Econometric Theory, **18**, 2002, 646–672. Printed in the United States of America.
DOI: 10.1017.S0266466602183046

INSTRUMENTAL VARIABLE INTERPRETATION OF COINTEGRATION WITH INFERENCE RESULTS FOR FRACTIONAL COINTEGRATION

FRANCESC MARMOL, ALVARO ESCRIBANO, AND FELIPE M. APARICIO
Universidad Carlos III de Madrid

In this paper we propose an alternative characterization of the central notion of cointegration, exploiting the relationship between the autocovariance and the cross-covariance functions of the series. This characterization leads us to propose a new estimator of the cointegrating parameter based on the instrumental variables (IV) methodology. The instrument is a delayed regressor obtained from the conditional bivariate system of nonstationary fractionally integrated processes with a weakly stationary error correction term. We prove the consistency of this estimator and derive its limiting distribution. We also show that, in the $I(1)$ case, with a semiparametric correction simpler than the one required for the fully modified ordinary least squares (FM-OLS), our fully modified instrumental variables (FM-IV) estimator is median-unbiased, a mixture of normals, and asymptotically efficient. As a consequence, standard inference can be conducted with this new FM-IV estimator of the cointegrating parameter. We show by the use of Monte Carlo simulations that the small sample gains with the new IV estimator over OLS are remarkable.

1. INTRODUCTION

The concept of cointegration was coined by Granger (1981, 1983) and Engle and Granger (1987) when the data generating process (DGP) was generated by integrated processes. The time series x_t is integrated of order d, denoted $x_t \sim I(d)$, if $\Delta^d x_t = (1 - B)^d x_t \sim I(0)$, where d is an integer number, B is the

We acknowledge financial support from the Spanish DGICYT contracts PB98-0026, PB98-0030, and SEC2001-08090, and the European Union contract TMR ERB-4061 PL 97-0994. We are very grateful to Peter C.B. Phillips, K. Tanaka, and two anonymous referees for helpful comments and suggestions. We are also indebted to M.A. Arranz for his help in carrying out the Monte Carlo simulations. Address correspondence to: Francesc Marmol, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, C/. Madrid 126, 28903 Getafe, Madrid, Spain; e-mail: fmarmol@est-econ.uc3m.es. Felipe M. Aparicio, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, C/. Madrid 126, 28903 Getafe, Madrid, Spain; e-mail: aparicio@est-econ.uc3m.es. Alvaro Escribano, Department of Economics, Universidad Carlos III de Madrid, C/. Madrid 126, 28903 Getafe, Madrid, Spain; e-mail: alvaroe@eco.uc3m.es.

 © 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

backshift operator, and $I(0)$ denotes a weakly stationary process. We can extrapolate the concepts of integration and cointegration to the *fractional case* where now d is not an integer but a real number. Loosely speaking, we say that a series x_t is fractionally integrated of order d, if $\Delta^d x_t \sim I(0)$, $d > -1$, with $\Delta^d = \sum_{k=0}^{\infty} \Gamma(k - d)B^k / \Gamma(k + 1)\Gamma(-d)$, $\Gamma(\cdot)$ denoting the gamma function. Correspondingly, two $I(d)$ fractionally integrated of order d processes, x_t, y_t, are said to be (fractionally) cointegrated if there exists a real number $\beta \neq 0$ such that $y_t - \beta x_t \sim I(d^*)$, $d^* < d$.

Phillips and Hansen (1990) and Kitamura and Phillips (1997) developed the asymptotic properties of instrumental variables (IV) estimates in multivariate cointegrating regressions. Those studies were completed with extensive Monte Carlo simulations by Kitamura and Phillips (1995). From Phillips and Hansen (1990) we learn that, in contrast with traditional theory for stationary time series, IV regressions are consistent even when the instruments are stochastically independent of the regressors, so that the instrument selection seems to be a problem of first magnitude. Kitamura and Phillips (1997) show that sometimes the finite sample bias and root mean square error of fully modified instrumental variables (FM-IV) are exceptionally high (in fact, higher than ordinary least squares [OLS] and crude IV) as a result of the occasional occurrence of extremely large estimation errors. As argued by the authors, these outliers are due to poor initial estimates obtained by the use of IV regressions in the first stage of the fully modified methodology. When the initial IV estimates are exceptionally poor due to, e.g., poor instruments, in the second stage the fully modified procedure can amplify the effects of these preliminary estimates.

Fractional cointegration with long memory errors has been recently considered by several authors, e.g., Cheung and Lai (1993), Dolado and Marmol (1996), Marinucci (1998), Robinson and Marinucci (1998, 1999), Jeganathan (1999), Davidson (2000), Hassler, Marmol, and Velasco (2000), and Kim and Phillips (2000). In particular, Kim and Phillips (2000) develop a fractional version of the FM regression method that is robust to the presence of long-range dependence in the regression errors and asymptotically optimal in certain models of fractional cointegration. Moreover, their estimator carries over reasonably well in finite samples and compares favorably with the narrow band frequency domain least squares procedure of Robinson and Marinucci (1998).

In this paper we introduce a characterization of (fractional) cointegration exploiting the linear measures of dependence and cross-dependence introduced by Aparicio and Escribano (1998). The definition is not model dependent and therefore can be extended to the nonlinear case along the lines of Aparicio and Granger (1995) and Aparicio and Escribano (1999). This characterization leads us to propose an estimator of the cointegrating parameter β based on the IV methodology. The instrument is a delayed replica of the regressor in the conditional model when, for simplicity, the DGP is assumed to be a bivariate system of nonstationary fractionally integrated processes with a weakly stationary error correction term. Because the instrument is only asymptotically uncorrelated

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

with the innovation term, the estimator will be called the *pseudo instrumental variable* (PIV) estimator. Therefore, our suggested instrument is by definition neither spurious, because it is always correlated with the corresponding regressor, nor a poor instrument, because the correlation of unit root processes tends to one asymptotically.

The paper is organized as follows. In Section 2 we set out the model and main assumptions. In Section 3 we motivate our IV estimator by introducing the alternative characterization of fractional cointegration. Section 4 derives the consistency and limiting distribution of the crude PIV estimator. Section 5 provides some Monte Carlo evidence about the small sample behavior of PIV relative to OLS. In Section 6 we derive the asymptotic properties of the so-called fully modified pseudo instrumental variables (FM-PIV) estimator, and show that this estimator is consistent and that its limiting distribution is median-unbiased, a mixture of normals, and asymptotically efficient so that inference can be conducted in the standard way. Section 7 provides a Monte Carlo study of the small sample behavior of FM-PIV relative to FM-OLS. Finally, some concluding remarks and extensions are given in Section 8. Mathematical proofs of the theorems are gathered in the Appendix.

2. PRELIMINARY DEFINITIONS AND ASSUMPTIONS

An interesting way of classifying time series is in terms of the memory of the underlying processes, and such a memory can be represented by means of some serial dependence measures. For stationary sequences, a series is said to have *long memory* if its spectral density is unbounded at the origin (correspondingly, if it has nonsummable autocovariances), in contrast with the *short memory* case, where the spectral density is finite and positive (correspondingly, with absolutely summable autocovariances).

A particularly important member of the family of the long memory processes is the *fractionally integrated process* (Granger and Joyeux, 1980; Hosking, 1981). See Beran (1994), Robinson (1994), and Baillie (1996) for overviews of the theory and major empirical applications of this family of processes.

A fractionally integrated process x_t of order d is stationary if and only if $d < \frac{1}{2}$ and nonstationary otherwise. Herein we are concerned with the nonstationary case because this is the most relevant range when dealing with cointegration. Furthermore, as a natural step and to keep things simple, we concentrate our attention on the simplest case, where the DGP is assumed to be generated by two cointegrated nonstationary fractionally integrated time series of order d, with an $I(0)$ error correction term and without any deterministic elements.

More specifically, in this paper we assume that the relevant DGP has a triangular representation

$$y_t = \beta x_t + u_{1t}, \tag{1}$$

$$\Delta^d x_t = u_{2t}, \quad d > \frac{1}{2}, \tag{2}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

where $u_t = (u_{1t}, u_{2t})'$ is generated by the linear process

$$u_t = \sum_{j=0}^{\infty} \Pi_j \varepsilon_{t-j}, \qquad u_t = 0 \quad \text{for } t \leq 0,$$

satisfying the following regularity conditions.

Assumption 1.

(i) The sequence of matrix coefficients $\{\Pi_j\}_{j=0}^{\infty}$ is 1-summable, with $\Pi_0 = I_2$, the (2×2) identity matrix.
(ii) The sequence of random vectors $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is independent and identically distributed (i.i.d.) with zero mean and covariance matrix $\Sigma > 0$, $E\|\varepsilon_1\|^{\zeta} < \infty$, some $\zeta > 4$.
(iii) $\max_i \sup_t E|\varepsilon_{it}|^{\zeta} < \infty$, $\zeta > \max\{2, 2/(2d-1)\}$.
(iv) $\text{rank}(\Sigma) = \text{rank}(\sum_{j=0}^{\infty} \Pi_j) = 2$.

Assumption 1 implies that the u_t process is (asymptotically) strictly stationary and ergodic with continuous spectral density given by

$$f_{uu}(\lambda) = \frac{1}{2\pi} \left(\sum_{j=0}^{\infty} \Pi_j \exp(ij\lambda) \right) \left(\sum_{j=0}^{\infty} \Pi_j \exp(ij\lambda) \right)^*$$

and (2×2) long-run covariance matrix

$$\Omega = 2\pi f_{uu}(0) = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix},$$

where the asterisk denotes simultaneous transposition and complex conjugation.

3. MOTIVATION OF THE ESTIMATOR

To motivate our estimator of the cointegrating parameter β in expression (1), we use the characterization of cointegration introduced by Aparicio and Escribano (1997). Let x_t, y_t be the two $I(d)$ time series of interest, $d > \frac{1}{2}$, and let $\gamma_{yx}(\tau, t) = \text{cov}(y_t, x_{t-\tau})$ represent the *cross-covariance function* of x_t, y_t, where we make explicit the time dependence in $\gamma_{yx}(\tau, t)$ to allow for some degree of heterogeneity in the series. Aparicio and Escribano (1997) suggest the use of the ratio

$$\frac{\gamma_{yx}(\tau, t)}{\gamma_x(\tau, t)} \tag{3}$$

as an alternative characterization of cointegration, where $\gamma_x(\tau, t) = \text{cov}(x_t, x_{t-\tau})$. The next result provides the theoretical justification.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

THEOREM 1. *Let* x_t, y_t *in* DGP *(1) and (2) with* $d > \frac{1}{2}, \tau = o(t)$. *Then,*

$$\lim_{\tau \to \infty} \frac{\gamma_{yx}(\tau, t)}{\gamma_x(\tau, t)} = \beta. \tag{4}$$

Theorem 1 implies that the rates of convergence of $\gamma_{yx}(\tau, t)$ and $\gamma_x(\tau, t)$ should be the same as τ increases without bound. Intuitively, the theorem states that, under cointegration, the remote past of x_t should be as useful as the remote past of y_t in long-run linear forecasting. Moreover, Marmol, Escribano, and Aparicio (1999) prove how result (4) also extends to cointegrated regressions with (stationary and nonstationary) long memory errors.

On the other hand, it is direct to prove in the spurious case where x_t, y_t are stochastically independent that

$$\lim_{\tau \to \infty} \frac{\gamma_{yx}(\tau, t)}{\gamma_x(\tau, t)} = 0, \tag{5}$$

for all τ, t, showing that the ratio (3) is consistent against this type of spurious alternatives.

4. ESTIMATION OF THE COINTEGRATING PARAMETER

Consider now the benchmark DGP given by expressions (1) and (2). In the previous section we motivated the use of the ratio (3) as an alternative characterization of cointegration. Yet note that for large values of τ this quotient would be nothing else but the expression of an IV estimator of the cointegrating parameter β, with instrument $x_{t-\tau}$. Therefore, Theorem 1 also provides theoretical justification for using the IV methodology to test for linear cointegration among fractionally integrated processes. However, before doing that and for comparative purposes, we include the limiting distribution of the OLS estimator of β, for a sample of size T:

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2}. \tag{6}$$

THEOREM 2. *Given the* DGP *(1) and (2) with* $d > \frac{1}{2}$ *and under Assumption 1, asymptotically, as* $T \to \infty$,

$$\hat{\beta}_{\text{OLS}} \xrightarrow{p} \beta, \tag{7}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

$$T^d(\hat{\beta}_{\mathrm{OLS}} - \beta) \Rightarrow \frac{\int_0^1 B_2(d,r)\, dB_1(r)}{\int_0^1 B_2^2(d,r)\, dr}, \qquad \text{if } d > 1, \tag{8}$$

$$T(\hat{\beta}_{\mathrm{OLS}} - \beta) \Rightarrow \frac{\int_0^1 B_2(r)\, dB_1(r) + \Delta_{21}}{\int_0^1 B_2^2(r)\, dr}, \qquad \text{if } d = 1, \tag{9}$$

$$T^{2d-1}(\hat{\beta}_{\mathrm{OLS}} - \beta) \Rightarrow \frac{\Delta_{21}(d)}{\int_0^1 B_2^2(d,r)\, dr}, \qquad \text{if } \frac{1}{2} < d < 1, \tag{10}$$

with $B(r) = (B_1(r), B_2(r))'$, $r \in [0,1]$, *being a vector Brownian motion with long-run covariance matrix* Ω, $B(d,r) = (B_1(d,r), B_2(d,r))'$ *a vector fractional Brownian motion given by the functional*

$$B(d,r) = \frac{1}{\Gamma(d)} \int_0^r (r-s)^{d-1} dB(s), \tag{11}$$

$\Delta_{21} = \sum_{k=0}^{\infty} E(u_{2,0} u_{1,k})$, *and* $\Delta_{21}(d) = \sum_{k=0}^{\infty} E(\Delta x_0 u_{1,k})$ *and where* $\xrightarrow{p}$ *and* $\Rightarrow$ *denote convergence in probability and weak convergence, respectively.*

Then, for all $d > \frac{1}{2}$, OLS is a consistent estimator of β. On the other hand, the presence of nuisance parameters in the limiting OLS distributions (8)–(10) prevents us from achieving an asymptotic mixture of normals. In the case $d = 1$, these nuisance parameters are given by Δ_{21} and ω_{21}, the (2,1)-element of the long-run covariance matrix Ω. The expression $\omega_{21} \neq 0$ implies that $B_1(r)$ and $B_2(r)$ are not long-run independent, giving rise to an *endogeneity bias*. $\Delta_{21} \neq 0$, in turn, causes the so-called *serial correlation* or *second-order bias*. When $d > 1$, expression (8) shows that the second-order bias is no longer present in the limiting OLS distribution. It remains, however, the corresponding endogeneity bias. When $\frac{1}{2} < d < 1$ the bias present in the limiting OLS distribution is now of second order.

Although none of these biases affect the consistency properties of $\hat{\beta}_{\mathrm{OLS}}$, they can be important in finite samples. In effect, these nuisance parameters, Δ_{21} and ω_{21}, produce a finite sample bias in mean and median, respectively. The limiting distribution is no longer either a mixture of normals or asymptotically efficient. These nuisance parameters invalidate the use of standard distributions for testing hypotheses about the cointegrating parameter β. This is in contrast with the special case where $\Delta_{21} = \Delta_{21}(d) = \omega_{21} = 0$ in (1) and (2). In this case, Dolado and Marmol (2001) show that the OLS limiting distribution of $\hat{\beta}_{\mathrm{OLS}}$ becomes

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

$$T^d(\hat{\beta}_{\mathrm{OLS}} - \beta) \Rightarrow \frac{\int_0^1 B_2(d,r)\, dB_1(r)}{\int_0^1 B_2^2(d,r)\, dr}, \tag{12}$$

for all $d > \frac{1}{2}$, being now median-unbiased, a mixture of normals, and asymptotically efficient, with the limiting distribution depending on nuisance parameters in a simple way that permits the construction of test statistics with asymptotic chi-square distributions under the null hypothesis.

Consider now the asymptotic behavior of the following new estimator of β in (1) and (2):

$$\hat{\beta}_{\mathrm{PIV}} = \frac{\sum_{t=\tau+1}^{T} y_t x_{t-\tau}}{\sum_{t=\tau+1}^{T} x_t x_{t-\tau}}, \tag{13}$$

called the *pseudo instrumental variable* (PIV) *estimator*, because $x_{t-\tau}$ is only an instrument of x_t in the standard sense (i.e., correlated with x_t and uncorrelated with u_{1t}) for $\tau \to \infty$. In finite samples, however, $x_{t-\tau}$ is not necessarily independent of the innovation u_{1t}, and consequently $x_{t-\tau}$, strictly speaking, is not an instrument. For fixed τ, PIV could be covered by the IV framework developed by Phillips and Hansen (1990) with $n_1 = n_2 = 1$ and $y_{3t} = x_{t-\tau}$ in their notation.

THEOREM 3. *Given the* DGP *(1) and (2) with* $d > \frac{1}{2}$ *and under Assumption 1, asymptotically, as* $T \to \infty, \tau \to \infty$, *and* $T^{-1}\tau \to 0$,

$$\hat{\beta}_{\mathrm{PIV}} \xrightarrow{p} \beta, \tag{14}$$

$$T^d(\hat{\beta}_{\mathrm{PIV}} - \beta) \Rightarrow \frac{\int_0^1 B_2(d,r)\, dB_1(r)}{\int_0^1 B_2^2(d,r)\, dr}, \qquad \text{if } d > 1, \tag{15}$$

$$T(\hat{\beta}_{\mathrm{PIV}} - \beta) \Rightarrow \frac{\int_0^1 B_2(r)\, dB_1(r)}{\int_0^1 B_2^2(r)\, dr}, \qquad \text{if } d = 1, \tag{16}$$

$$T^{2d-1}(\hat{\beta}_{\mathrm{PIV}} - \beta) \xrightarrow{p} 0, \qquad \text{if } \frac{1}{2} < d < 1. \tag{17}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

From this theorem we can make the following comments. First, as expected from Theorem 1, for all $d > \frac{1}{2}$, PIV is a consistent estimator of β. Second, when $\frac{1}{2} < d < 1$, $\hat{\beta}_{\text{PIV}}$ is $o_p(T^{1-2d})$ with a degenerate limiting distribution. Third, for $d > 1$, $\hat{\beta}_{\text{PIV}}$ and $\hat{\beta}_{\text{OLS}}$ have the same limiting distribution, i.e., nonstandard with endogeneity bias and not asymptotically efficient. Fourth, when $d = 1$, i.e., in the unit root case, $\omega_{21} \neq 0$ but $\Delta_{21} = 0$, eliminating one of the sources of the finite sample bias in the estimation of β. The estimator, however, is neither a mixture of normals nor asymptotically efficient, and standard inference remains invalid.

It is of interest for the Monte Carlo experiments to provide some insights on the finite sample performance of $\hat{\beta}_{\text{PIV}}$ relative to $\hat{\beta}_{\text{OLS}}$ and other competing estimators. Consider the following particular case of DGP (1) and (2):

$$y_t = \beta x_t + u_{1t}, \qquad u_{1t} = \varphi u_{1,t-1} + \varepsilon_{1t}, \qquad |\varphi| < 1, \tag{18}$$

$$\Delta x_t = \varepsilon_{2t}, \tag{19}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim NID\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \theta\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix} \right\}. \tag{20}$$

Under this setup, Gonzalo (1994) proves that

$$\begin{aligned} T(\hat{\beta}_{\text{OLS}} - \beta) \Rightarrow & \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \\ & \times \left\{ \left(\frac{\sigma_1}{1-\varphi} \right)(1-\theta^2)^{1/2} \int_0^1 B_2(r)dW_1(r) \right. \\ & \left. + \left(\frac{1}{1-\varphi} \right)\theta \frac{\sigma_1}{\sigma_2} \int_0^1 B_2(r)dB_2(r) + \left(\frac{1}{1-\varphi} \right)\theta\sigma_1\sigma_2 \right\}, \end{aligned} \tag{21}$$

$$\begin{aligned} T(\hat{\beta}_{\text{NLS}} - \beta) \Rightarrow & \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \\ & \times \left\{ \left(\frac{\sigma_1}{1-\varphi} \right)(1-\theta^2)^{1/2} \int_0^1 B_2(r)dW_1(r) \right. \\ & \left. + \left(\frac{1}{1-\varphi} \right)\left(\beta + \theta \frac{\sigma_1}{\sigma_2} \right) \int_0^1 B_2(r)dB_2(r) \right\}, \end{aligned} \tag{22}$$

and

$$T(\hat{\beta}_{\text{MLECM}} - \beta) \Rightarrow \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \left(\frac{\sigma_1}{1-\varphi} \right)(1-\theta^2)^{1/2} \int_0^1 B_2(r)dW_1(r) \right\}, \tag{23}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

where $W_1(r)$ is a standard Brownian motion independent of $B_2(r)$ and where $\hat{\beta}_{\mathrm{NLS}}$ and $\hat{\beta}_{\mathrm{MLECM}}$ stand for the nonlinear least squares and the maximum likelihood in fully specified error correction model (ECM) estimators of β, respectively. Also, from (16) it is not difficult to show that

$$T(\hat{\beta}_{\mathrm{PIV}} - \beta) \Rightarrow \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \times \left\{ \left(\frac{\sigma_1}{1-\varphi} \right)(1-\theta^2)^{1/2} \int_0^1 B_2(r)dW_1(r) + \left(\frac{1}{1-\varphi} \right)\theta \frac{\sigma_1}{\sigma_2} \int_0^1 B_2(r)dB_2(r) \right\}. \quad \textbf{(24)}$$

From expressions (21)–(24), the following comments can be readily deduced. First, with respect to the limiting distribution of the OLS estimator, note that it involves three different parts. The first one is a mixture of normals; the second one is a unit root term, $\{\int_0^1 B_2^2(r)dr\}^{-1}\int_0^1 B_2(r)dB_2(r)$, making the distribution asymmetrical, and the third one represents the serial correlation bias inducing a mean bias in the distribution. The second and third terms are due to the presence of ω_{21} and Δ_{21}, respectively.

Second, comparing the asymptotic distributions of OLS and NLS, we can see that Δ_{21} is no longer present in the limiting distribution of the latter estimator. However, as argued by Gonzalo (1994), the presence of the unit root term can make OLS perform better than NLS in finite samples. This is the case if β is large and θ is close to zero.

Third, the presence of the nuisance parameter ω_{21} in the limiting distribution of PIV gives rise again to the presence of the unit root term and the consequent median bias. Yet, as expected by looking at expressions (9) and (16), comparing the asymptotic distributions of PIV and OLS, the former clearly appears preferable for any value of the parameter space. On the other hand, comparing the asymptotic distributions of PIV and NLS, shows that NLS can be less asymmetric than PIV if, e.g., β is negative and θ positive.

Fourth, the limiting distribution of MLECM is a median-unbiased mixture of normals. In fact, it is asymptotically efficient as proved by Phillips (1991). Moreover, the remaining nuisance parameters are located in such a way that hypothesis tests can be conducted using standard asymptotic chi-square tests. Finally, it is clear that OLS, NLS, and PIV are no longer optimal estimators in the light of expression (23).

5. SMALL SAMPLE STUDY OF OLS AND PIV ESTIMATORS

To assess the relative finite sample performance of the OLS and PIV estimators, in this section we shall conduct a small Monte Carlo experiment. For this, the proposed DGP is an extension to the fractional case of the DGP used by

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

Banerjee, Dolado, Hendry, and Smith (1986), Engle and Granger (1987), Gonzalo (1994), and Arranz and Escribano (2000), among others.

The DGP is a bivariate cointegrating system of nonstationary fractionally integrated series of order d and has two equivalent representations. The first one is based on the cointegrating regression,

$$y_t = \beta x_t + u_{1t} \tag{25}$$

$$u_{1t} = (a - \beta)\Delta^d x_t + (1 + b)u_{1,t-1} + \eta_{1t}, \tag{26}$$

$$\Delta^d x_t = \varepsilon_{2t}, \tag{27}$$

where the errors η_{1t} and ε_{2t} are independent $NID(0, \sigma^2_{\eta_1})$ and $NID(0, \sigma^2_2)$ variables, respectively.

The second representation is based on the following ECM for fractionally integrated series:

$$\Delta y_t = \beta \Delta x_t + (a - \beta)\Delta^d x_t + b(y_{t-1} - \beta x_{t-1}) + \eta_{1t}, \tag{28}$$

$$\Delta^d x_t = \varepsilon_{2t}. \tag{29}$$

Notice that the ECM will only have the usual form, with variables only in first differences and levels, when the *common factor restriction*, $a = \beta$, is satisfied or in the unit root ($d = 1$) case. In effect, in this last case the ECM model becomes

$$\Delta y_t = a\Delta x_t + b(y_{t-1} - \beta x_{t-1}) + \eta_{1t}.$$

In general, as we will see in the Monte Carlo experiment, the values of the parameters a and b are important in terms of the small sample biases of the cointegrating vector. The reason is that the parameter a is related to the covariance of $(\varepsilon_{1t}, \varepsilon_{2t})'$, where $\varepsilon_{1t} = (a - \beta)\Delta^d + \eta_{1t}$. Thus, e.g., it is clear from equations (18)–(20) that $\theta = (a - \beta)\sigma_2/\sigma_1$ and $b = \varphi - 1$, so that the closer b is to 0, the higher the degree of autocorrelation is in the equilibrium errors u_{1t}. Therefore, the limiting distribution of the estimators of the long-run parameter β derived in equations (21)–(24) is affected by the parameter values assigned to $(a - \beta)$ and b in the experiment.

In the Monte Carlo simulations we consider the following parameter values: $\beta = 1$, $(a - \beta) = -2, -1, -0.5$, and 0 (common factor restriction), $b = -0.2$ and -0.5, $\sigma^2_{\eta_1} = 1$, and $\sigma^2_2 = 1$. The sample sizes are $T = 300$ and 500 observations, the long memory parameters of the variables are $d = 0.7$, 1, and 1.3, and the number of replications of each experiment is 20,000.

In Tables 1–5 the small sample bias of the OLS and PIV estimators is analyzed for different values of the lag τ of the IV $x_{t-\tau}$ and using five different statistics: mean bias, median bias, mean squared error (MSE), interquartile range (IQR), and the concentration probabilities $\Pr(|\hat{\beta} - \beta| \leq 0.05)$. Tables 1–3 report the simulation results obtained for $I(1)$ variables, whereas Tables 4 and 5 do so for fractionally integrated processes with $d = 0.7$ and 1.3, respectively.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 1. OLS and PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
−2.0	−.917E−01	−.673E−01	.154E−01	.831E−01	.367
−1.0	−.460E−01	−.337E−01	.471E−02	.507E−01	.652
−0.5	−.231E−01	−.168E−01	.202E−02	.381E−01	.811
0.0	−.302E−03	−.326E−03	.112E−02	.333E−01	.884
			$\tau = 1$		
−2.0	−.639E−01	−.497E−01	.783E−02	.620E−01	.498
−1.0	−.321E−01	−.248E−01	.283E−02	.424E−01	.744
−0.5	−.162E−01	−.126E−01	.158E−02	.358E−01	.843
0.0	−.305E−03	−.271E−03	.115E−02	.335E−01	.881
			$\tau = 2$		
−2.0	−.412E−01	−.349E−01	.410E−02	.510E−01	.632
−1.0	−.208E−01	−.177E−01	.193E−02	.387E−01	.803
−0.5	−.105E−01	−.914E−02	.138E−02	.350E−01	.858
0.0	−.317E−03	−.296E−03	.120E−02	.337E−01	.877
			$\tau = 3$		
−2.0	−.227E−01	−.228E−01	.278E−02	.496E−01	.714
−1.0	−.115E−01	−.121E−01	.163E−02	.388E−01	.830
−0.5	−.591E−02	−.634E−02	.134E−02	.351E−01	.864
0.0	−.326E−03	−.305E−03	.124E−02	.339E−01	.874
			$\tau = 4$		
−2.0	−.743E−02	−.131E−01	.294E−02	.537E−01	.730
−1.0	−.388E−02	−.756E−02	.171E−02	.401E−01	.834
−0.5	−.210E−02	−.407E−02	.140E−02	.356E−01	.863
0.0	−.321E−03	−.336E−03	.130E−02	.340E−01	.872
			$\tau = 5$		
−2.0	.510E−02	−.588E−02	.398E−02	.602E−01	.715
−1.0	.240E−02	−.383E−02	.201E−02	.426E−01	.827
−0.5	.105E−02	−.227E−02	.152E−02	.362E−01	.857
0.0	−.306E−03	−.339E−03	.136E−02	.341E−01	.867

Note: DGP is $\Delta y_t = a\Delta x_t - 0.2(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0,1)$, $\text{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 300, 20,000 replications.

Consider first the unit root case. Tables 1–3 report the finite sample behavior of the PIV estimator when $\tau = 0$ (OLS),1,…,5. On the one hand, when $(a - \beta) = 0$, i.e., when the regressor x_t is strongly exogenous with respect to β, OLS and PIV perform (as expected) equally well, with small mean and median biases, similar IQR, and high concentration probability in moderate sample sizes. On the other hand, OLS estimation biases increase with the absolute value of $(a - \beta)$. In particular, when $(a - 1) = 2$ and $b = -0.2$, the concentration probability of the OLS estimator is only about 36% when $T = 300$ and about 55% when $T = 500$.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 2. OLS and PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
−2.0	−.610E−01	−.449E−01	.669E−02	.541E−01	.551
−1.0	−.306E−01	−.223E−01	.202E−02	.326E−01	.803
−0.5	−.154E−01	−.112E−01	.849E−03	.246E−01	.915
0.0	−.204E−03	−.201E−03	.452E−03	.210E−01	.966
			$\tau = 1$		
−2.0	−.427E−01	−.331E−01	.336E−02	.402E−01	.688
−1.0	−.215E−01	−.166E−01	.119E−02	.273E−01	.878
−0.5	−.109E−01	−.837E−02	.648E−03	.228E−01	.940
0.0	−.221E−03	−.215E−03	.463E−03	.210E−01	.964
			$\tau = 2$		
−2.0	−.279E−01	−.234E−01	.171E−02	.326E−01	.815
−1.0	−.141E−01	−.120E−01	.784E−03	.247E−01	.926
−0.5	−.716E−02	−.613E−02	.552E−03	.221E−01	.952
0.0	−.241E−03	−.212E−03	.473E−03	.211E−01	.962
			$\tau = 3$		
−2.0	−.159E−01	−.155E−01	.107E−02	.305E−01	.888
−1.0	−.809E−02	−.821E−02	.629E−03	.240E−01	.945
−0.5	−.417E−02	−.426E−02	.519E−03	.220E−01	.957
0.0	−.257E−03	−.181E−03	.484E−03	.212E−01	.961
			$\tau = 4$		
−2.0	−.619E−02	−.906E−02	.104E−02	.326E−01	.898
−1.0	−.323E−02	−.521E−02	.626E−03	.247E−01	.948
−0.5	−.175E−02	−.280E−02	.525E−03	.222E−01	.958
0.0	−.266E−03	−.230E−03	.495E−03	.213E−01	.960
			$\tau = 5$		
−2.0	.173E−02	−.425E−02	.135E−02	.364E−01	.877
−1.0	.734E−03	−.276E−02	.710E−03	.261E−01	.939
−0.5	.233E−03	−.160E−02	.554E−03	.227E−01	.954
0.0	−.267E−03	−.252E−03	.507E−03	.215E−01	.958

Note: DGP is $\Delta y_t = a\Delta x_t - 0.2(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\text{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 500, 20,000 replications.

By contrast, the improvements obtained with the PIV estimator for the best value of τ in each case are quite impressive. All and all, PIV outperforms OLS in terms of biases, IQR, and concentration probabilities, and, what is more important, uniformly in τ, which means that the improvement obtained with the new IV estimator over OLS does not heavily rely on the particular value selected of τ. As we established in Theorem 3, the lag τ can increase with the sample size, and in Table 2 we observe that increasing the sample size from 300 to 500 makes the best τ increase from $\tau = 3$ to $\tau = 4$. Notice also that,

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 3. OLS and PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
−2.0	−.249E−01	−.180E−01	.115E−02	.221E−01	.884
−1.0	−.125E−01	−.898E−02	.346E−03	.132E−01	.976
−0.5	−.629E−02	−.451E−02	.144E−03	.992E−02	.996
0.0	−.968E−04	−.840E−04	.752E−04	.850E−02	.999
			$\tau = 1$		
−2.0	−.604E−02	−.590E−02	.170E−03	.123E−01	.997
−1.0	−.307E−02	−.315E−02	.100E−03	.960E−02	.999
−0.5	−.159E−02	−.165E−02	.827E−04	.878E−02	.999
0.0	−.104E−03	−.904E−04	.771E−04	.852E−02	.999
			$\tau = 2$		
−2.0	.351E−02	.164E−03	.283E−03	.165E−01	.981
−1.0	.170E−02	−.144E−03	.128E−03	.111E−01	.995
−0.5	.792E−03	−.173E−03	.901E−04	.919E−02	.998
0.0	−.113E−03	−.859E−04	.789E−04	.858E−02	.999
			$\tau = 3$		
−2.0	.835E−02	.303E−02	.506E−03	.202E−01	.955
−1.0	.412E−02	.126E−02	.184E−03	.125E−01	.988
−0.5	.200E−02	.562E−03	.105E−03	.963E−02	.997
0.0	−.119E−03	−.871E−04	.805E−04	.864E−02	.999
			$\tau = 4$		
−2.0	.108E−01	.438E−02	.669E−03	.223E−01	.938
−1.0	.536E−02	.193E−02	.226E−03	.134E−01	.984
−0.5	.262E−02	.936E−03	.117E−03	.997E−02	.996
0.0	−.116E−03	−.898E−04	.823E−04	.865E−02	.999
			$\tau = 5$		
−2.0	.121E−01	.504E−02	.776E−03	.235E−01	.927
−1.0	.601E−02	.228E−02	.254E−03	.139E−01	.981
−0.5	.295E−02	.108E−02	.126E−03	.101E−01	.995
0.0	−.109E−03	−.716E−04	.848E−04	.867E−02	.999

Note: DGP is $\Delta y_t = a\Delta x_t - 0.5(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\text{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 500, 20,000 replications.

ceteris paribus, the IQR decreases with the sample size and that it does not heavily depend on τ for fixed T and $(a - \beta)$.

When $b = -0.5$ the AR(1) coefficient of the equilibrium errors (u_{1t}) of equation (26) is 0.5. Therefore, the memory or temporal dependence is reduced, giving rise to lower estimation biases and lower required values of τ; see Table 3. Now the best results for a sample size of 500 observations are given by the PIV estimator with $\tau = 1$. OLS performs worse than PIV but better than the previous OLS when the adjustment coefficient was $b = -0.2$.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 4. OLS and PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
−2.0	−.354	−.312	.170	.298	.005
−1.0	−.177	−.155	.435E−01	.153	.081
−0.5	−.885E−01	−.760E−01	.119E−01	.847E−01	.326
0.0	.774E−04	.401E−04	.146E−02	.449E−01	.827
			$\tau = 1$		
−2.0	−.171	−.152	.398E−01	.139	.078
−1.0	−.854E−01	−.738E−01	.113E−01	.817E−01	.326
−0.5	−.427E−01	−.356E−01	.419E−02	.584E−01	.609
0.0	.117E−03	.627E−04	.182E−02	.483E−01	.797
			$\tau = 2$		
−2.0	−.632E−01	−.619E−01	.848E−02	.779E−01	.374
−1.0	−.316E−01	−.307E−01	.368E−02	.588E−01	.619
−0.5	−.158E−01	−.150E−01	.248E−02	.524E−01	.731
0.0	.555E−04	.290E−03	.208E−02	.504E−01	.776
			$\tau = 3$		
−2.0	−.328E−02	−.191E−01	.817E−02	.896E−01	.508
−1.0	−.161E−02	−.932E−02	.380E−02	.636E−01	.670
−0.5	−.777E−03	−.464E−02	.271E−02	.548E−01	.733
0.0	.584E−04	.352E−03	.235E−02	.515E−01	.763
			$\tau = 4$		
−2.0	.302E−01	.257E−02	.149E−01	.109	.499
−1.0	.152E−01	.105E−02	.573E−02	.717E−01	.655
−0.5	.766E−02	.530E−03	.343E−02	.586E−01	.720
0.0	.151E−03	.335E−03	.267E−02	.531E−01	.748
			$\tau = 5$		
−2.0	.489E−01	.140E−01	.219E−01	.122	.480
−1.0	.245E−01	.669E−02	.766E−02	.773E−01	.637
−0.5	.123E−01	.312E−02	.412E−02	.606E−01	.705
0.0	.122E−03	.353E−03	.298E−02	.543E−01	.737

Note: DGP is $\Delta y_t = \Delta x_t + (a - \beta)\Delta x_t^d - 0.5(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta^d x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\text{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, $d = 0.7$, sample size 300, 20,000 replications.

As expected with weakly exogenous variables for the cointegrating parameter, the improvement of PIV over OLS is marginal when the equilibrium errors are white noise, i.e., $b = -1$. In this case the best PIV estimator, again in terms of having the smallest biases and IQR, is obtained when $\tau = 1$ (results not reported but available from the authors upon request).

Table 4 shows the relative performance of PIV and OLS when the variables are fractionally integrated with $d = 0.7$. As suggested by equation (12), OLS and PIV have a nontrivial mean and median-unbiased limiting distribution under the common factor restriction $(a - 1) = 0$. In this particular case, OLS

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 5. OLS and PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (OLS)		
−2.0	.545E−03	−.204E−03	.377E−04	.554E−02	1.000
−1.0	.289E−03	−.136E−03	.190E−04	.362E−02	1.000
−0.5	.161E−03	−.697E−04	.143E−04	.297E−02	1.000
0.0	.332E−04	.275E−04	.128E−04	.272E−02	1.000
			$\tau = 1$		
−2.0	.292E−02	.815E−03	.781E−04	.678E−02	.997
−1.0	.148E−02	.357E−03	.293E−04	.411E−02	1.000
−0.5	.755E−03	.167E−03	.170E−04	.313E−02	1.000
0.0	.330E−04	.281E−04	.128E−04	.273E−02	1.000
			$\tau = 2$		
−2.0	.409E−02	.126E−02	.117E−03	.756E−02	.993
−1.0	.206E−02	.582E−03	.391E−04	.443E−02	.999
−0.5	.105E−02	.275E−03	.196E−04	.323E−02	1.000
0.0	.342E−04	.276E−04	.130E−04	.272E−02	1.000
			$\tau = 3$		
−2.0	.466E−02	.146E−02	.140E−03	.797E−02	.991
−1.0	.235E−02	.690E−03	.451E−04	.463E−02	.999
−0.5	.119E−02	.333E−03	.212E−04	.331E−02	1.000
0.0	.354E−04	.258E−04	.132E−04	.274E−02	1.000
			$\tau = 4$		
−2.0	.493E−02	.156E−02	.153E−03	.818E−02	.989
−1.0	.248E−02	.743E−03	.484E−04	.473E−02	.999
−0.5	.126E−02	.357E−03	.222E−04	.335E−02	1.000
0.0	.387E−04	.283E−04	.133E−04	.275E−02	1.000
			$\tau = 5$		
−2.0	.505E−02	.161E−02	.159E−03	.824E−02	.988
−1.0	.254E−02	.757E−03	.502E−04	.476E−02	.998
−0.5	.129E−02	.368E−03	.229E−04	.337E−02	1.000
0.0	.406E−04	.231E−04	.135E−04	.274E−02	1.000

Note: DGP is $\Delta y_t = \Delta x_t + (a - \beta)\Delta x_t^d - 0.2(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta^d x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\text{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, $d = 1.3$, sample size 300, 20,000 replications.

occasionally performs better than PIV as a result of a greater number of degrees of freedom. For $(a - 1) \neq 0$ the concentration probabilities of OLS never go over 32% and in most cases are below 1%, and the biases are in general very large, reaching a mean bias of 35% for $(a - 1) = -2$. However, when considering the PIV estimator, results change dramatically, with concentration probabilities about 75% and smaller IQR. Of course, this is what we expected in view of expression (17).

Finally, when d is larger than one, from expressions (8) and (15) we know that OLS and PIV have the same limiting distribution. Therefore in this case

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

both OLS and PIV estimators should perform equally well in moderate to large samples. Our Monte Carlo experiment strongly supports this claim (see Table 5).

6. ASYMPTOTIC EFFICIENCY OF PIV IN THE UNIT ROOT CASE

From the previous section we obtained a lack of optimality of PIV for all $d > \frac{1}{2}$. Moreover, with respect to OLS we found advantages in using PIV only in the unit root case. Thus, in the rest of the paper we will focus our attention on the DGP (1) and (2) in the particular case of $d = 1$.

To overcome the optimality problems of PIV in the unit root case, we suggest a semiparametric correction of this estimator such as the one originally proposed by Phillips and Hansen (1990). This procedure, known as *fully modified*, proceeds as follows. As a first step, define the following kernel estimators:

$$\hat{\omega}_{21} = \sum_{j=-k}^{k} \ell\left(\frac{j}{k}\right) T^{-1} \sum{}'(\Delta x_{t-j}\hat{u}_{1t}), \tag{30}$$

$$\hat{\omega}_{22} = \sum_{j=-k}^{k} \ell\left(\frac{j}{k}\right) T^{-1} \sum{}'(\Delta x_{t-j}\Delta x_t), \tag{31}$$

where $\sum'$ indicates summation over $1 \leq t, t-j \leq T$ and $\hat{u}_{1t}$ denotes the PIV residuals from equation (1), i.e., $\hat{u}_{1t} = y_t - \hat{\beta}_{PIV}x_t$.

The kernel function $\ell(\cdot): \Re \to [-1,1]$ is assumed to be a twice continuously differentiable even function with $\ell(0) = 1$, $\ell'(0) = 0, \ell''(0) \neq 0$, and $\ell(x) = 0$ for $|x| \geq 1$. Further, we also assume that any of the Parzen, quadratic spectral, or Tukey–Hanning kernels are used in the estimation of the elements of Ω.

With respect to the truncation or bandwidth parameter k, to conveniently characterize the rates of expansion of $k = k(T)$ as $T \to \infty$, we will use the expansion rate order symbol O_e defined in Phillips (1995). We said that $k = O_e(T^{\xi})$ if $k \sim \varsigma_T T^{\xi}$ as $T \to \infty$, where ς_T is slowly varying at infinity.

Following the same arguments as in Phillips (1995) for the OLS case, it can be shown that if $k, T \to \infty$ but $kT^{-1/2} \to 0$, then $\hat{\omega}_{2i} \xrightarrow{p} \omega_{21}$, $i = 1,2$. Hence, in terms of O_e, this implies that $k = O_e(T^{\xi})$ for some $\xi \in (0, \frac{1}{2})$. This will be our assumption about the bandwidth expansion rate of k as $T \to \infty$. It is worth mentioning that this expansion rate includes the optimal growth rate $k \sim cT^{1/5}$ (cf. Andrews, 1991), with c a constant, that applies when minimizing the asymptotic mean squared error of kernel estimates such as (30) and (31).

Now define the endogeneity bias-corrected disturbance

$$u_{1t}^{+} = u_{1t} - \omega_{12}\omega_{22}^{-1}u_{2t}, \tag{32}$$

which has zero coherence at the origin with u_{2t}, and its feasible counterpart

$$\hat{u}_{1t}^{+} = \hat{u}_{1t} - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\Delta x_t. \tag{33}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

Subtracting $\hat{\omega}_{12}\hat{\omega}_{22}^{-1}\Delta x_t$ from both sides of (1) yields

$$\hat{y}_t^+ = \beta x_t + \hat{u}_{1t}^+, \tag{1'}$$

where $\hat{y}_t^+ = y_t - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\Delta x_t$. Therefore, our proposed estimator, called *fully modified pseudo instrumental variable* (FM-PIV), has the expression

$$\hat{\beta}_{\mathrm{PIV}}^+ = \frac{\sum_{t=\tau+1}^{T} \hat{y}_t^+ x_{t-\tau}}{\sum_{t=\tau+1}^{T} x_t x_{t-\tau}}, \tag{34}$$

which contrasts with the formulation of the FM-OLS estimator by Phillips and Hansen (1990), $\hat{\beta}_{\mathrm{OLS}}^+$, say, given by

$$\hat{\beta}_{\mathrm{OLS}}^+ = \frac{\sum_{t=1}^{T} \hat{y}_t^+ x_t - T\hat{\Delta}_{21}^+}{\sum_{t=1}^{T} x_t^2}, \tag{34'}$$

where $\hat{y}_t^+ = y_t - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\Delta x_t$, $\hat{\Delta}_{21}^+ = \hat{\Delta}_{21} - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\hat{\Delta}_{22}$, and $\hat{\Delta}_{ij}, \hat{\omega}_{ij}, i,j = 1,2$, are constructed from equations (30) and (31) using the OLS residuals $\hat{u}_{1t} = y_t - \hat{\beta}_{\mathrm{OLS}}x_t$.

THEOREM 4. *Given the* DGP *(1) and (2) with* $d = 1$*, then, under Assumption 1 with* $k = O_e(T^{\xi})$ *for some* $\xi \in (0, \frac{1}{2})$ *and* $\tau = o(T^{1/4})$*, asymptotically, as* $T \to \infty$ *and* $\tau \to \infty$*,*

$$\hat{\beta}_{\mathrm{PIV}}^+ \xrightarrow{p} \beta, \tag{35}$$

and

$$T(\hat{\beta}_{\mathrm{PIV}}^+ - \beta) \Rightarrow \frac{\int_0^1 B_2(r)\, dB_{1.2}(r)}{\int_0^1 B_2^2(r)\, dr}, \tag{36}$$

where $B_{1.2}(r)$ *denotes a Brownian motion with long-run covariance* $\omega_{1.2} = \omega_{11} - \omega_{12}\omega_{22}^{-1}$ *and independent of* $B_2(r)$*.*

From equation (35) it follows that FM-PIV is a consistent estimator of β. Moreover, from equation (36) it is also median-unbiased, a mixture of normals, and asymptotically efficient (Phillips, 1991). In particular, and returning to the DGP (18)–(20), Gonzalo (1994) proves that in this case $\omega_{1.2}^{1/2} = \sigma_1(1-\theta)^{1/2}/(1-\varphi)$, and because $B_{1.2}(r) \equiv \omega_{1.2}^{1/2} W_1(r)$, it follows from equations (23) and

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

(24) that FM-PIV is asymptotically equivalent to MLECM and, thus, asymptotically efficient.

As a consequence, standard inference remains valid. In particular, consider the customary t-ratio of β,

$$t_\beta = \frac{\hat{\beta}^+_{\mathrm{PIV}} - \beta}{(\hat{\omega}_{1.2})^{1/2}\left(\sum_{t=1}^{T} x_t^2\right)^{-1/2}}. \tag{37}$$

Under the null hypothesis $H_0: \beta = \beta_0$, it is straightforward to prove that, asymptotically, $t_{\beta_0} \Rightarrow N(0,1)$. More general, the resulting test statistics from our FM-PIV estimator will have limiting chi-square distributions, thereby removing the obstacles to inference in cointegrated systems that were presented by the nuisance parameter dependencies in the PIV limiting distribution.

Finally, it is well known that, under Assumption 1,

$$T(\hat{\beta}^+_{\mathrm{OLS}} - \beta) \Rightarrow \frac{\int_0^1 B_2(r)\, dB_{1.2}(r)}{\int_0^1 B_2^2(r)\, dr} \tag{38}$$

as proved by Phillips and Hansen (1990), so that $\hat{\beta}^+_{\mathrm{PIV}}$ and $\hat{\beta}^+_{\mathrm{OLS}}$ are asymptotically equivalents. This fact, however, only holds for values of T arbitrarily large and, of course, does not exclude different behavior in small to medium samples. The performance in finite samples of both estimators is the topic of the next section.

7. FM-PIV AND FM-OLS ESTIMATORS: MONTE CARLO EVIDENCE

The Monte Carlo experiment was described in Section 5. Here we use the same DGP to study the small sample behavior of FM-PIV relative to FM-OLS. These estimators have been constructed as follows.

Step 1. Obtain the PIV estimator (equation (13)) for $\tau = 0$ (OLS), $1,\ldots,8$.
Step 2. Compute the residuals $\hat{u}_{1t} = y_t - \hat{\beta}_{\mathrm{PIV}} x_t$.
Step 3. Obtain $\hat{\omega}_{ij}$, $i,j = 1,2$, as in expressions (30) and (31). In our experiments, we used the quadratic spectral kernel and selected the bandwidth using the automatic procedure of Andrews (1991), after prewhitening with a first-order vector autoregression prior to kernel estimation, as suggested by Andrews and Monahan (1992).
Step 4. Obtain $\hat{y}_t^+ = y_t - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\Delta x_t$.
Step 5. Compute $\hat{\beta}^+_{\mathrm{PIV}}$ using expression (34).

Table 6 reports the results of the DGP given in equations (28) and (29) with parameter values $b = -0.2$ and a sample size of 300. FM-OLS performs better

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 6. FM-OLS and FM-PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (FM-OLS)		
−2.0	−.761E−01	−.551E−01	.111E−01	.718E−01	.458
−1.0	−.382E−01	−.275E−01	.364E−02	.461E−01	.710
−0.5	−.192E−01	−.138E−01	.176E−02	.367E−01	.831
0.0	−.288E−03	−.247E−03	.110E−02	.335E−01	.886
			$\tau = 1$		
−2.0	−.760E−01	−.550E−01	.111E−01	.716E−01	.459
−1.0	−.382E−01	−.274E−01	.365E−02	.461E−01	.710
−0.5	−.192E−01	−.138E−01	.177E−02	.367E−01	.830
0.0	−.298E−03	−.176E−03	.110E−02	.335E−01	.886
			$\tau = 2$		
−2.0	−.535E−01	−.410E−01	.589E−02	.547E−01	.582
−1.0	−.269E−01	−.204E−01	.236E−02	.398E−01	.781
−0.5	−.136E−01	−.104E−01	.147E−02	.351E−01	.854
0.0	−.311E−03	−.233E−03	.112E−02	.337E−01	.885
			$\tau = 3$		
−2.0	−.351E−01	−.291E−01	.325E−02	.458E−01	.697
−1.0	−.177E−01	−.149E−01	.173E−02	.370E−01	.825
−0.5	−.901E−02	−.769E−02	.134E−02	.346E−01	.863
0.0	−.313E−03	−.273E−03	.114E−02	.338E−01	.882
			$\tau = 4$		
−2.0	−.200E−01	−.197E−01	.227E−02	.439E−01	.763
−1.0	−.102E−01	−.103E−01	.152E−02	.367E−01	.846
−0.5	−.524E−02	−.542E−02	.133E−02	.346E−01	.866
0.0	−.317E−03	−.240E−03	.117E−02	.337E−01	.879
			$\tau = 5$		
−2.0	−.764E−02	−.122E−01	.231E−02	.467E−01	.781
−1.0	−.397E−02	−.672E−02	.157E−02	.377E−01	.849
−0.5	−.214E−02	−.351E−02	.138E−02	.350E−01	.866
0.0	−.301E−03	−.217E−03	.120E−02	.339E−01	.876
			$\tau = 6$		
−2.0	.253E−02	−.612E−02	.295E−02	.503E−01	.772
−1.0	.112E−02	−.374E−02	.177E−02	.391E−01	.842
−0.5	.409E−03	−.210E−02	.148E−02	.354E−01	.862
0.0	−.300E−03	−.280E−03	.124E−02	.339E−01	.873
			$\tau = 7$		
−2.0	−.934E−02	−.107E−01	.178E−02	.381E−01	.829
−1.0	−.482E−02	−.559E−02	.146E−02	.352E−01	.859
−0.5	−.256E−02	−.296E−02	.138E−02	.344E−01	.867
0.0	−.313E−03	−.267E−03	.128E−02	.341E−01	.872
			$\tau = 8$		
−2.0	−.102E−01	−.101E−01	.180E−02	.372E−01	.833
−1.0	−.524E−02	−.519E−02	.149E−02	.350E−01	.858
−0.5	−.277E−02	−.278E−02	.141E−02	.346E−01	.865
0.0	−.302E−03	−.269E−03	.134E−02	.343E−01	.868

Note: DGP is $\Delta y_t = a\Delta x_t - 0.2(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\mathrm{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 300, 20,000 replications.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

than crude OLS for all the parameter values of $(a - \beta)$ except in the common factor case, where both perform equally well.

As for the FM-PIV estimator defined in equation (34), the best results for $b = -0.2$ are for $\tau = 5$ or 6. For example the maximum mean bias with FM-PIV for $\tau = 5$ is 0.76% (with FM-OLS it is 7.6%) with corresponding concentration probability about 80% (about 46% in the FM-OLS case) and smaller IQR. Increasing the sample size to 500 observations only helps to confirm the uniform superiority of FM-PIV over FM-OLS (see Table 7). Taking $b = -0.5$, and therefore reducing the autocorrelation of the equilibrium errors, improves both FM-OLS and FM-PIV estimators (see Table 8). Nonetheless, even in this case the performance of the FM-PIV estimator dominates over FM-OLS, and the best lag of the instrumental variable $x_{t-\tau}$ is now reduced to $\tau = 2$.

8. CONCLUDING REMARKS

In this paper we have studied the estimation and inference properties of a recent characterization of the concept of cointegration proposed by Aparicio and Escribano (1997). Such a characterization exploited the relationship between the autocovariance and the cross-covariance functions of the series. This, in turn, led us to propose an estimator of the cointegrating parameter based on the IV methodology, where the instrument is the lagged regressor from the cointegrating equation.

We have shown here that the new PIV estimator is consistent for the cointegrating parameter β. Moreover, in the unit root $(d = 1)$ case, we showed that PIV is preferable to OLS, because the limiting distribution of the PIV estimator is free of the serial correlation bias that characterizes the OLS limiting distribution. The small sample results obtained by Monte Carlo simulations confirm the remarkable improvements of PIV over OLS. Nonetheless, even in this case, the PIV estimator does have an endogeneity bias in its limiting distribution that prevents this estimator from having a nuisance-parameter-free mixed normal limiting distribution.

We propose to eliminate this endogeneity bias by means of the fully modified methodology originally developed by Phillips and Hansen (1990). We have chosen this methodology because of its semiparametric nature and because of its good sampling behavior, as reported by Hansen and Phillips (1990), Phillips and Hansen (1990), and Kitamura and Phillips (1995), among others. After applying the new fully modified correction, the estimator obtained, called *fully modified pseudo instrumental variables* (FM-PIV), is asymptotically efficient, median unbiased, and follows a mixture of normals so that standard inference can be conducted. By Monte Carlo simulations we showed that, relative to the FM-OLS, our new FM-PIV obtains important improvements in terms of bias reductions.

Furthermore, from Kitamura and Phillips (1997) we know that efficient IV methods such as the generalize instrumental variable estimation (GIVE) meth-

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 7. FM-OLS and FM-PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (FM-OLS)		
−2.0	−.505E−01	−.365E−01	.479E−02	.465E−01	.638
−1.0	−.253E−01	−.182E−01	.155E−02	.297E−01	.846
−0.5	−.128E−01	−.913E−02	.734E−03	.237E−01	.929
0.0	−.183E−03	−.150E−03	.448E−03	.210E−01	.965
			$\tau = 1$		
−2.0	−.504E−01	−.365E−01	.479E−02	.464E−01	.639
−1.0	−.253E−01	−.182E−01	.155E−02	.297E−01	.846
−0.5	−.128E−01	−.913E−02	.734E−03	.237E−01	.929
0.0	−.182E−03	−.162E−03	.448E−03	.210E−01	.965
			$\tau = 2$		
−2.0	−.357E−01	−.272E−01	.251E−02	.355E−01	.759
−1.0	−.180E−01	−.137E−01	.983E−03	.258E−01	.901
−0.5	−.909E−02	−.692E−02	.598E−03	.223E−01	.946
0.0	−.199E−03	−.164E−03	.454E−03	.210E−01	.965
			$\tau = 3$		
−2.0	−.237E−01	−.196E−01	.135E−02	.294E−01	.857
−1.0	−.120E−01	−.100E−01	.697E−03	.236E−01	.936
−0.5	−.612E−02	−.510E−02	.532E−03	.219E−01	.955
0.0	−.220E−03	−.179E−03	.459E−03	.210E−01	.964
			$\tau = 4$		
−2.0	−.140E−01	−.134E−01	.877E−03	.271E−01	.913
−1.0	−.715E−02	−.693E−02	.582E−03	.230E−01	.950
−0.5	−.370E−02	−.363E−02	.509E−03	.217E−01	.958
0.0	−.231E−03	−.198E−03	.465E−03	.211E−01	.964
			$\tau = 5$		
−2.0	−.617E−02	−.836E−02	.812E−03	.282E−01	.925
−1.0	−.322E−02	−.450E−02	.572E−03	.233E−01	.952
−0.5	−.174E−02	−.244E−02	.514E−03	.219E−01	.958
0.0	−.234E−03	−.202E−03	.472E−03	.212E−01	.963
			$\tau = 6$		
−2.0	−.659E−02	−.781E−02	.681E−03	.250E−01	.940
−1.0	−.342E−02	−.406E−02	.543E−03	.224E−01	.955
−0.5	−.183E−02	−.219E−02	.510E−03	.216E−01	.958
0.0	−.233E−03	−.196E−03	.481E−03	.213E−01	.961
			$\tau = 7$		
−2.0	−.704E−02	−.719E−02	.631E−03	.235E−01	.944
−1.0	−.364E−02	−.373E−02	.535E−03	.220E−01	.956
−0.5	−.194E−02	−.204E−02	.512E−03	.215E−01	.957
0.0	−.233E−03	−.170E−03	.491E−03	.214E−01	.959
			$\tau = 8$		
−2.0	−.286E−02	−.468E−02	.647E−03	.238E−01	.946
−1.0	−.391E−02	−.343E−02	.541E−03	.220E−01	.955
−0.5	−.207E−02	−.183E−02	.518E−03	.217E−01	.957
0.0	−.230E−03	−.140E−03	.503E−03	.215E−01	.959

Note: DGP is $\Delta y_t = a\Delta x_t - 0.2(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\mathrm{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 500, 20,000 replications.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

TABLE 8. FM-OLS and FM-PIV estimators of β in equation $y_t = \beta x_t + u_{1,t}$

$a - \beta$	Mean bias	Median bias	MSE	IQR	$\Pr(\lvert\hat{\beta} - \beta\rvert \leq 0.05)$
			$\tau = 0$ (FM-OLS)		
−2.0	−.206E−01	−.142E−01	.978E−03	.220E−01	.904
−1.0	−.103E−01	−.715E−02	.390E−03	.160E−01	.968
−0.5	−.522E−02	−.363E−02	.242E−03	.142E−01	.987
0.0	−.113E−03	−.554E−04	.192E−03	.136E−01	.993
			$\tau = 1$		
−2.0	−.206E−01	−.141E−01	.985E−03	.220E−01	.904
−1.0	−.103E−01	−.715E−02	.392E−03	.161E−01	.968
−0.5	−.523E−02	−.364E−02	.243E−03	.142E−01	.987
0.0	−.128E−03	−.387E−04	.197E−03	.136E−01	.992
			$\tau = 2$		
−2.0	−.583E−02	−.529E−02	.268E−03	.151E−01	.986
−1.0	−.298E−02	−.274E−02	.216E−03	.140E−01	.991
−0.5	−.155E−02	−.146E−02	.204E−03	.138E−01	.992
0.0	−.138E−03	−.557E−04	.197E−03	.137E−01	.993
			$\tau = 3$		
−2.0	.170E−02	−.783E−03	.305E−03	.166E−01	.981
−1.0	.783E−03	−.538E−03	.231E−03	.144E−01	.988
−0.5	.326E−03	−.361E−03	.213E−03	.139E−01	.990
0.0	−.137E−03	−.788E−04	.196E−03	.137E−01	.993
			$\tau = 4$		
−2.0	−.107E−02	−.138E−02	.225E−03	.140E−01	.989
−1.0	−.597E−03	−.776E−03	.214E−03	.138E−01	.990
−0.5	−.360E−03	−.471E−03	.211E−03	.137E−01	.990
0.0	−.131E−03	−.100E−03	.199E−03	.137E−01	.993
			$\tau = 5$		
−2.0	.898E−03	−.229E−03	.249E−03	.143E−01	.986
−1.0	.396E−03	−.215E−03	.228E−03	.139E−01	.987
−0.5	.145E−03	−.200E−03	.223E−03	.138E−01	.988
0.0	−.113E−03	−.147E−03	.205E−03	.137E−01	.991
			$\tau = 6$		
−2.0	.191E−02	.320E−03	.272E−03	.146E−01	.983
−1.0	.902E−03	.241E−04	.241E−03	.140E−01	.986
−0.5	.397E−03	−.677E−04	.233E−03	.139E−01	.987
0.0	−.119E−03	−.128E−03	.211E−03	.138E−01	.990
			$\tau = 7$		
−2.0	−.182E−02	−.106E−02	.262E−03	.145E−01	.983
−1.0	−.973E−03	−.600E−03	.240E−03	.141E−01	.986
−0.5	−.547E−03	−.410E−03	.234E−03	.140E−01	.986
0.0	−.131E−03	−.139E−03	.220E−03	.139E−01	.989
			$\tau = 8$		
−2.0	−.151E−02	−.862E−03	.274E−03	.145E−01	.982
−1.0	−.814E−03	−.509E−03	.253E−03	.142E−01	.984
−0.5	−.464E−03	−.327E−03	.248E−03	.141E−01	.985
0.0	−.126E−03	−.150E−03	.232E−03	.140E−01	.987

Note: DGP is $\Delta y_t = a\Delta x_t - 0.5(y_{t-1} - \beta x_{t-1}) + \eta_{1t}$, with $\beta = 1$, $\Delta x_t = \epsilon_{2t}$, $\eta_{1t} \sim N(0,1)$, $\epsilon_{2t} \sim N(0, s^2)$, $s = 1$, $\mathrm{cov}(\eta_{1,t}, \epsilon_{2,t}) = 0$, sample size 300, 20,000 replications.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

ods (Sargan, 1988) and the generalized method of moments (GMM) (Hansen, 1982), after applying the corresponding fully modified transformation, are asymptotically more efficient than the FM-IV procedure with respect to the stationary components but asymptotically equivalent to FM-IV estimation with respect to the nonstationary components and, consequently, to our FM-PIV estimator. Therefore, it appears that no further modifications in the GIVE or GMM directions seem necessary in our setup.

For practical purposes, FM-PIV estimators require the specification of the kernel function, $\ell(\cdot)$, the bandwidth parameter, k, and the truncation lag, τ. Although issues of optimal choice of these parameters are beyond the scope of this paper, the following comments can provide some insights for the practical implementation of our estimator. First, from the relevant literature it seems that the choice of the kernel function is not so important as the choice of the bandwidth parameter. No essential differences have been found in the general qualitative features from using different kernels (see, e.g., Kitamura and Phillips, 1995).

Second, with respect to the choice of the bandwidth parameter, given that the expansion rate of the bandwidth parameter k includes the optimal growth rate $k \sim cT^{1/5}$, the method proposed by Andrews (1991), possibly after prewhitening with a first-order vector autoregressive filter prior to kernel estimation, as suggested by Andrews and Monahan (1992), seems to be a good choice.

Third, regarding the choice of τ, the hint from Theorem 4 is that the truncation lag τ of the instrumental variable should increase with the sample size T but as $o(T^{1/4})$. If the truncation lag is chosen too small, the tests could be biased. However, if the truncation point is too large, there could be a loss of efficiency. As usual, it appears that one of the best options could be to employ data dependent rules that incorporate the sample information, (see, e.g., Andrews, 1991) or to use some alternative standard order selection criteria. However, this is beyond the scope of this paper. In this sense, from the Monte Carlo simulations we obtained important bias reductions with FM-PIV over FM-OLS for small values of τ.

Those promising results suggest the extension of the analysis to more general multivariate contexts allowing for some cointegrating regressors, and we are actually doing research in this direction.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Aparicio, F.M. & A. Escribano (1997) Searching for Linear and Nonlinear Cointegration: A New Approach. Working paper 97-65, Universidad Carlos III de Madrid.

Aparicio, F.M. & A. Escribano (1998) Information-theoretic analysis of serial dependence and cointegration. *Studies in Nonlinear Dynamics and Econometrics* 3, 119–140.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

Aparicio, F.M. & C.W.J. Granger (1995) Nonlinear Cointegration and Some New Tests for Comovements. Working paper 95-15, University of California at San Diego.

Arranz, M.A. & A. Escribano (2000) Cointegration testing under structural breaks: A robust extended error correction model. *Oxford Bulletin of Economics and Statistics* 62, 23–52.

Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.

Banerjee, A., J.J. Dolado, D.F. Hendry, & G.W. Smith (1986) Exploring equilibrium relationships in econometrics through static models: Some Monte Carlo evidence. *Oxford Bulletin of Economics and Statistics* 48, 253–277.

Beran, J. (1994) *Statistics for Long Memory Processes*. New York: Chapman and Hall.

Cheung, Y.-W. & K.S. Lai (1993) A fractional cointegration analysis of purchasing power parity. *Journal of Business and Economic Statistics* 11, 103–112.

Davidson, J. (2000) A Model of Fractional Cointegration and Tests for Cointegration Using the Bootstrap. Preprint, Cardiff University.

Dolado, J.J. & F. Marmol (1996) Efficient Estimation of Cointegrating Relationships among Higher Order and Fractionally Integrated Processes. Working paper 9617, Bank of Spain.

Dolado, J.J. & F. Marmol (2001) Some Asymptotic Inference Results for Multivariate Long-Memory Processes. Preprint, Universidad Carlos III de Madrid.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.

Gonzalo, J. (1994) Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics* 60, 203–233.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121–130.

Granger, C.W.J. (1983) Cointegrated Variables and Error Correction Models. Manuscript, University of California, San Diego.

Granger, C.W.J. & R. Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–39.

Hansen, B.E. & P.C.B. Phillips (1990) Estimation and inference in models of cointegration: a simulation study. *Advances in Econometrics* 8, 225–248.

Hansen, L.P. (1982) Large sample properties of generalized methods of moments estimators. *Econometrica* 50, 1029–1054.

Hassler, U., F. Marmol, & C. Velasco (2000) Residual Log-Periodogram Inference for Long-Run Relationships. Preprint, Free University of Berlin.

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.

Jeganathan, P. (1999) On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583–621.

Kim, C. & P.C.B. Phillips (2000) Fully Modified Estimation of Fractional Cointegration Models. Preprint, Yale University.

Kitamura, Y. & P.C.B. Phillips (1995) Efficient IV estimation in nonstationary regression. *Econometric Theory* 11, 1095–1130.

Kitamura, Y. & P.C.B. Phillips (1997) Fully modified IV, GIVE, and GMM estimation with possibly nonstationary regressors and instruments. *Journal of Econometrics* 80, 85–123.

Marinucci, D. (1998) Band Spectrum Regression for Cointegrated Time Series with Long Memory Innovations. STICERD Discussion paper 353, London School of Economics.

Marinucci, D. & P.M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* 86, 103–120.

Marmol, F., A. Escribano, & F.M. Aparicio (1999) A New Instrumental Variable Approach for Estimation and Testing in Cointegrating Regressions. Working paper 99–19, Universidad Carlos III de Madrid.

Newey, K.N. & K.F. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.
Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
Robinson, P.M. (1994) Time series with strong dependence. In *Advances in Econometrics Sixth World Congress*, vol. 1. Cambridge: Cambridge University Press.
Robinson, P.M. & D. Marinucci (1998) Semiparametric Frequency Domain Analysis of Fractional Cointegration. STICERD Discussion paper 348, London School of Economics.
Robinson, P.M. & D. Marinucci (1999) Narrow-Band Analysis of Nonstationary Time Series. Preprint, London School of Economics.
Sargan, J.D. (1988) *Lectures on Advanced Econometrics*. New York: Basil Blackwell.

MATHEMATICAL APPENDIX

Proof of Theorem 1. From the definition of fractional cointegration, it follows that, when x_t, y_t are cointegrated and assuming $\tau \leq t$, $\tau = o(t)$, there exists a nonzero finite real number, β, such that

$$\lim_{\tau\to\infty} \frac{\gamma_{yx}(\tau,t)}{\gamma_x(\tau,t)} = \beta + \lim_{\tau\to\infty} \frac{\gamma_{zx}(\tau,t)}{\gamma_x(\tau,t)} = \beta + \lim_{\tau\to\infty} \frac{E(u_{1t}x_{t-\tau})}{E(x_t x_{t-\tau})}.$$

Thus, to prove the theorem, it suffices to show that the ratio

$$\frac{E(u_{1t}x_{t-\tau})}{E(x_t x_{t-\tau})} \tag{A.1}$$

vanishes as τ goes to infinity, $\tau = o(t)$. For this, consider the denominator,

$$E(x_t x_{t-\tau}) = E\left(\sum_{j=0}^{t-1} \psi_j(d) u_{2,t-j}\right)\left(\sum_{i=0}^{t-\tau-1} \psi_i(d) u_{2,t-\tau-i}\right), \tag{A.2}$$

where we have used the fact that $u_t = (u_{1t}, u_{2t})'$ is a linear process with $u_t = 0$ for $t \leq 0$ and with $\psi_j(d)$ given by the formal expansion $\Delta^{-d} = \sum_{i=0}^{\infty} \psi_i(d) B^i$. Hence, $E(x_t x_{t-\tau})$ is of the same order as

$$C \sum_{j=0}^{t-\tau-1} \psi_j(d)\psi_j(d) \propto C \sum_{j=0}^{t-\tau-1} j^{d-1}(j-\tau)^{d-1}, \tag{A.3}$$

where C denotes a generic positive constant and with the symbol $\propto$ denoting asymptotic equivalence, i.e., $a_j \propto b_j$ if $a_j/b_j \to 1$ as $j \to \infty$. The equivalence in (A.3) follows from using Stirling's approximation. Now,

$$C \sum_{j=0}^{t-\tau-1} j^{d-1}(j-\tau)^{d-1} \propto t^{2d-1} \int_0^1 v^{2d-1} dv, \tag{A.4}$$

using an integral approximation and the fact that $\tau = o(t)$. Consequently, $E(x_t x_{t-\tau})$ is $O_p(t^{2d-1})$. Therefore, because the numerator of expression (A.1) tends to zero exponentially fast, it follows that (A.1) dies out as τ goes to infinity. ■

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

Proof of Theorem 2. Weak convergence of partial sums of multivariate nonstationary fractionally integrated processes of order $d > \frac{1}{2}$ to $B(d,r)$ has been proved by Marinucci and Robinson (2000). Weak convergence to stochastic integrals of the form $\int_0^1 B_2(d,r)dB_1(r)$ is proved in Dolado and Marmol (2001). The rest of the results follow from a direct application of the continuous mapping theorem (CMT). ■

Proof of Theorem 3. Without any loss of generality, we will prove the theorem for $d \in (\frac{1}{2},\frac{3}{2})$, the extension to $d > \frac{3}{2}$ being straightforward. To prove the theorem, we proceed by parts. Hence, we will first analyze the denominator of

$$\hat{\beta}_{\mathrm{PIV}} - \beta = \frac{\sum_{t=\tau+1}^{T} u_{1t} x_{t-\tau}}{\sum_{t=\tau+1}^{T} x_t x_{t-\tau}}. \tag{A.5}$$

We have

$$\sum_{t=\tau+1}^{T} x_t x_{t-\tau} = \frac{1}{2}\sum_{t=\tau+1}^{T} x_t^2 + \frac{1}{2}\sum_{t=\tau+1}^{T} x_{t-\tau}^2 - \frac{1}{2}\sum_{t=\tau+1}^{T} (x_t - x_{t-\tau})^2. \tag{A.6}$$

On the one hand, $\sum_{t=\tau+1}^{T} x_t^2 = \sum_{t=1}^{T} x_t^2 - \sum_{t=1}^{[T\delta]} x_t^2$, where $\delta = \tau/T$ and $[\cdot]$ denotes the integer part operator. Then, from Dolado and Marmol (2001) and the CMT, asymptotically, when $T \to \infty$, τ fixed, we obtain

$$T^{-2d}\sum_{t=\tau+1}^{T} x_t^2 \Rightarrow \int_0^1 B_2^2(d,r)\,dr - \int_0^{\delta} B_2^2(d,r)\,dr,$$

and as $\tau \to \infty$ with $\delta \to 0$, yields

$$T^{-2d}\sum_{t=\tau+1}^{T} x_t^2 \Rightarrow \int_0^1 B_2^2(d,r)\,dr. \tag{A.7}$$

In the same manner, noting that $\sum_{t=\tau+1}^{T} x_{t-\tau}^2 = \sum_{t=1}^{T} x_t^2 - \sum_{t=T-\tau-1}^{T} x_t^2$, under Assumption 1 we obtain

$$T^{-2d}\sum_{t=\tau+1}^{T} x_{t-\tau}^2 \Rightarrow \int_0^1 B_2^2(d,r)\,dr. \tag{A.8}$$

Now rewrite x_t as $\Delta x_t = \eta_{2t}$ with $\Delta^{\varrho}\eta_{2t} = u_{2t}$, $d = 1 + \varrho$, $|\varrho| < \frac{1}{2}$, so that $x_t = x_{t-1} + e_{2t}(\tau)$, where $e_{2t}(\tau) = \sum_{j=0}^{\tau-1} \eta_{2,t-j}$. Then, $\sum_{t=\tau+1}^{T}(x_t - x_{t-\tau})^2 = \sum_{t=\tau+1}^{T} e_{2t}^2(\tau)$, but because

$$T^{-1}\tau^{-1}\sum_{t=\tau+1}^{T} e_{2t}^2(\tau) = T^{-1}\sum_{t=\tau+1}^{T}\eta_{2t}^2 + 2\sum_{i=1}^{\tau-1}\left(1 - \frac{i}{\tau}\right)T^{-1}\sum_{t=\tau+1}^{T}\eta_{2t}\eta_{2,t-i} + o_p(1),$$

it follows that

$$\sum_{t=\tau+1}^{T} e_{2t}^2(\tau) = O_p(T\tau^{2d-1}), \tag{A.9}$$

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press

and thus,

$$T^{-2d}\sum_{t=\tau+1}^{T} x_t x_{t-\tau} = \frac{1}{2}T^{-2d}\sum_{t=\tau+1}^{T} x_t^2 + \frac{1}{2}T^{-2d}\sum_{t=\tau+1}^{T} x_{t-\tau}^2 - \frac{1}{2}\delta^{2d-1}T^{-1}\tau^{1-2d}\sum_{t=\tau+1}^{T}(x_t - x_{t-\tau})^2 \Rightarrow \int_0^1 B_2^2(d,r)\,dr. \quad \textbf{(A.10)}$$

Consider now the numerator of (A.5),

$$\sum_{t=\tau+1}^{T} u_{1t}x_{t-\tau} = \sum_{t=\tau+1}^{T} u_{1t}x_t - \sum_{t=\tau+1}^{T} e_{2t}(\tau)u_{1t}. \quad \textbf{(A.11)}$$

With regard to the first term on the right-hand side of (A.11), from Dolado and Marmol (2001) it turns out that, as $T \to \infty, \tau \to \infty, \delta \to 0$,

$$T^{-d}\sum_{t=\tau+1}^{T} x_t u_{1t} \Rightarrow \int_0^1 B_2(d,r)\,dB_1(r) \quad \text{if } 1 < d < \frac{3}{2}, \quad \textbf{(A.12)}$$

$$T^{-1}\sum_{t=\tau+1}^{T} x_t u_{1t} \Rightarrow \int_0^1 B_2(r)\,dB_1(r) + \Delta_{21} \quad \text{if } d = 1, \quad \text{and} \quad \textbf{(A.13)}$$

$$T^{-1}\sum_{t=\tau+1}^{T} x_t u_{1t} \xrightarrow{p} \Delta_{21}(d) \quad \text{if } \frac{1}{2} < d < 1. \quad \textbf{(A.14)}$$

As regards the second term on the right-hand side of equation (A.11), note that, for $d \in (\frac{1}{2}, \frac{3}{2})$, $\sum_{t=\tau+1}^{T} e_{2t}(\tau)u_{1t} = \sum_{t=\tau+1}^{T} \eta_{2t}u_{1t} + \sum_{i=1}^{\tau-1}\sum_{t=\tau+1}^{T} \eta_{2,t-i}u_{1t}$, and then, when $T \to \infty, \tau \to \infty, \delta \to 0$,

$$T^{-1}\sum_{t=\tau+1}^{T} e_{2t}(\tau)u_{1t} \xrightarrow{p} \sum_{k=0}^{\infty} E(\eta_{2,0}u_{1k}) = \Delta_{21}(d), \quad \textbf{(A.15)}$$

which is well defined (cf. Dolado and Marmol, 2001), with $\Delta_{21}(1) = \Delta_{21}$. Theorem 3 now follows from collecting all the previous results and applying the CMT. ■

Proof of Theorem 4. We have $\sum_{t=\tau+1}^{T} \hat{u}_{1t}^{+}x_{t-\tau} = \sum_{t=\tau+1}^{T} u_{1t}x_{t-\tau} - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\sum_{t=\tau+1}^{T} u_{2t}x_{t-\tau}$. Proceeding as in the proof of Theorem 3, we get

$$T^{-1}\sum_{t=\tau+1}^{T} u_{2t}x_{t-\tau} = T^{-1}\sum_{t=\tau+1}^{T} u_{2t}x_t - T^{-1}\sum_{t=\tau+1}^{T} e_{2t}(\tau)u_{2t} \Rightarrow \int_0^1 B_2(r)\,dB_2(r), \quad \textbf{(A.16)}$$

provided that $\delta \to 0$ as $T \to \infty, \tau \to \infty$. Consequently,

$$T^{-1}\sum_{t=\tau+1}^{T} \hat{u}_{1t}^{+}x_{t-\tau} \Rightarrow \int_0^1 B_2(r)\,dB_{1.2}(r), \quad \textbf{(A.17)}$$

with $B_{1.2}(r) = B_1(r) - \omega_{12}\omega_{22}^{-1}B_2(r)$ so that $E(B_{1.2}(r)B_2(r)) = 0$ and thus is long-run independent.

Last, as regards the denominator of $\hat{\beta}_{PIV}^{+}$, note from the corresponding part of the proof of Theorem 3 that for expression (A.10) to hold, we can apply the results of Newey and West (1987) to the term $\sum_{t=\tau+1}^{T} e_{2t}^2(\tau)$ such that equation (A.9) also holds for $\tau = o(T^{1/4})$. The theorem finally follows from (A.10), (A.17), and the CMT. ■

https://doi.org/10.1017/S0266466602183046 Published online by Cambridge University Press