus large** de la matrice Q
- Pour $g = 1$, les deux tests sont identiques
- Notons que le test de trace a plus de **puissance** que le test $\lambda - \max$

Sous H_0 il y a toujours r relation de cointégration donc $\mathcal{L}_{H_0}^*$ somme sur r

La trace d'une matrice carré est donnée par la somme de ses coefficients diagonaux mais également par la somme de ses valeurs propres

La puissance d'un test est la probabilité de rejeter l'hypothèse nulle sachant que l'hypothèse nulle est incorrecte

Méthodologie d'application des tests de Johansen (1988)

- Les deux tests s'appliquent séquentiellement pour $r = 0, 1, 2, \dots$

\Rightarrow soit le test de trace pour un VECM tri-varié ($k = 3$)

$r = 0$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 0$ contre $H_1 : \text{rank}(\Pi) = 3$

\Rightarrow si on rejette H_0 , on rejette l'absence de cointégration

$r = 1$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 1$ contre $H_1 : \text{rank}(\Pi) = 3$

\Rightarrow si on rejette H_0 , le nombre de relation de cointégration > 1

$r = 2$ calculer $\mathcal{LR}_g^{(\tau)}$ sous $H_0 : \text{rank}(\Pi) = 2$ contre $H_1 : \text{rank}(\Pi) = 3$

\Rightarrow si on rejette H_0 , le VECM est un VAR stationnaire

\Rightarrow si on ne rejette pas H_0 , il y a 2 relations de cointégration

Plan

- 1 Théorie de la cointégration
- 2 Engle et Granger
- 3 Endogénéité

- 4 Analyse canonique
- 5 Johansen
- 6 Sum up

Ce qu'il faut retenir

- La cointégration permet d'éviter les régressions factices
 - ⇒ si une **combinaison linéaire** de séries $I(1)$ est $I(0)$, elles partagent un équilibre de long terme commun
- Il existe 2 principales représentations des systèmes cointégrés
 - la représentation triangulaire à la Phillips (1991)
 - la représentation vectorielle à correction d'erreur (VECM)
- Il existe 2 principales méthodologie d'estimation
 - l'approche en 2 étapes à la Engle et Granger (1987)
 - estimation de la relation de long terme : $\hat{\beta}$
 - test sur les résidus de la relation : $I(1)$ v.s. $I(0)$
 - l'approche en système à la Johansen (1991)
 - estimation du VECM par maximum de vraisemblance
 - test sur la matrice Π pour détecter les relations de long terme

- Engle, R. F., Granger, C. W. J. (1987). Co-integration and error correction : representation, estimation, and testing. *Econometrica*, 55(2), 251–276.
- Engle, R. F., Hendry, D., Richard, J. (1983). Exogeneity. *Econometrica*
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, 121–130.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 1551-1580.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control* 12, 231-254.
- Hansen, P. R. (2005). Granger's representation theorem : A closed-form expression for I (1) processes. *The Econometrics Journal*, 8(1), 23-38.
- Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton : Princeton university press.
- Kurozumi, E., Hayakawa, K., 2009. Asymptotic properties of the efficient estimators for cointegrating regression models with serially dependent errors. *Journal of Econometrics* 149, 118-135.
- Park, J. (1992). Canonical cointegrating regressions. *Econometrica*, 60(1), 119–143.
- Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica : Journal of the Econometric Society*, 283-306.
- Phillips, P., Hansen, B. (1990). Statistical inference in instrumental variables regression with I (1) processes. *The Review of Economic Studies*, 57, 99–125.
- Phillips, P., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. *Econometrica : Journal of the Econometric Society* 58, 165-193.

- Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric theory*, 7(1), 1-21.
- Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica : Journal of the Econometric Society*, 1035-1056.