

Cointegration in Fractional Systems with Deterministic Trends

P.M. Robinson* and F. Iacone
Department of Economics
London School of Economics
Houghton Street
London WC2A 2AE

February 11, 2004

Abstract

We consider a cointegrated system generated by processes that may be fractionally integrated, and by additive polynomial or generalized polynomial trends. In view of the consequent competition between stochastic and deterministic trends, we consider various estimates of the cointegrating vector and develop relevant asymptotic theory, including the situation where fractional orders of integration are unknown.

Keywords: Fractional cointegration; Deterministic trends; Ordinary least squares estimation; Generalized least squares estimation; Wald tests.

JEL classification: C22.

1 Introduction

Fractional cointegration analysis has developed on the basis that cointegrating relationships between nonstationary economic variables may exist without observable processes necessarily being unit root ($I(1)$) processes or cointegrating errors necessarily being short memory ($I(0)$) processes. Optimal rules of statistical inference based on the $I(1)/I(0)$ prescription lose their most desirable properties when the integration orders on which they are based are misspecified, while on the other hand methodology developed in the $I(1)/I(0)$ setting is not designed to detect fractional cointegration. A fractional cointegration analysis that allows integration orders to be unknown, and nests integer-valued orders as special cases, thus affords a substantial increase in flexibility. Though

*Corresponding author: Tel: +44-20-7955-7516, fax: +44-20-7831-1840/ E-mail: p.m.robinson@lse.ac.uk

the fractional cointegration literature is dwarfed by the $I(1)/I(0)$ literature, there are now a number of references dealing with theoretical issues (see e.g. Dolado and Marmol (1997), Chan and Terrin (1995), Jeganathan (1999), Kim and Phillips (2000), Velasco (2000), Hassler, Marmol and Velasco (2002)) and empirical applications (see e.g. Cheung and Lai (1993)).

The large $I(1)/I(0)$ literature focusses on many features of economic time series, in particular recognizing empirical evidence that the stochastic unit root trend frequently needs to be supplemented by a deterministic trend, such as one increasing linearly with time (see e.g. West (1988), Stock and Watson (1988), Johansen (1991), Hansen (1992), Perron and Campbell (1993), Campos *et al* (1996)). For empirical applications to the analysis of demand for money, see Hoffman and Rasche (1991), Stock and Watson (1993), and to the PPP/UIP relations see Johansen and Juselius (1992). A review of applications of cointegrated models with a background in economic theory is in Söderlind and Vredin (1996), and a particular treatment of cotrending, deterministic and stochastic cointegration is in Ogaki and Park (1997), who modelled the allocation of income in consumption of durable and non-durable goods.

On the other hand the fractional cointegration literature has mostly not allowed for deterministic trends. For example, this is the case in Robinson and Marinucci (2001) (hereafter RM) (see also Robinson and Marinucci, 2003) who study limiting behaviour of least squares and narrow-band least squares estimates of cointegrating coefficients, and in Robinson and Hualde (2003) (hereafter RH), who provide optimal methods for estimating a bivariate, purely stochastic, cointegrated system that allows for unknown integration orders and very general parametric modelling of stationary short memory, and show that Wald tests on the cointegrating coefficient have the asymptotic null χ^2 limit distribution found by, e.g., Phillips and Hansen (1990), Phillips (1991a,b), Johansen (1991), Hansen (1992) in the $I(1)/I(0)$ case. The present paper develops properties of estimates of the cointegrating coefficient in a bivariate model that either ignore or take account of additive deterministic trends.

To introduce some formal definitions, define the difference operator $\Delta = 1 - L$, where L is the lag operator, and, for any real d , $d \neq -1, -2, \dots$,

$$(1 - z)^{-d} = \sum_{j=0}^{\infty} a_j(d) z^j, \quad a_j(d) = \frac{\Gamma(j + d)}{\Gamma(d)\Gamma(j + 1)}, \quad (1)$$

with Γ denoting the gamma function. Denoting by v_t a covariance stationary process with zero mean and spectral density that is bounded and bounded away from zero at all frequencies, for any real-valued d we call w_t an $I(d)$ (stochastic) sequence if

$$w_t = \Delta^{-d} v_t^{\#}, \quad (2)$$

where, throughout, the $\#$ superscript has the meaning

$$v_t^{\#} = v_t 1(t > 0), \quad (3)$$

where $1(\cdot)$ is the indicator function. Thus (3) implies that $w_t = 0$, $t \leq 0$, so that even within the 'stationarity' region for d , $d < 1/2$, w_t is actually not covariance

stationary, though it may be thought of as 'asymptotically stationary' for such d . The truncation in (2) is introduced to cater for 'nonstationary' values, $d \geq 1/2$, where w_t would otherwise blow up.

We also introduce a corresponding class of deterministic sequences. We say w_t is a $J(d)$ (deterministic) sequence if

$$w_t \sim \frac{t^{d-1/2}}{\Gamma(d+1/2)}, \text{ as } t \rightarrow \infty, \quad |w_t - w_{t+1}| \leq Cw_t/t, \quad t \geq 1, \quad (4)$$

where throughout C denotes a finite generic constant and \sim indicates that the ratio between left- and right- sides tends to a finite positive constant. The definitions of $I(d)$ and $J(d)$ match in the sense that:

- (i) If w_t is $I(d)$ then $\Delta^c w_t^\#$ is $I(d-c)$; if w_t is $J(d)$ then $\Delta^c w_t^\#$ is $J(d-c)$.
- (ii) If $d > 0$, then: if w_t is $I(d)$, $Ew_t^2 \sim t^{2d-1}$ as $t \rightarrow \infty$; if w_t is $J(d)$, $w_t^2 \sim t^{2d-1}$ as $t \rightarrow \infty$.

We expect then, in a system containing $I(d)$ and $J(e)$ sequences, that (fractional) differencing has a similar effect on both and that there will be a tendency for an $I(d)$ component to dominate a $J(e)$ one for $d > e$, and vice versa when $d < e$. This is most simply seen in an additive model. We consider a bivariate observable sequence (x_t, y_t) given by the components model

$$x_t = g_{1t}(\phi_1) + h_{1t}(\delta_1), \quad t \geq 1, \quad (5)$$

$$y_t = g_{2t}(\phi_2) + h_{2t}(\delta_2), \quad t \geq 1, \quad (6)$$

$$g_{1t}(\phi_1) = h_{1t}(\delta_1) = g_{2t}(\phi_2) = h_{2t}(\delta_2) \equiv 0, \quad t \leq 0, \quad (7)$$

where $g_{it}(d)$ is $J(d)$ and $h_{it}(d)$ is $I(d)$, $i = 1, 2$ and the $g_{it}(\phi_i)$, $h_{it}(\delta_i)$ are unobservable sequences. Robinson and Marinucci (2000) discussed a fractional model of form (5)-(7). In particular they considered the asymptotic behaviour of the sample covariance matrix of (x_t, y_t) , based on observations at $t = 1, 2, \dots, n$, and of the averaged periodogram of (x_t, y_t) , being the normalized sum of the periodogram matrix of (x_t, y_t) over m Fourier frequencies close to the origin, namely $\lambda_j = 2\pi j/n$, $j = 1, 2, \dots, m$, where $1/m + m/n \rightarrow 0$ as $n \rightarrow \infty$; their stress is on cases where $\phi_i > 0$, $\delta_i > 1/2$, $i = 1, 2$, when the sample covariance matrix and the averaged periodogram have the same asymptotic behaviour (since they are equal when $m = n$, and low frequency components dominate the sample covariance matrix). Robinson and Marinucci (2000) described how stochastic or deterministic components dominate, depending on the values of the ϕ_i, δ_i . They also considered a (narrow-band) regression of y_t on x_t , in case the deterministic component of x_t dominates the error, and the stochastic component of the error dominates its deterministic component.

Regression relations also arise in case of cointegration. However, whereas in the case of Robinson and Marinucci (2000) the estimate of the regression coefficient is consistent and asymptotically normal due to the dominating effect of the deterministic component of x_t , consistency in cointegrated systems involving no

deterministic components is due to the domination of the stochastic component of the cointegrating error by that of the regressor. Here we look at implications of deterministic components. On the basis of (5)-(7), it is supposed that $\delta_1 = \delta_2 = \delta$ and there exists $\nu \neq 0$ such that $h_{2t}(\delta) - \nu h_{1t}(\delta)$ is $I(\gamma)$, $\gamma < \delta$.

Irrespective of whether or not we commence from the component model (5)-(7), we assume the following cointegrated system as data generating mechanism,

$$y_t = \nu x_t + \sum_{j=1}^{p_1} \mu_{1j} t^{\phi_{1j}-1/2} + u_{1t}(-\gamma), \quad (8)$$

$$x_t = \sum_{j=1}^{p_2} \mu_{2j} t^{\phi_{2j}-1/2} + u_{2t}(-\delta), \quad (9)$$

where for any sequence $\{v_t\}$, and any c , we use the notation of RH,

$$v_t(c) = \Delta^c v_t^\# . \quad (10)$$

In (8), (9), $u_t = (u_{1t}, u_{2t})'$, with the prime denoting transposition, is a jointly covariance stationary process with zero mean and spectral density matrix, $f(\lambda)$, satisfying

$$E(u_0 u_j') = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) d\lambda, \quad (11)$$

that is nonsingular and continuous at all frequencies;

$$\nu \neq 0, \quad (12)$$

$$\delta > \max(\gamma, \frac{1}{2}); \quad (13)$$

the ϕ_{ij} are real numbers satisfying

$$\phi_{11} > \dots > \phi_{1p_1} > 0, \quad (14)$$

$$\phi_{21} > \dots > \phi_{2p_2} > 0, \quad (15)$$

and also, for $0 \leq p_{11} \leq p_1$, $0 \leq p_{21} \leq p_2$,

$$\phi_{1p_{11}} > \gamma, \phi_{1,p_{11}+1} < \gamma, \quad (16)$$

$$\phi_{2p_{21}} > \delta, \phi_{2,p_{21}+1} < \delta; \quad (17)$$

for $p_{11} = p_1$ ($p_{21} = p_2$), the quantities $p_{11} + 1$ ($p_{21} + 1$) are not defined, so the second inequality of (16) ((17)) is irrelevant, while for $p_{11} = 0$ ($p_{21} = 0$) all ϕ_{1j} (ϕ_{2j}) are less than γ (δ), so the first inequality of (16) ((17)) is irrelevant. In (8), (9) an intercept term appears when $\phi_{1j} = 1/2$, $\phi_{2j} = 1/2$, respectively, while integer powers are also possible, but we allow for the ϕ_{ij} to be any real values satisfying (14)-(17).

The convention that powers of t be denoted $\phi_{ij} - 1/2$ rather than ϕ_{ij} is to enable convenient comparison with integration orders, as indicated by our definitions of $I(d)$ and $J(d)$ sequences. It is possible that one or more of the μ_{ij} are actually zero, though we do not know this, and we define

$$j^\dagger = \min \{j : \mu_{1j} \neq 0\}, \quad j^\ddagger = \min \{j : \mu_{2j} \neq 0\}. \quad (18)$$

We allow for this possibility because we wish (8), (9) to nest the working model used in estimation, and it is possible that one or more regressors $t^{\phi_{ij}-1/2}$ will be included in the latter whose coefficient is zero. For brevity write $\phi_{1\ddagger} = \phi_{1j\ddagger}$, $\phi_{2\ddagger} = \phi_{2j\ddagger}$, $\mu_{1\ddagger} = \mu_{1j\ddagger}$, $\mu_{2\ddagger} = \mu_{2j\ddagger}$.

The truncations in (8), (9) imply that $x_t = y_t = 0$, $t \leq 0$, and that x_t , y_t $t \geq 1$, have finite variance. Without truncation they would not be well-defined in mean square, since $\delta > 1/2$, while $\Delta^{-\gamma}u_{1t}$ is not well-defined in mean square when $\gamma \geq 1/2$. In particular, x_t , and thus y_t , have variance that increases with t (like $t^{2\delta-1}$). Note that the elements of $u_t^\#$ are $I(0)$ processes, while $u_{1t}(-\gamma)$ is an $I(\gamma)$ process and $u_{2t}(-\delta)$ is an $I(\delta)$ process.

Define, for $i = 1, 2$, the $p_{i1} \times 1$ vector sequences

$$g_i(t) = (t^{\phi_{i1}-1/2}, \dots, t^{\phi_{ip_{i1}}-1/2})', \quad t \geq 1, \quad (19)$$

$$= (0, \dots, 0)', \quad t \leq 0, \quad (20)$$

where the prime denotes transposition, and denote by μ_i , $i = 1, 2$, the $p_{i1} \times 1$ vectors whose j th elements are respectively μ_{ij} , $i = 1, 2$. In view of (18), (19) note that $\mu_1' g_1(t)$ is a $J(\phi_{1\ddagger})$ sequence and $\mu_2' g_2(t)$ is a $J(\phi_{2\ddagger})$ sequence.

The working models that are estimated are as follows.

Working Model I This is

$$y_t = \nu x_t + v_t, \quad (21)$$

where ν is estimated by non-intercept OLS, as if the errors v_t were serially uncorrelated random variables orthogonal to x_t . The main issues here are the effect of the misspecification error caused by neglecting the deterministic component in (8) and simultaneous equation bias due to (9).

Working Model II This is

$$y_t = \nu x_t + \mu_1' g_1(t) + v_t, \quad (22)$$

where ν and μ_1 are simultaneously estimated by multiple OLS, again as if the errors v_t were serially uncorrelated random variables orthogonal to x_t . Here, one issue is the robustness to possible omission of the component $s_{1t} = \sum_{j=p_{11}+1}^{p_1} \mu_{1j} t^{\phi_{1j}-1/2}$. Indeed the μ_{1j} , $p_{11} + 1 \leq j \leq p_1$, are not consistently estimable. Another is the simultaneous equations bias due to (9) and error autocorrelation due to (8).

Working Model III This is

$$y_t = \nu x_t + \mu_1' g_1(t) + v_{1t}, \quad (23)$$

$$x_t = \mu_2' g_2(t) + v_{2t}, \quad (24)$$

where ν , μ_1 and μ_2 are simultaneously estimated by a form of generalized least squares (GLS) as if $v_{1t} = u_{1t}(-\gamma)$, $v_{2t} = u_{2t}(-\delta)$, properly accounting for both autocorrelation and simultaneity, and with either γ and/or δ assumed known or suitably estimated in a side calculation. For our discussion of GLS we assume also that

$$\beta > \frac{1}{2}, \quad (25)$$

where $\beta = \delta - \gamma$. The requirement (25) includes the usual case of $I(0)$ cointegrating errors and $I(1)$ x_t ; the case $\beta \leq \frac{1}{2}$ leads to quite different asymptotics, see e.g. Jeganathan (1999), Hualde and Robinson (2001). The GLS method is an extension of that of RH, in which $\mu_1 = 0$, $\mu_2 = 0$ was correctly assumed *a priori*. Comparison with (8), (9) indicates that in fact $v_{1t} = s_{1t} + u_{1t}(-\gamma)$, $v_{2t} = s_{2t} + u_{2t}(-\delta)$, where $s_{2t} = \sum_{j=p_{21}+1}^{p_2} \mu_{2j} t^{\phi_{2j}-1/2}$, so that we have to show that ignoring s_{1t}, s_{2t} has no asymptotic effect; none of the μ_{ij} in these is consistently estimable. Thus robustness of RH's estimates to omission of such s_{it} will be implied.

We could replace the deterministic terms in (8), (9) by more general ones, and in particular could specify s_{1t} and s_{2t} in terms of bounds, rather than precise quantities, but we prefer the simple set-up afforded by (8), (9). While we allow for non-integer powers $\phi_{ij} - 1/2$, this is not on the basis of arguing that these are likely to be of great practical value (though they may turn out to provide improved approximations to some data), but rather because it affords a precise treatment of the competition between stochastic and deterministic trends. Indeed, though we mostly regard γ and δ as unknown, the ϕ_{ij} in $g_1(t)$ and $g_2(t)$, when these are included, are assumed known; there are difficulties with asymptotic theory for, say, OLS estimation of the ϕ_{ij} due to lack of uniform convergence of the objective function.

In the following three sections we discuss rates of convergence (if any) and asymptotic distributional properties of the estimates of working models I, II and III respectively. Proof details are left to appendices. Section 5 contains a small Monte Carlo study of finite sample performance, and Section 6 an empirical application to testing the PPP hypothesis on the basis of data for three US cities.

2 LEAST SQUARES ESTIMATES OF MODEL I

The OLS estimate of ν in working Model I (21) is

$$\bar{\nu} = M_{xy}/M_{xx}, \quad (26)$$

where for any column vector or scalar sequences a_t, b_t , $M_{ab} = \sum_{t=1}^n a_t b_t'$, and, throughout the paper, \sum_t denotes $\sum_{t=1}^n$. The estimate $\bar{\nu}$ has the advantage

of not requiring knowledge of δ and/or β . Under (8), (9) with $\mu_{ij} = 0$ for all i, j , RM showed under mild additional conditions that $\bar{\nu}$ is $n^{\min(2\delta-1, \beta)}$ -consistent (except when both $\delta > \beta$ and $2\delta - \beta = 1$ hold, in which case it is $(n^\beta / \log n)$ -consistent).

Denote by I_m the m -rowed identity matrix. We introduce

Assumption 1. *The process $u_t, t = 0, \pm 1, \dots$, has representation*

$$u_t = A(L) \varepsilon_t, \quad (27)$$

where

$$A(s) = I_2 + \sum_{j=1}^{\infty} A_j s^j, \quad (28)$$

and the A_j are 2×2 matrices such that :

$$(i) \quad \det \{A(s)\} \neq 0, \quad |s| = 1; \quad (29)$$

(ii) $A(e^{i\lambda})$ is differentiable in λ with derivative in $Lip(\eta)$, $\eta > \frac{1}{2}$;

and in addition, with $\|\cdot\|$ denoting the Euclidean norm:

(iii) the ε_t are independent and identically distributed vectors with mean zero, positive definite covariance matrix Ω , and $E \|\varepsilon_t\|^q < \infty$, $q \geq 4$, $q > 2/(2\delta - 1)$.

This is precisely Assumption 1 of RH, where it is discussed; it is a fairly general short memory specification for u_t , allowing weak convergence of fractional transforms, or central limit theorems, as required. In connection with this, denote by $W(r)$ the 2×1 vector Brownian motion with covariance matrix Ω , and define, for $d > 0$, (Type II-see Marinucci and Robinson, 1999b) fractional Brownian motion

$$W(r; d) = \int_0^r \frac{(r-s)^{d-1}}{\Gamma(d)} dW(s), \quad (30)$$

and then define $\bar{W}(r; d)$ and $\widetilde{W}(r; d)$ to respectively be the first and second elements of the vector $A(1)W(r; d)$.

Define

$$\Phi_1 = \int_0^1 \widetilde{W}(r; \delta)^2 dr, \quad (31)$$

$$\Phi_2 = \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} \widetilde{W}(r; \delta) dr, \quad (32)$$

$$\Phi_3 = \mu_{2\ddagger}^2 (2\phi_{2\ddagger})^{-1}, \quad (33)$$

and

$$\Psi_1 = \int_{-\pi}^{\pi} (1 - e^{i\lambda})^{-\gamma} (1 - e^{-i\lambda})^{-\delta} f_{12}(\lambda) d\lambda, \quad (34)$$

$$\Psi_2 = f_{12}(0) \sin \delta \pi, \quad (35)$$

where $f_{12}(\lambda)$ is the (1,2)th element of $f(\lambda)$, Ψ_3 is such that for $\gamma > 1/2$

$$\Psi_3 = \int_0^1 \overline{W}(r; \gamma) \widetilde{W}(r; \delta) dr, \quad (36)$$

and for $\gamma \leq 1/2$ Ψ_3 is an $O_p(1)$ random variable and

$$\Psi_4 = \begin{cases} \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} \overline{W}(r; \gamma) dr & \text{if } \gamma > 0, \\ \mu_{2\ddagger} \int_0^1 r^{\phi_{2\ddagger}-1/2} d\overline{W}(r) & \text{if } \gamma = 0, \end{cases} \quad (37)$$

$$\Psi_5 = \mu_{1\ddagger} \int_0^1 r^{\phi_{1\ddagger}-1/2} \widetilde{W}(r; \delta) dr, \quad (38)$$

$$\Psi_6 = \frac{\mu_{1\ddagger} \mu_{2\ddagger}}{\phi_{1\ddagger} + \phi_{2\ddagger}}. \quad (39)$$

Also define

$$\Phi_1^* = \Phi_1 1(\delta \geq \phi_{2\ddagger}), \quad (40)$$

$$\Phi_2^* = \Phi_2 1(\delta = \phi_{2\ddagger}), \quad (41)$$

$$\Phi_3^* = \Phi_3 1(\delta \leq \phi_{2\ddagger}), \quad (42)$$

and

$$\Psi_1^* = \Psi_1 1(\gamma + \delta < 1, \gamma + \phi_{2\ddagger} \leq 1, \phi_{1\ddagger} + \delta \leq 1, \phi_{1\ddagger} + \phi_{2\ddagger} \leq 1), \quad (43)$$

$$\Psi_2^* = \Psi_2 1(\gamma + \delta = 1, \gamma \geq \phi_{1\ddagger}, \phi_{2\ddagger} \leq \delta < 1), \quad (44)$$

$$\Psi_3^* = \Psi_3 1(\gamma + \delta > 1, \gamma \geq \phi_{1\ddagger}, \delta \geq \phi_{2\ddagger}), \quad (45)$$

$$\begin{aligned} \Psi_4^* &= \Psi_4 1(\gamma + \delta > 1, \phi_{1\ddagger} \leq \gamma, \delta \leq \phi_{2\ddagger}) \\ &\quad + \Psi_4 1(\gamma + \delta = 1 \text{ and } \gamma > 0, \phi_{1\ddagger} \leq \gamma, \delta < \phi_{2\ddagger}) \\ &\quad + \Psi_4 1(\gamma + \delta < 1, \phi_{1\ddagger} \leq \gamma, 1 \leq \phi_{2\ddagger} + \gamma), \end{aligned} \quad (46)$$

$$\begin{aligned} \Psi_5^* &= \Psi_5 1(\gamma + \delta > 1 \text{ or } \gamma = 0 \text{ and } \delta = 1, \phi_{1\ddagger} \geq \gamma, \phi_{2\ddagger} \leq \delta) \\ &\quad + \Psi_5 1(\gamma + \delta = 1 \text{ and } \gamma > 0, \gamma < \phi_{1\ddagger}, \phi_{2\ddagger} \leq \delta) \\ &\quad + \Psi_5 1(\gamma + \delta < 1, \phi_{2\ddagger} \leq \delta, 1 \leq \phi_{1\ddagger} + \delta), \end{aligned} \quad (47)$$

$$\begin{aligned} \Psi_6^* &= \Psi_6 1(\gamma + \delta > 1 \text{ or } \gamma = 0 \text{ and } \delta = 1, \phi_{1\ddagger} \geq \gamma, \phi_{2\ddagger} \geq \delta) \\ &\quad + \Psi_6 1(\gamma + \delta = 1 \text{ and } \gamma > 0, \phi_{1\ddagger} \geq \gamma, \phi_{2\ddagger} \geq \delta, \phi_{1\ddagger} + \phi_{2\ddagger} > 1) \\ &\quad + \Psi_6 1(\gamma + \delta < 1, \phi_{1\ddagger} \geq \gamma, \phi_{2\ddagger} \geq \delta, \phi_{1\ddagger} + \phi_{2\ddagger} \geq 1). \end{aligned} \quad (48)$$

Introduce the sequences

$$k_n = n^{\max(\delta, \phi_{2\ddagger})}, \quad (49)$$

$$\ell_n = n^{\max(1, \gamma + \delta, \gamma + \phi_{2\ddagger}, \delta + \phi_{1\ddagger}, \phi_{1\ddagger} + \phi_{2\ddagger})} + n \log n 1(\gamma + \delta = 1, \gamma > 0). \quad (50)$$

Theorem 1. *Let (8), (9), (12), (13) and Assumption 1 hold. Then as $n \rightarrow \infty$,*

$$(k_n^2/\ell_n)(\bar{\nu} - \nu) \Rightarrow \{\Psi_1^* + \Psi_2^* + \Psi_3^* + \Psi_4^* + \Psi_5^* + \Psi_6^*\} / \{\Phi_1^* + 2\Phi_2^* + \Phi_3^*\}, \quad (51)$$

where by “ \Rightarrow ” we mean convergence in the Skorohod J_1 topology.

We can deduce from Theorem 1 exact rates of convergence, if any, of $\bar{\nu}$ to ν . Let $O_e(\cdot)$ denote exact rate of convergence in probability. We have:

$$\bar{\nu} = \nu + O_e(n^{1-2\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma + \delta < 1 \text{ and } \delta + \phi_{1\ddagger} \leq 1 \quad (52)$$

$$\bar{\nu} = \nu + O_e(n^{1-2\delta} \log n), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma + \delta = 1 \text{ and } \gamma \geq \phi_{1\ddagger} \quad (53)$$

$$\bar{\nu} = \nu + O_e(n^{\gamma-\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma \geq \phi_{1\ddagger} \text{ and } \gamma + \delta > 1 \quad (54)$$

$$\bar{\nu} = \nu + O_e(n^{\phi_{1\ddagger}-\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger}, \phi_{1\ddagger} \geq \gamma \text{ and } \gamma + \delta > 1 \quad (55)$$

$$\text{or } \delta \geq \phi_{2\ddagger}, \phi_{1\ddagger} + \delta \geq 1 \text{ and } \gamma + \delta < 1$$

$$\bar{\nu} = \nu + O_e(n^{1-2\phi_{2\ddagger}}), \quad \text{if } \phi_{2\ddagger} \geq \delta, \gamma + \phi_{2\ddagger} < 1 \text{ and } \phi_{1\ddagger} + \phi_{2\ddagger} \leq 1 \quad (56)$$

$$\bar{\nu} = \nu + O_e(n^{\gamma-\phi_{2\ddagger}}), \quad \text{if } \phi_{2\ddagger} \geq \delta, \gamma \geq \phi_{1\ddagger} \text{ and } \gamma + \phi_{2\ddagger} > 1 \quad (57)$$

$$\bar{\nu} = \nu + O_e(n^{\phi_{1\ddagger}-\phi_{2\ddagger}}), \quad \text{if } \phi_{2\ddagger} \geq \delta, \phi_{1\ddagger} \geq \gamma \text{ and } \phi_{1\ddagger} + \phi_{2\ddagger} > 1. \quad (58)$$

It follows that $\bar{\nu}$ is not even consistent when

$$\phi_{1\ddagger} \geq \max(\delta, \phi_{2\ddagger}), \quad (59)$$

(see cases (55) and (58)) so that the deterministic trend in the cointegrating equation (8) dominates both the stochastic and deterministic trends in x_t . Otherwise, $\bar{\nu}$ is consistent. In case (55), with $\gamma < \phi_{1\ddagger} < \delta$, the deterministic trend in (8) worsens matters relative to the rates in RM, whereas in cases (52)-(54) the same rates are achieved, stochastic trends dominating. With $\phi_{2\ddagger} > \delta$, the dominating deterministic trend in x_t improves rates relative to RM, in (56) and (57), while in (58) both deterministic trends dominate and the rate simply reflects the extent to which x_t 's trend dominates that in (8).

We can deduce from Theorem 1 more precise results. These are complicated in case of equalities $\phi_{2\ddagger} = \delta$ and/or $\phi_{1\ddagger} = \gamma$, so we look only at strict inequalities. With $\gamma > \phi_{1\ddagger}$, $\delta > \phi_{2\ddagger}$ the limit distributions corresponding to (52), (53) and (54) are identical to those of RM Propositions 6.1, 6.2 and 6.5. In case (57), with $\gamma > \phi_{1\ddagger}$ and $\phi_{2\ddagger} > \delta$, stochastic trends dominate in (8) and deterministic ones in (9), and $n^{\phi_{2\ddagger}-\gamma}(\bar{\nu} - \nu)$ converges to the normal variates $\Phi_3^{-1}\Psi_4$ (cf. Robinson and Marinucci, 2000) while in (58) with $\phi_{1\ddagger} > \gamma$, $\phi_{2\ddagger} > \delta$, $n^{\phi_{2\ddagger}-\phi_{1\ddagger}}(\bar{\nu} - \nu)$ converges to the constant $\Phi_3^{-1}\Psi_6$. The familiar case in which x_t contains a unit root plus linear trend, and cointegrating errors that are $I(0)$ ($\gamma = 0$, $\delta = 1$, $\phi_{1\ddagger} = 0.5$, $\phi_{2\ddagger} = 1.5$) comes under (58).

3 LEAST SQUARES ESTIMATES OF MODEL II

The OLS estimate of $\nu_+ = (\nu, \mu_1)'$ in Model II (22) is

$$\tilde{\nu}_+ = M_{x_+x_+}^{-1} M_{x_+y}, \quad (60)$$

where $x_{+t} = (x_t, g_1'(t))'$. Define the sequence

$$m_n = \left\{ n^{\max(1, \gamma + \delta, \gamma + \phi_{2\ddagger})} + n \log n 1(\gamma + \delta = 1, \gamma \neq 0) \right\}, \quad (61)$$

and the matrix sequences

$$B_n = \begin{bmatrix} k_n & 0 \\ 0 & D_{1n}(0) \end{bmatrix}, \quad C_n = \begin{bmatrix} m_n/k_n & 0 \\ 0 & n^\gamma I_{p_{11}} \end{bmatrix}, \quad (62)$$

$$D_{in}(d) = \text{diag} \left\{ n^{\phi_{i1}+d}, \dots, n^{\phi_{ip_{i1}}+d} \right\}, i = 1, 2. \quad (63)$$

For $i, j = 1, 2$, define the $p_{i1} \times p_{j1}$ matrix $\Xi_{ij}(c, d)$, having (k, l) th element $(\phi_{ik} + \phi_{jl} - c - d)^{-1}$, and the $p_{11} \times 1$ vector Υ , having k th element $(\phi_{1k} + \phi_{2\ddagger})^{-1}$. Also define

$$\Phi_4 = \int_0^1 g_1(r) \widetilde{W}(r; \delta) dr, \quad (64)$$

$$\Phi_5 = \mu_{2\ddagger} \Upsilon, \quad (65)$$

$$\Phi_6 = \Xi_{11}(0, 0); \quad (66)$$

$$\Psi_7 = \int_0^1 \widetilde{W}(r; 1) d\overline{W}(r) + \sum_{j=-\infty}^0 \text{Cov}(u_{10}, u_{2j}), \quad \delta = 1, \quad (67)$$

$$= \int_0^1 \widetilde{W}(r; \delta) d\overline{W}(r), \quad \delta > 1; \quad (68)$$

$$\Psi_8 = \begin{cases} \int_0^1 g_1(r) \overline{W}(r; \gamma) dr & \text{if } \gamma > 0 \\ \int_0^1 g_1(r) d\overline{W}(r) & \text{if } \gamma = 0. \end{cases} \quad (69)$$

Let

$$\Phi_4^* = \Phi_4 1(\delta \geq \phi_{2\ddagger}), \quad (70)$$

$$\Phi_5^* = \Phi_5 1(\delta \leq \phi_{2\ddagger}), \quad (71)$$

and

$$\Phi = \begin{pmatrix} \Phi_1^* + \Phi_2^* + \Phi_3^* & \Phi_4^{*'} + \Phi_5^{*'} \\ \Phi_4^* + \Phi_5^* & \Phi_6 \end{pmatrix}, \quad (72)$$

and also

$$\Psi_1^{**} = \Psi_1 1(\gamma + \delta < 1, \gamma + \phi_{2\ddagger} \leq 1), \quad (73)$$

$$\Psi_2^{**} = \Psi_2 1(\gamma + \delta = 1, \phi_{2\ddagger} \leq \delta < 1), \quad (74)$$

$$\Psi_3^{**} = \Psi_3 1(\gamma + \delta > 1, \delta \geq \phi_{2\ddagger}, \gamma > 0), \quad (75)$$

$$\begin{aligned} \Psi_4^{**} &= \Psi_4 1(\gamma + \delta > 1, \delta \leq \phi_{2\ddagger}) \\ &\quad + \Psi_4 1(\gamma + \delta = 1 \text{ and } \gamma > 0, \delta < \phi_{2\ddagger}) \\ &\quad + \Psi_4 1(\gamma + \delta < 1 \text{ or } \gamma = 0 \text{ and } \delta = 1, \phi_{2\ddagger} + \gamma \geq 1), \end{aligned} \quad (76)$$

$$\Psi_7^{**} = \Psi_7 1(\gamma = 0, \delta \geq \max(\phi_{2\ddagger}, 1)), \quad (77)$$

and

$$\Psi = \begin{pmatrix} \Psi_1^{**} + \Psi_2^{**} + \Psi_3^{**} + \Psi_4^{**} + \Psi_7^{**} \\ \Psi_8 \end{pmatrix}. \quad (78)$$

To avoid multicollinearity in case x_t is dominated by an element of $g_2(t)$ we introduce

Assumption 2. *If $\delta < \phi_{2\ddagger}$ then $\phi_{2\ddagger} \neq \phi_{1k}$ for $k = 1, \dots, p_{11}$.*

Theorem 2. *Let (8), (9), (12), (13) and Assumptions 1 and 2 hold. Then as $n \rightarrow \infty$,*

$$C_n^{-1} \Phi B_n (\tilde{\nu}_+ - \nu_+) \Rightarrow \Psi.$$

The cumbersome norming (where indeed Φ can be stochastic) is to enable a neat, general statement, C_n and B_n not commuting with the non-diagonal matrix Φ , but we can readily deduce more comprehensible conclusions. Due to the sufficient accounting for deterministic trends in (8), $\tilde{\nu}$ is always consistent. We have the following cases, that we classify according to whether the stochastic or deterministic component of x_t dominates:

$$\begin{aligned} S &: \delta \geq \phi_{2\ddagger}; \\ D &: \delta \leq \phi_{2\ddagger}. \end{aligned}$$

S :

$$\tilde{\nu} - \nu = O_e(n^{1-2\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma + \delta < 1, \quad (79)$$

$$\tilde{\nu} - \nu = O_e(n^{1-2\delta} \log n), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma + \delta = 1, \gamma > 0, \quad (80)$$

$$\tilde{\nu} - \nu = O_e(n^{-1}), \quad \text{if } \gamma = 1 - \delta = 0, \phi_{2\ddagger} \leq 1, \quad (81)$$

$$\tilde{\nu} - \nu = O_e(n^{\gamma-\delta}), \quad \text{if } \delta \geq \phi_{2\ddagger}, \gamma + \delta > 1, \quad (82)$$

D :

$$\tilde{\nu} - \nu = O_e(n^{1-2\phi_{2\ddagger}}), \quad \text{if } \delta < \phi_{2\ddagger}, \gamma + \phi_{2\ddagger} \leq 1, \quad (83)$$

$$\tilde{\nu} - \nu = O_e(n^{-1}), \quad \text{if } \delta < \phi_{2\ddagger} = 1, \gamma = 0, \quad (84)$$

$$\tilde{\nu} - \nu = O_e(n^{\gamma-\phi_{2\ddagger}}), \quad \text{if } \delta \leq \phi_{2\ddagger}, \gamma + \phi_{2\ddagger} > 1. \quad (85)$$

When $\delta > \phi_{2\ddagger}$ the limit distributions corresponding to (79)-(82) are those of RM (see Propositions 6.1, 6.2, 6.3, 6.4 and 6.5). When $\phi_{2\ddagger} > \max(\delta, 1 - \gamma)$ we deduce that

$$\begin{bmatrix} n^{\phi_{2\ddagger} - \gamma} & 0 \\ 0 & D_{1n}(-\gamma) \end{bmatrix} (\tilde{\nu}_+ - \nu_+) \rightarrow_d \begin{bmatrix} \Phi_3 & \Phi'_5 \\ \Phi_5 & \Phi_6 \end{bmatrix}^{-1} \begin{bmatrix} \Psi_4 \\ \Psi_8 \end{bmatrix}, \quad (86)$$

where the right side is a multivariate normal vector. The familiar case $\gamma = 0$, $\delta = 1$, $\phi_{1\ddagger} = 0.5$, $\phi_{2\ddagger} = 1.5$ comes under (85).

4 GENERALIZED LEAST SQUARES ESTIMATES OF MODEL III

The estimates of Sections 2 and 3 are computationally convenient, especially as they both avoid the necessity of knowledge of memory parameters and do not require their estimation. However, even when they converge, their rates are not in general optimal, and their limit distributions are for the most part inconvenient for practical use.

To remedy these drawbacks we consider GLS estimation. For $c, d \geq 0$, define

$$z_t(c, d) = (y_t(c), x_t(d))', \quad (87)$$

$$w_t(c, d) = \begin{pmatrix} \Delta^c x_t & \Delta^c g'_1(t) & 0 \\ 0 & 0 & \Delta^d g'_2(t) \end{pmatrix}', \quad (88)$$

$$\nu_{++} = (\nu, \mu'_1, \mu'_2)'. \quad (89)$$

Thus (23), (24) can be written

$$z_t(\gamma, \delta) = w'_t(\gamma, \delta)\nu_{++} + v_t, \quad (90)$$

where now $v_t = (v_{1t}(\gamma), v_{2t}(\delta))'$. In RH the right hand side was simplified by the correct assumption that $\mu_1 = 0$, $\mu_2 = 0$ and $v_t = (u_{1t}, u_{2t})'$. Here we simultaneously estimate ν with μ_1 , μ_2 and show that s_{1t}, s_{2t} (defined after (25) in Section 1) exert negligible effect. As in RH we construct two kinds of estimate of GLS type that allow for flexible parametric modelling of $f(\lambda)$, that is, the autocorrelation structure of u_t , and that either depend on known γ, δ or allow substitution of estimates of these without affecting limiting distributional properties. One kind of estimate is 'time-domain', the other is 'frequency-domain', and the practitioner's choice between them is based on computational considerations and taste.

The time-domain estimate involves autoregressive (AR) transformation. From Assumption 1 u_t has an AR representation

$$B(L)u_t = \varepsilon_t, \quad (91)$$

with $B(s) = I_2 - \sum_{j=1}^{\infty} B_j s^j$, such that the B_j are unknown 2×2 matrices. We know functions $\Omega(h)$, $B_j(h)$, where $h \in \mathbb{R}^p$, $p \geq 1$, such that $\Omega = \Omega(\theta)$, $B_j \equiv B_j(\theta)$ for some $\theta \in \mathbb{R}$. Define $B(s; h) = I_2 - \sum_{j=1}^{\infty} B_j(h) s^j$ and then

$$\tilde{a}(c, d, h) = \sum_t \{B(L; h) w'_t(c, d)\}' \Omega(h)^{-1} \{B(L; h) z_t(c, d)\}, \quad (92)$$

$$\tilde{b}(c, d, h) = \sum_t \{B(L; h) w'_t(c, d)\}' \Omega(h)^{-1} \{B(L; h) w'_t(c, d)\}. \quad (93)$$

Each of the AR transformations is truncated since $w_t(c, d) = 0$, $z_t(c, d) = 0$, $t \leq 0$. Now write

$$\tilde{\nu}_{++}(c, d, h) = \tilde{b}(c, d, h)^{-1} \tilde{a}(c, d, h) \quad (94)$$

and consider

$$\tilde{\nu}_{++}(\gamma, \delta, \theta), \tilde{\nu}_{++}(\gamma, \delta, \hat{\theta}), \tilde{\nu}_{++}(\hat{\gamma}, \delta, \hat{\theta}), \tilde{\nu}_{++}(\gamma, \hat{\delta}, \hat{\theta}), \tilde{\nu}_{++}(\hat{\gamma}, \hat{\delta}, \hat{\theta}), \quad (95)$$

for given estimates $\hat{\gamma}$, $\hat{\delta}$, $\hat{\theta}$. These estimates of ν_{++} respectively cover the cases in which γ , δ and θ are all known, γ and δ are known but θ is not, only $\hat{\delta}$ is known, only γ is known, and γ , δ and θ is all unknown. Thus $\tilde{\nu}_{++}(\gamma, \delta, \hat{\theta})$ covers the familiar case where $\gamma = 0$, $\delta = 1$ is known, and u_t is, say, white noise or AR(1); $\tilde{\nu}_{++}(\hat{\gamma}, \delta, \hat{\theta})$ with $\delta = 1$ accepts the evidence of unit root behaviour suggested by a number of empirical studies of macroeconomic variables but allows for the possibility of long memory or mean-reversion in the cointegrating error; $\tilde{\nu}_{++}(0, \hat{\delta}, \hat{\theta})$ insists only that the cointegrating error has short memory, possibly white noise; and $\tilde{\nu}_{++}(\hat{\gamma}, \hat{\delta}, \hat{\theta})$ is completely agnostic.

When u_t is not a finite-degree AR process $\tilde{\nu}_{++}$ can still be computed, but the following frequency-domain estimates may be preferable, making use of the neat form of the spectral density matrix $f(\lambda)$ when u_t is a finite-degree moving average (MA) or autoregressive moving average (ARMA) sequence, or in some other models. Let $f(\lambda; h)$ be a known function of $\lambda \in (-\pi, \pi]$ and $h \in \mathbb{R}^p$, such that $f(\lambda; \theta) = f(\lambda)$, so

$$f(\lambda; h) = (2\pi)^{-1} B(e^{i\lambda}; h)^{-1} \Omega(h) B(e^{-i\lambda}; h)^{-1'}. \quad (96)$$

Define the discrete Fourier transforms

$$F_{w(c,d)}(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t w_t(c, d) e^{it\lambda}, \quad F_{z(c,d)}(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t z_t(c, d) e^{it\lambda}. \quad (97)$$

Put

$$a(c, d, h) = \sum_j F_{w(c,d)}(\lambda_j) f(\lambda_j; h)^{-1} F_{z(c,d)}(-\lambda_j), \quad (98)$$

$$b(c, d, h) = \sum_j F_{w(c,d)}(\lambda_j) f(\lambda_j; h)^{-1} F'_{w(c,d)}(-\lambda_j). \quad (99)$$

Define

$$\hat{\nu}_{++}(c, d, h) = b(c, d, h)^{-1} a(c, d, h). \quad (100)$$

Consider the frequency-domain variants of the five estimates (95),

$$\widehat{v}_{++}(\gamma, \delta, \theta), \widehat{v}_{++}(\gamma, \delta, \widehat{\theta}), \widehat{v}_{++}(\widehat{\gamma}, \delta, \widehat{\theta}), \widehat{v}_{++}(\gamma, \widehat{\delta}, \widehat{\theta}), \widehat{v}_{++}(\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}). \quad (101)$$

When $v_t = (v_{1t}, v_{2t})'$ is *a priori* white noise, $\widetilde{v}(c, d, h) \equiv \widehat{v}(c, d, h)$.

To handle the last four estimates in (95) and (101) we introduce the following further assumptions. Denote by Θ the compact set of all admissible values of $\widehat{\theta}$.

Assumption 3.

- (i) $f(\lambda; \theta) = f(\lambda)$.
- (ii) $f(\lambda; h)$ has determinant bounded away from zero on $([-\pi, \pi] \times \Theta)$.
- (iii) $f(\lambda; h)$ is boundedly differentiable in h on $([-\pi, \pi] \times \Theta)$, with derivative that is continuous in h at $h = \theta$ for all λ .
- (iv) $f(\lambda; \theta)$ is differentiable in λ , with derivative satisfying a Lipschitz condition of order greater than $1/2$ in λ .
- (v) $(\partial/\partial h)f(\lambda; h)$ is differentiable in λ at $h = \theta$, with derivative satisfying a Lipschitz condition of order greater than $1/2$ in λ .

Assumption 4.

- (i) There exists $K < \infty$ such that

$$|\widehat{\gamma}| + |\widehat{\delta}| \leq K, \quad (102)$$

and $\rho > \max(0, 1 - \beta)$ such that

$$\widehat{\gamma} = \gamma + O_p(n^{-\rho}), \quad \widehat{\delta} = \delta + O_p(n^{-\rho}). \quad (103)$$

- (ii)

$$\widehat{\theta} = \theta + O_p(n^{-\frac{1}{2}}), \quad \text{where } \theta \in \Theta. \quad (104)$$

These are identical to assumptions in RH, where they are discussed. However in view of the presence of deterministic trends in our model we need to say more about Assumption 4. In general it seems that if we carry out the procedure of Velasco and Robinson (2000) to estimate δ , but based on residuals from the OLS regression of x_t on $g_2(t)$, we will achieve the second part of (103) with $\rho = \frac{1}{2}$. As is often the case with justifying insertion of residuals in implicitly-defined estimation procedures, the proof is rather tedious so the details are not pursued. Alternatively, if all $\phi_{2j} - \frac{1}{2}$ in $g_2(t)$ are integers, the

use of Velasco and Robinson's (2000) procedure based on the raw x_t , but using a Kolmogorov taper of sufficiently high order, will exactly remove such a polynomial trend. Strictly, this does not require knowledge of ϕ_{21} , but rather of the largest value we might anticipate for ϕ_{21} . On the other hand, we might carry out the initial OLS estimation by including such arbitrarily large powers of t , so there may not be a great deal to choose between the two approaches. The higher the taper order, the greater the imprecision in estimation of δ , while inclusion of unnecessary regressors in the OLS approach is liable to have similar effect. The OLS approach has the advantage over tapering that non-integral powers of t may be employed, while tapering seems to be needed anyway to estimate large enough values of δ in the Velasco and Robinson (2000) approach. So far as the first part of (103) is concerned, Velasco and Robinson's (2000) approach can again be modified to estimate γ , with the use of residuals from the regression of y_t on x_t and $g_2(t)$ apparently necessary, while tapering is then unnecessary if $\gamma < \frac{1}{2}$ is anticipated. Recently, Chen and Hurvich (2003) have employed tapers in estimating the cointegrating coefficient in a fractional system with deterministic trends. However, they are concerned with estimating the relation between underlying (possibly nonstationary) stochastic components that are corrupted by deterministic trends. These (assumed to be polynomial in t) are handled by differencing, with tapering then employed in a modification of the narrow-band least squares regression estimate of Robinson (1994). Earlier, Robinson (1986) had employed tapering in reducing errors-in-variables bias in band-spectrum regression in which (contrary to Robinson (1994)) bands do not degenerate asymptotically and the processes have short memory.

Define

$$\Lambda_i(d) = \text{diag} \left\{ \frac{\Gamma(\phi_{i1} + 1/2)}{\Gamma(\phi_{i1} - d + 1/2)}, \dots, \frac{\Gamma(\phi_{ip_{i1}} + 1/2)}{\Gamma(\phi_{ip_{i1}} - d + 1/2)} \right\}, \quad i = 1, 2, \quad (105)$$

$$Q(r) = \begin{pmatrix} \widetilde{W}(r; \beta) \mathbf{1}(\delta \geq \phi_{2\ddagger}) + \mu_{2\ddagger} \frac{\Gamma(\phi_{2\ddagger} + 1/2)}{\Gamma(\phi_{2\ddagger} - \gamma + 1/2)} r^{\phi_{2\ddagger} - \gamma - 1/2} \mathbf{1}(\delta \leq \phi_{2\ddagger}) & 0 \\ \Lambda_1(\gamma) g_1(r) r^{-\gamma} & 0 \\ 0 & \Lambda_2(\delta) g_2(r) r^{-\delta} \end{pmatrix} \quad (106)$$

and introduce the matrix sequence

$$D_n = \begin{pmatrix} n^{\max(\delta, \phi_{2\ddagger}) - \gamma} & 0 & 0 \\ 0 & D_{1n}(-\gamma) & 0 \\ 0 & 0 & D_{2n}(-\delta) \end{pmatrix}. \quad (107)$$

Theorem 3. *Let (8), (9), (12), (25) and Assumptions 1-4 hold, with $q > 1/(2\beta - 1)$ in Assumption 1. Then, denoting by $\widehat{\nu}_{++}^*$ any of the estimates in (95) or (101), we have as $n \rightarrow \infty$,*

$$D_n (\widehat{\nu}_{++}^* - \nu_{++}) \Rightarrow \left\{ \int_0^1 Q(r) f(0)^{-1} Q'(r) dr \right\}^{-1} 2\pi \int_0^1 Q(r) B(1)' \Omega^{-1} dW(r). \quad (108)$$

As in RH we find that we can estimate ν , along with μ_1, μ_2 , as well without knowing γ and/or δ and/or θ as knowing them, so that efficiency of estimation of γ, δ and θ does not matter. When $\delta > \phi_{2\ddagger}$, we have precisely the same limit distribution and rate of convergence for $\hat{\nu}^*$ as the estimate of RH, which ignores the possibility of trends, and the same limit distribution. The distribution is changed when $\delta = \phi_{2\ddagger}$, but the rate of convergence is not. When $\delta < \phi_{2\ddagger}$ the rate is faster than in RH. When $\delta > \phi_{2\ddagger}$ there is mixed normal asymptotics as in RH, as there is also for $\delta = \phi_{2\ddagger}$, while when $\delta < \phi_{2\ddagger}$ the limit distribution is normal, as in the familiar case $\gamma = 0, \delta = 1, \phi_{1\ddagger} = 0.5, \phi_{2\ddagger} = 1.5$. Thus in all cases we can expect Wald tests on ν_{++} , for example tests on ν (such as $\nu = 1$ as in PPP testing) or on μ_1 (e.g. $\mu_1 = 0$, to test whether deterministic trends affect y_t only through x_t), to have standard, χ^2 , asymptotics. Our theory differs from that of Hansen (1992) (who assumed known $\gamma = 0, \delta = 1$ integer powers of t) in that he allowed x_t to be a vector and assumed its dimension greater than the number of deterministic regressors. This is impossible in our set-up. As in RH we do not allow x_t to be a vector because an interesting treatment of this case would entail, in our fractional setting, allowance for differing integration orders in the elements of x_t , and also two or more cointegrating relationships, possibly with different integration orders.

In general introduce the null hypothesis:

$$H_0 : E\nu_{++} = e, \quad (109)$$

where E is a given $q \times (1 + p_{11} + p_{21})$ matrix of rank $q < 1 + p_{11} + p_{21}$ and e is a given $q \times 1$ vector.

Corollary 4. Denoting by b^* any of the quantities $\tilde{b}(c, d, h), b(c, d, h)$, with $c = \gamma$ or $\hat{\gamma}$, $d = \delta$ or $\hat{\delta}$ and $h = \theta$ or $\hat{\theta}$, under (109) the Wald statistics

$$(E\hat{\nu}_{++}^* - e)' (Eb^{*-1}E')^{-1} (E\hat{\nu}_{++}^* - e) \rightarrow_d \chi_q^2, \quad \text{as } n \rightarrow \infty. \quad (110)$$

5 MONTE CARLO SIMULATIONS

We have encountered convergence rates and limit distributions that can vary substantially across both types of estimate and memory characteristics of the data generating process. In order to investigate how reliable a guide the asymptotic theory is in moderate-sized samples a small Monte Carlo study was carried out. We generated data from (8) and (9) for several combinations of stochastic and deterministic trends. Two different specifications for each equation were employed. For the cointegrating equation (8) these were:

$$DT1a : \quad p_1 = 1; \quad \mu_{11} = 1; \quad \phi_{11} = 0.5.$$

$$DT1b : \quad p_1 = 1; \quad \mu_{11} = \mu_{12} = 1; \quad \phi_{12} = 0.5, \phi_{11} = 1.5.$$

For the x equation (9) they were

$$DT2a : \quad p_2 = 1; \quad \mu_{21} = 1; \quad \phi_{21} = 1.5.$$

$$DT2b : \quad p_2 = 2; \quad \mu_{21} = \mu_{22} = 1; \quad \phi_{22} = 1.5, \phi_{21} = 2.5.$$

Thus $DT1a$ consists only of an intercept, while $DT1b$ is a time trend; $DT2a$ and $DT2b$ include no constants, the former consisting only of a linear term, the latter a linear and a quadratic. We employed all four combinations, $DTa = DT1a \times DT2a$, $DTb = DT1a \times DT2b$, $DTc = DT1b \times DT2a$ and $DTd = DT1b \times DT2b$. The stochastic component of the model was specified as follows. We took $A_j \equiv 0$, $j \geq 1$, in Assumption 1 where the covariance matrix Ω of Gaussian $u_t = \varepsilon_t$ was given by

$$\Omega = \begin{bmatrix} \tau & \rho\tau^{\frac{1}{2}} \\ \rho\tau^{\frac{1}{2}} & 1 \end{bmatrix} \quad (111)$$

for varying ρ , τ . Our stochastic trends for the cointegrating equation and the x equation were then determined by the following six choices of (γ, δ) :

$$STa: \quad (\gamma, \delta) = (0, 0.6)$$

$$STb: \quad (\gamma, \delta) = (0, 1.2)$$

$$STc: \quad (\gamma, \delta) = (0, 2)$$

$$STd: \quad (\gamma, \delta) = (0.4, 0.6)$$

$$STe: \quad (\gamma, \delta) = (0.4, 1.2)$$

$$STf: \quad (\gamma, \delta) = (0.4, 2)$$

Note that STd is not covered by the theory for our GLS estimate. We considered each of the 24 combinations of the 4 DT 's and 6 ST 's. Finally we took $\nu = 1$.

Tables 1-3 indicate rates of convergence of the various estimates of ν : $\bar{\nu}$, the OLS estimate in Section 2; the first element $\tilde{\nu}$ of the OLS estimate $\tilde{\nu}_+$ in Section 3; the first element $\hat{\nu}$ of the generic GLS estimate $\hat{\nu}_{++}^*$ in Section 4 (note indeed that in our case of white noise AR, (95) and (101) are identical). For comparison we include also the rates when there are no deterministic trends in either (8) or (9). Our design includes cases where such trends improve, leave unchanged, or reduce the rate for $\bar{\nu}$, or even make it inconsistent, and cases where rates for $\tilde{\nu}$ and $\hat{\nu}$ are either unchanged or improved. Nevertheless it would be possible to choose combinations that might seem more "interesting" in view of the various outcomes reported in Sections 2-4. Our choice is motivated by two factors. One is to enable comparison with the design of RH, who used precisely the same ST 's, and values of ρ and τ , with no DT 's, and computed $\bar{\nu}$ and the GLS estimate of Section 4 simplified by (correctly) assuming no deterministic trends. The other is that non-fractional powers of t in the DT 's seem rather typical of current macroeconomic practice. However our simulations fall very far short indeed of providing a comprehensive study, especially as behaviour will vary

with the μ_{ij} as well as the ϕ_{ij} , not to mention being affected by the presence of short memory autocorrelation.

We computed the OLS estimates $\bar{\nu}$ and $\tilde{\nu}$ described above, as well as infeasible and feasible GLS estimates of ν , namely $\hat{\nu}_I = \hat{\nu}(\gamma, \delta, \hat{\theta})$ and $\hat{\nu}_F = \hat{\nu}(\hat{\gamma}, \hat{\delta}, \hat{\theta})$ where $\hat{\nu}(c, d, h)$ is the first element of $\hat{\nu}_{++}(c, d, h)$ with $\hat{\gamma}$, $\hat{\delta}$ and $\hat{\theta}$ as follows.

Having obtained $\tilde{\nu}_+ = (\tilde{\nu}, \tilde{\mu}'_2)'$ we computed the $\tilde{v}_{1t} = y_t - \tilde{\nu}'_+ x_{+t}$ and $\tilde{v}_{2t} = x_t - \tilde{\mu}'_2 g_2(t)$. For given (c, d) define $\tilde{u}_t(c, d) = (\tilde{v}_{1t}(c), \tilde{v}_{2t}(d))'$. Since u_t is white noise, θ parameterizes only Ω , for which we employ the estimates $\tilde{\Omega} = \Omega_{(\gamma, \delta)}$ and $\hat{\Omega} = \Omega_{(\hat{\gamma}, \hat{\delta})}$, where $\Omega_{(c, d)} = n^{-1} \sum_t \tilde{u}_t(c, d) \tilde{u}'_t(c, d)$, the former referring to the case (γ, δ) known (i.e. $\hat{\nu}_I$), the latter to the case (γ, δ) is estimated by $(\hat{\gamma}, \hat{\delta})$ (i.e. $\hat{\nu}_F$). Here $\hat{\gamma}$ was the Whittle pseudo-maximum likelihood estimate of Velasco and Robinson (2000) applied to the series \tilde{v}_{1t} (without tapering the data). Likewise $\hat{\delta}$ was obtained by the same method, but applied to the first difference of \tilde{v}_{2t} , then adding back 1 when $\delta = 1.2$ or $\delta = 2$ (as an alternative to tapering, again as in the simulations of RH).

For the null hypothesis (109) we took $\nu = 1$, and in the Wald statistic (110) we computed W_I based on $\hat{\nu}_I$ and $b(\gamma, \delta, \hat{\theta})$ and W_F based on $\hat{\nu}_F$ and $b(\hat{\gamma}, \hat{\delta}, \hat{\theta})$, $\tilde{\theta}$ and $\hat{\theta}$ respectively denoting the vectors consisting of the three distinct elements of $\tilde{\Omega}$ and $\hat{\Omega}$. We employed sample sizes $n = 64, 125$ and 256 , with 1000 replications.

We present two groups of simulations: in the first we keep $\tau = 1$ and $\rho = 0.5$ fixed, and consider the 24 combinations of deterministic and stochastic trends, while in the second group we focus on $\delta = 0.6$, $\gamma = 0$ and study the effect of alternative combinations of τ and ρ .

In Tables 4 and 5 we present Monte Carlo bias (the difference between the average of the estimates and ν), and sample standard deviation for the first group of simulations. The stochastic component dominates in x_t in 4 of the 24 combinations considered, namely when $\delta = 2$ and $g_2(t) = t$ (for any γ , $g_1(t)$). In these situations the estimates have the same rate of convergence as in RH but the presence of $g_1(t)$ may affect the standard deviations of $\tilde{\nu}$, $\hat{\nu}_I$ and $\hat{\nu}_F$. This is apparently a minor issue when the other regressors have a very low order, as it is for DTa ($g_1(t) = 1$), where the bias, standard deviation and empirical sizes of $\tilde{\nu}$, $\hat{\nu}_I$ and $\hat{\nu}_F$ resemble those in RH, but under DTc , which has a larger order for $g_1(t)$, we observed larger dispersion.

In the remaining 20 situations the estimates are dominated by the deterministic component in x_t , and consistency under DTc fails for $\bar{\nu}$ while $\tilde{\nu}$ and GLS are not covered by the theory presented here. For the other three deterministic components, the faster rates of convergence result in much lower bias and standard deviation compared to the ones in RH. Contrary to the results in RH, moving from $\delta = 0.6$ to $\delta = 1.2$ (and to $\delta = 2$ too, when DTb and DTd , i.e. $g_2(t) = t + t^2$, are considered) does not greatly affect precision. Increasing γ from 0 to 0.4 has a visible, if rather small, effect, and again this is stable regardless of δ , so long as the deterministic component dominates in x_t .

In general the OLS estimates are subject to bias due to correlation between

u_{1t} and u_{2t} , and in $\bar{\nu}$ there is also the additional component due to omission of $g_1(t)$: both should induce a positive bias given the values of μ_1 and of ρ , but we find that this is small. The bias of $\tilde{\nu}$ is even smaller. Finally, we find that $\hat{\nu}_I$ is slightly more precise than $\hat{\nu}_F$, and that both are superior to $\tilde{\nu}$, but in all cases considered the standard deviation is within the same range.

In Table 6 we present empirical sizes, for nominal sizes $\alpha = 0.05$ and $\alpha = 0.10$, of $\hat{\nu}_I$ and $\hat{\nu}_F$. As in RH, those for $\hat{\nu}_I$ are fairly precise, but those for $\hat{\nu}_F$ tend to be too big, though the difference gets smaller the larger the sample is. The difference in rates of convergence does not seem to affect the pattern, which is rather stable among all combinations. It is interesting to observe that under *DTa*, *DTb* and *DTd*, the case $\delta = 0.6$, $\gamma = 0.4$ does not behave differently from the others, contrary to the simulations in RH.

In the second part of the simulation exercise, we consider the effects of alternative specifications of the matrix Ω . Notice that we have $E(u_{2t}) = 1$ and so altering τ affects the dispersion of u_{1t} , inverting the design of RH; due to the dominating deterministic component, changing the variance of u_{2t} would not affect the results in a relevant way. In fact, we already remarked that even increasing δ from 0.6 to 1.2 (and to 2 under *DTd*) for given γ has no major effect on variability, and this will be much more the case if δ is unchanged.

We present bias in Table 7, standard deviation in Table 8 and empirical size in Table 9. We set $\delta = 0.6$, $\gamma = 0$, and ran the simulations for $\rho = 0, .5, .75, -.5$, $\tau = .5, 1$ and 2 and all combinations of the deterministic components, but to facilitate readability we exclude *DTc*, for which Assumption 2 is not met, and $\rho = 0.5$, because the results are not much different than for $\rho = 0.75$. In Table 7 and in Table 9 we also exclude results for $\tau = 2$ and $\tau = 0.5$ because, as RH also found, they do not vary much, while in Table 8 we exclude *DTb* and *DTd* because the deterministic trend is so strong that the standard deviation is too small to indicate any effect.

The correlation ρ is a potential source of bias in OLS, but the effect, presented in Table 7, is minimal, mainly visible in small samples, and almost only for *DTa*, which has the lowest rate of convergence. In the simulations of RH a small fraction of the bias of $\tilde{\nu}$ passed to $\hat{\nu}_F$, but here the preliminary estimate of ν is so precise, due to the faster rate of convergence, that basically no bias is incurred at this stage. Altering τ only affects dispersion: in Table 8 sample variance increases with τ for given ρ and *DT*. The low impact of changes of ρ on precision is also important because it leaves empirical size nearly unaffected: it is not surprising that on average the best approximation to the nominal size is for $\rho = 0$, but even with the rather extreme $\rho = 0.75$ the effect on empirical size is often much less than 0.01.

6 Empirical analysis of the PPP hypothesis.

We apply our methodology to analyse the Purchasing Power Parity (PPP) hypothesis in three US cities: New York, Boston and Philadelphia. The PPP

hypothesis indicates that arbitrage should induce prices of the same items to be the same in different places, and to react together to shocks affecting one of the two: if y_t is the local price and x_t is price of the same good elsewhere converted to the local currency (both prices being in logarithms), then $\mu_{11} = 0$, $\nu = 1$ and $\gamma < 1$ in (8) and (9). Strictly speaking, cointegration is not prescribed, but in practise this is necessarily the case because it is generally acknowledged that $\delta \geq 1$ (a lower level of δ would imply antipersistence of inflation); empirical experience of positive inflation rates also suggests a deterministic component of order about $\phi_{2\ddagger} = 1.5$, i.e. a linear trend.

The intuition behind PPP is easy to grasp and in line with common sense. The implications for both economic theory and policy advice are important, because implicitly it can also be interpreted as indicating market integration, so it is not surprising that it has been widely analysed in the applied economic literature. The empirical failure of the PPP hypothesis, at least as a short run phenomenon, is well documented: it can be due to the use of price indices, rather than effective prices, their differences reflecting the difference in preferences of economic agents. In addition, the arbitrage effect may be reduced and delayed by the cost of actively searching on the neighbour market and of eventually shipping the good to the home market. The latter argument also suggests that some deviation should be allowed, at least in the short run, and the PPP model then quickly become a classical case-study for cointegration: Corbae and Ouliaris (1988) and Johansen and Juselius (1992) assuming $(\gamma, \delta) = (0, 1)$. They followed two different approaches: Corbae and Ouliaris tested for a unit root with a Dickey-Fuller statistic on the difference $y_t - x_t$, while Johansen and Juselius first estimated the potential cointegrating vector and then tested the restriction. In both the cases the joint restrictions $\nu = 1$, $\gamma = 0$ were rejected. The PPP hypothesis largely remains a puzzle in cointegration analysis, evidence being still dubious.

The restriction on γ imposed by the $(\gamma, \delta) = (0, 1)$ paradigm is stronger than economic theory implies, disregarding a wide class of mean- (or conditional-mean-) reverting processes that are indexed by a different γ . Allowing for $0 < \gamma < 1$, semiparametric fractional cointegration analysis of PPP was carried out by Cheung and Lai (1993), who estimated ν by OLS and γ by log periodogram regression. They discussed the existence of a cointegrating relation but did not test $\nu = 1$. The GLS methodology presented in the previous sections on the other hand allows for a more efficient estimation and testing of ν , but the trend must then be taken explicitly into account.

We employ logged quarterly price indices of Boston, New York and Philadelphia for 1950(1) through 2003 (3), so that $n = 215$. The data were collected from Datastream and have codes USCPBOMAF, USCPNYMAF, USCPPHMAF; the sampling frequency was intended to be monthly, but for several years the data were only collected every second or third month both for Boston and Philadelphia: for each city we produced quarterly data by averaging. Data were then normalised, dividing each series by the first observation, and logarithms are finally taken. The normalisation is introduced to make the series visually com-

parable: as seen in Figure 1, they are dominated by a long-term deterministic component, to the extent that the three paths cannot be clearly distinguished. Thus we also present, in Figure 2, annualised first differences, corresponding to inflation rate. Our assumption is that long term inflation remained stable over the whole period: phases of high inflation, such as those observed after the two oil shocks, are still possible, but on adopting a very long perspective they seem to be temporary (albeit still persistent) phenomena rather than substantial, non-reverting breaks. As Diebold and Inoue (2001) showed, allowing random, occasional breaks, and a long enough time perspective, long memory can be a convenient way to describe the data. On the other hand an analysis that takes explicit account of apparent structural breaks might be more satisfactory. Inspection of the figures, especially with respect to inflation, seems indeed to suggest a common stochastic trend, the three indices reacting to the most relevant events, including the oil shocks.

We analyse cointegration pairwise, denoting the three cases Bo-NY, Ph-NY and Ph-Bo for New York and Boston, New York and Philadelphia, and Boston and Philadelphia, respectively. Of course if two pairs are both cointegrated the third one will be too and it is then redundant, but considering all three is sensible, especially in the preliminary phase of semiparametric analysis, where the estimates are robust to model misspecification but inefficient. The distinction between explanatory and dependent variables has no econometric implications in our framework, but we take New York as x_t throughout, and Boston as x_t in the Ph-Bo model. The nominal size for the tests is set at 5%.

We first tested the usual $\gamma = 0$, $\delta = 1$ framework. Though inflation is sometimes modelled as an $I(1)$ process, it is usually taken to be $I(0)$, implying $\delta = 1$, as this is consistent with a monetary policy model in which the central bank aims to stabilise the growth rate of prices in the medium-run, and as it is often supported by empirical tests. We tested $\phi = 1$ by the augmented Dickey-Fuller test, allowing for a constant in the AR model in levels (thus a linear trend in the (9)), with an AR(4) selected by sequential testing. We applied the same procedure (without an intercept) to the first differences with $\delta = 2$ in mind but overall our results support $\delta = 1$. The vector AR for the Johansen procedure was determined by inspecting the sequential likelihood ratio test (LR), the Schwarz (SC) and the Hannan and Quinn (HQ) statistics, these pointing in general to 5 lags. Using the procedure of Johansen (1991) cointegration was rejected in all the three models.

Evidence against PPP is often interpreted as indication of a lack of integration between markets, and cannot be explained by trade or cultural barriers, regulations or exchange rate instability in the present case. Since naive inspection of the data is suggestive of cointegration we thus investigate whether this can emerge in a fractional framework. In particular, we first examine the existence of a cointegrating relationship, in such a way as to avoid the consequences of misspecification of high frequency behaviour. The robust, but inefficient, estimates of γ, δ, ν that are involved will also provide a comparison with the more efficient ones subsequently obtained. The results are presented in Table 10.

We estimated δ_x and δ_y , the orders of integration of x_t and of y_t , by means of the local Whittle estimates, $\tilde{\delta}_x$ and $\tilde{\delta}_y$, as described in Robinson (1995) but, in view of the anticipated nonstationarity we applied the method to first differences, then adding back 1. Any deterministic linear trend is thereby removed too. The bandwidth was $m = 0.24n^{4/5} = 16$, which is approximately MSE - optimal when the process is a FARIMA(1,d,0) with AR coefficient 0.5 (see Henry and Robinson (1996)). The estimates of δ were 1.53 (NY), 1.45 (Bo) and 1.40 (Ph). On the basis of the statistics \tilde{T}_{yx} and \bar{T}_{yx} of Robinson and Yajima (2001) the hypothesis $\delta_y = \delta_x$ was rejected at 5% for Ph-NY though the statistic depends on a trimming number, and if this is not large enough the rejection could be due to the presence of cointegration. In the other two cases equality of the orders was not rejected. We then computed the Hausman type statistics \bar{H}_y and \bar{H}_x for no-cointegration of Marinucci and Robinson (2001, equation (30)) rejecting in case of Ph-NY and of Ph-Bo but not for Bo-NY. This seeming inconsistency could reflect a type two error: in Marinucci and Robinson's Monte Carlo experiment the power of a comparable case was only about 50%. Note also that for Bo-NY the more efficient memory parameter estimate $\tilde{\delta}_*$ used in the test does not lie between the individual estimates.

For each of the two pairs, $\bar{\nu}$ was computed and in each case found to be very close to 1. No-intercept OLS accords with economic theory, and the preliminary Johansen analysis suggested that the level of persistence γ , even if less than δ , may be between between 1/2 and 1. The possibility of such a high γ would suggest estimating it from first difference of the OLS residuals $\bar{v}_{1t} = y_t - \bar{\nu}x_t$. On the basis of $\tilde{\gamma}$, the local Whittle estimate of γ , (25) is satisfied in each case but only barely in the case Bo-NY; point estimates were also below 1 for Ph-NY and Ph-Bo, again in accordance with economic theory, while it was just above 1 for Bo-NY.

Our semiparametric analysis suggests that the PPP model could be reasonable for the three cities in fractional context, the rejection of the $(\gamma, \delta) = (0, 1)$ version being due to the high persistence of deviations from the long-run relation. We then proceeded to a parametric analysis along the lines described in Section 4 with the aim of ultimately testing $\nu = 1$.

We consider the Ph-NY and Bo-NY relations only, the results appearing in Table 11. In (23) and (24) we took $g_1(t) = 0$, $g_2(t) = t$. Using the previously obtained estimates of γ , δ (pooled), ν , and denoting by $\tilde{\mu}_{21}$ the estimate of μ_{21} obtained by OLS regression of x_t on t , we formed the residuals $\tilde{v}_{1t} = y_t - x_t\bar{\nu}$, $\tilde{v}_{2t} = x_t - \tilde{\mu}_{21}t$, and then took $\tilde{\gamma}$ and $\tilde{\delta}$ fractional differences respectively, labelling them as $\bar{u}_{1t} = \tilde{v}_{1t}(\tilde{\gamma})$, $\bar{u}_{2t} = \tilde{v}_{2t}(\tilde{\delta})$. We then determined the structure $B(L; h)$ using the LR, SC and HQ procedures on the series \bar{u}_{1t} , \bar{u}_{2t} , concluding in favour of a vector AR(4). Since the coefficients of the second and third lags are small, we took them to be zero, representing the short term dynamics by, in effect, an AR(1) combined with a quarterly effect.

Now define, for given c, d , $\tilde{u}_t(c, d) = (\tilde{v}_{1t}(c), \tilde{v}_{2t}(d))'$ and $\tilde{\varepsilon}_t(c, d) = B(L; \tilde{\theta}(c, d))\tilde{u}_t(c, d)$, where $\tilde{\theta}(c, d)$ indicates the OLS estimates of the AR coefficients for given c, d , dropping $\tilde{u}_1(c, d)$ because this term is not differenced at all. We then take

$(\hat{\gamma}, \hat{\delta}) = \arg \min n^{-1} \sum_t \tilde{\varepsilon}_t(c, d) \tilde{\varepsilon}_t'(c, d)$, and $\hat{\theta} = \tilde{\theta}(\hat{\gamma}, \hat{\delta})$, $\hat{\Omega} = n^{-1} \sum_t \tilde{\varepsilon}_t(\hat{\gamma}, \hat{\delta}) \tilde{\varepsilon}_t'(\hat{\gamma}, \hat{\delta})$. Hypothesis on γ, δ, θ can be tested with a likelihood ratio statistic.

This analysis confirmed restrictions on the AR(4) model for u_t , while $\hat{\gamma}$ and $\hat{\delta}$ were similar to the semiparametric estimates. We strongly rejected the joint hypothesis that $\delta = 1, \gamma = 0$. On the other hand the hypothesis $\gamma = 1$ was rejected against the alternative $\gamma \neq 1$. We then applied the time domain GLS procedure to estimate ν : the estimates $\hat{\nu}$ were in both cases close to 1, and the null hypothesis was not rejected.

In Figure 3 and 4 we plotted the GLS residuals $y_t - \hat{\nu}x_t$ (where $\hat{\nu}$ is the GLS estimate of ν) and the restricted residuals $y_t - x_t$ (assuming $\nu = 1$), respectively. In both cases the residual series only occasionally cross 0 (the value that we assumed to be the mean the disturbances). The patterns of GLS and restricted residuals are very similar, but autocorrelation appears to be a little stronger for the latter: iterating the parametric procedure, the estimates $\hat{\gamma}$ using GLS residuals $y_t - \hat{\nu}x_t$ are 0.65 for the pair Ph-NY and 0.69 for the pair Bo-NY, while using the restricted residuals they are both 0.73.

As a final small exercise we investigated the role of the deterministic trend in increasing precision in the estimation of ν , bearing in mind the findings of the Monte Carlo exercise. We thus also computed the GLS estimate of RH, using detrended data. The estimates $((1, 0)(b^*)^{-1}(1, 0)')^{1/2}$ (see Corollary 4) were 0.0107 in case of Ph-NY, and 0.0151 in case of Bo-NY, whereas the $(b^*)^{-1/2}$ defined from RH for the detrended data were more than twice the size, being 0.0304 and 0.0415 respectively.

Appendix

Proof of Theorem 1 Define $g(t) = (t^{\phi_1 - \frac{1}{2}}, \dots, t^{\phi_r - \frac{1}{2}})'$, $D_{gn} = \text{diag}\{n^{\phi_1}, \dots, n^{\phi_r}\}$, for $0 < \phi_1 < \dots < \phi_r$. For $d \geq 0$ we have

$$n^{-d} (D_{gn}^{-1} \otimes I_2) \sum_t g(t) \otimes u_t(-d) \rightarrow_d \int_0^1 g(r) \otimes d(\overline{W}(r; d+1), \widetilde{W}(r; d+1))', \quad (112)$$

where \otimes denotes Kronecker product. For $d > \frac{1}{2}$ (112) follows from Theorem 1 of Marinucci and Robinson (2000) (hereafter MR), and the continuous mapping theorem. For $0 \leq d \leq \frac{1}{2}$ (112) follows from a central limit theorem; note that the right side of (112) is in any case a $2r$ -dimensional normal variate. We have

$$x_t = u_{2t}(-\delta) + s_{xt}, \quad (113)$$

where

$$s_{xt} = \sum_{j=1}^{p_2} \mu_{2j} t^{\phi_{2j} - \frac{1}{2}}. \quad (114)$$

From (112),

$$\begin{aligned} n^{-\phi_{2\ddagger} - \delta} \sum_t s_{xt} u_{2t}(-\delta) &= n^{-\phi_{2\ddagger} - \delta} \mu_{2\ddagger} \sum_t t^{\phi_{2\ddagger} - \frac{1}{2}} u_{2t}(-\delta) + o_p(1) \\ &\rightarrow_d \Phi_2. \end{aligned} \quad (115)$$

By Theorem 1 of MR and the continuous mapping theorem

$$n^{-2\delta} \sum_t u_{2t}^2(-\delta) \rightarrow_d \Phi_1, \quad (116)$$

and by integral approximation

$$\begin{aligned} n^{-2\phi_{2\ddagger}} \sum_t s_{xt}^2 &= n^{-2\phi_{2\ddagger}} \mu_{2\ddagger}^2 \sum_t t^{2\phi_{2\ddagger}-1} + o(1) \\ &\rightarrow \Phi_3. \end{aligned} \quad (117)$$

Thus

$$k_n^{-2} M_{xx} \rightarrow_d \Phi_1^* + 2\Phi_2^* + \Phi_3^*. \quad (118)$$

Next, the numerator of $\bar{\nu} - \nu$ is

$$a = \sum_t (s_{xt} + u_{2t}(-\delta))(s_{yt} + u_{1t}(-\gamma)) \quad (119)$$

where

$$s_{yt} = \sum_{j=1}^{p_1} \mu_{1j} t^{\phi_{1j}-\frac{1}{2}}. \quad (120)$$

Integral approximation gives

$$\begin{aligned} n^{-\phi_{1\ddagger}-\phi_{2\ddagger}} \sum_t s_{xt}s_{yt} &= n^{-\phi_{1\ddagger}-\phi_{2\ddagger}} \mu_{1\ddagger}\mu_{2\ddagger} \sum_t t^{\phi_{1\ddagger}+\phi_{2\ddagger}-1} + o(1) \\ &\rightarrow \Psi_6 \end{aligned} \quad (121)$$

and (112) gives

$$n^{-\gamma-\phi_{2\ddagger}} \sum_t s_{xt}u_{1t}(-\gamma) \rightarrow_d \Psi_4. \quad (122)$$

$$n^{-\delta-\phi_{1\ddagger}} \sum_t s_{yt}u_{2t}(-\delta) \rightarrow_d \Psi_5, \quad (123)$$

To deal with $b = \sum_t u_{1t}(-\gamma)u_{2t}(-\delta)$, we make use of results of RM. From Lemmas 4.1, 4.2 and 5.1 of RM (see also their Propositions 6.1 and 6.2) we have

$$n^{-1}b \rightarrow_p \Psi_1, \quad \text{if } \gamma + \delta < 1, \quad (124)$$

$$\frac{n^{-1}}{\log n}b \rightarrow_p \Psi_2, \quad \text{if } \gamma + \delta = 1, \gamma > 0. \quad (125)$$

From Lemmas 4.3, 4.5 and 5.1 of RM, $b = O_p(n^\delta)$ for $\gamma = 0, \delta \geq 1$, but since $\phi_{1\ddagger} > 0$ it follows that in this case $a = o_p(n^{\delta+\phi_{1\ddagger}})$ so Ψ_5 will dominate. Finally

$$n^{-\gamma-\delta}b \rightarrow_d \Psi_3, \quad \gamma + \delta > 1, \gamma > 0, \quad (126)$$

where Theorem 1 of MR and the continuous mapping theorem covers the case $\gamma > \frac{1}{2}$, and Lemmas 4.5 and 5.1 of MR the case $\gamma \leq \frac{1}{2}$ (RM discuss the problem of representation of Ψ_3 in this case). It follows that

$$\ell_n^{-1}a \rightarrow_d \Psi_1^* + \Psi_2^* + \Psi_3^* + \Psi_4^* + \Psi_5^* + \Psi_6^*, \quad (127)$$

noting that in case $\gamma + \delta = 1$, $\gamma > 0$, $n \log n$ dominates $n = n^{\gamma+\delta}$, and dominates $n^{\gamma+\phi_{2\ddagger}}$, $n^{\delta+\phi_{1\ddagger}}$, $n^{\phi_{1\ddagger}+\phi_{2\ddagger}}$ if and only if $\delta \geq \phi_{2\ddagger}$, $\gamma \geq \phi_{1\ddagger}$, $\phi_{1\ddagger} + \phi_{2\ddagger} \leq 1$ respectively.

Proof of Theorem 2 The proof that

$$B_n^{-1} M_{x_+x_+} B_n^{-1} \rightarrow_d \Phi \quad (128)$$

straightforwardly uses results employed in showing (112), and is omitted. We are left to consider

$$M_{x_+v} = \sum_t x_{+t} \{u_{1t}(-\gamma) + s_{1t}\} = \sum_t (u_{2t}(-\delta) + s_{xt}, g_1'(t))' \{u_{1t}(-\gamma) + s_{1t}\}. \quad (129)$$

By again employing results from the proof of Theorem 1,

$$D_{1n}^{-1}(\gamma) \sum_t g_1(t) u_{1t}(-\gamma) \rightarrow_d \Psi_8 \quad (130)$$

while

$$D_{1n}^{-1}(\gamma) \sum_t g_1(t) s_{1t} \rightarrow 0 \quad (131)$$

because, for $j > p_{11}$ and $i + 1 \dots p_{11}$

$$n^{-\gamma-\phi_{1i}} \sum_t t^{\phi_{1i}+\phi_{1j}} = O(n^{\phi_{1j}-\gamma}) = o(1) \quad (132)$$

because $\gamma > \phi_{1j}$ for $j > p_{11}$. Next,

$$\sum_t s_{xt} s_{1t} = O(n^{\phi_{2\ddagger}+\phi_{1,p_{11}+1}}) = o(n^{\phi_{2\ddagger}+\gamma}). \quad (133)$$

Then from the proof of Theorem 1,

$$m_n^{-1} \sum_t \{u_{2t}(-\delta) + s_{xt}\} u_{1t}(-\gamma) \rightarrow_d \Psi_1^{**} + \Psi_2^{**} + \Psi_3^{**} + \Psi_4^{**} + \Psi_7^{**} \quad (134)$$

with Φ_7^{**} coming from MR Theorem 1 and Kurtz and Protter (1991), see also MR Propositions 6.3 and 6.4 (unlike in Theorem 1, this contribution is not always dominated).

Further, we may obtain the joint result

$$C_n^{-1} B_n^{-1} M_{x_+v} \rightarrow_d \Psi. \quad (135)$$

From the commutativity properties of diagonal matrices

$$\tilde{\nu}_+ - \nu = B_n^{-1} (B_n^{-1} M_{x_+x_+} B_n^{-1})^{-1} C_n (C_n^{-1} B_n^{-1} M_{x_+v}) \quad (136)$$

so the proof is completed by application of (128) and (136).

Proof of Theorem 3 Consider the case of white noise u_t . From (8), (9), (90), (98) and (99) we can write

$$\hat{\nu}_{++}(\gamma, \delta, \theta) - \nu = b(\gamma, \delta, \theta)^{-1} \{e(\gamma, \delta, \theta) + f(\gamma, \delta, \theta)\}, \quad (137)$$

where

$$e(\gamma, \delta, \theta) = \sum_t w_t(\gamma, \delta, \theta) \Omega^{-1} u_t, \quad f(\gamma, \delta, \theta) = \sum_t w_t(\gamma, \delta, \theta) \Omega^{-1} \tilde{s}_t, \quad (138)$$

where $\tilde{s}_t = (\tilde{s}_{1t}, \tilde{s}_{2t})' = (\Delta^\gamma s_{1t}, \Delta^\delta s_{2t})'$. From Lemma 1 of Robinson (2003), with $v_t = t^c$,

$$v_t(d) = \frac{\Gamma(c+1)}{\Gamma(c-d+1)} t^{c-d} + O(t^{c-m-1}), \quad (139)$$

where m is the integer such that $d-1 < m \leq d$. Then

$$\begin{aligned} \sum_t \Delta^\gamma g_2'(t) \mu_2 \tilde{s}_{1t} &= O\left(\sum_t t^{\phi_{2\ddagger} - 2\gamma + \phi_{1,p_{11}+1} - 1}\right) \\ &= O\left(n^{\phi_{2\ddagger} + \phi_{1,p_{11}+1} - 2\gamma}\right), \end{aligned} \quad (140)$$

$$\begin{aligned} \sum_t \Delta^\gamma g_{1j}(t) \tilde{s}_{1t} &= O\left(\sum_t t^{\phi_{1j} - 2\gamma + \phi_{1,p_{11}+1} - 1}\right) \\ &= O\left(n^{\phi_{1j} + \phi_{1,p_{11}+1} - 2\gamma}\right), \quad j = 1, \dots, p_{11}, \end{aligned} \quad (141)$$

$$\begin{aligned} \sum_t \Delta^\delta g_{2j}(t) \tilde{s}_{1t} &= O\left(\sum_t t^{\phi_{2j} - \gamma - \delta + \phi_{1,p_{11}+1} - 1}\right) \\ &= O\left(n^{\phi_{2j} + \phi_{1,p_{11}+1} - \gamma - \delta}\right), \quad j = 1, \dots, p_{21}, \end{aligned} \quad (142)$$

and similarly $\sum_t \Delta^\gamma g_2'(t) \mu_2 \tilde{s}_{2t} = O\left(n^{\phi_{2\ddagger} + \phi_{2,p_{21}+1} - \gamma - \delta}\right)$, $\sum_t \Delta^\gamma g_{1j}(t) \tilde{s}_{2t} = O\left(n^{\phi_{1j} + \phi_{2,p_{21}+1} - \gamma - \delta}\right)$ for $j = 1 \dots p_{11}$, $\sum_t \Delta^\delta g_{2j}(t) \tilde{s}_{2t} = O\left(n^{\phi_{2j} + \phi_{2,p_{21}+1} - 2\delta}\right)$ for $j = 1 \dots p_{21}$.

On the other hand, much as in the proof of Theorem 1,

$$\sum_t u_{2t}(\gamma - \delta) \tilde{s}_{1t} = O_p\left(n^{\delta - 2\gamma + \phi_{1,p_{11}+1} + \frac{1}{2}}\right), \quad \sum_t u_{2t}(\gamma - \delta) \tilde{s}_{2t} = O_p\left(n^{-\gamma + \phi_{2,p_{21}+1} + \frac{1}{2}}\right). \quad (143)$$

It straightforwardly follows that

$$D_n^{-1} f(\gamma, \delta, \theta) \rightarrow 0. \quad (144)$$

From (139) and routine arguments

$$n^{\frac{1}{2}} D_{in}^{-1} (-d) \Delta^d g_i([rn]) \rightarrow \Lambda_i(d) \frac{g_i(r)}{r^d}, \quad r \in (0, 1], \quad d \geq 0, \quad i = 1, 2. \quad (145)$$

From this and MR Theorem 1

$$n^{\frac{1}{2}} D_n^{-1} w_{[rn]}(\gamma, \delta) \Rightarrow Q(r), \quad r \in (0, 1]. \quad (146)$$

Thus from the continuous mapping theorem and Kurtz and Protter (1991)

$$D_n^{-1} b(\gamma, \delta, \theta) D_n^{-1} \Rightarrow \int_0^1 Q(r) Q'(r) dr, \quad D_n^{-1} e(b, \delta, \theta) \Rightarrow \int_0^1 Q(r) dW(r) \quad (147)$$

to complete the proof for $\widehat{\nu}_{++}(\gamma, \delta, \theta)$. The application of Assumption 3 to prove the Theorem for the remaining quantities in (101), in the white noise u_t case, is straightforward, and thus omitted.

A good deal of the proof detail in RH is concerned with justifying the general short memory autocorrelation in u_t described in Section 4. It is clearly unnecessary to repeat this for our extended estimate, and it suffices merely to consider the implications for the deterministic components we have introduced. These are somewhat different from the treatment of stochastic trends. Consider the quantity

$$\widetilde{e}(\gamma, \delta, \theta) = \sum_j F_{\widetilde{w}(\gamma, \delta)}(\lambda_j) f(\lambda_j)^{-1} F_u(-\lambda_j), \quad (148)$$

where

$$F_{\widetilde{w}(\gamma, \delta)}(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t \widetilde{w}_t(\gamma, \delta) e^{it\lambda}, \quad F_u(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_t u_t e^{it\lambda}, \quad (149)$$

with $\widetilde{w}_t(c, d)$ defined like $w_t(c, d)$ in (85) but with $\Delta^c x_t$ replaced by $\Delta^c s_{xt}$. Denote by $\psi_L(\lambda) = \sum_{\ell=-L}^L \psi_\ell (1 - |\ell|/L) e^{-i\ell\lambda}$ the Cesaro sum, to L terms, of the Fourier series of $f(\lambda)^{-1}$. Define \widetilde{D}_n like D_n but with $n^{\max(\delta, \phi_{2z})}$ replaced by $n^{\phi_{2z}}$ and

$$\widehat{e}(\gamma, \delta, \theta) = \sum_j F_{\widetilde{w}(\gamma, \delta)}(\lambda_j) \psi_L(\lambda_j) F_u(-\lambda_j). \quad (150)$$

Then $\widetilde{D}_n^{-1} \{\widetilde{e}(\gamma, \delta, \theta) - \widehat{e}(\gamma, \delta, \theta)\}$ has mean zero and covariance matrix

$$\begin{aligned} & \frac{1}{2\pi n} \int_{-\pi}^{\pi} \left\{ \widetilde{D}_n^{-1} \sum_j F_{\widetilde{w}(\gamma, \delta)}(\lambda_j) \{f(\lambda_j)^{-1} - \psi_L(\lambda_j)\} \sum_s e^{i(\lambda_j - \lambda)s} \right\} f(\lambda) \\ & \times \left\{ \widetilde{D}_n^{-1} \sum_j F_{\widetilde{w}(\gamma, \delta)}(-\lambda_j) \{f(-\lambda_j)^{-1} - \psi_L(-\lambda_j)\} \sum_j e^{i(\lambda - \lambda_j)s} \right\}' d\lambda. \end{aligned} \quad (151)$$

Using the properties of the complex exponential function, this has norm bounded by a constant times

$$\sum_j \left\| \widetilde{D}_n^{-1} F_{\widetilde{w}(\gamma, \delta)}(\lambda_j) \{f(\lambda_j)^{-1} - \psi_L(\lambda_j)\} \right\|^2 \leq \varepsilon^2 \sum_j \left\| \widetilde{D}_n^{-1} F_{\widetilde{w}(\gamma, \delta)}(\lambda_j) \right\|^2 \quad (152)$$

for arbitrary $\varepsilon > 0$, on choosing L large enough and noting the continuity of $f(\lambda)^{-1}$. For any sequence c_t

$$n^{-1} \sum_j \left| \sum_t c_t e^{it\lambda_j} \right|^2 = \sum_t c_t^2, \quad (153)$$

so applying again (147), it straightforwardly follows that (152) = $O(\varepsilon^2)$. Thus

$$\widehat{e}(\gamma, \delta, \theta) = \sum_{\ell=-L}^L \sum'_t \widetilde{w}_t(\gamma, \delta) \psi_\ell \left(1 - \frac{|\ell|}{L}\right) u_{t-\ell} \quad (154)$$

$$+ \sum_{\ell=1}^L \sum_{t=1}^{\ell} \widetilde{w}_t(\gamma, \delta) \psi_\ell \left(1 - \frac{\ell}{L}\right) u_{t-\ell}, \quad (155)$$

$$+ \sum_{\ell=-L}^{-1} \sum_{t=n-\ell+1}^n \widetilde{w}_t(\gamma, \delta) \psi_\ell \left(1 + \frac{\ell}{L}\right) u_{t-\ell}, \quad (156)$$

where $\sum'_t = \sum_{1 \leq t, t-\ell \leq n}$. The sums over t in (155) and (156) include only ℓ terms, and, with L fixed and $n \rightarrow \infty$, will turn out to be dominated by (154). To deal with this, note from boundedness of f that for $c > 0$, $d > 0$ and any j ,

$$\sum_t \left\{ \Delta^d t^c - \frac{\Gamma(c+1)t^{c-d}}{\Gamma(c-d+1)} \right\} u_{t-j} \quad (157)$$

has mean zero and variance bounded by a constant times

$$\sum_t \left\{ \Delta^d t^c - \frac{\Gamma(c+1)t^{c-d}}{\Gamma(c-d+1)} \right\}^2 = O\left(\sum_t t^{2(c-m-1)}\right) \quad (158)$$

where m is as described after (139). Then (158) is $O(1)$ for $2c-2m < 1$, $O(\log n)$ for $2c-2m = 1$ and $O(n^{2c-2m+1})$ for $2c-2m > 1$. It follows that we may replace $\widetilde{D}_n^{-1} \widehat{e}(\gamma, \delta, \theta)$ by

$$\widetilde{D}_n^{-1} \sum_{\ell=-L}^L \sum'_t \widehat{w}_t(\gamma, \delta) \psi_\ell \left(1 - \frac{|\ell|}{L}\right) u_{t-\ell}, \quad (159)$$

where

$$\widehat{w}_t(\gamma, \delta) = \begin{bmatrix} \mu_{2\ddagger} \frac{\Gamma(\phi_{2\ddagger} + \frac{1}{2})}{\Gamma(\phi_{2\ddagger} - \gamma + \frac{1}{2})} t^{\phi_{2\ddagger} - \gamma - \frac{1}{2}} & \{\Lambda_1(\gamma) g_1(t)\}' t^{-\gamma} & 0 \\ 0 & 0 & \{\Lambda_2(\delta) g_2(t)\}' t^{-\delta} \end{bmatrix}', \quad (160)$$

the above arguments indicating that other contributions from the top left hand element of $\widehat{w}_t(\gamma, \delta)$ can be neglected. The asymptotic normality, for fixed L , of (159) follows standardly (see e.g. Hannan, 1970, Chapter 7), whence Bernstein's lemma, with $L \rightarrow \infty$, completes the central limit theorem for $\widetilde{D}_n^{-1} \widehat{e}(\gamma, \delta, \theta)$. The proof that we can neglect contributions from s_{1t} and s_{2t} follows much as above, as does, using also RH, the limiting behaviour of $D_n^{-1} b(\gamma, \delta, \theta) D_n^{-1}$. The proof that we can replace γ, δ, θ by $\widehat{\gamma}, \widehat{\delta}, \widehat{\theta}$ is lengthy but relies basically on RH and standard arguments to cope with the deterministic components. As in RH, the proof for the 'time - domain' estimates is similar but slightly simpler, and is omitted.

Acknowledgements

This research was supported by ESRC Grants R000235892 and R000239936. We thank two referees for constructive comments.

References

- [1] CAMPOS, J., ERICSSON, N.R., HENDRY, D.F., 1996. Cointegration in the presence of structural breaks. *Journal of Econometrics* 70, 187-220.
- [2] CHAN, N.H., TERRIN, N., 1995. Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Annals of Statistics* 23, 1662-1683.
- [3] CHEN, W.W., HURVICH, C.M., 2003. Estimating fractional cointegration in the presence of polynomial trends. *Journal of Econometrics* 117, 95-121.
- [4] CHEUNG, Y.W., LAI, K.S., 1993. A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economic Statistics* 11, 103-112.
- [5] CORBAE, D., OULIARIS, S., 1988. Cointegration and tests of purchasing power parity. *Review of Economics and Statistics* 70, 508-511.
- [6] DIEBOLD, F.X., INOUE, A., 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131-159.
- [7] DOLADO, J., MARMOL, F., 1997. Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes. Preprint.
- [8] HANSEN, B.E., 1992. Efficient estimation and testing of cointegration vectors in the presence of deterministic trends. *Journal of Econometrics* 53, 87-121.
- [9] HANNAN, E.J., 1970. *Multiple Time Series*. Wiley, New York.
- [10] HASSLER, U., MARMOL, F., VELASCO, C., 2002. Residual log-periodogram inference for long-run relationships. Preprint.
- [11] HENRY, M., ROBINSON, P.M., 1996. Bandwidth choice in Gaussian semi-parametric estimation of long-range dependence, in *Athens Conference on Applied Probability and Time Series Analysis, Vol. II: Time Series Analysis* (P. M. Robinson and M. Rosenblatt, eds), New York: Springer-Verlag, pp. 220-32.
- [12] HOFFMAN, D.L., RASCHE, R.H., 1991. Long run income and interest elasticities of money demand in the United States. *The Review of Economics and Statistics* 73, 665-674.
- [13] HUALDE, J., ROBINSON, P.M., 2001. Root-n-consistent estimation of weak fractional cointegration. Preprint.
- [14] JEGANATHAN, P., 1999. On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583-621.

- [15] JOHANSEN, S., 1991. Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.
- [16] JOHANSEN, S., JUSELIUS, K., 1992. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for the UK. *Journal of Econometrics* 53, 211-244.
- [17] KIM, C.S., PHILLIPS, P.C.B., 2000. Fully modified estimation of fractional cointegration models. Preprint.
- [18] KURTZ, T.G., PROTTER, P., 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035-1070.
- [19] MARINUCCI, D., ROBINSON, P.M., 2000. Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103-120.
- [20] MARINUCCI, D., ROBINSON, P.M., 2001. Semiparametric fractional cointegration analysis. *Journal of Econometrics* 105, 225-247.
- [21] OGAKI, M., PARK, J.Y., 1997. A cointegration approach to estimating preference parameters. *Journal of Econometrics* 82, 107-134.
- [22] PERRON, P., CAMPBELL, J.Y., 1993. A note on Johansen's procedure when trends are present. *Empirical Economics* 18, 777-789.
- [23] PHILLIPS, P.C.B., 1991a. Optimal inference in cointegrated systems. *Econometrica* 59, 283-306.
- [24] PHILLIPS, P.C.B., 1991b. Spectral regression for cointegrated time series. In W.A. Barnett, J. Powell and G. Tauchen (eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press, 413-435.
- [25] PHILLIPS, P.C.B., HANSEN, B.E., 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99-125.
- [26] ROBINSON, P.M., 1986. On the errors-in-variables problem for time series. *Journal of Multivariate Analysis* 19, 240-250.
- [27] ROBINSON, P.M., 1994. Semiparametric analysis of long memory time series. *Annals of Statistics* 22, 515-539.
- [28] ROBINSON, P.M., 1995. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630 - 1661.
- [29] ROBINSON, P.M., 2003. Efficiency improvements in inference on stationary and nonstationary fractional time series. Preprint.

- [30] ROBINSON, P.M., HUALDE, J., 2003. Cointegration in fractional systems with unknown integration orders. *Econometrica* 71, 1727-1766.
- [31] ROBINSON, P.M., MARINUCCI, D., 2000. The averaged periodogram for nonstationary vector time series. *Statistical Inference for Stochastic Processes* 3, 149-160.
- [32] ROBINSON, P.M., MARINUCCI, D., 2001. Narrow band analysis of non-stationary processes. *Annals of Statistics* 29, 947-986.
- [33] ROBINSON, P.M., MARINUCCI, D., 2003. Semiparametric frequency domain analysis of fractional cointegration. *Time Series with Long Memory* (P.M. Robinson, ed.). Oxford University Press, Oxford, 334-373.
- [34] ROBINSON, P.M., YAJIMA, Y., 2002. Determination of cointegrating ranking fractional systems. *Journal of Econometrics* 106, 217-241.
- [35] SÖDERLIND, P., VREDIN A., 1996. Applied cointegration analysis in the mirror of macroeconomic theory. *Journal of Applied Econometrics* 11, 363-381.
- [36] STOCK, J., WATSON, M.W., 1988. Testing for common trends. *Journal of the American Statistical Society* 83, 1097-1107.
- [37] STOCK, J., WATSON, M.W., 1993. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783-820.
- [38] VELASCO, C., 2000. Gaussian semiparametric estimation of fractional cointegration. Preprint.
- [39] VELASCO, C., ROBINSON, P.M., 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* 95, 1229-1243.
- [40] WEST, K.D., 1988. Asymptotic normality when regressors have a unit root. *Econometrica* 56, 1397-1417.

TABLE 1
CONVERGENCE RATES (POWERS OF n) OF $\bar{\nu}$

	STa	STb	STc	STd	STe	STf
0^*	0.2	1.2	2	0.2^\dagger	0.8	1.6
DTa	1	1	1.5	1	1	1.5
DTb	2	2	2	2	2	2
DTc	X	X	0.5	X	X	0.5
DTd	1	1	1	1	1	1

*: Refers to the case of no trend, as in RM.

\dagger : The rate is actually $n^{0.2}/\ln n$

X: Inconsistent

TABLE 2:
CONVERGENCE RATES (POWERS OF n) OF $\tilde{\nu}$

	STa	STb	STc	STd	STe	STf
DTa	1.5	1.5	2	1.1	1.1	1.6
DTb	2.5	2.5	2.5	2.1	2.1	2.1
DTc			2			1.6
DTd	2.5	2.5	2.5	2.1	2.1	2.1

Note: convergence rates not reported when Assumption 2 is not met.

TABLE 3:
CONVERGENCE RATES (POWERS OF n) OF $\hat{\nu}$

	STa	STb	STc	STd	STe	STf
DTa	1.5	1.5	2		1.1	1.6
DTb	2.5	2.5	2.5		2.1	2.1
DTc			2			1.6
DTd	2.5	2.5	2.5		2.1	2.1

Note: convergence rates not reported when Assumption 2 or $\beta > 0.5$ is not met.

TABLE 4
 MONTE CARLO BIAS FOR $\tau = 1$, $\rho = 0.5$, 1000 REPLICATIONS.

ST	DT	$n = 64$				$n = 128$				$n = 256$			
		$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$	$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$	$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$
a	a	.0235	.0009	.0000	.0000	.0117	.0003	.0000	.0000	.0059	.0001	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	1.0232	.2980	.1549	.1634	1.0112	.2515	.1145	.1201	1.0057	.2076	.0804	.0841
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
b	a	.0247	.0003	.0001	.0000	.0121	.0000	.0000	.0000	.0060	.0000	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	1.0912	.0176	-.0002	-.0040	1.0505	.0048	.0006	-.0007	1.0347	.0006	.0004	.0002
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
c	a	.0016	-.0017	-.0003	-.0006	.0005	-.0005	.0000	-.0001	.0001	-.0001	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	.0668	-.0057	-.0006	-.0020	.0481	-.0016	.0000	-.0003	.0193	-.0004	.0000	.0000
	d	.0195	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
d	a	.0237	.0010	.0001	.0001	.0116	.0001	-.0002	-.0001	.0058	.0001	.0001	.0001
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	1.0233	.4212	.3969	.3911	1.0112	.3907	.3657	.3622	1.0056	.3598	.3344	.3319
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
e	a	.0238	.0005	-.0003	-.0004	.0116	.0000	-.0004	-.0003	.0057	.0001	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	1.0903	.0761	.0194	.0224	1.0501	.0416	.0090	.0096	1.0344	.0240	.0043	.0051
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
f	a	.0027	-.0021	-.0001	-.0007	.0006	-.0008	.0000	-.0001	.0002	-.0002	.0001	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	.0678	-.0078	-.0003	-.0024	.0482	-.0030	-.0001	-.0006	.0193	-.0009	.0001	.0000
	d	.0195	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000

TABLE 5
 MONTE CARLO STANDARD DEVIATION FOR $\tau = 1, \rho = 0.5, 1000$
 REPLICATIONS.

ST	DT	$n = 64$				$n = 128$				$n = 256$			
		$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$	$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$	$\bar{\nu}$	$\tilde{\nu}$	$\hat{\nu}_I$	$\hat{\nu}_F$
a	a	.0031	.0069	.0060	.0062	.0011	.0024	.0021	.0022	.0004	.0008	.0007	.0007
	b	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	.0298	.1120	.1171	.1197	.0160	.0778	.0780	.0793	.0091	.0547	.0492	.0502
	d	.0001	.0004	.0004	.0004	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000
b	a	.0080	.0080	.0071	.0071	.0029	.0025	.0022	.0023	.0012	.0009	.0008	.0008
	b	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	.3691	.0512	.0462	.0492	.2614	.0223	.0204	.0213	.2163	.0088	.0083	.0085
	d	.0001	.0004	.0004	.0004	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000
c	a	.0132	.0041	.0032	.0033	.0047	.0012	.0009	.0010	.0016	.0002	.0002	.0002
	b	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	c	.5489	.0100	.0084	.0088	.4008	.0025	.0020	.0020	.2748	.0007	.0005	.0005
	d	.0014	.0004	.0004	.0004	.0005	.0001	.0001	.0001	.0002	.0000	.0000	.0000
d	a	.0137	.0189	.0168	.0171	.0067	.0087	.0076	.0077	.0031	.0040	.0035	.0035
	b	.0003	.0003	.0003	.0003	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000
	c	.0269	.1520	.1586	.1658	.0145	.1183	.1241	.1277	.0082	.0922	.0951	.0971
	d	.0003	.0009	.0008	.0008	.0001	.0002	.0002	.0002	.0000	.0000	.0000	.0000
e	a	.0143	.0214	.0176	.0183	.0065	.0092	.0073	.0075	.0029	.0042	.0034	.0034
	b	.0003	.0003	.0003	.0003	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000
	c	.3629	.0951	.0855	.0910	.2589	.0550	.0476	.0500	.2148	.0307	.0250	.0269
	d	.0002	.0009	.0008	.0008	.0001	.0002	.0002	.0002	.0000	.0000	.0000	.0000
f	a	.0153	.0087	.0075	.0076	.0052	.0034	.0026	.0027	.0017	.0009	.0008	.0008
	b	.0003	.0003	.0002	.0002	.0001	.0001	.0001	.0001	.0000	.0000	.0000	.0000
	c	.5489	.0189	.0166	.0168	.4005	.0062	.0054	.0055	.2747	.0019	.0016	.0017
	d	.0014	.0009	.0008	.0008	.0005	.0002	.0002	.0002	.0002	.0001	.0000	.0000

TABLE 6
EMPIRICAL SIZE OF W_I, W_F , FOR $\tau = 1, \rho = 0.5$, 1000 REPLICATIONS.

<i>ST</i>	<i>DT</i>	$\alpha = .05$						$\alpha = .10$					
		$n = 64$		$n = 128$		$n = 256$		$n = 64$		$n = 128$		$n = 256$	
		W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
<i>a</i>	<i>a</i>	.065	.186	.057	.134	.053	.095	.113	.255	.113	.191	.103	.148
	<i>b</i>	.070	.182	.055	.129	.053	.102	.122	.252	.108	.189	.108	.169
	<i>c</i>	.412	.474	.472	.513	.520	.535	.497	.554	.572	.608	.633	.646
	<i>d</i>	.068	.227	.052	.130	.051	.109	.113	.304	.095	.201	.106	.171
<i>b</i>	<i>a</i>	.062	.194	.058	.130	.057	.094	.112	.267	.114	.188	.102	.148
	<i>b</i>	.068	.191	.052	.127	.054	.096	.119	.248	.113	.178	.100	.164
	<i>c</i>	.060	.169	.058	.139	.047	.082	.121	.263	.126	.199	.090	.153
	<i>d</i>	.071	.228	.050	.136	.058	.111	.116	.297	.100	.202	.106	.168
<i>c</i>	<i>a</i>	.054	.199	.056	.151	.052	.101	.112	.272	.113	.215	.096	.168
	<i>b</i>	.068	.198	.059	.137	.048	.087	.114	.260	.112	.201	.095	.158
	<i>c</i>	.077	.227	.044	.147	.052	.120	.120	.302	.101	.230	.102	.189
	<i>d</i>	.072	.229	.045	.139	.051	.111	.119	.308	.096	.205	.109	.168
<i>d</i>	<i>a</i>	.071	.191	.061	.146	.053	.105	.117	.254	.116	.209	.108	.168
	<i>b</i>	.066	.182	.064	.139	.051	.099	.124	.268	.105	.204	.107	.167
	<i>c</i>	.889	.887	.960	.958	.990	.991	.925	.921	.973	.972	.995	.993
	<i>d</i>	.060	.219	.049	.139	.047	.106	.102	.291	.105	.197	.099	.169
<i>e</i>	<i>a</i>	.070	.192	.054	.135	.045	.101	.120	.259	.114	.199	.111	.161
	<i>b</i>	.069	.196	.059	.132	.045	.102	.122	.265	.102	.194	.114	.175
	<i>c</i>	.094	.228	.082	.176	.076	.137	.175	.305	.160	.242	.122	.215
	<i>d</i>	.060	.221	.054	.136	.054	.106	.109	.295	.095	.195	.103	.161
<i>f</i>	<i>a</i>	.071	.211	.058	.147	.043	.096	.127	.269	.105	.199	.096	.162
	<i>b</i>	.072	.204	.068	.145	.055	.112	.126	.268	.120	.213	.105	.172
	<i>c</i>	.068	.235	.056	.144	.041	.103	.124	.306	.114	.216	.097	.165
	<i>d</i>	.057	.236	.054	.144	.051	.105	.113	.309	.101	.199	.104	.172

TABLE 7
 MONTE CARLO BIAS FOR $\delta = 0.6$, $\gamma = 0$, DTa , DTb and DTd , $\tau = 1$,
 1000 REPLICATIONS

ρ	DT	$n = 64$				$n = 128$				$n = 256$			
		\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F	\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F	\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F
0	a	.0233	-.0001	.0000	.0000	.0116	.0000	.0000	.0000	.0058	.0000	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
0.75	a	.0237	.0015	.0000	.0000	.0118	.0004	.0000	.0000	.0059	.0001	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000
-0.5	a	.0230	-.0011	-.0001	.0000	.0116	-.0003	.0000	.0000	.0058	-.0001	.0000	.0000
	b	.0004	.0000	.0000	.0000	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	d	.0194	.0000	.0000	.0000	.0097	.0000	.0000	.0000	.0049	.0000	.0000	.0000

TABLE 8
 MONTE CARLO SD FOR $\delta = 0.6$, $\gamma = 0$, DTa , 1000 REPLICATIONS

τ	ρ	$n = 64$				$n = 128$				$n = 256$			
		\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F	\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F	\bar{v}	\tilde{v}	\hat{v}_I	\hat{v}_F
1	0	.0034	.0068	.0068	.0068	.0012	.0024	.0024	.0024	.0004	.0008	.0008	.0008
	0.75	.0030	.0069	.0048	.0052	.0011	.0024	.0017	.0018	.0004	.0009	.0006	.0006
	-0.5	.0036	.0067	.0060	.0061	.0013	.0023	.0021	.0021	.0005	.0008	.0007	.0007
2	0	.0047	.0096	.0097	.0097	.0017	.0034	.0034	.0034	.0006	.0012	.0012	.0012
	0.75	.0043	.0097	.0068	.0073	.0016	.0033	.0024	.0025	.0006	.0012	.0008	.0008
	-0.5	.0050	.0094	.0085	.0086	.0018	.0033	.0030	.0030	.0006	.0012	.0010	.0011
0.5	0	.0024	.0048	.0048	.0048	.0009	.0017	.0017	.0017	.0003	.0006	.0006	.0006
	0.75	.0020	.0049	.0034	.0037	.0007	.0017	.0012	.0012	.0003	.0006	.0004	.0004
	-0.5	.0027	.0047	.0042	.0043	.0009	.0016	.0015	.0015	.0003	.0006	.0005	.0005

TABLE 9
EMPIRICAL SIZE OF W_I , W_F FOR $\delta = 0.6$, $\gamma = 0$, DTa , DTb and DTd ,
 $\tau = 1$, 1000 REPLICATIONS.

ρ	DT	$\alpha = .05$						$\alpha = .10$					
		$n = 64$		$n = 128$		$n = 256$		$n = 64$		$n = 128$		$n = 256$	
		W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F	W_I	W_F
0	a	.062	.183	.061	.121	.058	.099	.115	.258	.110	.180	.105	.159
	b	.062	.190	.056	.121	.048	.091	.114	.248	.111	.172	.091	.148
	d	.067	.218	.052	.126	.057	.107	.119	.305	.100	.190	.107	.166
0.75	a	.062	.192	.059	.144	.053	.111	.115	.259	.110	.207	.111	.163
	b	.064	.179	.056	.124	.055	.104	.121	.245	.106	.197	.117	.163
	d	.064	.225	.044	.136	.047	.111	.115	.310	.094	.204	.101	.181
-0.5	a	.060	.196	.058	.122	.055	.103	.100	.258	.109	.193	.097	.158
	b	.055	.171	.046	.111	.046	.089	.101	.241	.089	.178	.088	.149
	d	.068	.245	.057	.143	.058	.114	.128	.312	.101	.196	.104	.167

Table 10 Empirical analysis of the PPP:
semiparametric analysis of fractional cointegration.

	$\tilde{\delta}_y$	$\tilde{\delta}_x$	$\tilde{\delta}_*$	$(\tilde{T}_{yx})^2$	$(\bar{T}_{yx})^2$	$(\bar{H}_y)^2$	$(\bar{H}_x)^2$	$\bar{\nu}$	$\tilde{\gamma}$
Ph-NY	1.40	1.53	1.35	4.83	0.56	0.26	4.00	0.99	0.82
Bo-NY	1.45	1.53	1.38	1.28	0.18	2.83	0.68	1.02	1.05
Ph-Bo	1.40	1.45	1.21	0.67	0.10	4.53	7.66	0.97	0.83

Note: critical values for χ_1^2 : 3.84 (5%) and 6.63 (1%).

Table 11 Empirical analysis of the PPP:
parametric analysis of fractional cointegration.

	LR_B	$\hat{\delta}$	$\hat{\gamma}$	$LR_{\{\delta=1, \gamma=0\}}$	$LR_{\{\gamma=1\}}$	$\hat{\nu}$	$\hat{\mu}_{21}$	$LR_{\{\nu=1\}}$
Ph-NY	9.69 [0.287]	1.42	0.66	21.50 [0.000]	9.83 [0.002]	0.98	0.0045	3.74 [0.053]
Bo-NY	4.51 [0.808]	1.43	0.69	35.87 [0.000]	9.85 [0.002]	1.02	0.0041	2.21 [0.137]

Note: LR_B is the likelihood ratio statistic for the hypothesis that the second and third lag are excluded; $LR_{\{\delta=1, \gamma=0\}}$, $LR_{\{\gamma=1\}}$, $LR_{\{\nu=1\}}$ are the the likelihood ratio statistics for the hypothesis $H_0 : \{\delta = 1, \gamma = 0\}$, $H_0 : \{\gamma = 1\}$, $H_0 : \{\nu = 1\}$; P-values are in brackets.

Figure 1 Normalised log prices, New York, Philadelphia, Boston;

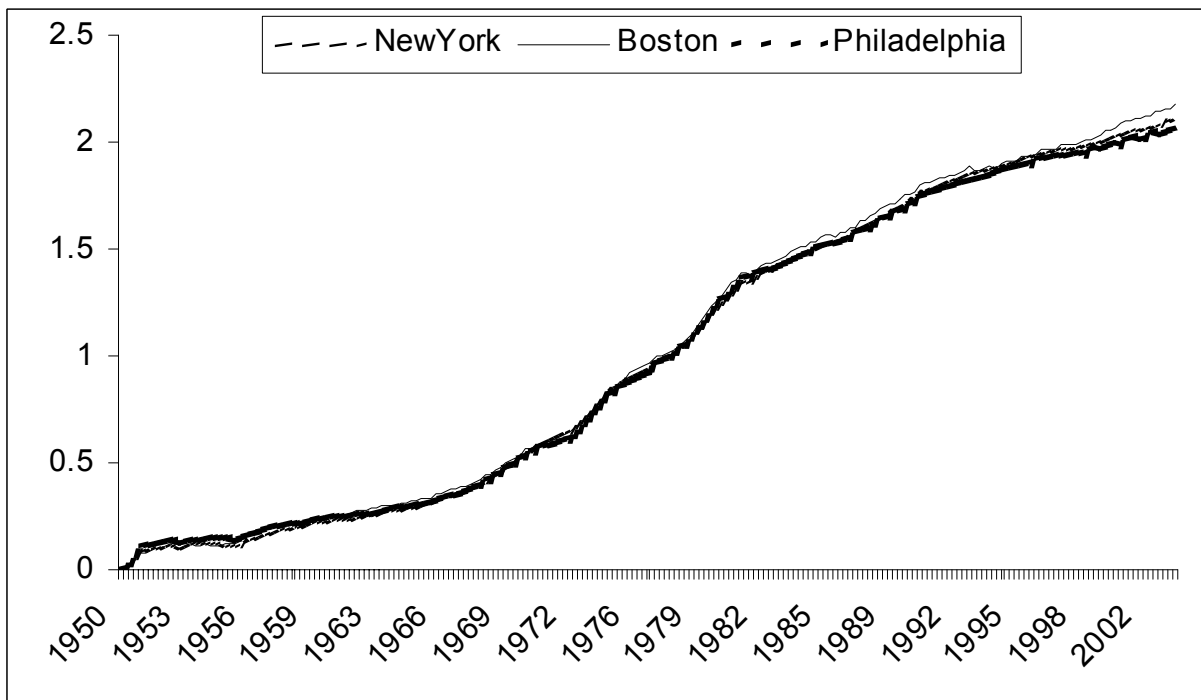


Figure 2 Inflation rate, New York, Philadelphia, Boston;
(rescaled first differences of normalised log prices).

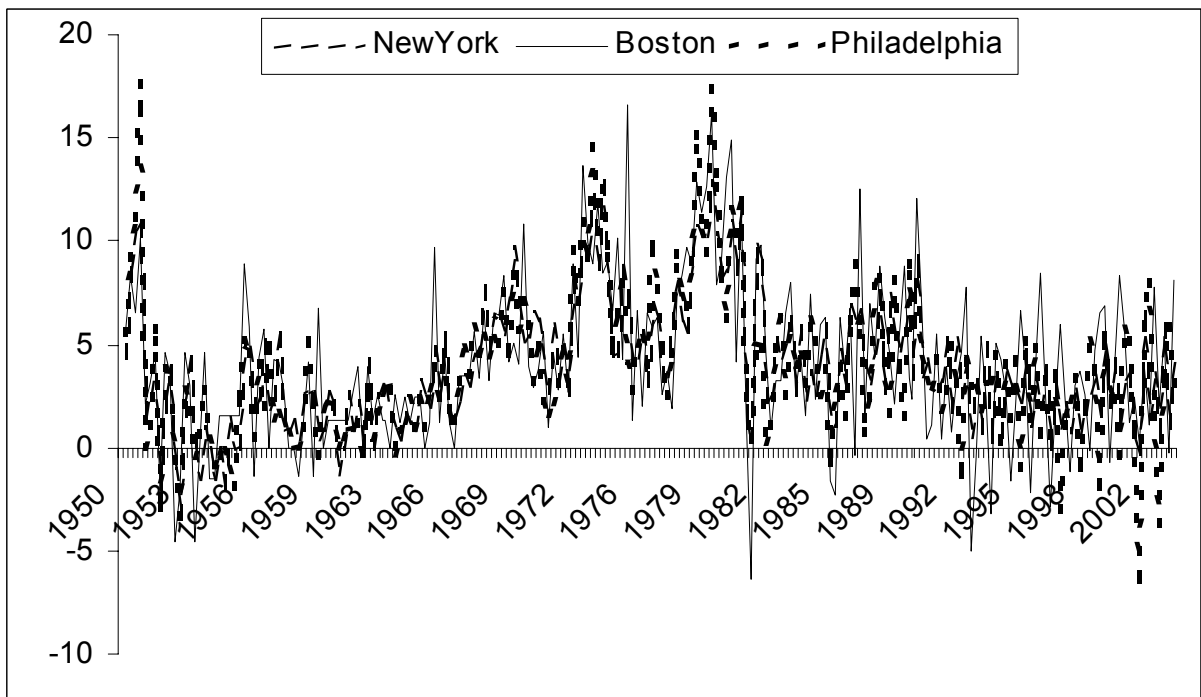


Figure 3 GLS cointegrating residuals $y_t - \hat{v}x_t$

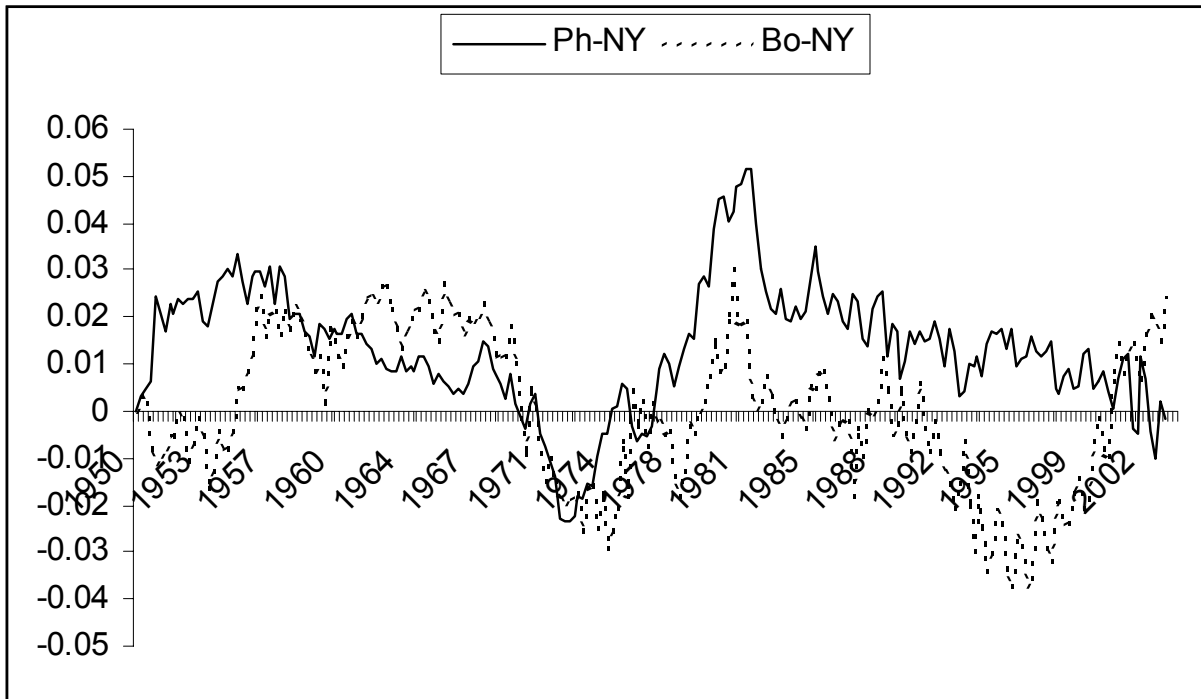


Figure 4 Restricted cointegrating residuals $y_t - x_t$

