

n° 9923

**Guide pratique
des séries non-stationnaires**

B. SALANIE¹

Mai 1999

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

¹ INSEE - D.E.S.E. - 15 Boulevard Gabriel Péri, 92245 Malakoff Cédex.

Cette version corrige quelques erreurs de détail de la précédente. Je remercie Pauline Givord, Guy Laroque, Françoise Maurel, Alain Monfort, Céline Prigent, l'éditeur et deux rapporteurs, et tout particulièrement Stéphane Grégoir et Fabrice Lenglart pour leurs commentaires.

Abstract

The purpose of this paper is to present traditional and more recent procedures for estimation and testing of systems that contain non-stationary variables. The aim throughout is not to prove asymptotic results, but rather to weigh the pros and cons of using each method according to the circumstances. This text attempts to show that more recent methods may be more convenient and/or robust than earlier procedures.

Keywords : non-stationarities, cointegration.

Résumé

L'objet de ce texte est de présenter les méthodes économétriques disponibles pour procéder à l'estimation et au test de systèmes comprenant des variables non-stationnaires, en insistant sur les contributions récentes qui fournissent des méthodes plus simples et/ou plus robustes que celles qui ont été développées au début des années quatre-vingt. L'approche utilisée est résolument pragmatique : aucun résultat asymptotique n'est démontré, mais on cherche à évaluer les avantages et inconvénients de l'utilisation de chaque procédure.

Mots-clé : non-stationnarité, cointégration.

Introduction

Soit X_t une série temporelle, c'est-à-dire une suite de variables aléatoires indexées par l'entier t . On dit qu'elle est stationnaire (du second ordre) quand ses deux premiers moments, soit les espérances

$$\mu_t = EX_t$$

et les autocovariances

$$\gamma_{tk} = \text{cov}(X_t, X_{t-k})$$

sont finies et indépendantes de t . C'est le cas notamment de toutes les séries engendrées par des ARMA. Visuellement, une telle série tend à retourner à sa moyenne quand elle s'en est écartée sous l'effet de chocs.

Considérons maintenant une marche aléatoire $X_t = \sum_{\tau=1}^t \varepsilon_\tau$, où les ε_t sont des bruits blancs forts (c'est-à-dire qu'ils sont indépendamment et identiquement distribués, de moyenne nulle et de variance σ^2 finie). On a alors $\mu_t = 0$ et (pour $t > k$) $\gamma_{tk} = \min(t, t-k)\sigma^2$, si bien que la série n'est pas stationnaire. En particulier, une marche aléatoire ne tend aucunement à revenir à sa moyenne : il n'y a pas de force de rappel. On notera aussi que pour $t > 0$, la variance de la série est

$$VX_t = \gamma_{t0} = t\sigma^2$$

et qu'elle explose donc au cours du temps. Ces deux propriétés sont typiques des séries qu'on appelle intégrées.

La marche aléatoire constitue de fait l'exemple de base d'une série $I(1)$, c'est-à-dire qu'il faut différencier une fois pour la rendre stationnaire (puisque $\Delta X_t = \varepsilon_t$ est à l'évidence stationnaire). On dit aussi qu'elle possède une racine unité. Cette dernière dénomination provient de la représentation AR fondamentale de la série :

$$A(L)X_t = \varepsilon_t$$

où $A(L)$ est un polynôme en l'opérateur retard L . Les racines du polynôme sont données par l'équation $A(z) = 0$ et, pour un processus autorégressif stationnaire, sont toutes de module supérieur à un. Dans le cas de la marche aléatoire, $A(L) = 1 - L$ a pour racine unique un, d'où le terme de racine unité. Dans tout ce guide, les seules séries non-stationnaires auxquelles nous nous intéresserons seront celles qui possèdent une racine unité ; on dit aussi qu'elles ont une tendance stochastique.

Pourquoi doit-on s'inquiéter des non-stationnarités ? Dans un article déjà ancien, Nelson et Plosser (1982) ont montré qu'on ne pouvait pas rejeter l'hypothèse de racine unité pour la plupart des séries macroéconomiques couramment utilisées. Or les techniques d'inférence statistique usuelles ne fonctionnent plus normalement quand les séries sont non-stationnaires. Ainsi, si on estime sur une marche aléatoire le modèle

$$X_t = \rho X_{t-1} + \varepsilon_t$$

pour lequel la vraie valeur de ρ est 1, l'application des résultats asymptotiques usuels suggérerait que l'estimateur $\hat{\rho}_T$ converge en $1/\sqrt{T}$ vers une loi normale :

$$\sqrt{T}(\hat{\rho}_T - 1) \longrightarrow N(0, \omega^2)$$

et que le test de Student de l'hypothèse $\rho = 1$ converge vers une loi normale centrée réduite. En fait, $(\hat{\rho}_T - 1)$ converge en $1/T$ vers une loi non-normale et le Student converge vers une autre loi non-normale, si bien que le test habituel n'est plus valide ici.

La démarche classique concernant des séries “explosives” était de les détrender¹ en utilisant le modèle

$$X_t = a + bt + u_t$$

ce qui est inadapté puisque la non-stationnarité passe alors dans la série détrendée u_t . On peut de fait démontrer que si X_t est une marche aléatoire, l'estimateur de b tend alors vers zéro, et son Student tend vers l'infini—ce qui peut donner l'impression trompeuse que le détrending a bien fonctionné.

La situation est encore pire dans le cadre d'un modèle de régression. A l'évidence, régresser une série non-stationnaire sur une série stationnaire n'a aucun sens, puisque la non-stationnarité passe alors dans les résidus de la régression, avec les conséquences que l'on imagine. Mais régresser une série non-stationnaire sur une autre série non-stationnaire ne peut être une bonne idée que s'il existe une combinaison linéaire de ces deux séries qui est elle-même stationnaire (on dit alors que les séries sont cointégrées). Dans le cas contraire, il faut s'attendre à des résultats très bizarres. L'exemple le plus spectaculaire est la régression fallacieuse (*spurious regression*) identifiée par Granger et Newbold (1974). Soient deux marches aléatoires

$$X_t = X_{t-1} + u_t \text{ et } Y_t = Y_{t-1} + v_t$$

que nous supposons indépendantes (au sens où u_t et v_t sont indépendants). La régression de Y_t sur X_t par

$$Y_t = a + bX_t + \varepsilon_t$$

produira un estimateur de b qui, loin de converger vers zéro (sa “vraie” valeur), converge vers une loi aléatoire non-dégénérée, tout comme le R^2 de la régression. Le seul signe que quelque chose ne tourne pas rond est le Durbin-Watson de la régression, qui tend vers zéro puisque la non-stationnarité passe dans le résidu ε_t .

Ces exemples démontrent clairement que lorsqu'on étudie des relations entre séries macroéconomiques, il convient de s'interroger sur leur éventuelle non-stationnarité et sur les relations de cointégration qui peuvent exister entre elles. L'objet de ce guide pratique est de rappeler les procédures classiques (dont certaines avaient été présentées par Maurel (1989) dans cette même revue) et de

¹J'emploie cet horrible anglicisme pour “ôter les tendances déterministes”.

présenter des procédures plus récentes qui sont parfois mieux adaptées. Aucun résultat ne sera démontré ; on mettra l'accent sur l'application pratique de ces méthodes. La section 1 s'intéressera aux tests de stationnarité et la section 2 à l'extraction d'une tendance et d'un cycle dans une série non-stationnaire. La section 3 présentera les tests de cointégration et la section 4 discutera diverses méthodes d'estimation de systèmes cointégrés. L'annexe 1 analyse les méthodes d'estimation de la variance de long terme d'une série stationnaire, qui jouent un rôle crucial dans beaucoup de ces procédures. L'annexe 2 donne quelques indications sur les procédures disponibles dans divers logiciels commerciaux.

Il existe toute une littérature, à la suite de Perron (1989) et Rappoport et Reichlin (1989), qui étudie les séries non-intégrées à tendance déterministe segmentée, dont un exemple est

$$X_t = a + RX_{t-1} + bt + c\max(t - t_0, 0) + \varepsilon_t$$

où $R < 1$, ε_t est un bruit blanc et la tendance déterministe de la série change de pente en t_0 (on peut penser par exemple au ralentissement de la croissance après le premier choc pétrolier). Il est facile de voir que pour une telle série, les procédures habituelles de test restent valides si la date de rupture t_0 est connue de l'économetre. On constate par ailleurs qu'il est très difficile de discriminer sur la base des données entre une tendance stochastique et une tendance déterministe segmentée. Le lecteur doit être prévenu que je ne traiterai pas de cette littérature. Dans tout l'article, les séries seront stationnaires ou auront une racine unité, avec éventuellement une tendance déterministe en sus dans les deux cas.

1 Tests de stationnarité

Les premiers tests proposés (et les plus utilisés encore aujourd'hui) prennent pour hypothèse nulle la non-stationnarité. Plus récemment, un test considérant la stationnarité pour hypothèse nulle a été étudié. On les présentera dans cet ordre.

1.1 Hypothèse nulle de non-stationnarité

1.1.1 Les tests de Dickey-Fuller et apparentés

L'ancêtre des tests de non-stationnarité est le test de Dickey-Fuller (DF). Il en existe trois principaux. Dans les trois cas, la procédure de test consiste à effectuer la régression

$$\Delta X_t = d_t + \rho X_{t-1} + u_t \quad (DF)$$

où d_t est une fonction déterministe du temps, et à examiner les valeurs des statistiques D_τ (le Student de l'hypothèse $\rho = 0$) et $D_\rho = T\hat{\rho}$. Il s'agit là d'un test unilatéral : on rejette la non-stationnarité quand D_ρ ou D_τ sont inférieurs à une valeur critique. Il existe trois choix pour d_t dans (DF) :

Table 1

Tests de Dickey-Fuller, $d_t = a$

Valeur critique à	1%	5%	10%
Test D_ρ			
$T = 25$	-17,2	-12,5	-10,2
$T = 50$	-18,9	-13,3	-10,7
$T = 100$	-19,8	-13,7	-11,0
$T = 250$	-20,3	-14,0	-11,2
$T = \infty$	-20,7	-14,1	-11,3
Test D_τ			
$T = 25$	-3,75	-3,00	-2,63
$T = 50$	-3,58	-2,93	-2,60
$T = 100$	-3,51	-2,89	-2,58
$T = 250$	-3,46	-2,88	-2,57
$T = \infty$	-3,43	-2,86	-2,57

- $d_t = 0$: ce cas, qui implique que la série est d'espérance nulle sous l'hypothèse alternative, n'a que peu d'intérêt en pratique et ne sera pas examiné ici.
- $d_t = a$: l'hypothèse nulle de ce test est $a = \rho = 0$. Les valeurs critiques sont indiquées en Table 1.
- $d_t = a + bt$: l'hypothèse nulle est cette fois $b = \rho = 0$. Les valeurs critiques sont indiquées en Table 2.

En pratique, on choisira $d_t = a$ quand la série ne paraît pas contenir de dérive, et $d_t = a + bt$ quand elle semble en contenir une (ce choix peut être déterminé par un examen visuel ou par un jugement économique a priori). On notera que les valeurs critiques à distance finie n'ont de sens que quand ε_t est effectivement normal. Si l'on n'en est pas sûr, il vaut mieux utiliser les valeurs critiques données pour $T = \infty$, qui sont valables asymptotiquement².

On peut également tester l'hypothèse jointe $b = \rho = 0$ par un test de Fisher dans le modèle où $d_t = a + bt$. Les valeurs critiques asymptotiques de ce test sont 8,27 à 1%, 6,25 à 5% et 5,34 à 10%. Si cette hypothèse n'est pas rejetée, on peut

²Les valeurs critiques des tests reprises dans tout cet article sont calculées par simulation. Elles peuvent donc différer de quelques pourcents selon les auteurs, en fonction des méthodes de simulation employées.

Table 2

Tests de Dickey-Fuller, $d_t = a + bt$

Valeur critique à	1%	5%	10%
Test D_ρ			
$T = 25$	-22,5	-17,9	-15,6
$T = 50$	-25,7	-19,8	-16,8
$T = 100$	-27,4	-20,7	-17,5
$T = 250$	-28,4	-21,3	-18,0
$T = \infty$	-29,5	-21,8	-18,3
Test D_τ			
$T = 25$	-4,38	-3,60	-3,24
$T = 50$	-4,15	-3,50	-3,18
$T = 100$	-4,04	-3,45	-3,15
$T = 250$	-3,99	-3,43	-3,13
$T = \infty$	-3,96	-3,41	-3,12

ensuite utiliser la régression

$$\Delta X_t = a + \varepsilon_t$$

pour tester la valeur de a (dont l'estimateur a cette fois les propriétés habituelles). Il faut toutefois se méfier des stratégies de test séquentielles : comme les tests effectués successivement n'ont aucune chance *a priori* d'être indépendants, on perd tout contrôle sur le niveau des tests. La stratégie de test suggérée par Jobert (1992) en est un exemple caricatural, qui conduit à enchaîner des tests sans savoir si la procédure complexe ainsi mise en œuvre a un niveau de 5%, de 10% ou de n'importe quoi d'autre.

L'hypothèse nulle d'une marche aléatoire pure (avec ou sans dérive) considérée jusqu'ici est peu réaliste ; il convient de prendre en compte la possibilité que ε_t soit autocorrélé. Pour ce faire, Dickey-Fuller ont proposé le test Augmented Dickey-Fuller (ADF), qui consiste à effectuer la régression

$$\Delta X_t = d_t + \sum_{i=1}^p \gamma_i \Delta X_{t-i} + \rho X_{t-1} + u_t \quad (ADF)$$

et à retenir la statistique D_τ et la statistique

$$AD_\rho = \frac{D_\rho}{1 - \sum_{i=1}^p \hat{\gamma}_i}$$

Bien entendu, le nombre de retards p doit être choisi pour éliminer l'autocorrélation des u_t . La méthode la plus simple consiste à choisir un p suffisamment grand et à tester la significativité des derniers retards en utilisant les tests habituels (les valeurs critiques usuelles s'appliquent dans ce cadre).

Les tests ADF sont des tests asymptotiques dont les valeurs critiques sont celles données dans les Tables 1 et 2 pour $T = \infty$. Plutôt que de recourir à la correction paramétrique implicite dans la régression augmentée de Dickey-Fuller (ADF), on peut utiliser une correction non-paramétrique suggérée par Phillips et Perron (1988). Il faut pour cela distinguer deux mesures de la variance des résidus estimés \hat{u}_t de (DF) :

- la variance de court terme $\sigma^2 = V u_t$, qu'on peut estimer par

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2$$

- la variance de long terme $s^2 = \lim_{T \rightarrow \infty} \frac{E(\sum_{t=1}^T u_t)^2}{T}$, qui prend en compte toutes les autocorrélations de u_t . L'annexe 1 explique comment estimer s^2 , par exemple en choisissant bien un nombre de retards l et en calculant

$$\hat{s}^2 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 + 2 \sum_{i=1}^l \left(1 - \frac{i}{l+1}\right) \frac{1}{T} \sum_{t=i+1}^T \hat{u}_t \hat{u}_{t-i}$$

Soient $k = \hat{\sigma}^2 / \hat{s}^2$ (qui vaut 1 asymptotiquement si u_t est un bruit blanc) et $\hat{\sigma}_\rho$ l'estimateur habituel de l'écart-type de $\hat{\rho}$; les statistiques proposées par Phillips et Perron sont

$$Z_\rho = D_\rho + \frac{T^2(k-1)\hat{\sigma}_\rho^2}{2k}$$

et

$$Z_\tau = \sqrt{k} D_\tau + \frac{T(k-1)\hat{\sigma}_\rho}{2\sqrt{k}}$$

Comme pour les tests ADF, leurs valeurs critiques asymptotiques peuvent être lues dans les Tables 1 et 2 à la ligne $T = \infty$. Notons que le test de Phillips-Perron a les avantages et les inconvénients habituels des procédures non-paramétriques : il est robuste aux erreurs de spécification de la dynamique des résidus, mais il est moins précis qu'une procédure paramétrique pour laquelle le modèle a été bien spécifié.

1.1.2 Les tests de Schmidt-Phillips

Les tests de Dickey-Fuller, on l'a vu, dépendent dans leur formulation de l'existence ou non d'une dérive sous l'hypothèse nulle. En fait, si on était sûr que X_t contient une dérive, il ne serait pas nécessaire de s'encombrer avec des tests non-standard : sous cette hypothèse, la tendance déterministe implicite dans

$$X_t = a + bt + \sum_{\tau=1}^t \varepsilon_\tau$$

Table 3

Tests de Schmidt-Phillips

Valeur critique à	1%	5%	10%
SP_ρ	-25,2	-18,1	-15,0
SP_τ	-3,56	-3,02	-2,75

domine la tendance stochastique, et le Student de $\rho = 0$ dans la régression de Dickey-Fuller (éventuellement augmentée) suit sa loi habituelle (asymptotiquement normale centrée réduite). Malheureusement, il est assez difficile de tester l'existence d'une dérive par des moyens standard. L'idéal serait de pouvoir recourir à des tests qui sont valables que la série contienne une dérive ou non. Schmidt et Phillips (1992) ont précisément proposé de tels tests. Ils commencent par détrender la série par

$$\hat{S}_t = X_t - \hat{\psi} - \hat{\xi}t$$

où $\hat{\xi} = (X_T - X_1)/(T - 1)$ et $\hat{\psi} = X_1 - \hat{\xi}$. On effectue ensuite une régression très semblable à celle de Dickey-Fuller avec $d_t = a$:

$$\Delta X_t = a + \rho \hat{S}_{t-1} + u_t$$

Les statistiques de test sont $SP_\rho = T\hat{\rho}$ et SP_τ (le Student de l'hypothèse $\rho = 0$). Il est généralement souhaitable de les corriger pour la présence d'autocorrélation. Pour ce faire, on effectue la régression auxiliaire

$$X_t = a + bt + \beta X_{t-1} + v_t$$

et on calcule le facteur $k = \hat{\sigma}^2/\hat{s}^2$, où σ^2 (resp. s^2) est la variance de court terme (resp. de long terme) des résidus estimés \hat{v}_t . Il suffit ensuite de diviser SP_ρ par k et SP_τ par \sqrt{k} . Les valeurs critiques (asymptotiques) de ces tests sont données dans la Table 3. On rejettera la non-stationnarité quand la statistique de test prend une valeur inférieure à ces valeurs critiques.

1.1.3 Le test d'Elliott-Rothenberg-Stock

Si le test de Schmidt-Phillips est aussi commode d'utilisation que les tests de Dickey-Fuller tout en offrant un paramétrage plus adapté, il partage avec eux un manque de puissance très gênant. Rappelons que dans ce contexte, la puissance du test est une fonction de la racine autorégressive R qui mesure la probabilité que le test rejette la non-stationnarité quand la vraie racine est R . Il est évidemment souhaitable que quand R est plus petite que un, la puissance soit aussi proche de

1 que possible. Tous les tests présentés ci-dessus ont bien sûr une puissance unité asymptotiquement ; en revanche, ils ont une puissance très faible à distance finie.

Fixons-nous le niveau du test, à 0,05 par exemple. Il est clair qu'à distance finie, tout test de non-stationnarité sera très peu puissant quand la véritable racine R est très proche de 1 : il sera toujours très difficile de distinguer un processus pour lequel $R = 1$ d'un autre pour lequel $R = 0,99$. Malheureusement, c'est également le cas pour des racines moins proches de l'unité, comme on le verra plus loin.

Afin de remédier autant que faire ce peut à ce manque de puissance, Elliott, Rothenberg, et Stock (1996) (ci-après ERS) ont présenté un test de non-stationnarité plus satisfaisant. Leur approche repose sur l'élimination des composantes déterministes par une régression sur les séries quasi-différenciées. Soit X_t la série concernée, qui comprend T observations. ERS proposent la démarche suivante :

1. Après examen de la série et/ou application d'*a priori* économiques, on décide de modéliser la série avec ou sans dérive bt
2. on définit la constante c par $c = 7$ dans le cas sans dérive et $c = 13.5$ dans le cas avec dérive et on calcule

$$\alpha = 1 - \frac{c}{T}$$

3. soit $z_t = 1$ dans le cas sans dérive et $z_t = (1, t)$ dans le cas avec dérive. On quasi-différencie les séries X_t et z_t par $\tilde{X}_1 = X_1$, $\tilde{z}_1 = z_1$ et pour $t > 1$,

$$\tilde{X}_t = X_t - \alpha X_{t-1}$$

et

$$\tilde{z}_t = z_t - \alpha z_{t-1}$$

4. soient β les coefficients estimés dans la régression de \tilde{X}_t sur \tilde{z}_t . On crée une nouvelle série détrendée

$$X_t^d = X_t - z_t \beta$$

5. on effectue ensuite la régression (éventuellement augmentée) de Dickey-Fuller avec $d_t = 0$

$$\Delta X_t^d = \rho X_{t-1}^d + \sum_{i=1}^p \gamma_i \Delta X_{t-i}^d + u_t$$

Le test proposé *ERS* est le Student de l'hypothèse $\rho = 0$ dans cette régression. Ses valeurs critiques asymptotiques sont données dans la Table 4.

Table 4

Test d'Elliott-Rothenberg-Stock

Valeur critique asymptotique à	1%	5%	10%
sans dérive	-2,58	-1,95	-1,62
avec dérive	-3,48	-2,89	-2,57

On remarquera que le calcul du test ERS est presque aussi simple que celui du test de Dickey-Fuller. Par ailleurs, sa puissance est nettement supérieure. Afin de l'illustrer, j'ai effectué une simulation de Monte-Carlo. Supposons qu'on dispose de trente années de données cohérentes (ce qui est à peu près le maximum que l'on puisse espérer en France) sur un processus engendré par

$$X_t = RX_{t-1} + \varepsilon_t$$

où ε_t est distribué suivant la loi normale centrée réduite et l'indice t repère des observations *trimestrielles*. X_t suit donc un $AR(1)$ de racine autorégressive R . On voit facilement que dans ces conditions, le processus annuel correspondant, défini par

$$Y_a = X_{4a-3} + X_{4a-2} + X_{4a-1} + X_{4a}$$

suit un $ARMA(1,1)$ dont la racine autorégressive est R^4 . On va s'intéresser à la puissance des tests de $R = 1$ quand le véritable R est inférieur à 1, c'est-à-dire que le processus est stationnaire. Pour ce faire, on simule 10000 échantillons de X_t et Y_a et on effectue le test ADF_τ et le test ERS en choisissant l'augmentation par un retard pour les données annuelles et par quatre retards pour les données trimestrielles (ces choix sont souvent retenus dans les études appliquées, et les conclusions qualitatives ne dépendent guère de l'ordre d'augmentation). La puissance d'un test est alors estimée par la proportion d'échantillons pour lesquels le test est inférieur à sa valeur critique à 5% et conduirait donc à rejeter la non-stationnarité.

Les résultats de cette expérience, recueillis dans la Table 5, suggèrent deux remarques importantes. La première est que l'utilisation de données trimestrielles plutôt qu'annuelles, si elle accroît par définition le nombre d'observations, n'améliore pas la puissance des tests de racine unité ; dans bien des cas, elle la réduit en fait. En effet, il faut mettre en regard de l'augmentation du nombre d'observations le fait que la racine autorégressive se rapproche de un quand la fréquence des observations croît : si elle est de R en données trimestrielles, elle est de R^4 en données annuelles. On peut démontrer que ces deux facteurs se compensent exactement dans de grands échantillons. A distance finie, l'utilisation de données annuelles peut même être préférable : il n'y a pas de règle générale.

La seconde remarque est que le test ERS est nettement plus puissant que le test de Dickey-Fuller. Considérons par exemple l'application de ces deux tests à

Table 5

Puissance des tests de racine unité

	Trimestriel		Annuel	
	ADF	ERS	ADF	ERS
$R = 0,99$	0,08	0,12	0,13	0,28
$R = 0,95$	0,14	0,41	0,22	0,60
$R = 0,9$	0,34	0,73	0,41	0,82
$R = 0,8$	0,73	0,92	0,68	0,94

des données trimestrielles quand $R = 0,9$; alors le test de Dickey-Fuller conduit à conclure en faveur de la non-stationnarité (et donc à tort) deux fois sur trois, alors que le risque n'est plus que d'un sur quatre pour le test ERS. On peut montrer en fait que le test ERS atteint presque l'enveloppe des tests localement optimaux (de Neyman-Pearson), si bien qu'on ne peut espérer découvrir un nouveau test qui soit plus puissant. Ce résultat doit conduire à une certaine prudence dans l'interprétation des tests supérieurs à la valeur critique, dans la mesure où les puissances du test ERS recueillies dans la Table 5 restent souvent assez éloignées de un.

Pour conclure cet examen des tests de non-stationnarité, le test de Dickey-Fuller n'a pas grand-chose pour le recommander. En termes de robustesse, le test de Schmidt-Phillips lui est bien supérieur ; et le test d'Elliott-Rothenberg-Stock est nettement plus puissant (on peut dire en gros que recourir aux tests de Dickey-Fuller plutôt qu'au test ERS revient à "perdre" la moitié des observations).

1.2 Hypothèse nulle de stationnarité

Kwiatkowski, Phillips, Schmidt, et Shin (1992) ont proposé le principal test fondé sur une hypothèse nulle de stationnarité. Son calcul est très simple. On commence par régresser X_t sur une constante (cas sans dérive) ou sur une constante et un trend (cas avec dérive). Soit \hat{e}_t le résidu estimé. On construit ensuite les sommes partielles

$$\hat{S}_t = \sum_{\tau=1}^t \hat{e}_\tau$$

et on estime la variance de long terme de \hat{e}_t par \hat{s}^2 . La statistique de test est alors

$$KPSS = \frac{1}{\hat{s}^2} \frac{\sum_{t=1}^T \hat{S}_t^2}{T^2}$$

En effet, si X_t est stationnaire, \hat{S}_t sera $I(1)$ et de moyenne nulle, si bien que

Table 6

Test de KPSS

Valeur critique asymptotique à	1%	5%	10%
sans dérive	0,739	0,463	0,347
avec dérive	0,216	0,146	0,119

$\sum_{t=1}^T \hat{S}_t^2$ sera d'ordre T^2 ; dans le cas contraire, $\sum_{t=1}^T \hat{S}_t^2$ explosera plus rapidement. Cette fois, on rejette la stationnarité quand $KPSS$ est supérieur aux valeurs critiques données dans la Table 6. Ce test donne souvent des réponses contradictoires avec les tests de non-stationnarité. Il semble être relativement fragile, dans la mesure où il repose sur une estimation précise de la variance de long terme s^2 . Ceci souligne encore une fois l'importance du choix de l'estimateur de s^2 (voir l'annexe 1).

2 Extraction tendance-cycle

Comme on l'a vu, la pratique traditionnelle du détrending pour séparer une série "explosive" en une tendance et un cycle est inadaptée au cas des séries possédant une racine unité. On présentera ici trois méthodes usuelles pour réaliser une extraction tendance-cycle, c'est-à-dire la décomposition

$$X_t = T_t + C_t$$

où C_t est stationnaire et la racine unité n'apparaît que dans T_t .

2.1 La méthode de Hodrick-Prescott

Hodrick et Prescott ont proposé de définir la tendance comme la série T_t qui minimise une somme pondérée du carré de l'écart entre la tendance et la série X_t et du carré de la seconde différence de la tendance, soit

$$\sum_{t=1}^T \left((X_t - T_t)^2 + \mu(T_{t+1} - 2T_t + T_{t-1})^2 \right)$$

Le premier terme de cette somme tend à rapprocher la tendance de la série, tandis que le second "pénalise" les variations de son taux de croissance.

La minimisation de cette somme revient en fait à définir $T_t = A(L)X_t$, où

$$A(L) = \frac{1}{1 + \mu(1 - L)^2(1 - F)^2}$$

(et F est l'opérateur avance L^{-1}). On voit facilement que la tendance obtenue est d'autant plus lisse que le coefficient μ est plus élevé; Hodrick et Prescott suggèrent de prendre $\mu = 1600$ pour des données trimestrielles.

2.2 La méthode de Beveridge et Nelson

Le filtre de Hodrick et Prescott a plusieurs inconvénients :

- Il est bilatéral au milieu de la série et unilatéral au début et à la fin, ce qui conduit à des "effets de bords" gênants.
- Le choix de μ est assez arbitraire mais a une influence certaine sur la forme de la tendance et du cycle, puisqu'il détermine les fréquences qui seront préservées dans l'opération.
- Le filtre retenu ne dépend pas des caractéristiques de la série.

Beveridge et Nelson (1981) ont proposé une méthode qui répond à ces trois objections. Elle part de la représentation de Wold (moyenne mobile infinie) de ΔX_t , soit

$$\Delta X_t = a + B(L)u_t$$

On décompose $B(L) = B(1) + (1 - L)C(L)$, ce qui donne³

$$\Delta X_t = a + B(1)u_t + C(L)\Delta u_t$$

soit en intégrant :

$$X_t = a_0 + at + B(1) \sum_{\tau=1}^t u_\tau + C(L)u_t$$

ce qui fait apparaître la décomposition en la tendance et le cycle $C(L)u_t$. On remarquera que la tendance est alors une marche aléatoire (à la dérive près).

Le grand avantage de cette décomposition⁴ est qu'elle a une définition intrinsèque : on montre facilement que

$$T_t = \lim_{k \rightarrow \infty} (E_t X_{t+k} - ak)$$

où E_t est l'espérance conditionnelle à l'information connue en t , si bien que la tendance est simplement le meilleur prédictor du niveau asymptotique de la série (après correction de la dérive). Son inconvénient majeur est que tendance et cycle sont parfaitement corrélés instantanément (en fait, la corrélation est de -1 dans le cas usuel où ΔX_t est positivement autocorrélé). Toutefois, cet inconvénient tend à s'atténuer pour les séries multivariées, auxquelles la méthode s'étend trivialement.

³A strictement parler, cette écriture n'est valide que quand $C(L)$ est de carré intégrable.

⁴Outre sa facilité d'emploi, puisqu'il suffit d'estimer la représentation de Wold de ΔX_t .

2.3 Les méthodes à composantes inobservables

Une autre classe de méthodes s'obtient en spécifiant directement les processus suivis par la tendance et le cycle et leur corrélation. On pourrait par exemple imposer que ΔT_t est (à la dérive près) un bruit blanc non-corrélé avec C_t et que C_t suit un processus $AR(2)$. Ce type de modèle a été popularisé notamment par Harvey ; il n'est malheureusement pas possible dans ce guide d'en donner plus qu'une vague idée. Dans l'exemple choisi, on voit assez facilement que X_t peut être modélisé comme un processus $ARIMA(2, 1, 2)$ contraint (il ne comprend que quatre paramètres au lieu des cinq habituels). Un tel modèle peut être estimé par le filtre de Kalman ou en estimant le processus non contraint et en appliquant dans une deuxième étape un estimateur de distance minimale tel que les moindres carrés asymptotiques.

Bien sûr, il existe toute une classe de modèles à composantes inobservables qui peuvent représenter une série donnée. Ce degré d'arbitraire est en fait inhérent au problème d'extraction tendance-cycle. Ainsi, de nombreux articles se sont intéressés au pourcentage de la variance de X_t qui vient de la tendance T_t . Quah (1992) a malheureusement montré que ce problème était mal posé : suivant la technique d'extraction retenue, ce pourcentage peut varier entre zéro et un. Il ne peut être estimé plus précisément qu'au prix de contraintes identifiantes sur le modèle utilisé.

3 Tests de cointégration

Un vecteur de n séries non-stationnaires $X_t = (X_t^1, \dots, X_t^n)$ est dit cointégré quand il existe un vecteur β tel que la série univariée $\beta'X_t$ est stationnaire⁵. A titre d'exemple, soient c_t le logarithme de la consommation et y_t le logarithme du revenu. Ces deux séries peuvent être considérées comme non-stationnaires, si bien que leur variance explose au cours du temps. En revanche, le taux d'épargne a des fluctuations bien moindres, et on peut donc espérer que $(c_t - y_t)$ soit stationnaire. Si tel est le cas, le théorème de représentation de Granger montre qu'on peut modéliser la fonction de consommation sous la forme d'un modèle à correction d'erreur (MCE) :

$$A(L)\Delta c_t = \alpha + B(L)\Delta y_t + \lambda(c_{t-1} - y_{t-1}) + u_t$$

qui ne contient que des variables stationnaires et peut donc être soumis aux techniques d'inférence statistique usuelles.

Il existe plusieurs grandes catégories de tests de cointégration ; je n'en présenterai que deux. La première repose sur des tests de stationnarité des résidus d'une régression entre les composantes de X_t (dite régression cointégrante). La seconde,

⁵Perron et Campbell (1992) présentent une intéressante revue de la littérature consacrée à la cointégration.

plus efficace mais aussi plus coûteuse, utilise l'estimation par le maximum de vraisemblance d'un VAR sur X_t .

3.1 Tests fondés sur les résidus

Supposons qu'on ait acquis la conviction (ou qu'on souhaite tester) qu'il existe une combinaison cointégrante de X_t où la variable X_t^1 a un coefficient non-nul. Pour déterminer les coefficients de cette combinaison, une idée naturelle consiste à effectuer la régression cointégrante

$$X_t^1 = d_t + \sum_{i=2}^n \beta_i X_t^i + u_t$$

où d_t est une fonction déterministe du temps. Si les variables sont effectivement cointégrées (avec un coefficient non-nul pour X_t^1), les estimateurs des β_i seront convergents, mais auront une distribution limite non-standard. En fait, ces estimateurs sont superconvergents⁶ (ils convergent en $1/T$ au lieu de $1/\sqrt{T}$), et ce même si les variables du membre de droite sont endogènes⁷. Comme l'ont montré Engle et Granger, cette superconvergence implique qu'on peut écrire le modèle à correction d'erreur à partir des résidus \hat{u}_t estimés, et que les coefficients de la dynamique de court terme ont alors des distributions limites standard.

On ne peut appliquer aux estimateurs des β_i les techniques d'inférence usuelles, mais on sait que toujours sous l'hypothèse de cointégration, les résidus u_t doivent être stationnaires. On peut donc tester la cointégration en testant la stationnarité des résidus de la régression cointégrante.

A priori, chacun des tests de stationnarité présentés dans la section 1 doit engendrer un test de cointégration. En pratique, le fait que les résidus \hat{u}_t soient estimés dans la régression cointégrante modifie les valeurs critiques des tests, qui doivent donc être recalculées. Ce travail n'a été fait que pour certains tests ; nous en présenterons deux ici.

Le premier repose sur le test ADF_τ ; il prend donc pour hypothèse nulle la non-stationnarité des résidus, soit la non-cointégration de X_t . Ses valeurs critiques sont données dans les Tables 7 et 8. On notera qu'elles sont asymptotiques et qu'elles dépendent du nombre de variables au membre de droite ($m = n - 1$), tout comme de la spécification de d_t dans la régression cointégrante. La régression de Dickey-Fuller augmentée utilisée ne contient ni constante ni trend ; c'est simplement

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

⁶Sous réserve que les variables du membre de droite ne soient pas elles-mêmes cointégrées, ce qui créerait de nouvelles complications.

⁷La corrélation entre les X_t^i (intégrées) et u_t (stationnaire) est en effet d'un ordre moindre que celles existant entre les X_t^i .

Table 7

Tests de non-cointégration, $d_t = a$

Valeur critique asymptotique à	1%	5%	10%
$m = 1$	-3,96	-3,37	-3,07
$m = 2$	-4,31	-3,77	-3,45
$m = 3$	-4,73	-4,11	-3,83
$m = 4$	-5,07	-4,45	-4,16
$m = 5$	-5,28	-4,71	-4,43

Table 8

Tests de non-cointégration, $d_t = a + bt$

Valeur critique asymptotique à	1%	5%	10%
$m = 1$	-4,36	-3,80	-3,52
$m = 2$	-4,65	-4,16	-3,84
$m = 3$	-5,04	-4,49	-4,19
$m = 4$	-5,36	-4,74	-4,46
$m = 5$	-5,58	-5,03	-4,73

où les \hat{u}_t sont les résidus estimés de la régression cointégrante. On rejette la non-cointégration si la valeur du Student de $\rho = 0$ est plus petite que la valeur critique idoine. On remarquera que ce test s'appuie sur le test de Dickey-Fuller, dont on a vu qu'il était peu puissant ; on risque donc de rejeter l'hypothèse de cointégration souvent alors même que les séries sont cointégrées.

Le second test, dû à Shin (1994), applique le test de stationnarité de KPSS aux résidus estimés de la régression cointégrante. L'hypothèse nulle est donc cette fois la cointégration. La régression cointégrante doit ici être augmentée de valeurs avancées et retardées des différences premières des variables du membre de droite, si bien qu'elle s'écrit

$$X_t^1 = d_t + \sum_{i=2}^n \beta_i X_t^i + \sum_{j=-q}^q \sum_{i=2}^n \lambda_i^j \Delta X_{t-j}^i + u_t$$

Comme nous le verrons plus loin, l'inclusion des termes ΔX_{t-j}^i permet de corriger de l'endogénéité des variables (X_t^2, \dots, X_t^n) et d'obtenir ainsi un estimateur efficace de la combinaison cointégrante. Soient encore \hat{u}_t les résidus estimés ; on calcule la statistique de KPSS exactement comme en section 1 et on rejette la

Table 9

Tests de cointégration, $d_t = a$

Valeur critique asymptotique à	1%	5%	10%
$m = 1$	0,533	0,314	0,231
$m = 2$	0,380	0,221	0,163
$m = 3$	0,271	0,159	0,121
$m = 4$	0,208	0,121	0,094
$m = 5$	0,158	0,097	0,075

Table 10

Tests de cointégration, $d_t = a + bt$

Valeur critique asymptotique à	1%	5%	10%
$m = 1$	0,184	0,121	0,097
$m = 2$	0,150	0,101	0,081
$m = 3$	0,126	0,085	0,069
$m = 4$	0,109	0,073	0,056
$m = 5$	0,087	0,061	0,050

cointégration si cette statistique prend une valeur supérieure aux valeurs critiques données dans les Tables 9 et 10. La valeur (raisonnable) retenue pour q ne semble pas influencer beaucoup les performances du test ; pour choisir q , la méthode la plus robuste consiste à partir d'une valeur suffisamment grande de q et à tester la nullité des derniers retards—ces tests ont la distribution asymptotique habituelle. En revanche, comme pour le test de KPSS, le choix de l'estimateur de la variance de long terme s^2 est crucial.

3.2 La méthode de Johansen

Les tests de cointégration présentés ci-dessus sont surtout utiles quand on a une idée assez précise des relations de cointégration possibles et/ou une raison particulière de se concentrer sur certaines d'entre elles. Dans le cas contraire, la première chose à faire est de tenter de déterminer le nombre de relations de cointégration entre les variables de $X_t = (X_t^1, \dots, X_t^n)$. La méthode la plus utilisée pour ce faire est due à Johansen (voir par exemple son livre, Johansen (1995)). Elle repose sur l'estimation d'un VAR sous forme de MCE par le maximum de

vraisemblance sous l'hypothèse de normalité des processus.

On écrit le modèle sous la forme

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + d_t + \varepsilon_t$$

en ayant déterminé le nombre de retards de manière à blanchir les résidus (encore une fois, ceci peut être fait avec les tests usuels en partant d'un k assez grand et en testant la nullité des derniers retards). Les équations du modèle sont a priori corrélées, c'est-à-dire que la matrice de variance-covariance Ω de ε_t n'est normalement pas diagonale. S'il y a r relations de cointégration dans le système, il est facile de voir que la matrice Π est de rang r . Elle peut donc s'écrire $\Pi = \alpha\beta'$, où α et β sont deux matrices (p, r) qui donnent respectivement une base de l'image de Π et de l'orthogonal de son noyau. Plus concrètement, α et β représentent respectivement les facteurs des termes de correction d'erreur dans chaque équation du VAR et les combinaisons cointégrantes : les termes de correction d'erreur dans l'équation i s'écrivent

$$\sum_{j=1}^r \alpha_{ij} \sum_{k=1}^p \beta_{kj} X_{t-1}^k$$

Bien entendu, ces deux matrices ne sont pas définies de manière unique : si K est une matrice (r, r) inversible, αK et $\beta(K')^{-1}$ engendrent la même matrice Π . Plus concrètement, toute combinaison linéaire de deux relations cointégrantes est elle-même cointégrante ; on ne peut donc estimer que l'espace engendré par les colonnes de β .

Comme toujours, la procédure de test sur le nombre de relations cointégrantes r dépend crucialement de la spécification du terme déterministe $d_t = a + bt$. Il y a une petite complication ici, dans la mesure où la tendance déterministe de d_t peut s'intégrer à la fois à la dynamique propre des séries et aux relations cointégrantes. Elle peut même engendrer des tendances quadratiques, ce qui ne paraît pas raisonnable. On peut montrer qu'il faut pour l'éviter imposer $b = \alpha b_0$ pour un vecteur r -dimensionnel b_0 . Nous nous placerons toujours sous cette hypothèse. Par ailleurs, on voit facilement que la contrainte $b = 0$ élimine les tendances déterministes dans les relations de cointégration, tandis que la contrainte $a = \alpha a_0$ interdit aux variables X_t elles-mêmes d'avoir des tendances déterministes. Ceci conduit à distinguer trois cas :

- **Cas 1** (a et b sont a priori différents de zéro mais $b = \alpha b_0$) : on croit qu'au moins une des séries de X_t a une tendance linéaire et qu'au moins une relation de cointégration en possède aussi une
- **Cas 2** (b est nul) : on admet une tendance linéaire dans au moins une série de X_t , mais pas dans les relations de cointégration (si bien qu'il y a à la fois cointégration des tendances stochastiques et des tendances déterministes).
- **Cas 3** (b est nul et $a = \alpha a_0$) : aucune série de X_t n'a de tendance linéaire.

Dans le **cas 1**, on définit les variables

$$Z_{0t} = \Delta X_t, Z_{1t} = \begin{pmatrix} X_{t-1} & t \end{pmatrix}, Z_{2t} = \begin{pmatrix} \Delta X_{t-1} & \cdots & \Delta X_{t-k+1} & 1 \end{pmatrix}$$

Soient R_{0t} le résidu (p -dimensionnel) de la régression de Z_{0t} sur Z_{2t} , et R_{1t} le résidu $((p+1)$ -dimensionnel) de la régression de Z_{1t} sur Z_{2t} . On définit quatre matrices S_{ij} pour $i, j = 0, 1$ par

$$S_{ij} = \frac{1}{T} \sum_{t=2}^T R_{it} R'_{jt}$$

et on construit la matrice $(p+1)$ -dimensionnelle $M = S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}$. Soient $\lambda_1^{(1)} > \cdots > \lambda_{p+1}^{(1)}$ ses valeurs propres ordonnées. La statistique de test de l'hypothèse qu'il y a au plus r relations de cointégration est

$$LR = -T \sum_{i=r+1}^p \ln(1 - \lambda_i^{(1)})$$

Elle est souvent appelée "statistique de la trace". Ses valeurs critiques (asymptotiques) ne dépendent que de $(p - r)$; elles sont données dans la Table 11. On rejetera l'hypothèse de r relations de cointégration quand la statistique de la trace prend une valeur supérieure à sa valeur critique. Johansen a également proposé une autre statistique (dite du λ_{max}) pour tester l'hypothèse de r relations de cointégration contre $(r+1)$ relations de cointégration; on n'en traitera pas ici, l'objectif n'étant pas de multiplier des tests qui risquent de donner des résultats contradictoires.

Dans le **cas 2**, il suffit de changer la définition de la variable Z_{1t} , qu'on prendra cette fois égale à X_{t-1} . La matrice M a maintenant p valeurs propres ordonnées $\lambda_1^{(2)} > \cdots > \lambda_p^{(2)}$, et la statistique de la trace s'écrit

$$LR = -T \sum_{i=r+1}^p \ln(1 - \lambda_i^{(2)})$$

La Table 12 donne les valeurs critiques du test dans le cas 2.

Enfin, dans le **cas 3**, les définitions des variables auxiliaires Z_{1t} et Z_{2t} deviennent

$$Z_{1t} = \begin{pmatrix} X_{t-1} & 1 \end{pmatrix} \text{ et } Z_{2t} = \begin{pmatrix} \Delta X_{t-1} & \cdots & \Delta X_{t-k+1} \end{pmatrix}$$

Soient $\lambda_1^{(3)} > \cdots > \lambda_p^{(3)}$ les p valeurs propres ordonnées de la matrice M ; la statistique de la trace s'écrit

$$LR = -T \sum_{i=r+1}^p \ln(1 - \lambda_i^{(3)})$$

Table 11

Test de r relations de cointégration, cas 1

Valeur critique asymptotique à	1%	5%	10%
$p - r = 1$	16,39	12,39	10,56
$p - r = 2$	30,65	25,47	22,95
$p - r = 3$	48,59	42,20	39,08
$p - r = 4$	70,22	62,61	58,96
$p - r = 5$	95,38	86,96	82,68
$p - r = 6$	124,61	114,96	110,00
$p - r = 7$	157,53	146,75	141,31
$p - r = 8$	194,12	182,45	176,13
$p - r = 9$	234,65	221,56	214,72
$p - r = 10$	278,80	264,23	257,08
$p - r = 11$	326,73	311,13	306,47
$p - r = 12$	377,54	361,07	356,39

Table 12

Test de r relations de cointégration, cas 2

Valeur critique asymptotique à	1%	5%	10%
$p - r = 1$	6,64	3,84	2,71
$p - r = 2$	19,69	15,34	13,31
$p - r = 3$	34,87	29,38	26,70
$p - r = 4$	53,91	47,21	43,84
$p - r = 5$	76,37	68,68	64,74
$p - r = 6$	102,95	93,92	89,37
$p - r = 7$	133,04	123,04	117,73
$p - r = 8$	166,95	155,75	149,99
$p - r = 9$	204,64	192,30	185,83
$p - r = 10$	246,17	232,60	228,55
$p - r = 11$	291,58	276,37	272,03
$p - r = 12$	339,64	323,93	319,21

Table 13

Test de r relations de cointégration, cas 3

Valeur critique asymptotique à	1%	5%	10%
$p - r = 1$	12,73	9,13	7,50
$p - r = 2$	24,74	19,99	17,79
$p - r = 3$	40,84	34,80	31,88
$p - r = 4$	60,42	53,42	49,92
$p - r = 5$	83,93	75,74	71,66
$p - r = 6$	111,38	101,84	97,17
$p - r = 7$	142,34	132,00	126,71
$p - r = 8$	177,42	165,73	159,74
$p - r = 9$	216,08	203,34	196,66
$p - r = 10$	258,31	244,56	237,35
$p - r = 11$	304,89	289,71	281,63
$p - r = 12$	354,32	338,10	333,26

La Table 13 donne les valeurs critiques du test dans le cas 3. On notera qu'on peut également tester l'hypothèse que $a = \alpha a_0$ (soit l'absence de tendances déterministes dans les X_t) en calculant les valeurs propres dans les cas 2 et 3 et en utilisant la statistique

$$-T \sum_{i=r+1}^p \ln \frac{1 - \lambda_i^{(3)}}{1 - \lambda_i^{(2)}}$$

qui suit asymptotiquement un $\chi^2(p - r)$.

Si on n'a pas rejeté l'hypothèse de r relations de cointégration, la méthode de Johansen permet également d'obtenir les estimateurs du maximum de vraisemblance de tous les paramètres du système. Les estimateurs de β sont simplement les r vecteurs propres de la matrice M correspondant à ses r plus grandes valeurs propres ; et les autres paramètres peuvent être estimés par les moindres carrés généralisés en remplaçant β par son estimateur dans le VAR. On peut également tester de nombreuses hypothèses sur la matrice β . Comme les estimateurs de β ont une distribution limite non-standard, les statistiques de test correspondantes requièrent encore des recherches de valeurs propres. On n'en mentionnera qu'une ici ; comme on le verra en section 4, il existe en effet des méthodes plus simples d'application.

Supposons donc qu'on se soit placé dans le cas $j = 1, 2, 3$ et qu'on souhaite tester l'hypothèse $\beta = H\phi$, où H est une matrice (p, s) donnée et ϕ une matrice (s, r) libre, avec $r < s < p$. Une telle hypothèse permet par exemple de tester que la variable X_t^1 n'apparaît dans aucune relation de cointégration (prendre

$s = p - 1$, la première ligne de H ne contenant que des zéros et les $(p - 1)$ dernières lignes égales à la matrice identité). On procède comme ci-dessus en remplaçant la matrice M par

$$N = (H'S_{11}H)^{-1}H'S_{10}S_{00}^{-1}S_{01}H$$

qui est une matrice (s, s) . Soient $\lambda_1^H > \dots > \lambda_s^H$ ses valeurs propres ordonnées. La statistique de test de $\beta = H\phi$ est

$$T \sum_{i=1}^r \ln \frac{1 - \lambda_i^H}{1 - \lambda_i^{(j)}}$$

et suit asymptotiquement un $\chi^2(r(p - s))$.

4 Estimation de systèmes cointégrés

Jusqu'à présent, tous les estimateurs de relations cointégrantes considérés avaient des distributions asymptotiques non-standard, ce qui complique notablement l'inférence statistique. En fait, il existe plusieurs méthodes qui permettent d'appliquer des procédures de test standard à ces estimateurs sous des hypothèses assez faibles.

4.1 La méthode de Stock et Watson

Dans bien des cas, on ne s'intéresse qu'à une relation de cointégration et on souhaite tester la valeur des coefficients de cette relation. Si par exemple on considère les logarithmes du revenu y_t et de la consommation c_t , il est tentant d'examiner une relation cointégrante

$$c_t = a + by_t + u_t$$

et de tester la stationnarité du taux d'épargne, soit l'égalité de b à un.

L'estimateur de b dans cette régression est superconvergent, mais l'endogénéité de y_t (c'est-à-dire la corrélation entre y_t et u_t) le rend inefficace⁸ et empêche de lui appliquer des tests standard. Plusieurs auteurs, dont Stock et Watson (1993), ont montré que pour y remédier, il suffit d'ajouter à la régression des valeurs retardées et avancées de Δy_t , soit

$$c_t = a + by_t + \sum_{i=-q}^q \gamma_i \Delta y_{t-i} + u_t$$

⁸Cette inefficacité n'est pas qu'un point théorique : des simulations de Monte-Carlo ont montré que l'estimateur "statique" ainsi défini peut avoir des biais très importants à distance finie.

comme nous l'avons déjà fait pour obtenir un estimateur efficace pour le test de cointégration de Shin. Si les résidus u_t ne sont pas autocorrélés, cette transformation simple suffit à faire que le Student t de l'estimateur de b suive asymptotiquement une loi normale centrée réduite. Dans le cas contraire, il faut le corriger par

$$t' = t \frac{\hat{\sigma}}{\hat{s}}$$

où σ^2 (resp. s^2) est la variance de court terme (resp. de long terme) des résidus estimés u_t .

La méthode de Stock et Watson se ramène donc à une correction paramétrique très simple. Je me contenterai d'ajouter ici qu'il existe d'autres méthodes plus complexes, développées notamment par Peter Phillips et ses coauteurs, qui parviennent au même résultat par des corrections non-paramétriques.

4.2 Le théorème de Sims, Stock et Watson

Sims, Stock, et Watson (1990) ont démontré une règle générale qui permet d'isoler des combinaisons de coefficients auxquelles s'appliquent les tests standard dans des VAR écrits en niveau, soit

$$X_t = \sum_{i=1}^k A_i X_{t-i} + u_t$$

Cette règle s'énonce ainsi :

Un test a une distribution asymptotique standard si on peut réécrire la régression (par un changement de variables linéaire et bijectif) de telle sorte qu'on teste en fait la nullité d'un ensemble de coefficients de variables stationnaires de moyenne nulle.

Dans la régression

$$Y_t = a_0 + a_1 t + b X_{t-1} + b' Y_{t-1} + \sum_{i=1}^q c_i \Delta X_{t-i} + \sum_{i=1}^q d_i \Delta Y_{t-i} + u_t$$

par exemple, soient μ_X et μ_Y les espérances de ΔX_t et ΔY_t . On peut réécrire la régression comme

$$Y_t = (a_0 + \sum_{i=1}^q (c_i \mu_X + d_i \mu_Y)) + a_1 t + b X_{t-1} + b' Y_{t-1} + \sum_{i=1}^q c_i (\Delta X_{t-i} - \mu_X) + \sum_{i=1}^q d_i (\Delta Y_{t-i} - \mu_Y) + u_t$$

si bien que les tests standard s'appliquent à toutes les combinaisons des coefficients c_i et d_i , ce qui permet de tester facilement le nombre de retards à inclure, comme nous l'avons déjà dit plusieurs fois. De même, on peut tester le nombre de retards d'un VAR de manière standard.

Cette règle s'applique également aux tests de causalité au sens de Granger. Dans un VAR à trois variables (X_t, Y_t, Z_t) par exemple, on peut tester la non-causalité de X vers Y sachant Z de manière standard si X est stationnaire ou si X et Y sont cointégrées. Les choses sont malheureusement plus compliquées si X et Y ne sont pas cointégrées.

Cette règle peut guider le choix d'une spécification. Considérons par exemple le modèle du cycle de vie avec anticipations rationnelles et sans contraintes de liquidité. Comme Hall l'a montré, avec une utilité quadratique et égalité du taux d'escompte et du taux d'intérêt,

$$E_{t-1}c_t = c_{t-1}$$

Il paraît donc raisonnable de tester la théorie de Hall en régressant Δc_t sur y_{t-1} et en testant la nullité du coefficient de y_{t-1} . En fait, cette régression conduit à un test non-standard. Il est bien plus rusé de régresser c_t sur c_{t-1} et y_{t-1} . En effet, si le taux d'épargne est stationnaire (soit $c_t - y_t = k$ à long terme), la régression

$$c_t = \alpha + \beta c_{t-1} + \gamma y_{t-1} + u_t$$

se réécrit

$$c_t = (\alpha - \gamma k) + (\beta + \gamma)c_{t-1} + \gamma(y_{t-1} - c_{t-1} + k) + u_t$$

et l'application de la règle de Sims-Stock-Watson montre que le test de $\gamma = 0$ est standard.

Une autre observation qui peut parfois être utile se rapporte au test de la nullité d'une combinaison linéaire de coefficients. Dans ce cas, le test est standard si la combinaison testée comprend au moins un coefficient dont le test asymptotique est standard (par exemple parce que la régression peut être réécrite de manière à le faire apparaître comme le coefficient d'une variable stationnaire de moyenne nulle).

4.3 La méthode d'Ahn et Reinsel

Ahn et Reinsel (1990) ont remarqué que si on connaît le nombre de relations de cointégration dans un vecteur de séries $X_t = (X_t^1, \dots, X_t^n)$, et si on peut normaliser à 1 le coefficient d'une variable différente dans chaque relation cointégrante, on peut obtenir très facilement des estimateurs auxquels s'appliquent des tests standard. On part encore une fois du VAR sous forme MCE

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + d_t + \varepsilon_t$$

Comme il y a r relations de cointégration, on peut écrire $\Pi = \alpha\beta'$ où α et β sont deux matrices (p, r) . Puisqu'à chaque relation cointégrante on peut associer

une variable différente qui a un coefficient non-nul dans cette relation, on peut normaliser β en écrivant (après avoir éventuellement renuméroté les variables)

$$\beta = \begin{pmatrix} I_r \\ \tilde{\beta} \end{pmatrix}$$

où I_r est la matrice identité d'ordre r . Une fois cette contrainte introduite dans le VAR, il suffit de l'estimer de manière efficace pour obtenir des estimateurs convergents des paramètres $(\alpha, \tilde{\beta}, \Gamma_i, d_t, \Omega)$ auxquels s'appliquent les tests standard, même si certains de ces estimateurs (ceux de $\tilde{\beta}$) sont en fait superconvergents.

Pour estimer le modèle de manière efficace, il faut prendre en compte l'existence du terme nonlinéaire $\alpha\tilde{\beta}$. Pour ce faire, Ahn et Reinsel proposent une procédure à deux étapes qui n'utilise que des estimations par les moindres carrés (ordinaires puis généralisés). Cette procédure est asymptotiquement équivalente au maximum de vraisemblance. En fait, il est probablement aussi simple d'estimer le modèle directement en maximisant sa vraisemblance, sans oublier bien entendu de laisser libre la corrélation des résidus des différentes équations.

Notons qu'on pourrait remplacer la matrice I_r dans ce qui précède par n'importe quelle matrice fixée K de plein rang r . Ceci suggère une stratégie d'estimation du modèle VAR en trois étapes :

1. on utilise la méthode de Johansen pour déterminer r
2. on détermine pour chaque relation cointégrante une série qui y a un coefficient non-nul et on lui attribue le coefficient un⁹
3. après avoir éventuellement modifié l'ordre des variables, on applique la normalisation de Ahn et Reinsel et on estime le VAR par le maximum de vraisemblance. On peut ensuite appliquer tous les tests standard que l'on veut.

A titre d'exemple, soient trois séries c_t , y_t et y_t^e qui représentent (les logarithmes de) la consommation, le PIB domestique et le PIB étranger. On suppose que le test de Johansen approprié indique que $r = 2$. Deux candidats naturels pour être stationnaires sont le taux d'épargne ($c_t - y_t$) et le décalage conjoncturel ($y_t - y_t^e$). On pourrait le tester directement, mais on peut aussi écrire le VAR sous forme à correction d'erreur. Pour ce faire, on peut normaliser les termes de correction d'erreur en

$$c_t + ay_t^e$$

et

$$y_t + by_t^e$$

⁹En principe, il faudrait procéder à un test formel dans cette étape. Malheureusement, ceci conduirait à effectuer r tests dont chacun est fondé sur une procédure d'estimation itérative, ce qui est vraiment très lourd. Là comme ailleurs, l'analyste devra utiliser son jugement économique.

si bien que la matrice β prend la forme

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ a & b \end{pmatrix}$$

où a et b sont des coefficients à estimer qui auront une distribution asymptotique standard, comme les coefficients des deux termes de correction d'erreur et toute la dynamique de court terme du modèle. La stationnarité de $(c_t - y_t)$ s'écrit alors $a = b$ et celle de $(y_t - y_t^e)$ se teste comme l'hypothèse $b = -1$.

5 Conclusion

Ce bref exposé a cherché à démontrer que la littérature sur les non-stationnarités a considérablement avancé ces dernières années et que de nouvelles méthodes d'estimation et de test sont disponibles qui permettent d'analyser les données de manière plus simple et/ou plus efficace.

En ce qui concerne les tests de stationnarité, on a vu que les tests de Dickey-Fuller, qui sont de loin les plus utilisés dans les études appliquées, ont maintenant des concurrents très sérieux. Il semble clairement préférable de recourir à des tests de Schmidt-Phillips (plus commodes d'utilisation) ou d'Elliott-Rothenberg-Stock (nettement plus puissants). Par ailleurs, lorsque c'est l'hypothèse nulle de stationnarité qui paraît s'imposer, il peut être souhaitable d'utiliser le test de KPSS.

Malheureusement, les tests de cointégration ont été moins étudiés. Si l'hypothèse nulle est la non-cointégration, le test de Dickey-Fuller sur les résidus estimés reste un choix naturel. Dans le cas contraire, on pourra recourir au test de Shin.

L'intérêt principal de la méthode de Johansen est qu'elle permet de déterminer le nombre de relations cointégrantes dans un système multivarié. En revanche, les tests sur les coefficients de ces relations n'y sont pas très commodes d'utilisation. Si on a de bonnes raisons a priori de soupçonner qu'il y a r relations cointégrantes dans le système, il est sans doute préférable d'utiliser directement la méthode d'Ahn et Reinsel. Dans le cas contraire, on pourra recourir à la procédure en trois étapes présentée plus haut. En tout état de cause, il est évident que la méthode de Johansen n'a qu'un intérêt limité dans un système non-stationnaire bivarié, où il ne peut exister qu'une relation cointégrante dont l'existence peut être testée plus simplement.

Plus généralement, et compte tenu du faible nombre de points dont nous disposons en France (pas plus d'une cinquantaine d'années au mieux), il paraît illusoire de se reposer entièrement sur les procédures de test, qui seront forcément peu puissantes¹⁰. La qualité de ces tests n'est pas suffisante pour justifier

¹⁰La taille des tests dans des échantillons finis peut également s'éloigner beaucoup du niveau

une approche rigoriste qui suivrait entièrement leurs résultats—qui se révèlent d'ailleurs souvent contradictoires. La pratique courante qui consiste à appliquer test après test à des séries finalement assez courtes dans l'espoir illusoire d'obtenir des conclusions tranchées relève en fait de l'acharnement thérapeutique.

Malheureusement, il n'est pas possible pour autant de négliger les non-stationnarités : il faut bien prendre en compte le fait que la distribution asymptotique des tests subit une discontinuité quand on s'approche de la racine unité. La position adoptée ici est qu'il convient de combiner une interprétation souple et distanciée des tests et l'utilisation raisonnée d'*a priori* économiques. *In fine*, il est inévitable que la modélisation soit guidée par des intuitions issues de la théorie économique. Ceci suggère d'attaquer ces problèmes par un angle Bayésien¹¹, mais c'est une autre histoire...

usuel. En statistique, ce type de problème est habituellement traité par le bootstrap ; malheureusement, l'application du bootstrap à des séries non-stationnaires recèle des surprises. En particulier, il n'est pas démontré qu'elle permette une meilleure approximation des propriétés des tests que l'approximation asymptotique.

¹¹Ce qui permet également de lisser la discontinuité à l'approche de la racine unité.

Annexe 1 : estimation de la variance de long terme

Plusieurs des procédures exposées dans le texte requièrent l'estimation de la variance de long terme¹² d'une série stationnaire de moyenne nulle u_t . Rappelons que cette variance est donnée par

$$s^2 = \lim_{T \rightarrow \infty} \frac{E(\sum_{t=1}^T u_t)^2}{T}$$

qui prend en compte toutes les autocorrélations de u_t .

Pour estimer s^2 , on peut adopter une approche paramétrique et modéliser u_t par un processus ARMA, soit

$$\Phi(L)u_t = \Theta(L)v_t$$

On montre alors facilement que

$$s^2 = \left(\frac{\Theta(1)}{\Phi(1)} \right)^2 \sigma_v^2$$

Cette approche paramétrique a l'inconvénient d'imposer des restrictions fortes à la fonction d'autocorrélation de u_t . La plupart des auteurs ont donc préféré recourir à une approche non-paramétrique qui part de l'observation que si $\gamma_j = Eu_t u_{t-j}$ est l'autocovariance d'ordre j de u_t , alors

$$s^2 = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j$$

Cette formule suggère d'estimer les autocovariances par

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^T u_t u_{t-j}$$

et s^2 par

$$\hat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k\left(\frac{j}{S_T}\right) \hat{\gamma}_j$$

où k est un "noyau" qui sous-pondère les autocovariances d'ordre élevé—qui seront forcément mal estimées—et S_T est la "fenêtre" de ce noyau.

Le choix de noyau le plus fréquent est le noyau de Bartlett,

$$k(x) = \max(1 - x, 0)$$

¹²Ou, ce qui revient au même, de la densité spectrale en 0.

Andrews (1991) a suggéré d'utiliser le noyau QS (spectral quadratique), qui est un peu plus compliqué :

$$k(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$$

On peut montrer que ces deux noyaux donnent toujours des estimateurs positifs de s^2 (même si QS ne garde pas un signe constant). Le problème est de choisir la fenêtre S_T de manière à estimer s^2 le plus précisément possible. Il ne s'agit pas là simplement d'un raffinement technique : l'estimateur de s^2 obtenu peut beaucoup varier avec S_T .

On démontre que le choix optimal de S_T est donné par les formules suivantes :

– pour le noyau de Bartlett :

$$S_T = 1.1447(\alpha T)^{1/3}$$

– pour le noyau QS :

$$S_T = 1.3221(\beta T)^{1/5}$$

où α et β sont deux constantes qui dépendent de la forme de la densité spectrale de u_t en zéro et sont donc inconnues a priori. Une façon simple de les approximer est de calculer $\hat{\rho}_1$, l'autocorrélation du premier ordre de u_t , et d'utiliser

$$\alpha = \frac{4\hat{\rho}_1^2}{(1 - \hat{\rho}_1^2)^2} \text{ et } \beta = \frac{4\hat{\rho}_1^2}{(1 - \hat{\rho}_1)^4}$$

Andrews montre également qu'avec le choix optimal de fenêtre, le noyau QS donne un estimateur de s^2 dont l'erreur quadratique moyenne tend plus vite vers zéro qu'avec le noyau de Bartlett ; il semble donc préférable d'utiliser le noyau QS. Cette question ne semble toutefois pas tranchée en pratique. Newey et West (1994) présentent des simulations qui suggèrent que le choix de noyau n'est en fait pas très important. Ils proposent également une manière un peu différente de choisir la fenêtre pour le noyau de Bartlett. Elle consiste à quasi-différencier la série par

$$\tilde{u}_t = u_t - \hat{\rho}_1 u_{t-1}$$

à estimer les autocovariances $\tilde{\gamma}_j$ de \tilde{u}_t , à calculer les quantités

$$s_0 = \tilde{\gamma}_0 + 2 \sum_{j=1}^n \tilde{\gamma}_j \text{ et } s_1 = 2 \sum_{j=1}^n j \tilde{\gamma}_j$$

où

$$n = 4 \left(\frac{T}{100} \right)^{2/9}$$

puis à déterminer la fenêtre en utilisant $\alpha = (s_1/s_0)^2$. On obtient in fine un estimateur de la variance de long terme de \tilde{u}_t qui doit être divisé par $(1 - \hat{\rho}_1)^2$ pour donner un estimateur de s^2 . Ouf!

En tout état de cause, ces formules montrent clairement (et des simulations confirment) que quand ρ_1 est proche de un, l'estimateur de s^2 risque d'être très mauvais. C'est pourquoi les procédures de KPSS et de Shin (où u_t est stationnaire sous l'hypothèse nulle et a une racine unité sous l'alternative) peuvent se révéler fragiles. On notera également que dans le cas polaire où les autocorrélations estimées sont proches de zéro, il est sans doute préférable d'identifier variance de long terme et variance de court terme.

Annexe 2 : fonctions disponibles dans les logiciels

L'objectif de cette annexe est de donner quelques précisions sur les procédures disponibles dans les principaux logiciels économétriques. Le lecteur doit être prévenu que l'auteur de cet article préfère écrire ses propres procédures dans le langage de programmation matriciel GAUSS et n'a donc que peu d'expérience directe des logiciels cités ; ce qui va suivre s'appuie surtout sur les *software reviews* du *Journal of Applied Econometrics*.

Certains logiciels ne disposent que du minimum vital : des tests de Dickey-Fuller et les tests de cointégration d'Engle-Granger. C'est notamment le cas de TSP. EViews offre aussi une implémentation minimale de la méthode du maximum de vraisemblance de Johansen (avec détermination du rang r , mais pas de test sur les vecteurs de cointégration β). RATS comprend trois tests de racine unité (Dickey-Fuller, Phillips-Perron et un test bayésien dû à Sims) ; il possède également un module optionnel appelé CATS qui est le logiciel de référence pour les méthodes de Johansen, mais qui se limite à ces méthodes. PC-Give et son alter ego PC-Fiml proposent à la fois les tests de Dickey-Fuller et la méthode de Johansen.

Le logiciel le plus complet semble en fait être COINT, qui est un module optionnel de GAUSS. COINT offre toute une batterie de tests de racine unité et de tests de cointégration (classiques ou bayésiens), ainsi que la méthode de Johansen. Il s'interface également de manière parfaitement transparente avec GAUSS.

On notera que certaines méthodes récentes (par exemple le test de racine unité d'Elliott-Rothenberg-Stock) ne semblent pas encore avoir été intégrées à ces logiciels. Par ailleurs, il est connaissance commune que des méthodes a priori identiques donnent parfois des résultats différents selon les logiciels¹³. La méfiance s'impose, et peut justifier mon biais personnel en faveur de GAUSS, qui possède par ailleurs une flexibilité dont sont dépourvus certains des logiciels cités.

¹³Le lecteur qui en douterait peut utilement se reporter à McCullough et Vinod (1998).

Il existe encore assez peu de manuels qui exposent ces méthodes de manière assez complète. Le meilleur d'entre eux me paraît être Hamilton (1994). On peut aussi citer Banerjee, Dolado, Galbraith, et Hendry (1993), qui est toutefois moins systématique. Les articles cités sont malheureusement assez techniques en général (sauf les plus anciens : Nelson et Plosser (1982), Beveridge et Nelson (1981), Granger et Newbold (1974)), dans la mesure où ils consacrent tous une bonne partie de leur exposé à la démonstration des propriétés statistiques des procédures proposées, qui n'a que peu d'intérêt pour le praticien. Je recommande toutefois la lecture de Perron et Campbell (1992), et, pour les plus courageux, celle des deux chapitres de Stock (chapitre 46) et Watson (chapitre 47) dans le volume IV du *Handbook of Econometrics*.

Références

- AHN, S., ET G. REINSEL (1990) : "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813-823.
- ANDREWS, D. (1991) : "Heteroskedastic and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- BANERJEE, A., J. DOLADO, J. GALBRAITH, ET D. HENDRY (1993) : *Cointegration, Error Correction, and the Econometric Analysis of Nonstationary Data*. Oxford University Press.
- BEVERIDGE, S., ET C. NELSON (1981) : "A New Approach to the Decomposition of Economic Time Series into Permanent and Transient Components, with Particular Attention to Measurement of the Business Cycle," *Journal of Monetary Economics*, 7, 151-174.
- ELLIOTT, G., T. ROTHENBERG, ET J. STOCK (1996) : "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813-836.
- GRANGER, C., ET P. NEWBOLD (1974) : "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111-120.
- HAMILTON, J. (1994) : *Time Series Analysis*. Princeton University Press.
- JOBERT, T. (1992) : "Tests de racine unitaire : une stratégie et sa mise en œuvre," Discussion Paper Ecomath 9244, Paris I.
- JOHANSEN, S. (1995) : *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- KWIATKOWSKI, D., P. PHILLIPS, P. SCHMIDT, ET Y. SHIN (1992) : "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159-178.
- MAUREL, F. (1989) : "Modèles à correction d'erreur : l'apport de la théorie de la cointégration," *Economie et prévision*, 88-89, 105-125.

- MCCULLOUGH, B., ET H. VINOD (1998) : "The Numerical Reliability of Econometric Software," *Journal of Economic Literature*, à paraître.
- NELSON, C., ET C. PLOSSER (1982) : "Trends and Random Walks in Macroeconomic Time Series : Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139-162.
- NEWKEY, W., ET K. WEST (1994) : "Automatic Lag Selection and Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.
- PERRON, P. (1989) : "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.
- PERRON, P., ET J. CAMPBELL (1992) : "Racines unitaires en macroéconomie : le cas multidimensionnel," *Annales d'Economie et de Statistique*, 27, 1-50.
- PHILLIPS, P., ET P. PERRON (1988) : "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335-346.
- QUAH, D. (1992) : "The Relative Importance of Permanent and Transitory Components : Identification and Some Theoretical Bounds," *Econometrica*, 60, 107-118.
- RAPPOPORT, P., ET L. REICHLIN (1989) : "Segmented Trends and Nonstationary Time Series," *Economic Journal*, 99, 168-177.
- SCHMIDT, P., ET P. PHILLIPS (1992) : "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, 287-297.
- SHIN, Y. (1994) : "A Residual-based Test of the Null of Cointegration Against the Alternative of Non-cointegration," *Econometric Theory*, 10, 91-115.
- SIMS, C., J. STOCK, ET M. WATSON (1990) : "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113-144.
- STOCK, J., ET M. WATSON (1993) : "A Simple Estimator of Cointegrated Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.