Université de Montréal
Faculté des arts et des sciences
Département de sciences économiques

CAHIER 9534

ESTIMATION AND INFERENCE IN NEARLY UNBALANCED, NEARLY COINTEGRATED SYSTEMS

Serena NG[1] and Pierre PERRON[1]

[1] Département de sciences économiques, Université de Montréal and Centre de recherche et développement en économique (C.R.D.E.).

August 1995

C.P. 6128, succursale Centre-ville
Montréal (Québec) H3C 3J7

Télécopieur (FAX) : (514) 343-5831
Courrier électronique (E-Mail) : econo@tornade.ERE.Umontreal.CA

RÉSUMÉ

Cet article s'attarde au rôle de la normalisation dans un contexte d'estimation par moindres carrés des vecteurs de coïntégrations. Il est démontré, par le biais d'un exemple empirique et d'expériences de Monte Carlo, que lorsqu'une des variables I(1) retourne fortement vers sa moyenne, comme c'est le cas lorsqu'il y a une grande composante moyenne mobile négative ou encore lorsque le ratio signal-bruit est faible, certaines normalisations vont donner des estimés de moindres carrés qui ont de mauvaises propriétés en échantillons finis et qui sont très inconsistants dans un cadre asymptotique local bien défini, alors que des normalisations alternatives peuvent donner des estimés avec de très bonnes propriétés. Le choix de la variable dépendante a aussi des implications pour les résidus basés sur des tests de racine unitaire pour la coïntégration. On utilise deux modèles asymptotiques locaux pour fournir une explication théorique de ces résultats. On suggère de classer la densité spectrale à fréquence zéro des séries en différence première comme point de repère en pratique pour déterminer quelle variable dépendante devrait être utilisée.

Mots clés : racine unitaire, coïntégration, normalisation.

ABSTRACT

This paper considers the role of normalization in least squares estimations of cointegrating vectors. It is shown, by means of an empirical example and Monte Carlo simulations, that when one of the I(1) variables has a strong tendency for mean reversion, as would be the case when there is a large negative moving-average component or a low signal-to-noise ratio, some normalizations will yield least squares estimates that have very poor finite sample properties and are outright inconsistent in a well-defined local asymptotic framework, while alternative normalizations can yield estimates with very good properties. The choice of the regressand also has implications for residual-based unit root tests for cointegration. We use two local asymptotic models to provide theoretical explanations to these results. Ranking the spectral density at frequency zero of the first-differenced series is suggested as a practical guideline to determine which variable to use as the regressand.

Key words : unit root, cointegration, normalization.

1. Introduction

Cointegration is an important concept. It provides a tight analytical framework for analyzing the comovements of variables at low frequencies. A convenient way to obtain consistent estimates of cointegrating vectors is least squares estimation. This paper is concerned with the robustness of the static least squares estimator in single equation estimations of cointegrating vectors and of the role of normalization in estimating these cointegrating relationships.

A static cointegrating regression consists of regressing a variable belonging to the cointegrated system on the contemporaneous values of the remaining variables in the system, where all variables are known to be or can be tested as being $I(1)$. An important finding, due to Stock (1987), is that the least squares estimator for this regression is super-consistent. That is to say, it converges at a rate of T. However, it has also been shown that serial correlation in some series and/or correlation among the variables in the system will, in general, induce asymptotic bias, asymmetry, and nuisance parameters to the limiting distribution of the least squares estimator. See Phillips and Hansen (1990) among others. Accordingly, the least squares estimator is sub-optimal relative to fully efficient estimators such as the FIML estimator of Johansen (1991), and the Dynamic Ordinary Least Squares [DOLS] estimator of Saikkonen (1991) and Stock and Watson (1993).

In spite of the inefficiency of the least squares estimator, its properties are still worthy of investigation for several reasons. First, the least squares estimator provides the theoretical basis for the construction of more efficient estimators. Indeed, the fully-modified estimator [FM-OLS] of Phillips and Hansen (1990) and the Canonical Cointegrating Regression [CCR] of Park (1992) are built upon consistency of the static OLS estimator. Second, super-consistency of the static OLS estimator implies that the estimates should be reasonably precise. This provides a rationale for using the least squares residuals from the static regression as the basis of tests for cointegration. See, for example, Phillips and Ouliaris (1990). Properties of the least squares estimator have direct implications for the size and power of tests for cointegration to the extent that they affect the properties of the least squares residuals. Third, least squares estimation of the cointegrating vector provides an estimate of "equilibrium error" for use in subsequent estimations of error correction models. Least squares estimation of the long-run cointegrating relationships therefore affects the estimated dynamics of the cointegrated system.

Estimations of cointegrating vectors require the practitioner to take a stand on normalization. In most instances, this means deciding which variable to put on the left hand side as the regressand. This has not been seen as an issue of much consequence as the conventional

wisdom holds that while the normalization is known to imply different point estimates for elements of the cointegrating vector (except in the unrealistic situation when the R^2 of the regression is unity), it is not thought to have implications for the properties of the estimator in large samples. In particular, all normalizations yield super-consistent estimates. Accordingly, the regressions are usually normalized in a way to facilitate economic interpretation. However, few (if any) studies have examined if and when normalization affects the precision of the estimates. Our analysis in subsequent sections suggest that it does. We show, using a bivariate model, that the least squares estimator can yield drastically different point estimates of the cointegrating vector depending on the normalization. Section 2 suggests two data generating processes to be used for analysis and presents simulation results which highlight the parameter space where the estimation problem will arise. Section 3 provides a theoretical explanation for the simulation results by means of local asymptotic analyses. Section 4 studies the properties of residual based unit root tests in this context. Some observations on alternative estimators of cointegrating vectors are discussed in Section 5. We conclude with practical guidelines for inference and estimation.

We shall use the after-tax Fisher equation to give a synopsis of the problem. The Fisher equation is defined as

$$(1 - \tau)i = r + \pi^e,$$

where τ is the average marginal tax rate, i is the nominal interest rate, π^e is the expected rate of inflation, and r is the real interest rate which is assumed to be a constant. The expected rate of inflation is unobserved, and replacing it by the actual rate of inflation will induce an errors-in-variable problem. Least squares estimation will yield estimates that are inefficient and possibly suffer from simultaneity bias. But as long it can be shown that i and π are I(1) variables and that they are cointegrated, consistency of the estimates is implied by standard asymptotic results. To show this, we need to review the unit root and cointegration tests used in this study, as well as estimators of the spectral density at frequency zero used to construct these tests.

1.1 Preliminaries

Throughout this analysis, we use three statistics to test for the presence of a unit root, namely, the Phillips and Perron (1988) $Z_{\alpha\mu}$ test, its modified variant $MZ_{\alpha\mu}$, and the $t_{\rho\mu}$ statistic of Said and Dickey (1984). These statistics are also used for testing the null hypothesis of no cointegration. A constant term is included in the regressions where appropriate.

The Said and Dickey (1984) $t_{\rho\mu}$ statistic is constructed from an augmented autoregression

with a constant and k lagged first-differences of the data. The statistic tends to overreject the unit root hypothesis when there is a large negative MA component in Δy_t. However, as discussed in Ng and Perron (1995), the size of the test is more robust if we use a data dependent method to select the truncation lag, k. In this analysis, k is selected using a general to specific procedure to test for the significance of the last lag beginning from the largest lag, $kmax$. [See Ng and Perron (1995) for details].

The Phillips-Perron $Z_{\alpha\mu}$ test is defined as $T(\hat{\alpha} - 1) - (s^2 - s_u^2)(2T^{-2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2)^{-1}$, where $\tilde{y}_t$ are the demeaned data, $s_u^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2$, and $\hat{u}_t$ are the residuals from the first-order autoregression

$$y_t = \mu + \alpha y_{t-1} + u_t. \tag{1}$$

The estimate of the spectral density at frequency zero of u_t, denoted s^2, is constructed by applying a Quadratic kernel to the sample covariances of the regression residuals. The bandwidth is selected using Andrews' (1991) automatic procedure based on an AR(1) approximation. Note that to ensure consistency of the unit root and cointegration tests, the estimated residuals from (1) must be used instead of the (demeaned) first-differences of the data [See Phillips and Ouliaris (1990)].

The $MZ_{\alpha\mu}$ statistic is relatively new and warrants some explanation. Perron and Ng (1994) discuss a class of tests proposed by Stock (1990) which can be viewed as modifications of tests of Phillips-Perron and Sargan-Bhargarva. The $MZ_{\alpha\mu}$ statistic, a member of this class, is defined as

$$MZ_\alpha = (\tilde{y}_T^2 - T\,s^2)(2T^{-1}\sum_{t=1}^{T}\tilde{y}_t^2)^{-1}.$$

It can be viewed as a modified Phillips-Perron test because it can be written as

$$MZ_{\alpha\mu} = Z_{\alpha\mu} + T(\hat{\alpha} - 1)^2/2,$$

where $\hat{\alpha}$ is the least squares estimate of α in the autoregression (1). Even though $\hat{\alpha}$ converges to 1 at rate T under the standard asymptotic framework, Nabeya and Perron (1994) showed that $\hat{\alpha}$ is inconsistent for α when the noise function of the unit root series has an AR or MA coefficient local to -1.

The properties of the least squares estimator imply that the $MZ_{\alpha\mu}$ statistic is asymptotically equivalent to $Z_{\alpha\mu}$ since $\hat{\alpha}$ is consistent for $\alpha = 1$. However, as shown in Perron and Ng (1994), $Z_{\alpha\mu}$ will diverge to $-\infty$ regardless of the choice of the spectral density estimator when the root of Δy_t is local to unity, and a kernel estimator based on the estimated residuals from (1) will speed up the rate of divergence. The problem originates from the fact that $\hat{\alpha}$ is not consistent for α in these local cases, and the estimated residuals give inconsistent

3

estimates of the autocovariances. For this reason, kernel based $Z_{\alpha\mu}$ tests are severely over-sized as Schwert (1989) and many others have shown using simulations. The $MZ_{\alpha\mu}$ statistic achieves an exact size that is quite close to the nominal size because the modification factor, $T(\hat{\alpha}-1)^2/2$, offsets the explosive terms in the unmodified statistic, i.e. $Z_{\alpha\mu}$.

It is important to clarify that $MZ_{\alpha\mu}$ achieves small size distortions in the boundary cases only if a consistent estimator of the non-normalized spectral density at frequency zero is used. An estimator that we have investigated in some detail is the autoregressive spectral density estimator, first used by Stock (1990). It is defined as

$$s_{AR}^2 = \sigma_{ek}^2 / (1 - \sum_{i=1}^{k} \hat{b}_i)^2, \qquad (2)$$

where values for $\hat{b}_i$ and σ_{ek}^2 are obtained from the regression:

$$\Delta y_t = c_0 + b_0 y_{t-1} + \sum_{j=1}^{k} b_j \Delta y_{t-j} + e_{tk}. \qquad (3)$$

Note that the lagged level y_{t-1} is needed to ensure that the statistic is bounded above zero under the alternative hypothesis of stationarity. Otherwise, the estimate converges to zero too fast and the test $MZ_{\alpha\mu}$ is inconsistent.

The kernel and s_{AR}^2 estimators are both consistent when the roots of Δy_t are bounded away from the unit circle, and the two estimators should give very similar estimates under standard conditions. Note that the estimator formulated according to (2) and (3) is consistent for the true spectral density function of the *first-differences* of the data only under the null hypothesis of a unit root because the estimates are bounded above zero even under the alternative of stationarity. In spite of this, it is still superior to kernel based spectral density estimators that use the residuals from (1) from the point of view of unit root tests. To see why this is the case, note that for an ARMA process Δy_t satisfying $A(L)\Delta y_t = c + B(L)e_t$, the spectral density at frequency zero is defined as $f_{\Delta y}(0) = \sigma_e^2 B(1)^2/A(1)^2$. Thus, $f_{\Delta y}(0) = 0$ if y_t is stationary. However, if y_t is I(1) but the moving average component is close to -1, then $B(1)$ and hence $f_{\Delta y}(0)$ will be close to zero. When the error process is highly negatively autocorrelated, $f_{\Delta y}(0)$ will also be small because $A(1)$ is large. As discussed earlier, $\hat{\alpha}$ is inconsistent in both cases in the limit, and it is in precisely these two cases that kernel estimators based on estimated residuals from (1) will severely over estimate $f_{\Delta y}(0)$. The s_{AR}^2 estimator does not depend on $\hat{\alpha}$ and has been shown by Perron and Ng (1994) to consistently estimate $f_{\Delta y}(0)$ even when Δy_t has an MA or an AR coefficient local to -1. This has desirable implications for unit root tests.

4

From the above discussion, it becomes clear that kernel estimators based on the estimated residuals and s_{AR}^2 based upon (3) are constructed with consistency of the unit root tests in mind. However, better estimates are possible if one's interest is in $f_{\Delta y}(0)$ itself. For kernel based estimators, the demeaned first-differences of the data can be used in place of the estimated residuals from (1). For the autoregressive spectral density estimator, one can obtain more efficient estimates using an autoregression of the form (3) without the inclusion of the lagged level. As we will see later, this statistic plays an important role in our analysis.

We now summarize the salient implications of this subsection:

- The $Z_{\alpha\mu}$ and $MZ_{\alpha\mu}$ tests are asymptotically equivalent except when the noise function of the series exhibits large negative serial correlation. In that case, $Z_{\alpha\mu}$ diverges to $-\infty$ but $MZ_{\alpha\mu}$ is $O_p(1)$. Of the three tests considered, the size of $Z_{\alpha\mu}$ is most distorted, and the size of $MZ_{\alpha\mu}$ is more robust to a large negative MA component than $t_{\rho\mu}$.

- The spectral density at frequency zero of Δy_t is small when there is negative serial correlation in the residuals.

With this backdrop, we now turn to our empirical example.

1.2 Testing the Fisher Equation

To test the Fisher equation, we use monthly observations from 1954:1 to 1993:11 for the U.S. taken from CITIBASE (479 observations). The interest rate is the three month yield on treasury bills (FYGM3) at annual rates, inflation is the annualized quarterly change in the consumer price index ($\pi_t = 4\log(punew_t/punew_{t-3})$).

We first examine the properties of the data and the results are reported in Table 1. All three unit roots tests conclude that there is a unit root in the nominal interest rate (i_t). However, while $Z_{\alpha\mu}$ strongly rejects the presence of a unit root in the inflation rate (π_t), $MZ_{\alpha\mu}$ and $t_{\rho\mu}$ cannot reject the unit root hypothesis. Such inconsistencies across unit root tests are symptomatic of negative residual serial correlation in view of the discussion of the previous subsection.

The spectral density at frequency zero used in the unit root tests, when normalized, are in the range of unity for i_t according to both the kernel and the s_{AR}^2 estimates [see columns (1) and (3) in Table 1]. But, whereas the normalized Quadratic kernel spectral density for π_t is around one, it is much smaller and is closer to zero according to s_{AR}^2. These discrepancies in the spectral density estimates reinforce our conjecture (based on the discrepancy between $Z_{\alpha\mu}$ and $MZ_{\alpha\mu}$) that there is negative serial correlation in the inflation series.

5

The more efficient estimates of the spectral density function based on the first-differences of the data are shown in Columns 2 and 4. The Quadratic and autoregressive spectral estimates are 1.243 and .906 respectively for interest rate. For the inflation series they are .312 and .253 respectively. The small values of the latter estimates again suggest negative residual serial correlation. It is of interest to note the large discrepancy between the residuals based kernel estimate (i.e. those used in the unit root tests) and the kernel estimate based on the first-differenced data for the inflation series. This discrepancy reflects the poor properties of $\hat{\alpha}$, of which negative residual serial correlation is a possible cause.

The conclusion that one would draw from these results is that the interest rate series is unambiguously $I(1)$. The more robust $MZ_{\alpha\mu}$ and $t_{\rho\mu}$ tests suggest that π is also $I(1)$ but there is strong negative serial correlation which induces a strong force for the series to revert to its mean. Indeed, estimations of ARIMA models for π_t reveal the significance of negative moving average lags. The best model selected by the AIC criteria suggests π_t is an $ARMA(1,3)$ with an autoregressive coefficient that is almost one and with the moving-average coefficient at the third lag in the neighborhood of -.8, far outweighing the sum of the positive moving average coefficients at lags one and two.

'Granted the result that both i and π are $I(1)$, we then proceed to estimate the Fisher equation and test if the two series are cointegrated. The Fisher equation makes no suggestion as to whether empirical tests of the relationship should use i or π as the regressand, so that without a strong a priori reasoning, it is equally legitimate to use π as the regressand as it is to use i. Assuming an average marginal tax rate in the U.S. of around .3 over the sample, one would expect a regression of i on π to yield an estimate of $1/(1 - \tau) > 1$. If we regress π on i, we would expect a regression coefficient of 0.7. Furthermore, the estimates from the two equations should be (approximately) the reciprocal of each other. Table 2 reports the estimation results.

The estimated coefficient from the equation with i as the regressand is quite different from our prior as it falls short of unity. The coefficient from the equation with π as the regressand suggests a marginal tax rate of .22, which seems plausible. But, the two estimated coefficients are evidently not the reciprocal of one another. One might argue that inefficiency of the static least squares estimator might be the source of the problem. However, discrepancies remain even when the equations are estimated with an efficient estimator such as the DOLS. The second through fourth columns of Table 2 augment the static least squares regression with four, eight, and twelve leads and lags of the regressor. For the equation with i on the left hand side, estimation by DOLS raises the point estimate of the coefficient, but it continues to fall short of one. For the equation with π on the left hand side, the additional regressors made

6

practically no change to the static least squares estimates. Furthermore. the coefficients from the two regressions look like they are identical rather than being the reciprocal of one another. Not only do the regressions give puzzling evidence about the empirical Fisher relationship, the results from the various residual based tests for cointegration are just as confusing. The Phillips-Perron statistic always reject the null hypothesis of no-cointegration, the Said-Dickey statistic suggests no-cointegration, while the MZ_α statistic gives mixed results depending on which variable is used as the regressand.

The above example suggests that the choice of normalization can potentially yield dramatically different point estimates on coefficients of economic interest. As will become clear, one of the two estimates has very poor properties. The key to finding the appropriate normalization lies in the spectral density at frequency zero of the first-differences of the regressand relative to those of the first-differences of the regressors. The rest of this analysis provides a formal framework for analyzing the issues raised, with the aim of providing practical recommendations for which variable to use as the regressand.

2. When Might Normalization Matter?

In this section, we first present the two data generating processes used, and then report simulations to illustrate the nature of the problem. Unless noted otherwise, all simulations are performed using 1,000 replications. The programs are written in C with routines from Press, Teukolsky, Vetterling and Flannery (1992) running under IRIX 5.2 on an SGI system.

2.1 The Data Generating Processes

There are many ways to model a cointegrated system. One possibility is the common trend framework of Stock and Watson (1988). Consider a bivariate model

$$
\begin{aligned}
x_t^{\cdot} &= \gamma_1^{\cdot}\mu_t + e_{1t}^{\cdot}, & e_{1t}^{\cdot} &\sim (0, \sigma_1^2), \\
y_t^{\cdot} &= \gamma_2^{\cdot}\mu_t + e_{2t}^{\cdot}, & e_{2t}^{\cdot} &\sim (0, \sigma_2^2), \\
\mu_t &= \mu_{t-1} + v_t, & v_t &\sim (0, \sigma_v^2).
\end{aligned}
$$

The variables x_t and y_t are driven by a common stochastic trend μ_t, as well as stationary innovations $e_{1t}^{\cdot}$ and $e_{2t}^{\cdot}$ that are serially and mutually uncorrelated. Furthermore, $e_{1t}^{\cdot}$ and $e_{2t}^{\cdot}$ are uncorrelated with v_t at all leads and lags by assumption. A non-zero drift can be added to μ_t, but is omitted without loss of generality. Letting $x_t = x_t^{\cdot}/\sigma_1$, $y_t = y_t^{\cdot}/\sigma_2$, $e_{1t} = e_{1t}^{\cdot}/\sigma_1$, $e_{2t} = e_{2t}^{\cdot}/\sigma_2$, $\gamma_1 = \gamma_1^{\cdot}/\sigma_1$, and $\gamma_2 = \gamma_2^{\cdot}/\sigma_2$, we have:

$$
\text{DGP1:} \quad x_t = \gamma_1 \mu_t + e_{1t}, \qquad e_{1t} \sim (0, 1),
$$

7

$$y_t = \gamma_2 \mu_t + e_{2t}, \qquad e_{2t} \sim (0,1),$$
$$\mu_t = \mu_{t-1} + v_t, \qquad v_t \sim (0, \sigma_v^2). \tag{4}$$

The variables x_t and y_t are now of unobserved components form as in Clark (1987). It is observationally equivalent to an ARIMA model for a large range of correlations. It can be written as

$$\Delta x_t = \gamma_1 v_t + e_{1t} - e_{1t-1} = u_t^x + \theta_x u_{t-1}^x,$$
$$\Delta y_t = \gamma_2 v_t + e_{2t} - e_{2t-1} = u_t^y + \theta_y u_{t-1}^y, \tag{5}$$

where θ_x is such that $\theta_x(1 + \theta_x^2)^{-1} = \sigma_1^2[\gamma_1^2 \sigma_v^2 + 2\sigma_1^2]^{-1}$ and θ_y is such that $\theta_y(1 + \theta_y^2)^{-1} = \sigma_2^2[\gamma_2^2 \sigma_v^2 + 2\sigma_2^2]^{-1}$. The model can also be parameterized as

$$\Delta x_t = \gamma_1 v_t + e_{1t} - e_{1t-1} = \tilde{u}_t^x,$$
$$y_t = \frac{\gamma_2}{\gamma_1} x_t - \frac{\gamma_2}{\gamma_1} e_{1t} + e_{2t} = \frac{\gamma_2}{\gamma_1} x_t + \tilde{u}_t^y, \tag{6}$$

where $E(\tilde{u}_t^x \tilde{u}_t^y) \neq 0$. Because of the absence of exogeneity between x_t and y_t, OLS estimation is, in principle, sub-optimal asymptotically regardless of the choice of the normalization.

The above parameterization of DGP1 shows that it belongs to the class of triangular models analyzed in Phillips (1991):

$$\text{DGP2:} \quad \Delta x_t = u_{1t}, \qquad u_{1t} = e_{1t} + \theta_x e_{1t-1} + \sigma_{12} e_{2t},$$
$$y_t = \beta x_t + u_{2t}, \qquad u_{2t} = e_{2t} + \theta_y e_{2t-1} + \sigma_{21} e_{1t}, \tag{7}$$

where $E(u_{1t} u_{2t})$ can be zero.

We begin our analysis with DGP1. We will then show that the results are not unique to the particular parameterization of the data. Suppose it is known that there is a cointegrating vector in the bivariate system. Substituting out the common trend in x_t and y_t of DGP1, we can either write

$$y_t = \beta_x x_t + e_{2t} - \beta_x e_{1t}, \qquad \beta_x = \gamma_2/\gamma_1, \tag{8}$$
$$\text{or} \quad x_t = \beta_y y_t + e_{1t} - \beta_y e_{2t}, \qquad \beta_y = \gamma_1/\gamma_2. \tag{9}$$

Since standard asymptotic results show T consistency for the coefficients of both regressions, (8) and (9) form equally legitimate basis for estimating the cointegrating vector.

2.2 Simulation Results

To examine the precision of the estimates from both regressions. we simulate DGP1 with $\sigma_v^2 = 1$. The coefficient γ_2 is set to one and we vary γ_1 over the range .01 and 5. Table 3 reports the simulation results for sample sizes of 50, 200 and 500. Turning first to the results from regressing x_t on y_t (Table 3.a), we see that estimates of the true regression coefficient, $\gamma_1 = \beta_y$, is precisely estimated over the range of parameter values considered and the accuracy of the estimates increases as the sample size increases. There is one feature of the results that is noteworthy. Since by construction, y_t is not weakly exogenous for x_t, the least squares estimator for β_y and all test statistics associated with it should have non-standard limiting distributions. Upon examining the upper and lower five percent critical values of the empirical distribution of the t statistic on β_y, we find that for large values of γ_1, the empirical distribution of the t statistic is indeed being shifted to the left of the normal distribution. However, as the value of γ_1 falls, the distribution approaches normality. In the extreme case when $\gamma_1 = .01$, the upper and lower five percent critical values are practically the same as those from the normal distribution, even with a moderate sample size of 200.

The picture is very different when we regress y_t on x_t (see Table 3.b). The true value of β_x is $1/\gamma_1$, which we also report for convenience. For large values of γ_1, β_x tends to be downward biased when T=50, but the estimates are reasonably accurate. The precision of the estimates starts to deteriorate when γ_1 falls below unity. When $\gamma_1 = .2$, the least squares estimator is severely biased downwards. For example, the mean of $\hat{\beta}_x$ is only 1.101 when $\beta_x = 5$ at T=50. Although the bias is reduced as T increases, there is still a substantial discrepancy between the true value of 5 and the average estimated value of 3.483 at T=500. As γ_1 diminishes and hence β_x increases, one might expect $\hat{\beta}_x$, although biased, to at least increase with the true coefficient. However, the simulations reveal that as $\gamma_1 \to 0$ and hence $\beta_x \to \infty$, β_x tends towards 0 rather than increases with β_x. Curiously, the t statistic associated with $\hat{\beta}_x$ appears to diverge to $-\infty$ as $\gamma_1 \to 0$ but approaches normality as γ_1 increases.

The above simulation results clearly illustrate the fact that the choice of the regressand can severely affect the precision of the estimates. How do these results relate to our empirical example of the Fisher equation? We want to suggest that x_t should be treated as inflation and y_t as the interest rate. We also want to suggest that the estimate from a regression of π_t on i_t is to be trusted. To justify these interpretations, we first note that regressions of y_t on x_t become problematical when γ_1 is small. Also recall that DGP1 implies Δx_t is an MA(1)

process with parameter θ_x, where

$$\frac{\theta_x}{(1 + \theta_x^2)} = \frac{\sigma_1^2}{\gamma_1^2 \sigma_v^2 + 2\sigma_1^2}. \tag{10}$$

Since $\gamma_1 = \gamma_1^*/\sigma_1$, $\gamma_1 \to 0$ if either $\gamma_1^* \to 0$, or if $\sigma_1^2 \to \infty$. The former corresponds to the case when the common trend, μ_t, is a weak driving force of x_t. The latter corresponds to the case of large variability in the idiosyncratic noise of x_t, so large that even though these innovations are stationary, their impact dominates that of the stochastic trend in x_t.

The value of γ_1 affects the econometric analysis involving x_t because it follows from (10) that as $\gamma_1 \to 0$, $\theta_x \to -1$. Such a process, referred to as *nearly integrated nearly white noise* by Nabeya and Perron (1994), has a strong tendency to be mean reverting. This feature is inherent in our inflation series. As mentioned earlier, a moving average coefficient at lag 3 of -0.8 is found in the inflation series. The issue, of course, is not so much the size of the MA coefficient on a particular lag, but the sum of the coefficients at all lags. Although DGP1 allows only an MA(1) in the noise function of Δx_t, it nevertheless encompasses the feature that is of interest.[1]

One way to judge if the simulated values of x_t indeed behave like a nearly integrated nearly white noise process is to examine the size of the unit root tests. As mentioned earlier, $Z_{\alpha\mu}$ is oversized relative to $t_{\rho\mu}$, and more so relative to $MZ_{\alpha\mu}$. In Table 4, we report the exact size of the three statistics for the nominal size of five percent.[2] At very small values of γ_1, all tests reject the unit root hypothesis, as expected. The parameter space of interest is when γ_1 falls short of unity. A γ_1 of 0.5 translates into a moving average coefficient in Δx_t of -0.6. Previous work by Schwert (1989) and Perron and Ng (1994) have documented that unit root tests suffer from size distortions even when θ is -.5, and it is not surprising that we also detect some size distortions. The size problem becomes severe when $\gamma_1 = .1$. $Z_{\alpha\mu}$ almost always rejects the unit root hypothesis, while $t_{\rho\mu}$ rejects it almost half the time. These, of course, are cases when $f_{\Delta y}(0)$ is small. By contrast, $MZ_{\alpha\mu}$, while still oversized, is not as distorted even under the conditions considered.

If large negative residual serial correlation is indeed the reason why the estimates have properties that depend on the normalization, then we should observe similar results from DGP2 given by (7). We simulate DGP2 letting the noise function in Δx_t be a moving average process. Given that $y_t = \beta x_t + u_{2t}$, it follows that $\Delta y_t = \beta \Delta x_t + \Delta u_{2t}$. Since u_{2t} is itself

---

[1]Perron (1994) suggests that a negative moving-average component should be present in the inflation series if the monetary authorities react to offset inflationary/disinflationary pressures that are inconsistent with an inflation target for the path of the price level. This makes inflation strongly mean reverting.

[2]In the simulations, the bandwidth is selected using Andrews' (1991) automatic procedure using an AR(1) approximation. The value of $kmax$ in $t_{\alpha\mu}$ is 5, and for s_{AR}^2 in $MZ_{\alpha\mu}$ is 10.

10

a stationary moving-average process, Δu_{2t} is over differenced. Thus, if Δx_t has a negative moving-average component, Δy_t will inherit a moving average component that has an even stronger tendency for mean reversion. In other words, the moving-average component in y_t is more negative than that in x_t. In light of the results from DGP1, one might then expect that if the data were generated by DGP2, using y_t as the regressand will give more precise estimates because it has a smaller spectral density at frequency zero than x_t.

Figure 1 provides a summary of the results for DGP2 assuming $\sigma_{12} = \sigma_{21} = 0$ for the case $\beta = 1$. Least squares regressions of y_t on x_t give very accurate estimates regardless of the values of θ_x and θ_y. However, when we use x_t as the regressand, the estimates, while invariant to the values of θ_y, are severely downward biased when θ_x is negative and the biases are larger the closer θ_x is to -1. For example, when $\theta_x = -.8$, β is estimated to be below .5 when we use x_t as the regressand, half the true value.

The picture that emerges from the simulations of the two DGP's points to the following general observation. If the first difference of one of the series has negative serial correlation (or a small spectral density at frequency zero), estimates based upon an equation with that series as the regressand always have better properties than those normalized on other variables. The next section provides a theoretical rationale for this result.

3. Local Asymptotic Analysis

This section uses local asymptotic analyses to explain why the static least squares estimator experiences substantial downward bias for certain normalizations only. The first subsection analyzes DGP1, and the second subsection focuses on DGP2.

3.1 DGP1: γ_1 local to 0

A notable feature of the simulation results reported for DGP1 is that the least squares estimator is severely downward biased when γ_1 is small and when we use y_t as the regressand. We therefore parameterize

$$\gamma_1 = c/\sqrt{T}, \qquad c \neq 0,$$

where we recall that DGP1 is the unobserved components model given by (4). The parameter γ_1 tends to zero as $T \to \infty$ at rate $\sqrt{T}$. We also recall that $\gamma_1 \to 0$ if the noise to signal ratio is large ($\sigma_1 \to \infty$) or if the common trend component is small ($\gamma_1^* \to 0$). This in turn implies that the moving average component (θ_x) in (5) is

$$\theta_x = -1 + \delta/\sqrt{T} \to -1 \quad \text{as} \quad T \to \infty,$$

11

for some non-centrality parameter $\delta > 0$. This local parameterization of θ_x has been used in Nabeya and Perron (1994) and Perron and Ng (1994) to analyze the local asymptotic properties of I(1) processes with MA(1) noise functions. Extending the "nearly integrated nearly white noise" terminology in the univariate case to the present multivariate model, y_t and x_t can be said to have a "nearly cointegrated nearly unbalanced" relationship when γ_1 is local to zero. We begin our analysis with the following lemma.

Lemma 1 : *(Sample Moments) Let x_t and y_t be generated by (4) with $\gamma_1 = c/\sqrt{T}$. Let $W_v(r)$ be a Wiener process defined on $C[0.1]$. Then as $T \to \infty$:*

   i. $T^{-1} \sum_{t=1}^{T} x_t^2 \Rightarrow c^2 \sigma_v^2 \int_0^1 W_v(r)^2 dr + 1$;

   ii. $T^{-2} \sum_{t=1}^{T} y_t^2 \Rightarrow \gamma_2^2 \sigma_v^2 \int_0^1 W_v(r)^2 dr$;

   iii. $T^{-3/2} \sum_{t=1}^{T} y_t x_t \Rightarrow c\gamma_2 \sigma_v^2 \int_0^1 W_v(r)^2 dr$.

The proof to the lemma is standard and is omitted. Part (ii) of the lemma is the usual result for an $I(1)$ process and follows from the fact that y_t is invariant to local variations in γ_1. However, this is not the case with x_t, which is a white noise process in the limit and $\sum_{t=1}^{T} x_t^2$ is $\mathcal{O}_p(T)$. A consequence of this slower rate of normalization is that the sample moment for x_t^2 is influenced by σ_1^2 (normalized to 1) in the limit. The properties of the least squares estimator from the two normalizations are then immediate from Lemma 1. These results are summarized in the following theorem.

Theorem 1 : *Let x_t and y_t be generated by (4) with $\gamma_1 = c/\sqrt{T}$. Let $\hat{\beta}_y$ be the least squares estimate from a regression of x_t on y_t, and let $\hat{\beta}_x$ be the least squares estimate from a regression of y_t on x_t. Let $W_1(r)$ be a Weiner process independent of $W_v(r)$. As $T \to \infty$:*

   From a regression of x_t on y_t with $\beta_y = \gamma_1/\gamma_2$,

   1. $T^{1/2} \hat{\beta}_y \Rightarrow c/\gamma_2 = T^{1/2} \beta_y$;

   2. $T(\hat{\beta}_y - \beta_y) \Rightarrow \dfrac{\int_0^1 W_v(r) dW_1(r)}{\gamma_2 \sigma_v \int_0^1 W_v(r)^2 dr}$;

   3. $t_{\hat{\beta}_y} \Rightarrow N(0,1)$.

   From a regression of y_t on x_t with $\beta_x = \gamma_2/\gamma_1$,

   1. $T^{-1/2} \hat{\beta}_x \Rightarrow \dfrac{c\gamma_2 \sigma_v^2 \int_0^1 W_v(r)^2 dr}{c^2 \sigma_v^2 \int_0^1 W_v(r)^2 dr + 1}$;

   2. $T^{-1/2}(\hat{\beta}_x - \beta_x) \Rightarrow \dfrac{-\gamma_2/c}{c^2 \sigma_v^2 \int_0^1 W_v(r)^2 dr + 1}$;

3. $T^{-1/2}t_{\widehat{\beta}_x} \Rightarrow \dfrac{-1}{c\sigma_v\left(\int_0^1 W_v(r)^2 dr\right)^{1/2}}.$

Remarks:

1. The main result of the theorem is that in this local framework, $\widehat{\beta}_y$ is consistent but $\widehat{\beta}_x$ is not. Absent the influence of σ_1^2 (normalized to 1) on the sample moment of x_t^2, $T^{-1/2}\widehat{\beta}_x$ would have converged to γ_2/c, the true regression coefficient. However, the presence of $\sigma_1^2 = 1$ in the denominator of the estimator induces a downward bias to the estimator. This explains the simulation results that regressions with y_t as the regressand yield estimates that can be far below the true value.

2. Note that $T^{-1/2}\widehat{\beta}_x$ approaches 0 when c is either very large or very small. Our local asymptotic analysis shows that c appears in the numerator and the denominator of the limiting distribution of the normalized least squares estimator. Hence, $\widehat{\beta}_x \to 0$ when c is very large (and hence $T^{-1/2}\beta_x$ is small) or when c is small (and hence $T^{-1/2}\beta_x$ is large). This feature can be seen from the simulations as noted earlier.

3. In the standard asymptotic framework with γ_1 fixed, the distribution of $t_{\widehat{\beta}_y}$ is nonstandard since weak exogeneity between x_t and y_t is not satisfied in the regressions. Obtaining asymptotic normality of the test statistics is the motivation for the modifications to $t_{\widehat{\beta}_y}$ introduced by Phillips and Hansen (1990). However, as stated in the theorem, $\widehat{\beta}_y$ is asymptotically mixed normal and $t_{\widehat{\beta}_y}$ is asymptotically standard normal even though the exogeneity assumption is violated in finite samples. The reason for this result is that as $\beta_y \to 0$, y_t becomes weakly exogenous for the innovations in x_t in the regression with x_t as the regressand. However, this is not the case when y_t is used as the regressand; those regression residuals have a non-trivial correlation with the innovations driving x_t. In consequence, the t statistic has non-standard properties even when suitably standardized.

4. The theorem suggests that it is desirable from the point of view of both estimation and hypothesis testing to use as regressand the variable that is "less integrated". In other words, the variable whose spectral density at frequency zero is the smallest. There are two intuitive reasons why this works well. The first is that using the "more integrated" variables as regressors amounts to putting the variables with more variability in the right hand side of the regression. As is well known, the greater are the variations in the regressors, the more precise are the parameter estimates. The second rationale can be seen with reference to the two regression specifications given in (8) and (9). Given

13

that $\gamma_1 \rightarrow 0$, $\beta_x \rightarrow \infty$ and $\beta_y \rightarrow 0$ asymptotically, the regression error in (8) with y_t as the regressand has a variance that diverges at the same rate as γ_1 approaches 0. The regression noise in (9) with x_t as the regressand is, on the other hand, invariant to γ_1 and has finite variance in the limit. The choice of normalization is obvious viewed in this light.

3.2 DGP2: θ_x local to -1

Recall that DGP2 is the triangular representation of a cointegrated system:

$$\Delta x_t = u_{1t}, \quad u_{1t} = e_{1t} + \theta_x e_{1t-1} + \sigma_{12} e_{2t};$$
$$y_t = \beta x_t + u_{2t}, \quad u_{2t} = e_{2t} + \theta_y e_{2t-1} + \sigma_{21} e_{1t}.$$

To rationalize the results of Figure 1, our local asymptotic framework lets $\theta_x \rightarrow -1$ in the limit. That is,

$$\theta_x = -1 + \delta/\sqrt{T}, \ \delta > 0.$$

A distinct feature of DGP2 is that the variables in the system can be weakly exogenous with respect to each other. In other words, σ_{12} and σ_{21} may or may not be zero. As we will now show, the properties of the least squares estimator depend on this assumption.

Lemma 2 : *(Sample Moments) Let x_t and y_t be generated by (7) and let $W_1(r)$ and $W_2(r)$ be independent Wiener processes defined on the space $C[0,1]$.*

If $\sigma_{12} \neq 0$:

1. $T^{-2} \sum_{t=1}^{T} x_t^2 \Rightarrow \sigma_{12}^2 \int_0^1 W_2(r)^2 dr$;

2. $T^{-2} \sum_{t=1}^{T} y_t^2 \Rightarrow \beta^2 \sigma_{12}^2 \int_0^1 W_2(r)^2 dr$;

3. $T^{-2} \sum_{t=1}^{T} x_t u_{2t} \rightarrow 0$;

4. $T^{-2} \sum_{t=1}^{T} x_t y_t \Rightarrow \beta \sigma_{12}^2 \int_0^1 W_2(r)^2 dr$.

If $\sigma_{12} = 0$:

1. $T^{-1} \sum_{t=1}^{T} x_t^2 \Rightarrow \sigma_1^2 + \delta^2 \sigma_1^2 \int_0^1 W_1(r)^2 dr$;

2. $T^{-1} \sum_{t=1}^{T} y_t^2 \Rightarrow \beta^2 \sigma_1^2 (1 + \delta^2 \int_0^1 W_1(r)^2 dr) + (2\beta\sigma_{21} + \sigma_{21}^2)\sigma_1^2 + (1 + \theta_y^2)\sigma_2^2$;

3. $T^{-1} \sum_{t=1}^{T} x_t u_{2t} \rightarrow \sigma_{21} \sigma_1^2$;

14

4. $T^{-1} \sum_{t=1}^{T} x_t y_t \Rightarrow \beta \sigma_1^2 (1 + \delta^2 \int_0^1 W_1(r)^2 dr) + \sigma_{21} \sigma_1^2$.

Results for the case $\sigma_{12} = 0$ are straightforward applications of the results in Nabeya and Perron (1994). Of note is that the rates of normalization for the case $\sigma_{12} \neq 0$ is higher than for the case $\sigma_{12} = 0$. The reason is that x_t is driven by two partial sums when $\sigma_{12} \neq 0$: one relating to its own innovations, and one relating to the innovations in the y_t process. Therefore even though the partial sum of innovations in x_t induces mean-reversion, this effect is dominated by the stochastic trend consisting of integrated innovations on y_t. Accordingly, x_t behaves like a strictly $I(1)$ variable when $\sigma_{12} \neq 0$. The properties of the least squares estimator can now be summarized in the following theorem.

Theorem 2 : Let x_t and y_t be generated by (7). Let $\hat{\beta}_y$ be the least squares estimator with x_t as the regressand, and let $\hat{\beta}_x$ be the least squares estimator with y_t as the regressand.
If $\sigma_{12} \neq 0$:

1. $\hat{\beta}_y \Rightarrow 1/\beta = \beta_y$;

2. $\hat{\beta}_x \Rightarrow \beta = \beta_x$.

If $\sigma_{12} = 0$:

1. $\hat{\beta}_y \Rightarrow \dfrac{\beta(1 + \delta^2 \int_0^1 W_1(r)^2 dr) + \sigma_{21}}{\beta^2(1 + \delta^2 \int_0^1 W_1(r)^2 dr) + (2\beta\sigma_{21} + \sigma_{21}^2) + (1 + \theta_y^2)\sigma_2^2/\sigma_1^2} \neq 1/\beta \equiv \beta_y$;

2. $\hat{\beta}_x \Rightarrow \beta + \dfrac{\sigma_{21}}{(1 + \delta^2 \int_0^1 W_1(r)^2 dr)} \neq \beta \equiv \beta_x$.

Remarks:

1. An interesting aspect of Theorem 2 is that the issue of normalization is irrelevant when $\sigma_{12} \neq 0$ as far as consistency of the least squares estimator is concerned. The intuition is essentially that both $\dot{x}_t$ and y_t have the same order of variability as standard $I(1)$ variables in the limit. In particular, both variables are dominated by the stochastic trend built upon innovations in y_t, even though the DGP looks superficially as though innovations in x_t were the source of variation. Although cross correlation among variables of a cointegrated system induces nuisance parameters which invalidate classical inference, $\sigma_{12} \neq 0$ has the unusual but desirable effect of retaining the I(1) nature of the series, allowing standard asymptotic results to apply to this local asymptotic framework.

15

2. In a regression of y_t on x_t when $\sigma_{12} = 0$, $\widehat{\beta}_x \to \beta = \beta_x$ if $\sigma_{21} = 0$. We are back to the result of the standard $I(0)$ asymptotic framework, where the least square estimator is susceptible to simultaneity bias. The result stated in the above theorem generalizes to cases where the innovations are correlated at different lags, and the extent of the bias is independent of the value of β. The least squares bias will persist unless x_t is orthogonal to the regression error, a condition that is determined by whether $\sigma_{21} = 0$.

3. From a regression of x_t on y_t, $\widehat{\beta}_y$ does not converge to $1/\beta \equiv \beta_y$ even if $\sigma_{21} = 0$. The reason is that in this case, y_t inherits the nearly integrated nearly white noise property of x_t, and its sample moments also require a smaller rate of normalization. Stationary innovations in y_t which would otherwise have converged to zero in the standard asymptotic framework have a non-trivial effect in this local asymptotic framework, as seen from the sample moment of y_t^2 in Lemma 2. This in turn generates a downward bias on the least squares estimator. In general, the bias in $\widehat{\beta}_y$ is a function of σ_{21} and the true value of β_y. However, it can be shown that holding σ_{21} fixed, the larger is β, the more precise are the estimates $\widehat{\beta}_y$ and $\widehat{\beta}_x$ for β_y and β_x irrespective of the value of σ_{21}. Simulations confirm this to be the case.

4. Under the assumption that $\sigma_{12} = 0$, y_t and x_t both behave like stationary processes in the limit. Accordingly, the least squares estimator does not converge at the fast rate of T as would be the case with strictly integrated variables, but at the slower rate of $\sqrt{T}$ as in the case of stationary variables. An implication of Theorem 2 is that there is a discontinuity in the limiting distribution of the least squares estimator at $\sigma_{12} = 0$. However, as we show in Figure 2, there is substantial leakage around $\sigma_{12} = 0$ in finite samples in the sense that when σ_{12} is greater than but close to zero, the precision of the estimates are still affected. This is so even when the sample size is as large as 500.

5. The limiting distributions of the estimators are of little practical interest and their expressions are henceforth omitted. Suffice it to mention that when $\sigma_{12} \neq 0$, neither $T(\widehat{\beta}_y - \beta_y)$ nor $T(\widehat{\beta}_x - \beta_x)$ is mixture normal, both are non-centrally located, and are influenced by all the nuisance parameters in the model. Accordingly, although both estimators are super-consistent, test statistics associated with them cannot be used for inference.

6. From a practical standpoint, normalizing the cointegrating regression on the variable with the smallest spectral density at frequency zero (in this case y_t) is still the preferred choice. When $\sigma_{12} \neq 0$, the issue of normalization is irrelevant and hence the

16

prescription can do no worse than alternative normalizations. When $\sigma_{12} = \sigma_{21} = 0$, using y_t as the regressand is the only normalization that can yield asymptotically unbiased estimates. The results reported in Figure 1 supports this claim. When $\sigma_{12} = 0$ but $\sigma_{21} \neq 0$, estimators such as the DOLS and the FM-OLS will give asymptotically unbiased and efficient estimates provided y_t is the regressand. This is because these estimators provide the proper cure for the problem with the OLS estimator by rendering the regression error uncorrelated with the regressors. However, σ_{21} is not the only source of bias in a regression with x_t as the regressand. Thus, even fully efficient estimators will have limited ability in improving the properties of the estimates. Simulations (available on request) confirm these results.

7. DGP1 is a special case of DGP2 with $\sigma_{12} = 0$ and $\sigma_{21} \neq 0$ for a given sample size. It would seem, by implication of Theorem 2, that the static least squares estimator will be biased whether normalized on x_t or y_t. However, Theorem 1 suggests that a consistent estimator can be obtained when normalized on x_t. To understand these apparently contradicting results, we need to clarify the assumptions underlying the two local asymptotic models. In the first model, parameterizing γ_1 to be local to zero has two effects. It induces a (non-invertible) negative moving average component to the noise function of x_t, and it changes the correlation between the regression innovation and the regressor. As mentioned earlier, the regression innovations with x_t as the regressand are weakly exogenous for y_t in the limit. Under DGP2, θ_x is parameterized to be local to -1 holding σ_{12} and σ_{21} fixed. The results of the two theorems therefore hold under similar but not identical conditions.

4. Residuals Based Tests for Cointegration

Residuals based tests for cointegration are valid in the standard asymptotic framework because the least squares estimator is super-consistent. However, to the extent that the choice of normalization affects the properties of the least squares residuals, the size and power of tests of the null hypothesis of no cointegration might also be affected. To see if this is the case, we use the least squares estimates of cointegrating vectors reported in the Section 2 to construct the estimated residuals. Three statistics, Z_α, MZ_α, and t_ρ, are then used to test for the presence of a unit root in these residuals using the 5 percent critical values in Phillips and Ouliaris (1990). Accordingly, the power of the tests is the frequency that the statistics reject a unit root in the estimated residuals. The size of the tests is constructed as follows. For DGP1, γ_2 in (4) is set to 0 and e_{2t} is specified as a random walk independent

17

of μ_t. Thus, x_t and y_t are both $I(1)$ but are not cointegrated. Similarly, for DGP2, an independent random walk is added to u_{2t} in (7). In this setting, u_{1t} and u_{2t} can still be correlated depending on the values of σ_{12} and σ_{21}. Without loss of generality, we report results for $T = 200$. Results pertaining to DGP1 are given in Table 5 and those for DGP2 are given in Table 6.[3]

The least squares residuals under DGP1 with y_t as the regressand [see (8)] are:

$$\hat{u}_{xt} = (\beta_x - \hat{\beta}_x)x_t + e_{2t} - \beta_x e_{1t}, \tag{11}$$

where $\beta_x = \gamma_2/\gamma_1$. Theorem 1 shows that $\hat{\beta}_x$ is an inconsistent estimator for β_x if $\gamma_1 \to 0$ as T increases. Then

$$\Delta \hat{u}_{xt} = (\beta_x - \hat{\beta}_x)\gamma_1 v_t + \Delta e_{2t} - \hat{\beta}_x \Delta e_{1t}$$

is a stationary and invertible process even as T increases. Accordingly, the least squares autoregression in $\hat{u}_{xt}$ will yield a regression coefficient of one. This evidence of a unit root will lead us to conclude that there is no cointegration even though x_t and y_t shares the common trend μ_t. Simulations reported in Table 5 suggest that MZ_α and t_ρ will suffer from this problem, whereas Z_α is more likely to conclude correctly that there is cointegration when γ_1 is small. The discrepancy between Z_α and the other two tests is again due to the presence of a negative moving-average component in $\Delta \hat{u}_{xt}$, leading Z_α to over-reject the unit root hypothesis.

The least squares residuals from a regression with x_t as the regressand are similarly defined as

$$\hat{u}_{yt} = (\beta_y - \hat{\beta}_y)y_t + e_{1t} - \beta_y e_{2t}, \tag{12}$$

with $\beta_y = \gamma_1/\gamma_2$. Since $\hat{\beta}_y$ is a consistent estimator for β_y, $\sum_{t=1}^{T} \hat{u}_{yt}$ is $O_p(T^{1/2})$, which is indicative of a process that is stationary or $I(0)$. An autoregression in $\hat{u}_{xt}$ will yield a least squares coefficient that is away from one in the limit, and the null hypothesis of no cointegration is rejected. This is consistent with the Monte-Carlo results (see Table 5B).

Under DGP2, the regression residuals are:

$$\hat{u}_{xt} = (\beta_x - \hat{\beta}_x)x_t + u_{2t} \tag{13}$$

when normalized on y_t, and

$$\hat{u}_{yt} = (\beta_y - \hat{\beta}_y)y_t - \beta_y u_{2t} \tag{14}$$

---

[3]In constructing the $Z_{\alpha\mu}$ test we used a Quadratic kernel with a bandwidth selected using the automatic procedure of Andrews (1991) based on an $AR(1)$ approximation. For the statistic $MZ_{\alpha\mu}$, $kmax = 4$ in constructing s_{AR}^2, and $kmax=5$ in constructing t_ρ.

18

when normalized on x_t. Given super-consistency of $\hat{\beta}_x$ and $\hat{\beta}_y$ when $\sigma_{12} \neq 0$, both series have partial sums that are $O_p(T^{1/2})$. Unit root tests on these residuals are then consistent following the analysis of Phillips and Ouliaris (1990).

When $\sigma_{12} = 0$, the results in Table 6 suggest that the residuals based cointegration tests will also reject the null hypothesis of no cointegration. The reason, in this case, is that although the least squares estimator is biased and inconsistent, both x_t and y_t are stationary series in the limit. The two sets of residuals defined above have partial sums that are $O_p(T^{1/2})$, and standard tests once again reject the null hypothesis of a unit root. However, neither series has a stochastic trend in the limit, and it might be more appropriate to think of x_t and y_t as sharing common features in the sense of Engle and Kozicki (1993).

5. Some Observations on Alternative Estimators

Asymptotically efficient methods of estimating cointegrating vectors have been proposed in the literature. Does least squares bias of the type examined here extends to these estimators? This section sheds some light on this issue.[4]

5.1 Least Squares Based Methods

We consider three estimators: the FM-OLS of Phillips and Hansen (1990), the CCR of Park (1992), and the DOLS of Saikkonen (1991) and Stock and Watson (1993). All three are least squares based estimators that are asymptotically efficient, and all require an explicit choice of a regressand. The FM-OLS and CCR estimates of $\partial i / \partial \pi$ from the Fisher equation are .61 and .97 when normalized on i_t, but are .75 and .81 when normalized on π_t. These are in the same range as the OLS and the DOLS estimates presented in Table 1, suggesting that the normalization problem also applies to the FM-OLS and the CCR.

To verify the properties of these efficient estimators, simulations were conducted for DGP1 defined in (4) and DGP2 defined in (7) using the same parameterizations as presented in Tables 3.a and 3.b. Each efficient estimator necessitates the choice of a truncation parameter, k. This is the truncation lag of the kernel estimator of the long-run variance in the case of FM-OLS and CCR, and the number of leads and lags of first differences of the regressor in the case of DOLS. As will be discussed below, the results are robust to the choice of the kernel but are more sensitive to the choice of k.[5]

---

[4] The simulations in this section are done using Gauss 3.2 in conjunction with Coint 2.0 of Ouliaris and Phillips (1994).

[5] The Parzen window is used in the FM-OLS and the CCR. The same kernel is used to construct standard errors associated with the DOLS estimator.

The distributions of these fully efficient estimates are qualitatively similar, but generally closer to the true values for both normalizations than the OLS estimates.[6] This result is not surprising since the OLS estimates, though consistent, are biased and inefficient. Of interest is that the estimates remain noticeably inferior when normalized in one direction. In other words, there is a "good" and a "bad" normalization, the latter being y_t on x_t in the case of DGP1 and x_t on y_t in the case of DGP2. To illustrate the problem, we report results in Table 7.a for DGP1 with $\gamma_1 = .2$, and DGP2 for $\sigma_{12} = 0$ and $\sigma_{12} = .5$, both with $\beta = 1$. When the variable with the smaller normalized spectral density is put on the left hand side the estimates are always more precise. For example, $\hat{\beta}_y$ in DGP1 is .19 on average, close to the true value of .2. However, when normalized on y_t, the average FM-OLS, CCR and DOLS estimates are at best 2.5, 3.5 and 4.7 respectively, with the true value of the coefficient being 5. For DGP2, the estimates from both normalizations are less precise when $\sigma_{12} = 0$ as expected, but even when $\sigma_{12} \neq 0$, the estimates are always better when normalized on y_t.

An observation of note from Table 7.a is that the FM-OLS and the CCR estimates cannot be improved upon by varying the choice of k. However, the DOLS estimates for the "bad" normalization can be made more precise when the lag length is sufficiently large. For example, the average of the DOLS estimates improves from 3.2 to 4.7 as the lag length increases from 1 to 8. Indeed, it can be shown that increasing the value of k at an appropriate rate in the DOLS regression restores consistency of the DOLS estimator when the "bad" normalization is used. Interestingly, it is solving the problem related to the moment matrix of (strongly mean reverting) regressors as indicated in Lemma 1 that makes the DOLS estimates superior to the OLS estimates, and not one of valid conditioning and weak exogeneity that the estimator was intended for. The choice of k is evidently an important issue here, as it needs to be large enough to remove the bias in the OLS estimator on the one hand but without inducing excessive variations to the estimator on the other. Simulations suggest that for practical sample of sizes between 100 and 200, a choice of k between four and six will be sufficient to bring $\hat{\beta}_x$ from an unacceptable range of around 2 under OLS to a more acceptable neighborhood of around 4.5 when $\beta_x = 1/\gamma_1 = 5$. However, for DGP2, k needs to be slightly larger depending on whether or not $\sigma_{12} = 0$.

It should be noted that while increasing k can improve the precision of the DOLS estimates when the "bad" normalization is used, the t statistic associated with the coefficient on x_t is still not well approximated by the standard normal distribution. This can be seen from Table 7.b. The upper and lower 5% critical values are still far from those of the normal distribution even as k increases. The departure from normality of the t statistic associated

---

[6]The full set of estimates are available on request.

with CCR and FM-OLS from the "bad" normalization is even more apparent. Thus, from the point of view of inference, it remains the case that using x_t in DGP1 and y_t in DGP2 as the regressand is the preferred choice because the spectral density at frequency zero of the first difference of these variables is smallest.

Our focus has been on least squares based estimators with the regressand expressed in level form. It is important to emphasize that our recommendation of using as regressand the variable whose spectral density of the differenced series is smallest is based on analyses of regression models of this class. There are evidently other methods of estimating cointegrating vectors. For example, the regressand can be expressed in first-differenced form as in the case of non-linear ECM.[7] Evaluating the many alternative estimators is outside the scope of the present analysis, and we have not explored whether our criteria for choosing the regressand will generalize. We have nevertheless considered a method of estimating cointegrating vectors outside an OLS based framework, and to this we now turn.

5.2 The Johansen Method

For most methods of estimating cointegrating vectors, "normalization" is taken to mean the choice of the regressand. An exception is the reduced-rank regression approach of Johansen (1991), wherein normalization is taken to mean the choice of the length of the eigenvectors. This is sometimes referred to as an empirical normalization. The Johansen procedure does not pretest for the presence of a unit root, and it analyzes all the variables as a system. Because of these fundamental differences with other estimators that belong to the LAMN class, it is of interest to ask if the Johansen approach is rid of the problem being analyzed.

Simulations are again conducted for the parameterizations of DGP1 and DGP2 used earlier, and selected for discussion (without loss of generality) are results from one parameterization of DGP1 and two of DGP2. In Table 8, we report the pair of eigenvectors (normalized on x_t) for various number of lagged first differences used in the reduced rank regression. In each case, we tabulate the frequency distribution of the cointegrating rank, r, chosen by the Trace and Max-λ statistics, noting that if $r = 1$, the first eigenvector is the cointegrating vector. Also reported are the upper and lower 5% critical values of the Wald-type statistic for testing if the estimated cointegrating vector equals the null value.[8]

As we can see from the results, conditional on $r = 1$, the first eigenvector gives quite precise point estimates of the true cointegrating relationship in all the cases considered, though the Wald test tends to over-reject the null hypothesis. This is reminiscent of the size

---

[7]See Gonzalo (1994) for a review of these methods.

[8]The statistic has a χ^2 distribution with 1 degree of freedom. See Johansen (1991) p. 1564 for details.

21

distortions in the t test associated with the three fully efficient estimators when the "bad" normalization is used. Using the 10 percent critical values in Osterwald-Lenum (1992), the frequency that the Trace statistic concludes $r = 1$ is between .8 and .9 for DGP1, and also for DGP2 with $\sigma_{21} \neq 0$, noting in passing that r is never chosen to be zero. However, for DGP2 with $\sigma_{12} = 0$, that is, when both y_t and x_t are nearly integrated nearly white noise, the size of the statistics is distorted. The size problem is more severe the shorter the number of lags. With 4 lags, the statistics report $r = 1$ with a frequency of only 0.4, and finds $r = 2$ in the remaining cases. The exact size of the test remains above 50% for a nominal size of 10% when $\sigma_{12} = .25$. This is the leakage problem around $\sigma_{12} = 0$ discussed earlier.

A finding that $r=2$ in these bivariate DGPs is confounding because if there are n I(1) series, there cannot be more than $n - 1$ cointegrating vectors. A finding that $r = 2$ implies that both series are stationary. Some intuition to this result can be gained by examining the properties of the second eigenvector. Consider DGP2 with $\sigma_{12} = 0$, the case when both series are strongly mean reverting. As seen from Table 8, the first eigenvector is the correct estimate of the cointegrating vector. The second is the unit vector that selects x_t. Since y_t is weakly exogenous for x_t by construction, the eigenvector optimally puts zero weight on the redundant regressor. Simulations of more complex parameterizations of DGP1 also found the statistics to under report $r = 1$ when the second eigenvector is $(1,0)$. In those cases, applying a zero weight on the decisively I(1) process will also give a stationary combination of the variables, although the combination is economically uninteresting. From our analysis, size distortions with the Trace and Max-λ statistics arise when the second eigenvector is of the $(1,0)$ type. Whether the statistics will select one or two cointegrating vectors will depend on how strong are the unit root components relative to the mean-reverting components, and on the causal structure underlying the variables. In general, the stronger is the unit root component in both series, the less chance there is of finding two cointegrating vectors and vice versa.

Do size distortions in testing for the cointegrating rank arise in practice? Returning to the Fisher equation example, the Trace and Max-λ statistics are 29.17 and 22.03 for the null hypothesis that $r = 0$ against the alternative $r = 1$, and 7.15 for both statistics for the hypothesis that $r = 1$ against the alternative that $r = 2$.[9] While we can decisively reject the absence of any cointegrating vector, we reject the presence of one cointegrating vector in favor of two cointegrating vectors at the 10% level but not at the 5% level. The

---

[9] The 90 percent critical values are 15.66 and 6.5 for the Trace statistic; and 12.91 and 6.5 for Max-λ taken from Osterwald-Lenum (1992).

$r = 2$ scenario is therefore of empirical relevance.[10] However, if one had pretested for the presence of a unit root, one would have dismissed the possibility that $r = 2$ since it implies both variables are I(0). One would then obtain the unique cointegrating vector (normalized on π_t) of $[1, -.87]$, in line with the fully-efficient least squares estimates based on the π_t normalization. This, however, is based upon results from pretesting for a unit root, and in this regard, $MZ_{\alpha\mu}$ still plays a useful role.

6. Concluding Comments

Estimations of cointegrating vectors are by now standard practice in dynamic analysis of time series data. This paper questions two conventional wisdoms on least squares estimation of cointegrating vectors. First, we show that although the least squares estimator achieves strong consistency in the standard asymptotic framework, it is not so when a regressor has a large negative moving average component that is modeled as local to -1. Second, we show that normalization in one direction can yield estimates that are consistent but not the other. These results extend to least squares based fully efficient estimators.

Using an unobserved components framework, we show that a negative moving average component can arise when a variable in the cointegrating system has a weak correlation with the common trend, or when the idiosyncratic noise of the variable dominates its total asymptotic variance. Such a variable has properties of an I(1) process, but also exhibits a strong tendency to revert to its mean. For this reason, standard tests have trouble determining whether or not a unit root is present in the series. A non-exhaustive investigation suggests inflation and inventories are integrated processes with large negative moving-average components. Gali (1992) also presented evidence for a negative moving component in the noise function of many macroeconomic time series, and Franses and Haldrup (1994) discussed how negative moving average errors can arise as a result of additive outliers. The issues analyzed in the present paper are therefore not merely of theoretical interest, but are also issues of practical concern.

Our theoretical and empirical investigations lead to the following practical guideline. The cointegration analyses should begin with tests for the presence of a unit root in the variables, preferably using a test that is robust to negative MA errors such as the MZ_α.

---

[10]Using DGP1 to calibrate the data to the time series properties of i_t and π_t shows that the Trace and Max-λ statistics indeed suffer from size distortions. Specifically, γ_1^* is .6, and γ_2^* is 1; $(1 - .6L)e_{1t}^* = v_{1t}^*$, v_{1t}^* is $N(0, (1.25)^2)$, $(1 - .6L)e_{2t}^* = (1 + .25L)v_{2t}^*$, $v_{*2t} \sim N(0, (.5)^2)$, and v_t is $N(0, (.5)^2)$. The normalized spectral densities for the two series are .2 and 1.2 respectively in 100 simulations. The order of the variables in the VECM is y_t, and then x_t. The Trace and Max-λ statistics based on 4 lags find $r = 1$, 80% of the times, and $r = 2$, 20% of the times using a test with nominal size 10%.

23

Step two is to rank the variables by their estimated normalized spectral density at frequency zero, preferably using estimators that are more efficient than those used in the unit root tests. For instance, one can use a kernel based estimator with the demeaned first-differences of the data, or an autoregressive spectral density estimator based on the autoregression (3) without the lagged level. Step three is to choose the variable with the smallest estimated normalized spectral density at frequency zero as the regressand. The steps proposed applies to other least squares based estimators of the cointegrating vector.

There are other instances when a regression can be nearly unbalanced. For example, if the error process of an I(1) series has a large autoregressive root, so that the series is close to being twice integrated. Accordingly, its spectral density at frequency zero is very large. Although we have not provided a theoretical analysis to such cases, simulations reveal that the least squares estimator is also more biased when normalized in one direction. However, using the variable with the smallest spectral density at frequency zero as the regressand still gives more precise estimates.

Our focus has been on estimating a single cointegrating vector. As discussed in Hargreaves (1994), ordinary least squares (and its more efficient variants) can also be used to estimate multiple cointegrating vectors from a single equation with multiple regressors. In that methodology, the second through r cointegrating vectors are constructed to be orthogonal to the first. The choice of the regressand is even more important in that context because all cointegrating vectors are normalized on the same regressand. Our recommendation of comparing the values of the spectral density function will be especially useful.

24

References

Andrews, D. W. K. (1991), Heteroskedastic and Autocorrelation Consistent Matrix Estimation, *Econometrica* **59**, 817–854.

Clark. P. K. (1987), The Cyclical Component of U.S. Economic Activity, *Quarterly Journal of Economics* **102**, 798–814.

Engle. R. F. and Kozicki, S. (1993), Testing for Common Features, *Journal of Business and Economic Statistics* **11**, 369–379.

Franses, P. H. and Haldrup, D. N. (1994), The Effects of Additive Outliers on Tests of Unit Roots and Cointegration, *Journal of Business and Economic Statistics* **12**, 471–478.

Gali, J. (1992), How Well Does the IS-LM Model Fit Postwar U.S. Data?, *Quarterly Journal of Economics* **CVII**, 709–738.

Gonzalo. J. (1994), Five Methods to Estimate Cointegratng Vectors, *Journal of Econometrics* **60**, 203–234.

Hargreaves, C. (1994), A Review of Methods of Estimating Cointegrating Relationships, *in* C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press.

Johansen, S. (1991), Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **59**, 1551–80.

Nabeya, S. and Perron, P. (1994), Local Asymptotic Distribution Related to the AR(1) Model with Dependent Errors. *Journal of Econometrics* **62**, 229–264.

Ng, S. and Perron, P. (1995), Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag, *Journal of the American Statistical Association* **90**, 268–281.

Osterwald-Lenum, M. (1992), A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics, *Oxford Bulletin of Economics and Statistics* **54**, 461–472.

Ouliaris, S. and Phillips, P. C. B. (1994), *COINT 2.0: GAUSS Procedures for Cointegrated Regressions*, Predicta Software.

Park, J. Y. (1992), Canonical Cointegrating Regressions, *Econometrica* **60**, 19–143.

Perron, P. (1994), Non-stationarities and Nonlinearities in Canadian Inflation, *Economic Behaviour and Policy Choice under Price Stability*, Bank of Canada.

Perron, P. and Ng, S. (1994), Useful Modifications to Unit Root Tests with Dependent Errors and their Local Asymptotic Properties, CRDE Discussion Paper 3194, Université of Montréal.

Phillips, P. C. B. (1991), Optimal Inference in Cointegrated Systems, *Econometrica* **59**, 283–306.

Phillips, P. C. B. and Hansen, B. E. (1990), Statistical Inference in Instrumental Variables Regression with I(1) Processes, *Review of Economic Studies* **57**, 99–125.

Phillips, P. C. B. and Ouliaris, S. (1990), Asymptotic Properties of Residual Based Tests for Cointegration, *Econometrica* **58**, 165–193.

Phillips, P. C. B. and Perron, P. (1988), Testing for a Unit Root in Time Series Regression, *Biometrika* **75**, 335–346.

Press, W. H., Teukolsky, S., Vetterling, W. and Flannery, B. (1992), *Numerical Recipies in C*, second edn, Cambridge University Press, Cambridge.

Said, S. E. and Dickey, D. A. (1984), Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order, *Biometrika* **71**, 599–607.

Saikkonen, P. (1991), Asymptotically Efficient Estimation of Cointegrating Regressions, *Econometric Theory* **7**, 1–21.

Schwert, G. W. (1989), Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics* **7**, 147–160.

Stock, J. H. (1987), Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors, *Econometrica* **55**, 1035–1056.

Stock, J. H. (1990), A Class of Tests for Integration and Cointegration, manuscript, Harvard University.

Stock, J. H. and Watson, M. W. (1988), Testing for Common Trends, *Journal of the American Statistical Association* **83**, 1097–1107.

Stock, J. H. and Watson, M. W. (1993), A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems, *Econometrica* **61**:4, 783–820.

Table 1: Analysis of the Fisher Equation

Unit Root tests

	$Z_{\alpha\mu}$	$MZ_{\alpha\mu}$	t_μ	$s^2(QS)/s^2_u$		$s^2(AR)/s^2_u$	
				(1)	(2)	(3)	(4)
i	-9.63	-8.07	-2.10	1.254	1.243	1.014	0.906
π	-40.31	-11.07	-2.54	1.160	0.312	0.306	.253

Notes: (1) The quadratic kernel-based estimate $s^2(QS)$ is constructed using the least-squares residuals from the first-order autoregression (1). Andrews' (1991) method to select the bandwidth is applied using an AR(1) approximation. These estimates are used in the construction of $Z_{\alpha\mu}$. (2) The quadratic kernel-based estimate $s^2(QS)$ is constructed using the demeaned first-differences of the data; the bandwidth is selected as in (1) except that an MA(3) approximation is used for the automatic bandwidth selection in the case of the inflation rate. (3) The estimate $s^2(AR)$ is constructed from the autoregression (2) using k = 8. These estimates are used in the construction of $MZ_{\alpha\mu}$. (4) The estimate $s^2(AR)$ is constructed from an autoregression similar to equation (2) but without the lagged level, the truncation lag is set to k = 8. In all cases s^2_u refers to the sample variance of the first-differences of the data.

Table 2: Cointegration Analysis

	OLS	DOLS(4)	DOLS(8)	DOLS(12)	Z_α	MZ_α	t_α
i on π	.609	.719	.739	.750	-33.75	-20.90	-2.01
π on i	.785	.788	.806	.798	-61.02	-33.04	-2.43

Table 3.A

Regression: $x_t = c + \beta_y\, y_t + u_t$, $\quad \beta_y = \gamma_1/\gamma_2$

DGP: $\mu_t = \mu_{t-1} + v_t$,
$\qquad x_t = \gamma_1 \mu_t + e_{1t}$,
$\qquad y_t = \gamma_2 \mu_t + e_{2t}$,
$\qquad (v, e_1, e_2) \sim (\underline{0}, I)$, $\gamma_2 = 1$.

OLS Estimates

$\gamma_1 = \beta_y$ T	.01	.05	.10	.20	.50	1.0	1.5	2.0	5.0
50	.004	.043	.088	.169	.418	.840	1.268	1.686	4.243
200	.009	.047	.095	.190	.476	.953	1.425	1.898	4.755
500	.010	.049	.098	.196	.490	.979	1.469	1.958	4.898

Empirical Distribution of t_β

	.01	.05	.10	.20	.50	1.0	1.5	2.0	5.0
T=50 $t_{.05}$ $t_{.95}$	-1.738 1.550	-1.760 1.554	-1.756 1.380	-2.016 1.186	-2.859 .533	-3.848 -.127	-4.154 -.226	-4.407 -.430	-4.864 -.536
T=200 $t_{.05}$ $t_{.95}$	-1.611 1.617	-1.917 1.409	-1.857 1.412	-2.216 1.127	-3.004 .534	-3.961 .059	-4.526 -.260	-4.777 -.054	-5.103 -.705
T=500 $t_{.05}$ $t_{.95}$	-1.763 1.562	-1.747 1.473	-1.892 1.378	-2.273 1.090	-3.112 .360	-4.252 -.120	-4.671 -.460	-5.089 -.516	-5.231 -.779

Table 3.B

Regression: $Y_t = c + \beta_x x_t + u_t, \quad \beta_x = \gamma_2/\gamma_1$

OLS Estimates

γ_1 / β_x	.01 / 100	.05 / 20	.10 / 10	.20 / 5	.50 / 2	1.0 / 1	1.5 / .75	2.0 / .5	5.0 / .2
T = 50	.094	.423	.762	1.101	1.219	.842	.613	.477	.199
200	.332	1.503	2.189	2.479	1.679	.948	.651	.494	.200
500	.858	3.208	3.939	3.483	1.853	.979	.660	.497	.200

Empirical Distribution of t_β

	.01 / 100	.05 / 20	.10 / 10	.20 / 5	.50 / 2	1.0 / 1	1.5 / .75	2.0 / .5	5.0 / .2
$t_{.05}$ T=50	357.93	-73.005	-35.902	-18.762	-7.244	-3.572	-2.667	-2.266	-1.724
$t_{.95}$	-75.979	-15.571	-7.952	-4.487	-1.719	.001	.772	1.111	1.528
$t_{.05}$ T=200	-385.23	-79.416	-40.356	-20.474	-8.360	-4.088	-2.744	-2.315	-1.686
$t_{.95}$	-74.954	-16.280	-9.111	-5.515	-1.985	-.156	.592	1.071	1.607
$t_{.05}$ T=500	-407.73	-78.834	-41.303	-20.529	-8.447	-4.390	-2.905	-2.314	-1.747
$t_{.95}$	-77.677	-16.788	-10.571	-6.043	-2.110	-.104	.676	.986	1.500

Table 4: Size of Unit Root Tests on x, T=200.

DGP: $\mu_t = \mu_{t-1} + v_t$,

$x_t = \gamma_t \mu_t + e_t$,

$(v, e_t) \sim (\underline{0}, I)$

γ_1	.01	.05	.10	.20	.50	1.0	1.5	2.0	5.0
Z_α	1.0	1.0	1.0	.986	.589	.209	.104	.070	.051
MZ_α	.958	.691	.350	.191	.109	.086	.064	.052	.059
t_ρ	1.0	.828	.491	.261	.097	.088	.070	.063	.061

Table 5: Power (Size) of Rejecting the Null Hypothesis of No Cointegration in DGP1, T=200.

A. Using Residuals from Regression of x on y

γ_1	.01	.05	.10	.20	.50	1.0	1.5	2.0	5.0
Z_α	1.000 (1.000)	1.000 (1.000)	1.000 (.985)	1.000 (.814)	1.000 (.300)	1.000 (.070)	1.000 (.034)	1.000 (.021)	1.000 (.013)
MZ_α	.999 (.997)	.994 (.800)	.994 (.421)	.987 (.155)	.966 (.033)	.920 (.018)	.898 (.019)	.864 (.008)	.833 (.019)
t_ρ	1.000 (1.000)	1.000 (.890)	1.000 (.545)	1.000 (.213)	.990 (.035)	.962 (.010)	.955 (.015)	.949 (.013)	.923 (.008)

B. Using Residuals from Regression of y on x

γ_1	.01	.05	.10	.20	.50	1.0	1.5	2.0	5.0
Z_α	.025 (.001)	.051 (.002)	.353 (.031)	.806 (.077)	.999 (.039)	1.0 (.018)	1.0 (.009)	1.0 (.008)	1.0 (.011)
MZ_π	.006 (.001)	.003 (.003)	.005 (.002)	.025 (.006)	.552 (.011)	.938 (.013)	.971 (.012)	.998 (.015)	.996 (.016)
t_ρ	.004 (.001)	.002 (.002)	.003 (.002)	.031 (.002)	.577 (.012)	.947 (.011)	.982 (.003)	.998 (.008)	.999 (.010)

DGP: $\mu_t = \mu_{t-1} + v_t$, $x_t = \gamma_1 \mu_t + e_{1t}$, $y_t = \gamma_2 \mu_t + e_{2t}$, $(v, e_1, e_2) \sim$ i.i.d. $(0, I)$, $\gamma_2 = 1$. size is constructed by setting γ_2 to zero and letting $e_{2t} = e_{1t}$, $v_t \sim N(0,1)$, $e \sim$ ...

31

Table 6

Power (size) of Rejecting the Null Hypothesis of No Cointegration in DGP2; $\beta=1$; $T=200$.

A. Using Residuals from Regression of x on Y

θ_x		-0.80	-0.50	-0.20	0.00	0.20	0.50	0.80
Z_α	$\sigma_{12}=0$	1.000 (.574)	1.000 (.354)	1.000 (.190)	1.000 (.186)	1.000 (.199)	1.000 (.190)	1.000 (.191)
	$\sigma_{12}=.5$	1.000 (.964)	1.000 (.438)	1.000 (.169)	1.000 (.151)	1.000 (.165)	1.000 (.180)	1.000 (.197)
MZ_α	$\sigma_{12}=0$	0.921 (.108)	0.976 (.094)	0.976 (.072)	0.987 (.071)	0.999 (.081)	0.998 (.088)	0.997 (.077)
	$\sigma_{12}=.5$	0.999 (.332)	0.999 (.102)	0.998 (.070)	0.998 (.065)	0.998 (.087)	0.999 (.080)	0.999 (.086)
t_r	$\sigma_{12}=0$	0.895 (.119)	0.964 (.093)	0.997 (.082)	0.999 (.066)	1.000 (.084)	1.000 (.092)	1.000 (.078)
	$\sigma_{12}=.5$	1.000 (.093)	1.000 (.101)	1.000 (.077)	1.000 (.074)	1.000 (.078)	1.000 (.068)	1.000 (.073)

B. Using Residuals from Regression of y on x

θ_x		-0.80	-0.50	-0.20	0.00	0.20	0.50	0.80
Z_α	$\sigma_{12}=0$	1.000 (.386)	1.000 (.299)	1.000 (.248)	1.000 (.217)	1.000 (.221)	1.000 (.219)	1.000 (.217)
	$\sigma_{12}=.5$	1.000 (.324)	1.000 (.280)	1.000 (.233)	1.000 (.232)	1.000 (.234)	1.000 (.228)	0.996 (.211)
MZ_α	$\sigma_{12}=0$	1.000 (.053)	1.000 (.092)	0.999 (.088)	0.998 (.082)	1.000 (.089)	1.000 (.092)	1.000 (.087)
	$\sigma_{12}=.5$	0.998 (.084)	0.999 (.076)	1.000 (.086)	1.000 (.093)	1.000 (.101)	1.000 (.078)	1.000 (.090)
t_r	$\sigma_{12}=0$	1.000 (.052)	1.000 (.095)	1.000 (.082)	1.000 (.076)	1.000 (.087)	1.000 (.091)	1.000 (.086)
	$\sigma_{12}=.5$	1.000 (.093)	1.000 (.080)	1.000 (.086)	1.000 (.098)	1.000 (.095)	1.000 (.082)	1.000 (.088)

DGP: $\Delta x_t = u_{1t}$; $u_{1t} = e_{1t} + \theta_1 e_{1t-1} + \sigma_{12} e_{2t}$; $Y_t = \beta x_t + u_{2t}$; $e_{1t} \sim N(0,1)$, $u_{2t} \sim N(0,1)$, ... $\sim N(0,1)$ to $u_{...}$

Table 7.A Average Estimates of some Asymptotically Efficient Estimators

		k = 2			k = 4			k = 8		
		DOLS	CCR	FMOLS	DOLS	CCR	FMOLS	DOLS	CCR	FMOLS
DGP1	$\beta_x=5$	4.198	3.428	2.713	4.530	3.569	2.581	4.757	3.575	2.529
	$\beta_y=.2$	.198	.196	.191	.199	.198	.190	.201	.199	.192
DGP2 $\sigma_{12}=0$	$\beta_x=1$	.995	.994	.996	.998	.999	1.000	.999	.998	.999
	$\beta_y=1$	.863	.766	.678	.907	.785	.662	.951	.796	.661
DGP2 $\sigma_{12}=.5$	$\beta_x=1$	1.010	1.005	1.007	1.003	1.006	1.008	1.006	1.008	1.011
	$\beta_y=1$	.962	.935	.931	.931	.943	.931	.996	.941	.938

Table 7.B Upper and Lower 5% of the t-statistic

		k = 2						k = 4						k = 8					
		DOLS		CCR		FMOLS		DOLS		CCR		FMOLS		DOLS		CCR		FMOLS	
DGP1	$\beta_x=5$	-15.6	.35	-12.94	-2.07	-16.17	-2.21	-11.69	2.26	-9.12	-1.41	-12.21	-1.34	-9.86	5.47	-7.36	-.98	-10.30	-.92
	$\beta_y=.2$	-1.80	1.66	-1.94	1.39	-1.95	1.39	-1.66	1.61	-1.89	1.64	-1.96	1.63	-1.79	1.83	-1.94	1.72	-2.00	1.72
DGP2 $\sigma_{11}=0$	$\beta_x=1$	-1.79	1.54	-1.85	1.64	-1.93	1.68	-1.87	1.73	-1.87	1.77	-2.06	1.91	-1.78	1.68	-1.84	1.73	-1.98	1.89
	$\beta_y=1$	-6.49	-.02	-9.35	-1.82	-12.06	-2.45	-4.91	.87	-8.25	-1.31	-11.34	-1.84	-3.71	1.66	-7.23	-.85	-10.30	1.30
DGP2 $\sigma_{11}=.5$	$\beta_x=1$	-.40	.98	-1.02	1.41	-.99	1.53	-.48	.63	-.93	1.38	-.88	1.58	-.47	.54	-.92	1.67	-.84	1.93
	$\beta_y=1$	-4.99	.12	-5.52	-.57	-6.45	-.35	-2.26	.49	-4.40	-.0002	-5.21	-.19	-1.20	.78	-4.13	-.29	-5.21	-.18

Table 8 Statistics Related to the Johansen procedure

		k = 2		k = 4		k = 8	
DGP1	eigenvec	(1, -.199)	(1, 5.862)	(1, -.198)	(1, -2.720)	(1, -.201)	(1, -6.424)
	χ^2 (.05,.95)	.0048	4.52	.0042	5.53	-.0081	7.71
	Tr, Maxλ	.801	.801	.883	.883	.878	.882
DGP2 $\sigma_{12}=0$	eigenvec	(1, -1.118)	(1, .004)	(1, -1.016)	(1, .044)	(1, -1.012)	(1, .169)
	χ^2 (.05,.95)	.0090	10.99	.0059	9.92	.0089	9.03
	Tr, Maxλ	.051	.051	.362	.362	.769	.770
DGP2 $\sigma_{11}=.5$	eigenvec	(1, -1.0)	(1, -.310)	(1, -1.0)	(1, -.947)	(1, -1.0)	(1, 1.547)
	χ^2 (.05,.95)	.0013	2.41	.0018	2.25	.0034	3.46
	Tr, Maxλ	.547	.547	.814	.814	.890	.890

Note: k is the number of lagged first differences used in the reduced rank regressions. For each k, the first row labelled "eigenvec" reports the eigenvectors $(1,-\beta)$, normalized on x, from solving the reduced-rank equations. The row labelled "χ^2" reports the 5% and 95% critical values of the Wald statistic for tesing the null hypothesis that $\beta=\beta^*$, where β^* is the value of β under the null hypothesis. The row labelled "Tr, Max-λ" are the frequencies that the Trace and Max-λ statistics conclude r=1, the true value.
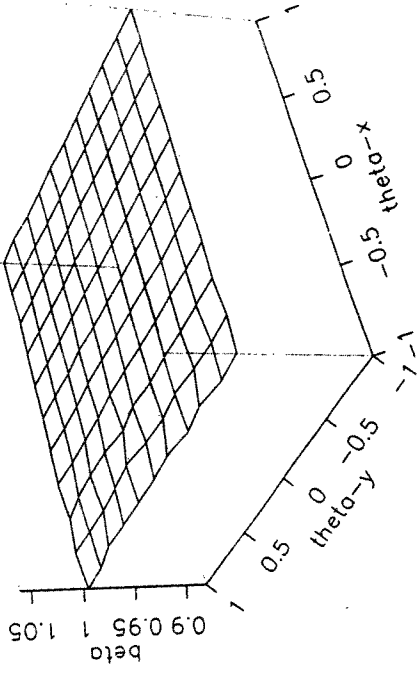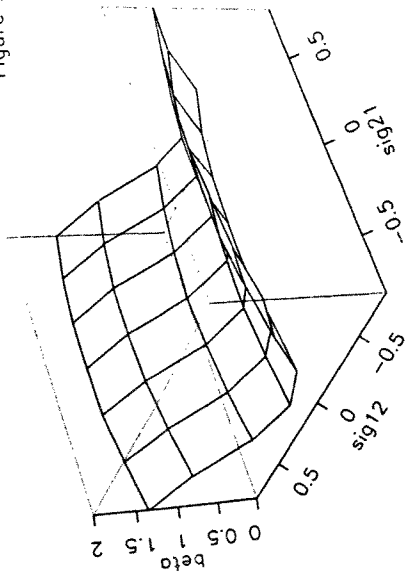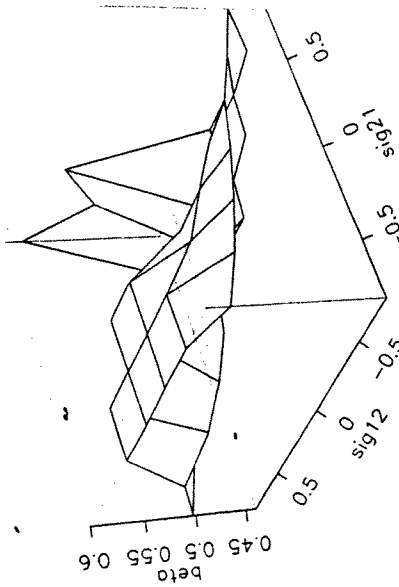
OLS: x on y
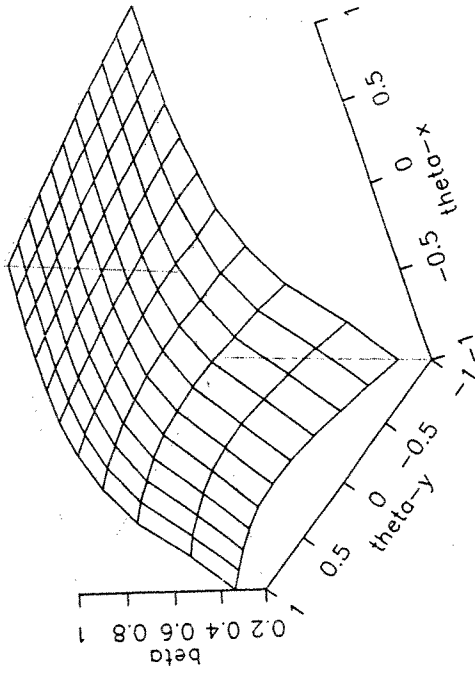
OLS: y on x

Figure 1: DGP2

35
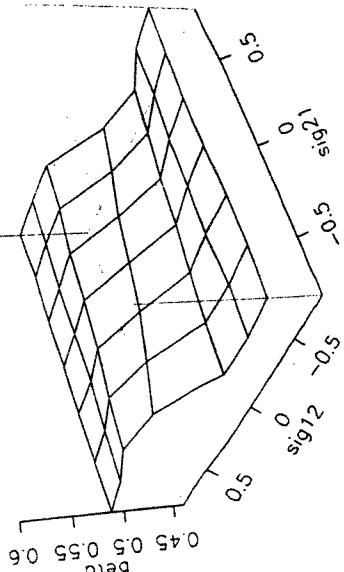
Figure 2: DGP2

y on x

x on y

Beta=2.0 theta−x=−0.8, theta−y=0 T=200

Beta=2.0 theta−x=−0.88, theta−y=0 T=500
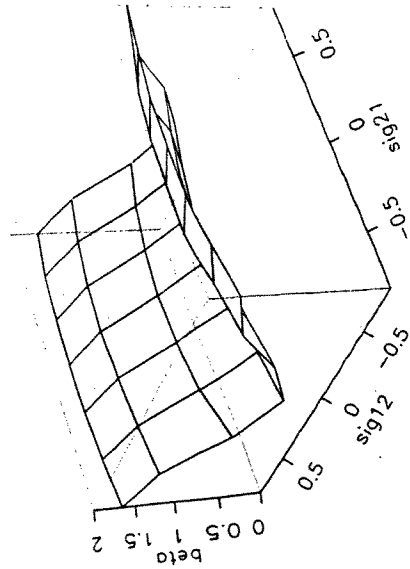
Cahiers de recherche (Discussion papers)
1994 à aujourd'hui (1994 to date)

Si vous désirez obtenir un exemplaire, vous n'avez qu'à faire parvenir votre demande et votre paiement (5 $ l'unité) à l'adresse ci-haut mentionnée. / To obtain a copy ($ 5 each), please send your request and prepayment to the above-mentioned address.

9401 : Mercenier, Jean et Bernardin Akitoby, "On Intertemporal General-Equilibrium Reallocation Effects of Europe's Move to a Single Market", janvier 1994, 41 pages.

9402 : Gauthier, Céline et Michel Poitevin, "Using Ex Ante Payments in Self-Enforcing Risk-Sharing Contracts", février 1994, 38 pages.

9403 : Ghysels, Eric et Joanna Jasiak, "Stochastic Volatility and Time Deformation : an Application of Trading Volume and Leverage Effects", février 1994, 37 pages.

9404 : Dagenais, Marcel G. et Denyse L. Dagenais, "GMM Estimators for Linear Regression Models with Errors in the Variables", avril 1994, 33 pages.

9405 : Bronsard, C., Fabienne Rosenwald et Lise Salvas-Bronsard, "Evidence on Corporate Private Debt Finance and the Term Structure of Interest Rates", avril 1994, 42 pages.

9406 : Dinardo, John, Nicole M. Fortin et Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992 : A Semiparametric Approach", avril 1994, 73 pages.

9407 : Campbell, Bryan et Jean-Marie Dufour, "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter", avril 1994, 32 pages.

9408 : Bollerslev, Tim et Eric Ghysels, "Periodic Autoregressive Conditional Heteroskedasticity", mai 1994, 29 pages.

9409 : Cardia, Emanuela, "The Effects of Government Financial Policies : Can We Assume Ricardian Equivalence?", mai 1994, 42 pages.

9410 : Kollmann, Robert, "Hidden Unemployment : A Search Theoretic Interpretation", mai 1994, 9 pages.

9411 : Kollmann, Robert, "The Correlation of Productivity Growth Across Regions and Industries in the US", juin 1994, 14 pages.

9412 : Gaudry, Marc, Benedikt Mandel et Werner Rothengatter, "Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models within a Quasi-Direct Format (QDF)", juin 1994, 64 pages.

9413 : Gaudry, Marc et Alexandre Le Leyzour, "Improving a Fragile Linear Logit Model Specified for High Speed Rail Demand Analysis in the Quebec-Windsor Corridor of Canada", août 1994, 39 pages.

9414 : Lewis, Tracy et Michel Poitevin, "Disclosure of Information in Regulatory Proceedings", juillet 1994, 38 pages.

9415 : Ambler, Steve, Emanuela Cardia et Jeannine Farazli, "Export Promotion and Growth", août 1994, 41 pages.

9416 : Ghysels, Eric et Haldun Sarlan, "On the Analysis of Business Cycles Through the Spectrum of Chronologies", août 1994, 37 pages.

9417 : Martel, Jocelyn et Timothy C.G. Fisher, "The Creditors' Financial Reorganization Decision : New Evidence from Canadian Data", août 1994, 21 pages.

9418 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Entrance Quotas and Admission to Medical Schools : A Sequential Probit Model", septembre 1994, 26 pages.

9419 : Cannings, Kathy, Claude Montmarquette et Sophie Mahseredjian, "Major Choices : Undergraduate Concentrations and the Probability of Graduation", septembre 1994, 26 pages.

9420 : Nabeya, Seiji et Pierre Perron, "Approximations to Some Exact Distributions in the First Order Autoregressive Model with Dependent Errors", septembre 1994, 40 pages.

9421 : Perron, Pierre, "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", octobre 1994, 50 pages.

9422 : Vogelsang, Timothy J. et Pierre Perron, "Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time", novembre 1994, 57 pages.

9423 : Ng, Serena et Pierre Perron, "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag", décembre 1994, 41 pages.

9424 : Perron, Pierre, "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors", décembre 1994, 37 pages.

9425 : Ghysels, Eric et Pierre Perron, "The Effect of Linear Filters on Dynamic Time Series with Structural Change", décembre 1994, 35 pages.

9426 : Boyer, Marcel, Jean-Jacques Laffont, Philippe Mahenc et Michel Moreaux, "Sequential Location Equilibria Under Incomplete Information", décembre 1994, 38 pages.

9427 : Perron, Pierre et Serena NG, "Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties", décembre 1994, 41 pages.

9428 : Garcia, René et Pierre Perron, "An Analysis of the Real Interest Rate Under Regime Shifts", décembre 1994, 42 pages.

9501 : Boyer, Marcel et Jean-Jacques Laffont, "Environmental Risks and Bank Liability", janvier 1995, 46 pages.

9502 : Margolis, David. N., "Firm Heterogeneity and Worker Self-Selection Bias Estimated Returns to Seniority", décembre 1994, 29 pages.

9503 : Abowd, John M., Francis Kramarz et David N. Margolis, "High-Wage Workers and High-Wage Firms", janvier 1995, 73 pages

9504 : Cardia, Emanuela et Steve Ambler, "Indexation Lags and Heterodox Stabilization Programs", janvier 1995, 29 pages.

9605 : Garcia, René et Huntley Schaller, "Are the Effects of Monetary Policy Asymmetric?", février 1995, 42 pages.

9506 : Parent, Daniel, "Survol des contributions théoriques et empiriques liées au capital humain", février 1995, 70 pages.

9507 : Parent, Daniel, "Wages and Mobility : The Impact of Employer-Provided Training", février 1995, 34 pages.

9508 : Parent, Daniel, "Industry-Specific Capital and the Wage Profile : Evidence from the NLSY and the PSID", février 1995, 21 pages.

9509 : Parent, Daniel, "Matching, Human Capital, and the Covariance Structure of Earnings", février 1995, 54 pages.

9510 : Garcia, René, "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", mars 1995, 50 pages.

9511 : Garcia, René, Annamaria Lusardi et Serena Ng, "Excess Sensitivity and Asymmetries in Consumption : An Empirical Investigation", mars 1995, 26 pages.

9512 : Sprumont, Yves, "An Axiomatization of the Pazner-Schmeidler Rules in Large Fair Division Problems", mars 1995, 26 pages.

9513 : Ghysels, Eric, Lynda Khalaf et Cosmé Vodounou, "Simulation Based Inference in Moving Average Models", mars 1995, 10 pages.

9514 : Ng, Serena, "Looking for Evidence of Speculative Stockholding in Commodity Markets", mars 1995, 25 pages.

9515 : Ng, Serena et Huntley Schaller, "The Risky Spread, Investment, and Monetary Policy Transmission : Evidence on the Role of Asymmetric Information", mars 1995, 26 pages.

9516 : Ng, Serena, "Testing for Homogeneity in Demand Systems when the Regressors are Non-Stationary", mars 1995, 26 pages.

9517 : Ghysels, Eric, Clive W.J. Granger et Pierre L. Siklos, "Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?", mars 1995, 34 pages.

9518 : Ghysels, Eric, Alastair Hall et Hahn S. Lee, "On Periodic Structures and Testing for Seasonal Unit Roots", mars 1995, 45 pages.

9519 : Sprumont, Yves, "On the Game-Theoretic Structure of Public-Good Economies", mars 1995, 21 pages.

9520 : Charles, Sandra, François Vaillancourt et Nicolas Marceau, "The Impact of Decentralization on Growth and Democracy : A Note", mars 1995, 13 pages.

9521 : Sprumont, Yves, "Balanced Egalitarian Redistribution of Income", mars 1995, 17 pages.

9522 : Bronsard, Camille, Lise Salvas-Bronsard et Alain Trognon, "On the Residual Dynamics Implied by the Rational Expectations Hypothesis", mars 1995, 18 pages.

9523 : Campbell, Bryan et Eric Ghysels, "An Empirical Analysis of the Canadian Budget Process", mars 1995, 30 pages.

9524 : Ghysels, Eric, Alain Guay et Alastair Hall, "Predictive Tests for Structural Change with Unknown Breakpoint", avril 1995, 29 pages.

9525 : Ghysels, Eric, "On Stable Factor Structures in the Pricing of Risk", avril 1995, 37 pages.

9526 : Kollmann, Robert, "Mark Up Fluctuations in U.S. Manufacturing and Trade : Empirical Evidence Based on A Model of Optimal Storage", avril 1995, 18 pages.

9527 : Laberge-Nadeau, Claire, Georges Dionne, Urs Maag, Denise Desjardins, Charles Vanasse et J.-M. Ékoé, "Medical Conditions and the Severity of Commercial Motor Vehicle (CMV) Drivers' Road Accidents", mai 1995, 26 pages.

9528 : Dionne, Georges, Manuel Artis et Montserrat Guillén, "On the Repayment of Personal Loans Under Asymmetrical Information : A Count Data Model Approach", mai 1995, 32 pages.

9529 : Ruge-Murcia, Francisco J., "Government Expenditure and the Dynamics of High Inflation", juin 1995, 45 pages.

9530 : Persaran, M. Hashem et Francisco J. Ruge-Murcia, "A Discrete-Time Version of Target Zone Models with Jumps", juin 1995, 50 pages.

9531 : Dionne, Georges et N. Fombaron, "Non-Convexities and Efficiency of Equilibria in Insurance Markets with Asymmetric Information", juin 1995, 22 pages.

9532 : Bronsard, C., P. Michel et L. Salvas-Bronsard, "Singular Demand Systems as an Instrument of Scientific Discovery", juillet 1995, 17 pages.

9533 : Dionne, G. et C. Fluet, "Incentives in Multi-Period Regulation and Procurement : A Graphical Analysis", juillet 1995, 46 pages.

9534 : Ng, Serena et Pierre Perron, "Estimation and Inference in Nearly Unbalanced, Nearly Cointegrated Systems", août 1995, 38 pages.