Introduction

- Relation entre des variables stationnires en différences (variables dites DS ou I(1))
- En général, la regression linéaire simple entre deux variables non stationnaires peut qualifiée de "regression fallacieuse" [1]
- Des exceptions lorsque l'on a une relation de cointégration entre des variables stationnaires.
- cointégration : existence d'une relation "stable" entre deux variables non stationnaires

1. En anglais, "spurious regression".

Introduction. Exemple 1 : la parité de pouvoir d'achat

- Condition de parité de pouvoir d'achat absolue
- Egalité des prix (exprimés dans la même devise) d'un même bien vendu dans le pays et à l'étranger en l'absence de barrières au commerce international :

$$P_t = E_t P_t^*$$

- P_t est le prix du bien domestique exprimé en monnaie nationale
- P_t^* est le prix du bien étranger exprimé en monnaie étrangère
- E_t est le taux de change.
- En passant en logarithme, la relation s'écrit :

$$\begin{aligned} ln(P_t) &= ln(E_t) + ln(P_t^*) \\ \Rightarrow p_t &= e_t + p_t^* \end{aligned}$$

Introduction. Exemple 1 : la parité de pouvoir d'achat

- La relation d'équilibre $p_t = e_t + p_t^*$ n'est pas forcément satisfaite à chaque date. On ajoute un terme d'erreur u_t supposé stationnaire qui représente l'écart temporaire par rapport à l'équilibre de long terme.
- La relation $p_t = e_t + p_t^* + u_t$ est une relation de cointégration (à la condition que les variables p_t, e_t p_t^* soient stationnaires en différence).

Introduction Exemple 2 : "pair trading"

- Une application en finance
- On forme un portefeuille de deux actions qui sont liées entre elles.
- L'objectif est de tirer profit des situations où leur prix relatif s'éloigne de sa valeur d'équilibre, en prenant une position
  - longues sur l'action sous-évaluée
  - courte sur l'action sur-évaluée
- On peut espérer obtenir un profit en dénouant la position en période de convergence du spread vers sa valeur d'équilibre.

Introduction Exemple 2 : "pair trading"

- Important de modéliser le spread, de détecter les situations où il y a un écart par rapport à la relation de long terme et de mesurer la vitesse d'ajustement vers cette équilibre.
- On note p_{1t} et p_{2t} le prix des actions 1 et 2 à la date t. On suppose que ces prix sont DS (ou I(1)).
- La relation d'équilibre de long terme est $p_{1t} - \beta p_{2t} = \alpha + u_t$
- Cet équilibre de long terme définit une relation de cointégration.

Cointégration entre deux variables

- Relation de cointégration entre deux variables
- On considère deux variables stationnaires en différence x_{1t} et x_{2t}.
- On dira qu'il existe une relation de cointégration entre x_{1t} et x_{2t} s'il existe une relation linéaire stationnaire entre ces deux variables :

$$x_{1t} - \alpha - \beta x_{2t} = u_t, \text{ avec } u_t \sim I(0)$$

- On peut aussi écrire que :

$$x_{1t} = \alpha + \beta x_{2t} + u_t, \text{ avec } u_t \sim I(0)$$

- Intuition : x_{1t} et x_{2t} contiennent toutes les deux une marche aléatoire
- Ces deux marches aléatoires sont liées : les deux séries ne peuvent s'éloigner durablement l'une de l'autre.

Généralisation à n variables

- On considère un vecteur de n variables stationnaires en différence :

$$X_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{nt} \end{pmatrix}$$

- Il existe une relation de cointégration entre ces n variables s'il existe un vecteur $\beta = (\beta_1, \beta_2, ..., \beta_n)'$ tel que :

$$\beta' X_t = \sum_{i=1}^{n} \beta_i x_{it} = u_t \text{ avec } u_t \sim I(0)$$

Propriétés des relations de cointégration

- Une relation de cointégration est interprétée comme une relation de long terme entre les variables x_{it}. Elle n'est pas vérifiée exactement à chaque date mais l'est en moyenne.
- Une relation de cointégration est définie à un scalaire multiplicatif près. On peut donc égaliser à 1 l'un des coefficients sans affecter la relation de cointégration.
- Si l'on considère n variables DS, le nombre de relations de cointégration noté r peut varier 0 entre et *n-1* : $0 \leq r \leq n - 1$.
- Dans le cas où le nombre de relations de cointégration est égal à n $r = n$, ceci signifie que les n variables x_{it} sont en fait stationnaires.

Test de cointégration en deux étapes de Engle et Granger.

- Première étape : estimation de la relation de cointégration
- On estime par les MCO la relation linéaire :

$$x_{1t} = \hat{c} + \hat{\beta}_2 x_{2t} + \cdots + \hat{\beta}_n x_{nt} + \hat{u}_t$$

- On choisit de normaliser à 1 le coefficient de x_{1t}.
- On aurait pu choisir n'importe quelle autre variable parmi $x_{2t}, ..., x_{nt}$ comme variable expliquée.

Test de cointégration en deux étapes de Engle et Granger.

- Seconde étape : test de l'hypothèse de cointégration
- On veut savoir si le résidu $\hat{u}_t = x_{1t} - \hat{c} + \hat{\beta}_2 x_{2t} - \cdots - \hat{\beta}_n x_{nt}$ est stationnaire
- Application du test de Dickey-Fuller augmenté à $\hat{u}_t$:

$$\Delta \hat{u}_t = \rho \hat{u}_{t-1} + \sum_{i}^{p} \gamma_i \Delta \hat{u}_{t-i} + e_t$$

- Les résidus sont nuls en moyenne
- pas nécessaire d'inclure une constante ou une tendance déterministe.
- Hypothèses du test

$$\begin{cases} H_0 : \rho = 0 \Rightarrow \text{pas de relation de cointegration} \\ H_a : \rho < 0 \Rightarrow \text{une relation de cointegration} \end{cases}$$

Test de cointégration en deux étapes de Engle et Granger.

- Estimation de la regression ADF par les MCO
- Statistique du test $t_{\hat{\rho}}$.
- On rejette l'hypothèse d'absence de relation de cointégration si la probabilité critique du test est inférieure au niveau du risque de première espèce choisi.
- Remarque : d'autres tests de cointégration
- Test de cointégration de Johansen
- permet de vérifier l'existence de plus d'une relation de cointégration, lorsque l'on a plus de deux variables.