Introduction à la cointégration

Gilbert Colletaz

3 mars 2020

Résumé

... en construction ..

Table des matières

1 La Cointégration

Cette propriété concerne les séries non stationnaires, et au sein de celles-ci, les séries intégrées. Le concept a été présenté par Granger en 1981 [1] et re-précisé en 1987 par Engle et Granger [2]. Il a intéressé les économistes en raison de la connection existant entre les équations d'équilibre de leurs modèles et cette propriété de cointégration. Par ailleurs, le théorème de représentation de Granger qui s'applique aux séries cointégrées a rationalisé la structure des modèles à correction d'erreur dans lesquels la dynamique d'une variable intégrée va dépendre à la fois de l'état de la conjoncture et des ajustements vers un équilibre de long terme.

1.1 Définition

Soient deux séries x_t et y_t intégrées d'ordre d. S'il existe deux constantes, α et β telles que leur combinaison linéaire, $z_t = \alpha x_t + \beta y_t$, est intégrée d'ordre inférieur à d alors on dira que x_t et y_t sont cointégrées et $(\alpha, \beta)^\top$ est appelé vecteur cointégrant.

Ce sera le cas par exemple si $x_t \sim I(1)$, $y_t \sim I(1)$ et qu'il existe un couple (α, β) tels que $\alpha x_t + \beta y_t \sim I(0)$ ou bien encore si $x_t \sim I(2)$, $y_t \sim I(2)$ avec (α, β) tels que $\alpha x_t + \beta y_t \sim I(1)$ ou $I(0)$. Dans la suite de ce cours nous ne traiterons pas des variables intégrées d'ordre 2 : seules des non stationnaires I(1) seront considérées, et elles seront donc cointégrées si on peut construire une combinaison linéaire qui soit I(0).

1. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, pp. 121-130, 1981.
2. Co-Integration and Error Correction : Representation, Estimation, and Testing, *Econometrica* Vol. 55, No. 2, , pp. 251-276, 1987.

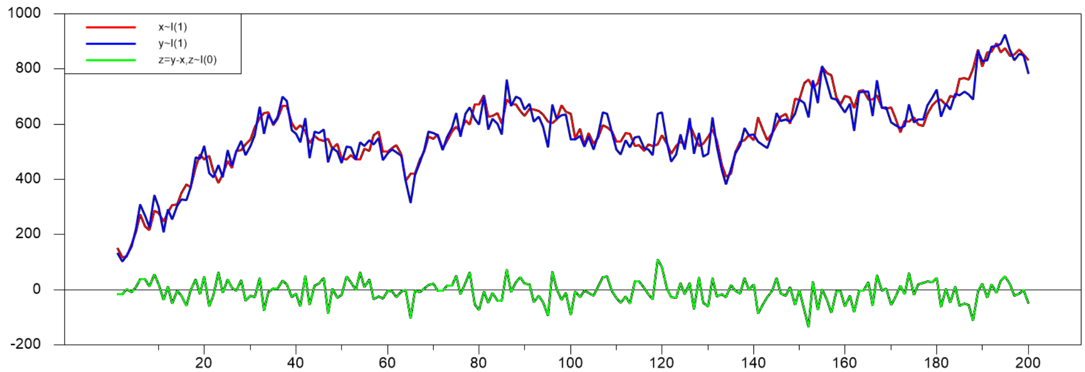

Figure 1 – Exemple de deux séries I(1) cointégrées.

Soit par exemple

$$\begin{aligned}
&x_t = x_{t-1} + u_t,\\
&u_t \sim i.i.d\ \mathcal{N}(0, \sigma_u^2),\ \text{avec } \sigma_u^2 \text{ une constante positive},\\
&y_t = x_t + v_t,\\
&v_t \sim i.i.d\ \mathcal{N}(0, \sigma_v^2),\ \text{avec } \sigma_v^2 \text{ une constante positive},\\
&z_t = y_t - x_t
\end{aligned}$$

Ici x_t qui est une marche au hasard est intégrée d'ordre 1, y_t qui est une marche au hasard augmentée d'une composante stationnaire est également intégrée d'ordre 1, mais $z_t = y_t - x_t = v_t$ est intégrée d'ordre 0. Ainsi, $(x_t, y_t)^\top$ est constitué de variables cointégrées de vecteur cointégrant $(-1, 1)^\top$. La Figure 1 reproduit une simulation sur 200 observations du système d'équation précédent.

Remarques :

— Toute transformation linéaire d'un vecteur cointégrant est également cointégrante. En effet, si (α, β) est cointégrant, alors $\alpha x_t + \beta y_t \sim I(0)$, mais alors pour toute constante c, la combinaison linéaire réalisée avec la transformée $(c\alpha, c\beta)$ vérifie $c\alpha x_t + c\beta y_t = c(\alpha x_t + \beta y_t)$ et est donc bien I(0) : $(c\alpha, c\beta)^\top$ est cointégrant.

— Dans un cadre bivarié, et à ses transformées linéaires près, le vecteur cointégrant est unique. Preuve par l'absurde : soient deux variables x et y intégrées d'ordre 1 telles que $z_1 = \alpha_1 x_t + \beta_1 y_t \sim I(0)$ et un autre vecteur cointégrant, non proportionnel au premier tel que $z_2 = \alpha_2 x_t + \beta_2 y_t \sim I(0)$. Cette dernière propriété, si elle est vraie, assure que $-(\alpha_1/\alpha_2)[\alpha_2 x_t + \beta_2 y_t]$ est également I(0). L'addition de z_1 et de $-(\alpha_1/\alpha_2)z_2$, qui sont toutes deux I(0), donnerait une variable également I(0) définie par $(\beta_1 - (\alpha_1/\alpha_2)\beta_2)y_t$ ce qui impliquerait que y_t est I(0) or elle est I(1).

1.2 Cointégration et équations d'équilibre de long terme

Pour beaucoup, les théories économiques ne sont capables que d'énoncer des propriétés relatives à des équilibres de long terme. Afin d'illustrer le lien existant entre ce type d'équilibre et le concept de cointégration nous allons prendre comme premier exemple la théorie de la Parité des Pouvoirs d'Achat.

1. L'exemple de la PPA

   Vous savez que sous un certain nombre d'hypothèses simplificatrices (pas de coût de transport, pas de risque, etc...) on doit vérifier la loi du prix unique : le prix d'un bien exprimé en une monnaie quelconque doit être égal à son prix exprimé en une monnaie de référence, la conversion se faisant au moyen du taux de change courant. Ainsi, si $P_€$ est le prix en euros, $P_\$$ le prix en dollars et s le taux de change tel que 1 euro $=$ s dollars, il vient avec cette loi du prix unique : $P_\$ = s\, P_€$, ce qui nous donne une équation à 3 inconnues. Considérant que les prix sont plus visqueux que les taux de change, la théorie de la PPA considère que cette équation détermine le taux de change d'équilibre pour des niveaux donnés de prix, *i.e.* :

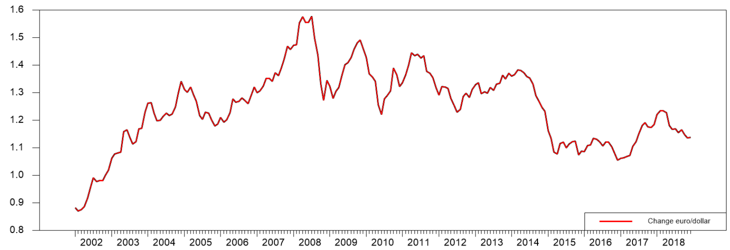

FIGURE 2 – Taux de change euro-dollar, janvier 2002 - décembre 2018

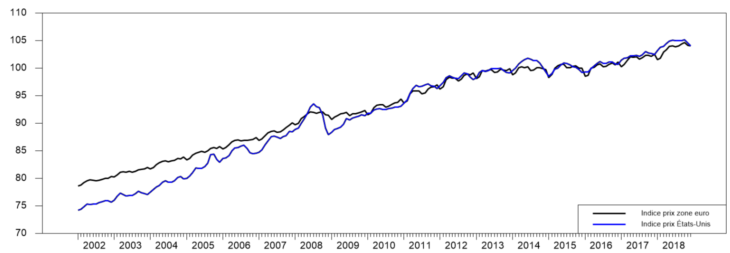

FIGURE 3 – Indices harmonisés des prix à la consommation, janvier 2002 - décembre 2018

$$s_t = \frac{P_{\$t}}{P_{€t}} \tag{1}$$

Dans la suite du texte nous avons utilisé la moyenne mensuelle du change entre l'euro et le dollar US pour représenter s_t. Les prix sont les indices harmonisés des prix à la consommation base 100 en 2015, observés mensuellement pour les États-Unis d'une part et la zone euro [3] d'autre part. Toutes les variables sont prises dans la base de données *Eurostat* sur la période janvier 2002 à décembre 2018. Leurs trajectoires sont représentées dans les Figures 2 et 3.

Avant de continuer, il convient de relever une difficulté dans l'application directe de l'équation de la PPA : les prix étant des indices, le résultat de la division de l'un part l'autre ne peut pas être directement comparé à la valeur du change. Par exemple, sur l'année 2015 ce rapport vaut 1. Si on les avait mesurés en base 100 en 2002, leur rapport aurait été égal à 1 cette année là et ceci indépendamment de la valeur du taux de change. En d'autres termes, si on observe la valeur exacte du change, celle du rapport des indices est arbitraire et il faut introduire une contrainte pour donner un sens à leur comparaison. Nous allons imposer par la suite que la PPA est vérifiée en moyenne sur toute la période d'observation.

Pour prendre en compte cette contrainte, il suffit de construire une série de prix relatifs ajustés telle que sa moyenne sur les 17 années couvertes par les données soit égale à celle du change. En effet, avec les deux moyennes :

$$\bar{s} = \frac{1}{204}\sum_{t=1}^{204} s_t, \text{ et } \overline{PR_{€\$}} = \frac{1}{204}\sum_{t=1}^{204} \frac{P_{\$t}}{P_{€t}}$$

on peut définir une série de prix relatifs ajustés comme :

$$\widetilde{PR}_{€\$,t} = \left(\bar{s}/\overline{PR_{€\$}}\right)\frac{P_{\$t}}{P_{€t}}$$

Par construction les variables s_t et $\widetilde{PR}_{€\$,t}$ possèdent la même moyenne, $\bar{s}$: la PPA est donc bien vérifiée en moyenne de 2002 à 2018, avec une valeur d'équilibre qui à toute date t serait donc égale à $\widetilde{PR}_{€\$,t}$. La Figure 4 présente les évolutions de ces deux variables.

Le premier enseignement de ce graphique est que la théorie de la PPA n'est pas valide : les deux courbes n'étant

3. Constituée de 12 pays au début de la période d'observation, de 13 pays à partir de 2007, de 15 en 2008, 16 en 2009, 17 en 2011, 18 en 2014, et 19 en 2015.

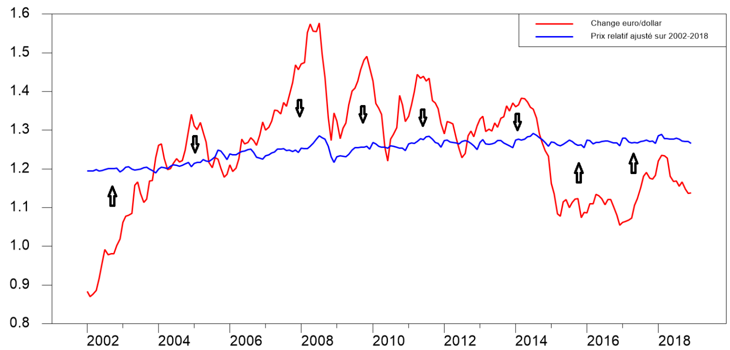

FIGURE 4 – change euro-dollar et prix relatifs ajustés, PPA vérifiée en moyenne de 2002 à 2018

pas superposées, l'équation (1) n'est pas vérifiée. Supposons que l'on veuille cependant défendre cette théorie de la PPA en montrant que même si elle n'est pas vérifiée à court terme, elle peut toutefois être utile sur des horizons plus longs. Pour cela, on interprète le niveau d'équilibre qu'elle définit, non plus comme une cible de court terme mais comme un attracteur de moyen terme pour le change courant.
On peut avec cette interprétation relire cette Figure 4 : de 2002 à 2004 sur le marché des changes l'euro est sous-évalué [4] puisque le change est inférieur à son prix d'équilibre. On devrait donc attendre un retour du change vers son attracteur et donc une appréciation de l'euro. C'est exactement ce qui est observé. De 2004 à la mi-2005, le change est passé au-dessus du prix d'équilibre défini par la PPA et on devrait attendre une correction allant vers une dépréciation de la monnaie européenne, ce qui intervient à la fin de 2005. Sur 2006-2007 on observe une forte appréciation de l'euro dont la valeur passe très largement au-dessus de ce qui serait son prix d'équilibre. Si celui-ci est un attracteur, on doit s'attendre à une correction à la baisse et elle intervient effectivement au début de l'année 2008. Ce type de raisonnement peut être mené sur la totalité de la période et sur ce graphique, le sens des flèches indique la direction attendue du change à une date donnée lorsque le prix d'équilibre de la PPA est une cible de moyen terme pour le niveau courant du change. Avec cette logique, on pourrait tirer deux enseignements :

(a) La théorie de la PPA ne serait pas valide à court terme, mais elle peut être utile à toute date pour construire des prévisions en donnant le sens attendu du change courant dans les mois qui suivent,

(b) Si le change a une tendance à retourner vers sa valeur d'équilibre c'est qu'il ne peut s'en écarter sans limite. Ce serait le cas avec un écart qui serait I(1), mais alors on ne pourrait pas défendre l'idée d'une tendance du change courant à retourner vers $\widetilde{PR}_{€\$,t}$: la PPA ne définirait même pas un attracteur à moyen ou long terme, le change évoluant indépendamment des prix relatifs. A contrario, la stationnarité de l'écart devient une condition nécessaire pour la validité de la PPA à moyen-long terme. Au final, si par exemple les séries de change et de prix relatifs sont I(1), cette version de la théorie de la PPA exige pour être valide qu'elle définisse avec les prix relatifs une cible de long terme telle que l'écart entre le change et ces prix relatifs soit I(0).

2. La leçon à retenir de l'exemple précédent est simple : dès lors qu'une théorie exprime une quantité d'équilibre qui est intégrée d'ordre 1, elle ne peut être valide que si l'écart entre la valeur courante de la variable concernée et la valeur d'équilibre est I(0). La justification de ce résultat est l'existence d'une force de rappel vers la position d'équilibre. En son absence, l'écart est également I(1) et l'équilibre en question ne définit rien de particulier pour la variable concernée : la théorie en question n'est pas validée. Ce raisonnement va être justifié par le théorème de représentation de Granger.

1.3 Théorème de représentation de Granger et modèle à correction d'erreur

On énonce ce théorème sans le démontrer. Nous reviendrons sur ce point ultérieurement lorsque nous traiterons de l'approche de Johansen.

1. Énoncé : Soient deux variables x_t et y_t intégrées d'ordre 1 et cointégrées, *i.e.* il existe un couple de constantes (α, β) tel que $z_t = \alpha x_t + \beta y_t$ est I(0) alors leurs variations sont gouvernées par les équations suivantes :

4. De façon équivalente on peut dire que le dollar est surévalué sur cette première sous période.

$$\Delta x_t = b_x z_{t-1} + \sum_{i=1}^{p} a_i \Delta x_{t-i} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + u_{x,t},$$
$$\Delta y_t = b_y z_{t-1} + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}$$

avec au moins un des deux coefficients b_x ou b_y non nul.
Vous noterez déjà que toutes les variables de ce système sont I(0).

2. Signification : Avec la section précédente, nous savons que derrière la cointégration peut se cacher une relation d'équilibre de long terme entre les variables. Afin de préciser la signification du théorème, nous allons arbitrairement supposer que x définit un équilibre pour y, soit :

$$y_t = f_0 + f_1 x_t + e_t \tag{2}$$

où $f_0 + f_1 x_t$ constitue la position d'équilibre, et e_t l'écart entre la valeur courante de y et cette cible. Cette relation d'équilibre et la relation de cointégration sont évidemment liées puisque :

$$\alpha x_t + \beta y_t = z_t \Leftrightarrow y_t = -(\alpha/\beta) x_t + (1/\beta) z_t \tag{3}$$

et donc $f_1 = -(\alpha/\beta)$ et $e_t = (1/\beta) z_t$. Comme il y a cointégration, z_t est I(0) et en conséquence l'écart à l'équilibre, e_t, doit également être I(0) alors que les variables x, y ainsi que la cible sont I(1) : on retrouve bien des caractéristiques énoncées dans la section précédente. On peut maintenant remplacer le terme z_{t-1} du théorème de représentation par son expression en termes d'écart à la position d'équilibre e_{t-1}. Il vient immédiatement :

$$\Delta x_t = \delta_x e_{t-1} + \sum_{i=1}^{p} a_i \Delta x_{t-i} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + u_{x,t}, \text{ avec } \delta_x = b_x \beta \tag{4}$$
$$\Delta y_t = \delta_y e_{t-1} + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}, \text{ avec } \delta_y = b_y \beta \tag{5}$$

A ce stade, on voit que les variations en t de x ou de y sont expliquées par deux éléments distincts :

— d'une part l'écart à la position d'équilibre de long terme, c'est l'élément correspondant à la force de rappel à laquelle on faisait référence précédemment,
— d'autre part la conjoncture représentée par les variations passées des deux variables x et y.

Par ailleurs vous devez également comprendre l'exigence d'au moins un coefficient b_x ou b_y non nul dans le théorème de représentation :
— Si les deux sont nuls, cela signifie que l'écart à l'équilibre, e_{t-1}, n'intervient pas dans l'évolution de l'une et de l'autre variables : elles évoluent indépendamment l'une de l'autre au regard de la cible d'équilibre, et sans force de rappel il ne peut y avoir cointégration entre les I(1).

C'est en raison de la présence du terme en e_{t-1} que ce type de système est qualifié de modèle à correction d'erreur. Remarquez également qu'en son absence, c'est à dire s'il n'y a pas cointégration entre les deux variables I(1) alors la modélisation adaptée est simplement un VAR d'ordre p sur les différences premières Δx_t et Δy_t.

Enfin, dans le cadre d'un MCE et dans les cas simples on peut encore préciser une condition de stabilité de l'équilibre. En effet, en présence de cointégration :
— Si δ_x est nul dans (4) alors $\delta_y \neq 0$ dans (5) : la variable x est exogène par rapport à l'écart à l'équilibre de long terme e_t et donc, lorsque y_{t-1} n'est pas égal à sa cible on attend un mouvement correcteur dans la seule trajectoire de y en t. Plus précisément si $e_{t-1} > 0$, i.e. d'après l'équation (2), si y_{t-1} est au-dessus de sa cible alors y doit baisser en t pour aller dans la direction de l'équilibre et donc Δy_t doit être négatif : le coefficient δ_y doit être négatif. Symétriquement, lorsqu'en $t-1$ la variable y est inférieure à sa position d'équilibre, alors l'écart $e_{t-1} < 0$ ce qui doit provoquer en t une variation positive de y qui doit alors se rapprocher de sa cible, comme précédemment le coefficient δ_y doit être négatif. Cette condition de signe sur δ_y assure donc la stabilité de l'équilibre dans le cas où x définit la cible d'équilibre et où seul y s'ajuste.

— On vous laisse vérifier que dans le cas où $\delta_y = 0$ dans (5) , et que donc $\delta_x \neq 0$ dans (4), alors seul x s'ajuste et il est préférable de définir la relation d'équilibre comme $x_t = f_0 + f_1 y_t + e_t$, plutôt que comme $y_t = f_0 + f_1 x_t + e_t$. Si on opère ce renversement alors un raisonnement identique au précédent montre que la stabilité de l'équilibre exige $\delta_x < 0$.

— les choses sont évidemment plus complexes lorsque le déséquilibre est susceptible d'affecter les trajectoires des deux variables, i.e. lorsque $\delta_x \neq 0$ et $\delta_y \neq 0$. Ici, contrairement aux deux configurations précédentes, aucune variable n'est exogène relativement à l'écart d'équilibre e_t : les trajectoires des deux variables et de la cible d'équilibre vont être affectées.

1.4 Intégration, cointégration et choix de modélisation

On rappelle que dans cette introduction, les variables que l'on considère sont intégrées au maximum à l'ordre 1. Sous cette condition, et compte tenu des acquis des sections précédentes, il est déjà possible d'énoncer des règles qui vont guider le choix de la modélisation dynamique à appliquer à un vecteur de k variables $y_1, y_2, \ldots, y_k$.

1. Si toutes les variables sont I(0), alors un processus VAR standard est adapté, soit dans un cadre bivarié [5] :

$$x_t = \delta_x e_{t-1} + \sum_{i=1}^{p} a_i x_{t-i} + \sum_{i=1}^{p} b_i y_{t-i} + u_{x,t}$$
$$y_t = \delta_y e_{t-1} + \sum_{i=1}^{p} c_i x_{t-i} + \sum_{i=1}^{p} d_i y_{t-i} + u_{y,t}$$

2. Si toutes les variables sont I(1) non cointégrées, alors un VAR standard sur leurs différences premières qui sont I(0) doit être utilisé,

$$\Delta x_t = \sum_{i=1}^{p} a_i \Delta x_{t-i} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + u_{x,t}$$
$$\Delta y_t = \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}$$

3. Si toutes les variables sont I(1) cointégrées, alors, d'après le théorème de représentation, un modèle à correction d'erreur doit être ajusté,

$$\Delta x_t = \delta_x e_{t-1} + \sum_{i=1}^{p} a_i \Delta x_{t-i} + \sum_{i=1}^{p} b_i \Delta y_{t-i} + u_{x,t}$$
$$\Delta y_t = \delta_y e_{t-1} + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}, \text{ avec } e_t \text{ proportionnel à } z_t \text{ définit par}$$
$$z_t = \alpha x_t + \beta y_t \sim I(0)$$

sachant également que dans toutes les écritures précédentes, $(u_x, u_y)^\top$ est un bruit blanc.

Vous connaissez déjà les tests de racine unitaire à la Dickey-Fuller, l'important maintenant est donc de disposer d'un test de cointégration.

2 Comment tester la cointégration : de Dickey-Fuller à Engle-Granger

Supposons que nous soyons en présence de deux séries I(1), la section précédente souligne que la question à laquelle on doit répondre pour la conduite à suivre est celle de leur cointégration ou de leur non cointégration. On doit donc disposer d'un test de cointégration. Deux cas de figure doivent alors être considérés : soit on connaît a priori le vecteur cointégrant, et donc la variable sur laquelle est portée la question, soit on ne connaît pas ce vecteur et il est nécessaire de l'estimer. Dans ce dernier cas, la variable sur laquelle on s'interroge est $\alpha x_t + \beta y_t$ alors qu'on ne disposera au mieux que de $\hat{\alpha} x_t + \hat{\beta} y_t$. Dans le premier cas, on est placé dans des conditions qui sont exactement celles du test de Dickey-Fuller : on observe une

5. Nous verrons ultérieurement comment ces équations se modifient lorsque le vecteur étudié est composé de plus de deux variables.

variable et on s'interroge sur la présence d'un trend stochastique dans son niveau. Dans le second, on ne dispose au mieux que d'un estimateur de la variable d'intérêt et c'est le test d'Engle-Granger qui devra être réalisé. Nous allons illustrer ces propos au moyen de deux exemples.

2.1 Le vecteur cointégrant est connu a priori

Reprenons l'exemple de la PPA considéré dans l'une des sections précédentes. La question est de savoir si l'écart entre le change courant et le prix relatif d'équilibre est I(0), soit :

$$H0 : s_t - \widetilde{PR}_{€\$,t} \sim I(1), \text{versus}$$
$$H1 : s_t - \widetilde{PR}_{€\$,t} \sim I(0)$$

Dans ce cadre bivarié on sait que le vecteur cointégrant, s'il existe, est unique et égal à (1-1). La série sur laquelle on s'interroge est donc parfaitement connue dès lors que le change courant et la valeur cible sont observés. On est donc placé dans le cadre d'application du test de Dickey-Fuller : il s'agit de savoir si on ne rejette pas la présence d'un trend stochastique dans une série donnée a priori, c'est à dire sans qu'il soit besoin de réaliser une quelconque estimation.

Cependant, indépendamment de ce point essentiel, notre exemple amène une difficulté qui lui est propre : on sait que les valeurs de $\widetilde{PR}_{€\$,t}$ sont arbitraires. Ici on a supposé que la PPA était vérifiée en moyenne sur la période 2002-2018 ; si on avait exigé qu'elle soit vérifiée en 2002, nous aurions simplement multiplié notre série par une constante de sorte que les moyennes de s_t et des prix relatifs soient identiques en 2002. Si on on décide tout aussi arbitrairement qu'elle devait être valide en 2018, nous aurions multiplié la série actuelle par une constante assurant l'égalité des moyennes du change et des prix relatifs en 2018, etc... En d'autres termes, la série des prix d'équilibre n'est connue qu'à une transformée linéaire près : quelle que soit la constante positive c, la série $c \times \widetilde{PR}_{€\$,t}$ est tout aussi raisonnable que la série $\widetilde{PR}_{€\$,t}$ elle-même. Elle change simplement la base sur laquelle on impose que la PPA est vérifiée. Dans cet exercice, la question est donc de savoir si nous pouvons statuer sur la non validité de la PPA indépendamment de la série de prix relatifs utilisée, *i.e.* quelle que soit la valeur (positive) de c. Il peut sembler que la réponse est négative puisque si $s_t - \widetilde{PR}_{€\$,t}$ est I(0) alors, en raison de l'unicité du vecteur cointégrant, $s_t - c \times \widetilde{PR}_{€\$,t}$ ne peut être I(0) si $c \neq 1$. En d'autres termes, si la PPA n'est pas rejetée pour une certaine base des prix relatifs, elle devrait théoriquement l'être pour toute autre base.
Fort heureusement en passant aux logarithmes on lève cette difficulté : comme $\log(s_t) - \log(c \times \widetilde{PR}_{€\$,t}) = log(s_t) - \log(\widetilde{PR}_{€\$,t}) - \log(c)$, et que l'ajout d'une constante ne modifie pas l'ordre d'intégration, on vérifie évidemment que, pour tout $c > 0$ l'ordre d'intégration de $\log(s_t) - \log(c \times \widetilde{PR}_{€\$,t})$ est le même que celui de $\log(s_t) - \log(\widetilde{PR}_{€\$,t})$: cet ordre ne dépend plus du choix de la base.

Afin de réaliser le test de non validité de la PPA, il suffit donc de prendre les logarithmes des variables d'intérêt, s_t et $\widetilde{PR}_{€\$,t}$, pour ensuite, et ce qui est le plus important ici, leur appliquer un vecteur cointégrant connu a priori égal à (1,-1) afin de créer la variable observée $\log(s_t) - \log(\widetilde{PR}_{€\$,t})$ sur laquelle il suffit de mener un test de Dickey-Fuller.
En procédant ainsi, avec $e_t = \log(s_t) - \log(\widetilde{PR}_{€\$,t})$, on obtient :

$$\Delta e_t = \underset{(0.00153)}{0.0006} - \underset{(0.0143)}{0.04326}\, e_{t-1} + \underset{(0.067)}{0.2602}\Delta e_{t-1} + \hat{u}_t$$

Ce qui donne une statistique $\tau_\mu = -3.03$. En conséquence à 5% de risque avec une valeur critique de l'ordre de -2.86 nous rejetons la présence d'une racine unitaire : l'écart serait I(0), le change et les prix relatifs cointégrés et nous ne pourrions donc pas rejeter la validité de la PPA à moyen terme sur la période étudiée.

2.2 Le vecteur cointégrant doit être estimé

Nous avons toujours deux variables x et y intégrées d'ordre 1 mais ici, contrairement à la section précédente, on ne connaît pas les coefficients α et β qui assurent que $\alpha x_t + \beta y_t$ est intégrée d'ordre 0. On s'interroge donc sur l'existence d'une relation d'équilibre qui correspondrait à l'équation (2), soit pour mémoire :

$$y_t = f_0 + f_1 x_t + e_t \tag{2}$$

Les développements précédents ont conclu en particulier que cette relation d'équilibre n'existe que si l'écart à la cible est I(0). Dans la dernière équation, la cible est définie par x_t, et l'écart en question est donc $e_t = y_t - f_0 - f_1 x_t$, c'est à dire le résidu de la régression linéaire de y_t sur x_t.

Vous devez alors comprendre aisément la logique du test de cointégration proposé par Engle et Granger qui s'effectue en deux étapes :

1. réaliser une estimation OLS de (2) pour récupérer les résidus empiriques $\hat{e}_t$,
2. réaliser un test de racine unitaire semblable techniquement au test de Dickey-Fuller sur cette série de résidus empiriques.

La règle de décision est également simple : ne pas rejeter la présence d'une racine unitaire dans $\hat{e}_t$ implique d'admettre que le résidu est I(1) c'est à dire qu'il n'y a pas cointégration entre les deux variables. A contrario, rejeter la présence d'une racine unitaire est favorable à l'hypothèse d'un résidu qui serait I(0) et donc à la l'existence d'une relation d'équilibre. Cette simplicité apparente dissimule en fait un ensemble de difficultés que nous allons discuter. On abordera notamment les questions suivantes :

— dans la première étape, on peut se demander quelles sont les propriétés des OLS appliqués à des variables I(1) ?
— dans la deuxième étape, si la mise en oeuvre du test de présence d'une racine unitaire d'Engle-Granger est identique à celle de Dickey-Fuller pour autant, les deux statistiques possèdent-elles la même distribution sachant qu'ici la variable utilisée n'est pas observée, comme chez Dickey-Fuller, mais est le résultat d'une estimation préalable ?
— dans ce cadre bivarié, deux régressions peuvent être réalisées à la première étape : expliquer y_t par x_t comme dans (2), ou expliquer x_t par y_t mais alors les conclusions de la seconde étape ne risquent-elles pas de différer selon le choix fait à la première ?

2.2.1 OLS sur variables I(1) : régressions spécieuses ou super-convergence ?

1. Régressions sur des stationnaires en différence indépendantes : le cas des régressions spécieuses

   Une des premières publications ayant attiré l'attention sur les difficultés que rencontre l'utilisation des OLS sur des variables I(1) est due à Granger et Newbold [6]. Dans un travail essentiellement empirique, utilisant des séries simulées, ils font remarquer que les régressions de variables I(1) indépendantes entre elles amènent trop souvent à des conclusions erronées, les statistiques et tests usuels laissant souvent croire à l'existence de dépendances pourtant inexistantes . Ces régressions qu'ils qualifient de spécieuses (*spurious regressions*) ont comme caractéristiques :

   — d'avoir des statistiques de Fisher conduisant à des pourcentages de rejet de l'hypothèse de nullité des coefficients très largement supérieurs aux seuils de risque de première espèce habituellement utilisés,
   — de donner des statistiques de student trop souvent significatifs, souffrant donc des mêmes problèmes que les F,
   — d'afficher des R^2 souvent supérieurs à 50%,
   — d'avoir des statistiques de Durbin-Watson souvent proches de zéro, signalant une autocorrélation des résidus.

   Ils soulignent également que ces résultats fallacieux sont exacerbés lorsque le nombre d'explicatives augmente. Pour illustration, la table 1 présente une partie de leurs résultats. Notez qu'il faudra attendre 1986 et un travail de Phillips [7] pour que l'on ait les explications et la justification théorique des résultats de Granger-Newbold. Dans ce travail de Phillips on retiendra deux ensembles de résultats importants :

   (a) Lorsque le nombre d'observations T tend vers l'infini alors les OLS appliqués sur des I(1) indépendantes conduisent d'une part à ce que les t de student ne suivent plus une loi de student mais tendent vers l'infini et d'autre part, à ce que la statistique DW converge vers zéro. Remarquez que l'augmentation de la taille des échantillons ne fait pas disparaître les difficultés, au contraire, elle les exacerbe.
   (b) L'emploi de la parade "classique" sur séries non stationnaires consistant soit à régresser les variables en question sur un trend déterministe pour récupérer les résidus censés être alors stationnaires, ou l'inclusion dans la régression d'un trend déterministe ne résout rien : dans le premier cas on retrouve des régressions spécieuses entre les résidus et le problème est simplement déplacé mais non résolu, dans le second cas, les statistiques de student continuent de diverger avec T interdisant l'emploi des tests d'hypothèses usuels [8].

   Une solution pourrait paraître évidente : si le problème des spurious regression est dû au fait que les séries sont non stationnaires car intégrées, on pourrait évidemment penser à les différencier pour se ramener à des séries stationnaires

6. Granger, C. W. J., and Paul Newbold, Spurious Regressions in Econometrics, *Journal of Econometrics* 2 :111-120, 1974.
7. .Phillips, P.C.B., Understanding spurious regressions in econometrics, *Journal of Econometrics*, 33, 311-340, 1986.
8. L'intuition de ces résultats est qu'un trend déterministe n'équivaut pas à un trend stochastique. La régression sur un trend déterministe ne parvient pas à éliminer le stochastique qui est à l'origine des difficultés : la cause des difficultés ne disparaissant pas, ses conséquences sont toujours présentes.

nombre d'explicatives	% de rejets de la nullité des coefficients par Fisher	$\bar{R}^2$ moyen	% $\bar{R}^2 > 0.7$
1	76	0.26	5
2	78	0.34	8
3	93	0.46	25
4	95	0.55	34
5	96	0.59	37

TABLE 1 – Résultats d'Engle-Granger (1974) obtenus sur les niveaux de variables I(1) indépendantes

sur lesquelles les OLS retrouveraient des propriétés satisfaisantes. Il s'agit en fait d'une solution dangereuse : elle est bonne si les séries initiales sont indépendantes, elle est mauvaise si les séries sont liées par une relation d'équilibre, *i.e.* si elles sont cointégrées. Pour comprendre la difficulté, considérons successivement chacun des cas :

(a) soit $x_t = x_{t-1} + u_t$, $y_t = y_{t-1} + v_t$ et (u_t, v_t) un couple de bruits blancs indépendants. On est donc en présence de deux marches au hasard indépendantes : d'une part la régression $y_t = \beta_0 + \beta_1 x_t + e_t$ est spécieuse avec une valeur de β_1 théoriquement nulle, d'autre part son résidu e_t est I(1). En revanche sur les séries prises en différences premières, qui sont alors stationnaires, il vient : $\Delta y_t = \beta_1 \Delta x_t + \Delta e_t$, soit $v_t = \beta_1 u_t + r_t$ et ici toutes la variables sont stationnaires. Cette dernière régression devrait fournir des R^2 proches de zéro et un pourcentage de rejets de l'hypothèse de nullité de β_1 se rapprochant du seuil de risque nominal [9].

(b) On suppose maintenant toujours des séries x_t et y_t intégrées d'ordre 1 mais liées par une relation d'équilibre, soit $y_t = f_0 + f_1 x_t + e_t$ avec donc $e_t \sim I(0)$ et $E[e_t] = 0$. Sur les différences premières il vient $\Delta y_t = f_1 \Delta x_t + r_t$ avec $E[r_t] = 0$. Supposons pour simplifier que x ne varie pas, $x_t = x_0$ pour tout t, que $e_t = 0$ pour tout t à l'exception de $e_1 = 1$. Ce choc unitaire fait dévier y de sa position d'équilibre à la date 1 : $y_1 = f_0 + f_1 x_0 + 1 = y_{eq} + 1$. On sait que ce choc étant stationnaire, ses effets sont transitoires et donc la trajectoire à venir de y va ramener cette variable vers son niveau d'équilibre y_{eq}. Cependant quelqu'un qui ne connaîtrait que l'équation en différences et ayant observé l'augmentation d'une unité de y en $t = 1$ anticiperait les variations futures sur la base de $\Delta y_{t+h} = f_1 \Delta x_{t+h}$, soit $\Delta y_{t+h} = 0$ avec nos hypothèses : il ne pourrait prévoir la baisse à venir ramenant y sur y_{eq}. [10].

Pour clore cette section, retenez qu'Engle et Granger énonce une règle pratique de détection d'une régression spécieuse : on doit être particulièrement vigilant lorsque la statistique de Durbin-Watson est inférieure au R^2 de la régression réalisée sur des variables stationnaires en différences.

2. Régressions sur des variables cointégrées : la super-convergence

Ayant discuté des difficultés rencontrées lorsqu'on régresse des variables intégrées non stationnaires indépendantes entre elles, on peut maintenant se demander ce qu'il advient lorsque ces variables ne sont plus indépendantes, et plus précisément lorsqu'elles sont cointégrées. Dans ce qui suit, nous avons donc deux variables x_t et y_t qui sont $I(1)$ et deux constantes α et β telles que $z_t = \alpha x_t + \beta y_t$ est $I(0)$
Un premier résultat, dû à Stock [11] est que les estimateurs des OLS d'une régression de la forme $y_t = f_0 + f_1 x_t + e_t$ convergent vers les vraies valeurs, et ils le font avec un vitesse d'ordre T et non pas en $\sqrt{T}$ comme dans le cas standard. En d'autres termes, ils convergent plus vite que ne le font les estimateurs lorsque les hypothèses usuelles des OLS, et notamment la stationnarité des variables utilisées, sont vérifiées. On dit qu'ils sont super-convergents [12]. Notez cependant que cette propriété de super-convergence n'empêche pas que les statistiques de Fisher et de student ne possèdent pas leurs distributions habituelles : on ne peut pas utiliser les tests d'hypothèses usuels sur les coefficients f de cette régression [13]. Elle permet cependant d'obtenir plusieurs résultats utiles :

9. D'ailleurs Granger et Newbold, sur les séries simulées ayant conduit aux résultats donnés dans la table 1 mais prises en différences premières affichent alors des R^2 en moyenne de l'ordre de 1%, et des pourcentages de rejet de l'hypothèse de nullité des coefficients compris entre 8% et 10% pour un seuil de risque nominal de 5%.

10. L'intuition de ce résultat est simple : la cointégration impose une contrainte d'équilibre sur le niveau des variables intégrées. En passant aux différences on perd le niveau et donc la relation de long terme qui lie les variables en question.

11. Stock, J.H., Asymptotic Properties of Least Squares Estimation of Cointegrating Vectors, *Econometrica*, 55, 1035-1056, 1987

12. On peut avoir l'intuition de ce résultat en considérant une régression de la forme $y_t = f_1 x_t + e_t$, les résidus des OLS sont donnés par $\hat{e}_t = y_t - \hat{y}_t = y_t - \hat{f}_1 x_t = (f_1 - \hat{f}_1)x_t + e_t$. L'erreur que les OLS cherchent à minimiser a deux composantes : l'erreur théorique e_t et l'écart entre le vrai coefficient et son estimation pondérée par x_t. Dans la minimisation de la somme des carrés des erreurs le poids de $f_1 - \hat{f}_1$ devient relativement plus important lorsque x est non stationnaire.

13. Nous reviendrons ultérieurement sur les propriétés des tests usuels dans les VAR sur séries stationnaires en différence cointégrées ou non.

(a) Les estimateurs OLS sont super-convergents aussi bien pour la régression de y_t sur x_t que pour celle de x_t sur y_t : ceci résout le problème de l'indétermination du choix de la régression à effectuer lors de la première étape du test d'Engle-Granger. Il faut cependant se rappeler que cette propriété est valide asymptotiquement : sur petits échantillons des biais peuvent apparaître notamment si le R^2 de la régression de cointégration s'éloigne de l'unité.

(b) Une autre conséquence est de faciliter l'estimation des modèles à correction d'erreur. Considérons ainsi les équations (2) et (5) vues précédemment :

$$y_t = f_0 + f_1 x_t + e_t \tag{2}$$

$$\Delta y_t = \delta_y e_{t-1} + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t} \tag{5}$$

La première est une régression de cointégration et s'il y a cointégration, les estimateurs de f_0 et f_1 convergent à la vitesse T. Dans la seconde, qui correspond au modèle à correction d'erreur, toutes les variables sont $I(0)$ et les estimateurs OLS convergent à la vitesse usuelle $\sqrt{T}$. Compte-tenu de ces différences dans les vitesses de convergence, on peut considérer pour l'estimation des δ_y, c_i et d_i, que les coefficients f_0 et f_1 ont convergé et sont donc des constantes lorsqu'on les intègre dans (5) [14]. Pour estimer le MCE, il suffira d'estimer (2) dans une première étape et de remplacer e_{t-1} dans (5) par les résidus empiriques $\hat{e}_t = y_t - \hat{f}_0 - \hat{f}_1 x_t$. On ajustera donc à la seconde étape :

$$\Delta y_t = \delta_y (y_{t-1} - \hat{f}_0 - \hat{f}_1 x_{t-1}) + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}, \text{ soit encore,}$$

$$\Delta y_t = \delta_y \, \hat{e}_{t-1} + \sum_{i=1}^{p} c_i \Delta x_{t-i} + \sum_{i=1}^{p} d_i \Delta y_{t-i} + u_{y,t}$$

14. Plus précisément, Engle et Granger montrent que les estimateurs des coefficients du MCE ont les mêmes propriétés que le vecteur cointégrant soit connu ou estimé, *i.e.* que l'on utilise e_{t-1} ou $\hat{e}_{t-1}$ dans l'estimation du MCE.