Abstract

In this paper we propose an alternative characterization of the central notion of cointegration, exploiting the relationship between the autocorrelation and the cross-correlation functions of the series. This characterization lead us to propose a new estimator of the cointegrating parameter based on the instrumental variables methodology. The instrument is a delayed replica of the regressor variable in a 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 with a semi-parametric correction the estimator for the unit root case is median-unbiased, a mixture of normals and asymptotically efficient. As a consequence, standard inference can be conducted with this fully modified instrumental variable estimator of the cointegrating parameter.

Key Words

Short memory; long memory; cointegration; instrumental variables; fully modified OLS estimation.

*Mármol, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain, e-mail: fmarmol@est-econ.uc3m.es; Escribano, Department of Economics, Universidad Carlos III de Madrid, Spain, e-mail: alvaroe@eco.uc3m.es; Aparicio, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain, e-mail: aparicio@est-econ.uc3m.es. We acknowledge financial support from the Spanish DGICYT contract PB95-0298 and the European Union contract TMR ERB-4061 PL 97-0994.
1. INTRODUCTION

Many time series of interest exhibit important changes in the mean and the variance. Such a series are often said to be integrated, since it is possible to simulate the most important features in their patterns with sums of an increasing number of weakly dependent random variables. More specifically, we said that a 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 and $I(0)$ denotes a short memory process.

In some cases, the changes in mean behavior can be correlated across series. Pairs of series which exhibit a common stochastic trend are said to be cointegrated. The concept of cointegration was coined by Granger (1981, 1983) and later developed by Engle and Granger (1987) when the data generating process (DGP) was generated by integrated processes.

However, there are no a priori reasons for not extrapolating the concepts of integration and cointegration to the fractional case where now $d$ is not an integer but a real number. Loosely speaking, we said that a series $x_t$ is fractionally integrated of order $d$, if $\Delta^d x_t \sim I(0)$ with $d$ being a real number. Correspondingly, two $I(d)$ fractionally integrated of order $d$ processes, $x_t, y_t$, are said to be (fractionally) cointegrated if there exist a real number $\beta \neq 0$ such that the error term $z_t = (y_t - \beta x_t) \sim I(d_e), d_e < d$. In practice many real time series seem to be well approximated by fractionally rather than purely integrated processes (cf., e.g., Baillie, 1996), but still there are only few theoretical and empirical results on fractional cointegration.

In this paper we propose a new characterization of fractional cointegration, exploiting linear measures of dependence and cross-dependence as introduced by Aparicio and Escribano (1999a) based on the nonlinear counterpart of Aparicio and Escribano (1999b). This characterization lead us to propose an estimator for the cointegrating parameter $\beta$ based on the instrumental variables methodology, where the instrument is a delayed replica of the regressor variable in the conditional model, and where for simplicity the DGP is assumed to be a bivariate system of nonstationary fractionally integrated processes with a weakly stationary $I(0)$ error correction term. Since the
instrument is only asymptotically uncorrelated with the innovation term, the proposed estimation method will be called a pseudo instrumental variable (PIV) estimator.


Herein we prove that PIV is a consistent estimator of the cointegrating parameter $\beta$. Moreover, in the $d = 1$ unit root case we show that PIV is preferable to ordinary least squares (OLS), because the limiting distribution of the PIV estimator is free of the serial correlation bias that characterizes the OLS limiting distribution. 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 Park and Phillips (1988, 1989) and Phillips and Hansen (1990). We have chosen this methodology because of its semi-parametric nature and because of its good sampling behavior reported by Hansen and Phillips (1990), Phillips and Hansen (1990), Hargreaves (1994), Kitamura and Phillips (1995) and Haug (1998). 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.

Phillips and Hansen (1990) were the first that developed the asymptotic properties of IV estimates in multivariate cointegrating regressions. Later on, this study was 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. From Kitamura and Phillips (1995) we learn that sometimes the bias and root mean square error of FM-IV are exceptionally high (in fact, higher than OLS and crude IV) due to 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, for instance, poor instruments, in the second stage the fully-modified procedure can amplify the effect of poor preliminary estimates. With our estimator we provide a solution to the instrument selection problem. Our suggested instrument is by definition neither spurious, since it is always correlated with the corresponding regressor, nor a poor instrument, since the correlation of unit root processes tend asymptotically to one.

On the other hand, from Kitamura and Phillips (1995) we also learn that efficient IV methods such as the generalize instrumental variable estimation (GIVE) methods (Sargan, 1988) and the generalized method of moments (GMM) (Hansen, 1982) after their corresponding fully-modified transformation are asymptotically more efficient that the FM-IV procedure with respect to stationary components but that they are asymptotically equivalent to FM-IV (and, consequently, to our FM-PIV) estimation with respect to nonstationary components. This is nothing else that a direct consequence of the asymptotic OLS/GLS equivalence in nonstationary models that was shown by Phillips and Park (1988). Therefore, it appears that no further modifications in the GIVE or GMM directions seem necessary in our set-up.

The paper is organized as follows. In Section 2 we define several measures of linear dependence as well as the underlying assumptions. In Section 3 we motivate our IV estimator by introducing an alternative characterization of fractional cointegration. We prove that this alternative characterization is consistent against spurious alternatives. Section 4 derives the consistency and limiting distribution of the crude PIV estimator, whereas in Section 5 we derive the asymptotic properties of the FM-PIV estimator. It turns out that this estimator is consistent and its limiting distribution is median-unbiased, a mixture of normals and asymptotically efficient so that inference can be conducted in the standard way. Finally, some concluding remarks are given in Section 6. Mathematical proofs of the theorems are gathered in an 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. In the linear case, the standard measure is the correlation function (ACF), say $\rho_x(\tau, t)$, for the series $x_t$, which may not necessarily be (wide-sense) stationary. By defining $\rho_x(\tau, t)$ as

\begin{equation}
\rho_x(\tau, t) = \frac{\text{cov}(x_t, x_{t-\tau})}{\text{var}(x_{t-\tau})}, \tag{1}
\end{equation}

an early definition of memory in time series has to do with the concept of mean reversion. A process $x_t$ is said to be mean reverting if $\forall t$,

\begin{equation}
limit_{\tau \to 0} \rho_x(\tau, t) = 0. \tag{2}
\end{equation}

According to this definition, when the process is not mean reverting $\lim_{\tau \to 0} \rho_x(\tau, t) > 0$, and thus any two infinitely distant variables from the process are still correlated. However, even for a mean reverting process the memory span can be very large in the sense that its decay is very slow as $\tau$ grows. This motivates the distinction between short and long memory. A process $x_t$ is said to have short memory if $\forall t, \exists b_t < \infty$ such that

\begin{equation}
limit_{n \to \infty} \sum_{\tau=-n}^{n} |\rho_x(\tau, t)| = b_t. \tag{3}
\end{equation}

Otherwise, if $\lim_{n \to \infty} \sum_{\tau=-n}^{n} |\rho_x(\tau, t)|$ is nonfinite, then the process $x_t$ is said to have long memory. Processes that are not mean reverting must necessarily have long memory. On the contrary, processes that are mean reverting can be either long or short memory, depending on the rate at which their ACF vanishes with increasing lags.

A particular case of short memory time series are the well-known ARMA($p,q$) processes,

\begin{equation}
\Phi(B)z_t = \Theta(B)e_t, \tag{4}
\end{equation}

---

1 See Aparicio and Granger (1995) and Aparicio and Escribano (1999b) for an extension to the nonlinear case based on the concept of mutual information.
2 The usual concept of autocorrelation is in the stationary case equivalent to the $\rho_x(\tau)$ defined in (1). For convenience, as will become clear later on, in this paper we use expression (1).
where $\Phi(B), \Theta(B)$ are polynomials in the lag operator $B$ of orders $p, q$, respectively, with all their roots lying outside the unit circle, with no common factors and with $\varepsilon_i$ being a zero mean white noise process. A stationary and invertible ARMA process has autocorrelations which are geometrically bounded, i.e., for large $\tau$ and some constant $c$,
\[ |\rho_2(\tau, t)| = |\rho_2(\tau)| \leq c \zeta^{-\tau}, \quad 0 < \zeta < 1 \]
and is hence a short memory process.

On the other hand, within the family of the long memory processes, a very important member are the so-called integrated processes. A time series $x_t$ is said to be integrated of order $d$, in short, $x_t \sim I(d)$, if it has long memory and $d$, the memory or persistence parameter, is the smallest real number such that $z_t = \Delta^d x_t$ is a short memory process with finite and positive spectral density. Correspondingly, then, a short memory process is denoted $I(0)$. When $d = 1, 2, 3, \ldots$, $\Delta^d = (1 - B)^d$, and $z_t$ has an ARMA representation as in expression (4), then $x_t$ belongs to the well-known family of the ARIMA($p,d,q$) processes. These processes are not mean reverting and have infinite memory, i.e., the effect on an innovation $\varepsilon_i$ on the present and future behavior of the series always remains. In turn, when $d$ is not an integer but a real number, then the series is said to be fractionally integrated of order $d$. If the short memory component has an ARMA representation, then $x_t$ is said to be an ARFIMA($p,d,q$) process, and now
\[ \Delta^d = 1 - dB + \frac{d(d-1)B^2}{2!} - \frac{d(d-1)(d-2)B^3}{3!} + \ldots, \]
for any real $d > -1$. This expansion, in turn, admits a synthetic expression for $d > 0$ by introducing the gamma and hypergeometric functions,
\[ \Delta^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)B^k}{\Gamma(k+1)\Gamma(-d)} = F(-d,1,1;B). \]

Fractionally integrated processes were originally introduced by Granger and Joyeux (1980) and Hosking (1981). See Beran (1994), Robinson (1994) and Baillie (1996) for recent overviews of the theory and major empirical applications of this family of processes. A fractionally integrated process $x_t$ of order $d$ is both stationary and invertible if and only if $-\frac{1}{2} < d < \frac{1}{2}$. For $0 < d < \frac{1}{2}$ the process is stationary with long memory, its autocorrelations are all positive and decay at a hyperbolic rate, i.e.,
\[ \rho_x(\tau, t) = \rho_x(\tau) \approx c \tau^{2d-1}, \quad \text{for large } \tau, \]
so that its spectral density will be unbounded at low
frequencies. If \( d \leq 0 \) the process has absolutely summable autocorrelations and is therefore stationary. When \( d < 0 \) all its autocorrelations, excluding lag zero, are negative and decay hyperbolically to zero. In this situation, the process is said to be \textit{antipersistent}, and it can be shown that its spectral density is finite but zero at zero frequency so that an antipersistent process is not short memory according to our previous definition. On the other hand, the process is nonstationary if \( d \geq \frac{1}{2} \). Yet, while in this case the process is not covariance stationary, for \( d < 1 \) it is nevertheless mean reverting. Lastly, when \( d \geq 1 \) the process is both nonstationary and not mean reverting.

In this paper we are interested in the estimation and testing of linear cointegration among nonstationary fractionally integrated processes. From Granger (1981, 1983) and Engle and Granger (1987), we said that a set of fractionally integrated processes of order \( d \geq \frac{1}{2} \) are \textit{(fractionally) cointegrated of order \( (d, b) \)}, if there exists (at least) a linear combination among them having memory parameter \( d - b, d \geq b > 0 \).

Notice from this definition that herein we are only concerned with the nonstationary case\(^3\), since this is the most relevant range when dealing with cointegration. Moreover, to avoid invertibility problems we shall further assume that the memory parameter \( d \) belongs to the set \( \partial = \{ d \in \mathbb{R} \mid d \geq \frac{1}{2}, d \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \} \). This is done without loss of generality in the sense that the set \( \{ \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \} \) has Lebesgue measure zero. Furthermore, as a natural step and in order to keep things simple, we concentrate our attention on the simplest bivariate case where the data generating process \((DGP)\) is assumed to be generated by two cointegrated fractionally integrated time series of order \( d \in \partial \), with an \( I(0) \) error correction term and without any deterministic elements. Suitable modifications to include the multivariate case with quite general deterministic processes in the data-generating mechanism are rather direct.

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

\begin{align}
(5) \quad y_t &= \beta x_t + z_t, \quad \beta \in \mathbb{R} \setminus \{0\}, \\
(6) \quad \Delta^d x_t &= u_{2t}, \quad d \in \partial, \\
(7) \quad \Delta^d z_t &= u_{1t}, \quad d \geq d, \geq 0, 
\end{align}

\(^3\) See Robinson (1994) and Robinson and Hidalgo (1997) for the stationary case.
with \( u_t = (u_{1t}, u_{2t})' \) satisfying the following regularity conditions.

**ASSUMPTION 1.** Let \( u_t = (u_{1t}, u_{2t})' \) be generated by the linear process

\[
    u_t = \sum_{j=0}^{\infty} \Pi_j e_{t-j}, \quad \varepsilon_t = 0 \text{ for } t \leq 0,
\]

where the sequence of random vectors \( \varepsilon_t = (e_{1t}, e_{2t})' \) is i.i.d. \((0, \Sigma)\), \( \Sigma > 0 \),
\( E(\varepsilon_t, \varepsilon_t | \ldots, \varepsilon_{t-2}, \varepsilon_{t-1}) \leq c \) (a.s.) for some constant \( c > 0 \) and the sequence of matrix coefficients \( \{\Pi_j\}_{j=0}^{\infty} \) is 1-summable. Further, assume that

\[
    \max \sup_{t} E|\varepsilon_t|^{\sigma+1} < \infty,
\]

where \( \sigma > 0 \) and \( g = 2 \) if \( d > \frac{1}{2} \), \( g = 4 \) if \( \frac{1}{2} \leq d < \frac{1}{2} \) a \( g = 8(1-d)/(2d-1) \) if \( d < \frac{1}{2} \).

Hence, throughout this paper we shall allow \( u_t \) be generated by the linear process

\[
    u_t = \sum_{j=0}^{\infty} \Pi_j e_{t-j}. \quad \text{This general class of stationary } I(0) \text{ processes includes all stationary and invertible } ARMA \text{ processes and is therefore of wide applicability. Further, Assumption 1 implies that the } u_t \text{ process is strictly stationary and ergodic with continuous spectral density given by}
\]

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

and \((2 \times 2)\) long-run covariance matrix \( \Omega = 2\pi f_{uu}(0) = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \).

3. MOTIVATION OF THE NEW ESTIMATOR

In order to motivate our estimator of the cointegrating parameter \( \beta \) in expression (5), we use the characterization of cointegration introduced by Aparicio and Escribano (1999a). For this, let \( x_t, y_t \) be the two \( I(d) \) time series of interest, and let \( \rho_{yx}(\tau; t) \) represent the cross-correlation function (CCF) of \( x_t, y_t \), defined as in (1) by

\[
    \rho_{yx}(\tau; t) = \frac{\text{cov}(y_{t+\tau}, x_{t})}{\text{var}(x_{t+\tau})}.
\]
Once again, we make explicit the time dependence in $\rho_{yx}(\tau,t)$ to allow for some degree of heterogeneity in the series.

**THEOREM 1.** The time series $x_t, y_t$ in the DGP (5)-(7) are fractionally cointegrated of order $(d, d - d_z)$ if and only if $\forall t$,

$$\lim_{\tau \to \infty} \frac{\rho_{yx}(\tau,t)}{\rho_x(\tau,t)} = \beta.$$  

If the series are not fractionally cointegrated, i.e., in the spurious case where $d = d_z$ and $x_t, y_t$ are stochastically independent, then

$$\lim_{\tau \to \infty} \frac{\rho_{yx}(\tau,t)}{\rho_x(\tau,t)} = 0.$$  

Theorem 1 implies that the rates of convergence of $\rho_{yx}(\tau,t)$ and $\rho_x(\tau,t)$ should be the same as $\tau$ increases without bound. Intuitively, the theorem states that, under cointegration, the remote past of $y_t$ should be as useful as the remote past of $x_t$ in long-run linearly forecasting $x_t$.

It is worth noting that condition (9) in Theorem 1 needs not be checked in the limit for most practical cases when we are looking for cointegration. For example, suppose $x_t, y_t \sim I(1)$ and $z_t$ a sequence of i.i.d. random variables independent of $x_t$. In this case, $\rho_{yx}(\tau,t)/\rho_x(\tau,t) = \beta$ for all $\tau, t$. In general, however, the constancy of this ratio will only take place for $\tau$'s beyond some value.

On the other hand, expression (10) shows that the ratio is consistent against spurious alternatives.

**Example 1.** To illustrate the statements in Theorem 1, consider the following linear common factor model,

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \beta \\ 1 \end{pmatrix} w_t + \begin{pmatrix} v_t \\ \xi_t \end{pmatrix},$$

(11)
\( \beta \neq 0 \), with \( w_t = w_{t-1} + \eta_t \) and where \((\eta_t, v_t, \xi_t)\) are independent sequences of zero mean \( i.i.d. \) random variables. Let \( P' = (1, -\beta) \). Thus the cointegrating relationship is obtained as

\[
(12) \quad z_t = P'(y_t, x_t) = y_t - \beta x_t,
\]

where \( z_t = v_t - \beta \xi_t \sim i(0) \).

In this case, after some algebra one obtains

\[
\rho_x(\tau, t) = \frac{(t-\tau)\sigma^2}{(t-\tau)\sigma^2 + \sigma_x^2},
\]

which clearly converges to 1 as \( \tau \to \infty \) for any \( t \). Similarly, it is easy to show that \( \rho_y(\tau, t) \to \beta \) as \( \tau \to \infty \) for any \( t \). Thus \( \rho_y(\tau, t)/\rho_x(\tau, t) \to \beta \) as \( \tau \to \infty \) for any \( t \), and since by assumption \( \beta \neq 0 \), we conclude that the series \( x_t, y_t \) are cointegrated.

**Example 2.** Consider now the pair of non-cointegrated series,

\[
\begin{align*}
(13) \quad x_t &= w_t + \xi_t, \\
(14) \quad y_t &= q_t + v_t,
\end{align*}
\]

where \( q_t, w_t \) are zero mean independent \( i(d) \) series, and where \( v_t, \xi_t \) are zero mean \( ARMA \) processes independent of \( q_t, w_t \). Thus, for any number \( \beta \) we can write \( y_t = \beta x_t + z_t \) with \( z_t = q_t - \beta w_t + v_t - \beta \xi_t \) so that \( \text{cov}(y_t, x_{t-\tau}) = \text{cov}(v_t, \xi_{t-\tau}) \).

Consequently, in this spurious case, the covariance \( \text{cov}(y_t, x_{t-\tau}) \) will tail off exponentially as \( \tau \) grows to infinity for any \( t \). On the contrary, \( \text{cov}(x_t, x_{t-\tau}) \) will decay hyperbolically with growing \( \tau \). Thus, the limit in expression (10) exists and it is equal to (the true parameter) zero, as expected.

### 4. ESTIMATION OF THE COINTEGRATING PARAMETER

Consider now the benchmark \( DGP \) given by expressions (5)-(7) in the particular case where \( d_z = 0 \) so that \( z_t = u_{it} \). In the previous section we motivated the use of the ratio

\[
(15) \quad \frac{\rho_y(\tau, t)}{\rho_x(\tau, t)} = \frac{E(y_t x_{t-\tau})}{E(x_t x_{t-\tau})},
\]
as an alternative characterization of the central notion of cointegration. Yet, notice that for large values of $r$ this quotient would be nothing else but the expression of an IV estimator of the cointegrating parameter $\beta$, with instrument $x_{t-r}$. 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 $\hat{\beta}_{OLS}$ of $\beta$, for a sample of size $T$,

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

(16)

**Theorem 2.** Given the data generating process (5)-(7) with $d \in \mathbb{R}$, $d_z = 0$ and under Assumption 1, asymptotically, as $T \to \infty$,

(17) $\hat{\beta}_{OLS} \overset{p}{\to} \beta$, for all $d \in \mathbb{R}$,

(18) $T^2(\hat{\beta}_{OLS} - \beta) \Rightarrow \begin{cases} \frac{1}{\int_0^1 B_2^2(d,r)dr}, & \text{if } d > 1, \\ \frac{1}{\int_0^1 B_2^2(d,r)dr}, & \text{if } d = 1, \\ \frac{\Delta_{21}}{\int_0^1 B_2^2(r)dr}, & \text{if } d < 1, \end{cases}$

(19) $T(\hat{\beta}_{OLS} - \beta) \Rightarrow \begin{cases} \frac{1}{\int_0^1 B_2^2(d,r)dr}, & \text{if } d > 1, \\ \frac{1}{\int_0^1 B_2^2(r)dr}, & \text{if } d = 1, \\ \frac{\Delta_{21}(d)}{\int_0^1 B_2^2(r)dr}, & \text{if } d < 1, \end{cases}$

(20) $T^{d^{-1}}(\hat{\beta}_{OLS} - \beta) \Rightarrow \begin{cases} \frac{1}{\Gamma(d)\int_0^1 (r-s)^{d-1}dB(s)} dB(r), & \text{if } d > 1, \\ \frac{1}{\Gamma(d)\int_0^1 (r-s)^{d-1}dB(s)} dB(r), & \text{if } d = 1, \\ \frac{\Delta_{21}(d)}{\Gamma(d)\int_0^1 (r-s)^{d-1}dB(s)} dB(r), & \text{if } d < 1, \end{cases}$

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

$$B(d,r) = \frac{1}{\Gamma(d)\int_0^1 (r-s)^{d-1}dB(s)}.$$
\[ \Delta_{21} = \sum_{k=0}^{\infty} E(u_{2,0} u_{1,k}), \quad \Delta_{21}(d) = \sum_{k=0}^{\infty} E(\Delta x_{0} u_{1,k}), \text{ and where } \xrightarrow{p}, \xrightarrow{w}, \text{ denote convergence in probability and weak convergence, respectively.} \]


Then, for all \( d \in \partial \), OLS is a consistent estimator of \( \beta \). On the other hand, the presence of nuisance parameters in the limiting OLS distributions (18)-(20) prevents achieving an asymptotic mixture of normals. In the \( d = 1 \) case, these nuisance parameters are given by \( \Delta_{21} \) and \( \omega_{21} \), the \((2,1)\)-element of the long-run covariance matrix \( \Omega \). \( \omega_{21} \neq 0 \) 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 (18) 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 \( \hat{\beta}_{OLS} \), they can be important in finite samples. In effect, these nuisance parameters, \( \Delta_{21} \) and \( \omega_{21} \), produce a finite sample bias in mean and median, respectively. The limiting distribution is no longer neither a mixture of normals nor asymptotically efficient. They invalidate the use of standard distributions for testing hypothesis about the cointegrating parameter \( \beta \). This is in contrast with the especial case where \( \omega_{21} = \Delta_{21} = 0 \), i.e., when \( x_t \) is strongly exogenous with respect to \( \beta \) in (5)-(7). In this case the limiting distribution is median-unbiased, a mixture of normals and asymptotically efficient, with the limiting distribution depending on nuisance parameters in a simple way which 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 \( \beta \) in (5)-(7),

\[ \hat{\beta}_{PIV} = \frac{\sum_{t=T+1}^{T} y_t x_{t-r}}{\sum_{t=T+1}^{T} x_t x_{t-r}}, \]
called *pseudo instrumental variable (PIV)* estimator, since \( x_{t-r} \) is only an instrument of \( x_t \) in the standard sense (i.e., correlated with \( x_t \) and uncorrelated with \( u_t \)) for \( \tau \to \infty \). In finite samples, however, \( x_{t-r} \) is not necessarily independent of the innovation \( u_t \) and consequently \( x_{t-r} \), strictly speaking, is not an instrument. Notice also that for fix \( \tau \), PIV could be covered by the IV framework developed by Phillips and Hansen (1990) with \( n_1 = n_2 = 1 \) and \( y_{3t} = x_{t-r} \) in their notation.

**THEOREM 3.** Given the DGP (5)-(7) with \( d \in \partial \), \( d_z = 0 \) and under Assumption 1, asymptotically, as \( T \to \infty \), \( \tau \to \infty \) and \( T^{-1}\tau \to 0 \),

\[
\hat{\beta}_{PIV} \xrightarrow{p} \beta, \quad \text{for all } d \in \partial, \tag{22}
\]

\[
T^d \left( \hat{\beta}_{PIV} - \beta \right) \xrightarrow{d} \frac{\int_1^1 B_2(d,r)dB_1(r)}{\int_0^1 B_1^2(d,r)dr}, \quad \text{if } d > 1, \tag{23}
\]

\[
T \left( \hat{\beta}_{PIV} - \beta \right) \xrightarrow{d} \frac{\int_0^1 B_2(r)dB_1(r)}{\int_0^1 B_2^2(r)dr}, \quad \text{if } d = 1, \tag{24}
\]

\[
T^{2d-1} \left( \hat{\beta}_{PIV} - \beta \right) \xrightarrow{d} 0, \quad \text{if } d < 1. \tag{25}
\]

From this theorem the following comments are in order. First, as expected, for all \( d \in \partial \), PIV is a consistent estimator of \( \beta \). Second, when \( \frac{1}{2} < d < 1 \), \( \hat{\beta}_{PIV} \) is \( o_p \left(T^{1-2d}\right) \) with a degenerate limiting distribution. Third, for \( d > 1 \), \( \hat{\beta}_{PIV} \) and \( \hat{\beta}_{OLS} \) have the same limiting distribution, i.e., nonstandard with endogeneity bias and no 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 \( \beta \). The estimator, however, is neither a mixture of normals nor asymptotically efficient, and standard inference remains invalid.
Example 2 (continued). Consider the asymptotic behavior of $\hat{\beta}_{PIV}$ for the data generating process of example 2. We have

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

so that

$$\begin{align*}
\hat{\beta}_{PIV} - \beta &= \frac{\sum_{t=1}^{T} (w_{t+1} + \xi_{t+1}) (q_t - \beta w_t + v_t - \beta \xi_{t+1})}{\sum_{t=1}^{T} (w_t + \xi_t) (w_{t+1} + \xi_{t+1})} - \beta + o_p(1) \\
&= \frac{\sum_{t=1}^{T} q_t w_{t+1}}{\sum_{t=1}^{T} w_t w_{t+1}} + o_p(1).
\end{align*}$$

Now as $T \to \infty$, $\tau \to \infty$ and $T^{-1} \tau \to 0$ it can be proved that

$$\hat{\beta}_{PIV} \Rightarrow \frac{1}{\int_{0}^{1} B_q(d,r)B_u(d,r)dr},$$

where $B_q(d,r), B_u(d,r)$ denote the corresponding fractional Brownian motions associated with $q_t$ and $w_t$, respectively. It turns out to be the same limiting distribution as that of $\hat{\beta}_{OLS}$ in the spurious case with nonstationary fractionally integrated process. See, e.g., Marmol (1998).

Lastly, to close this section we will provide some insights on the finite sample performance of $\hat{\beta}_{PIV}$ relative to $\hat{\beta}_{OLS}$ and other competing estimators. For this, consider the following particular case of DGP (5)-(7)

(26) $y_t = \beta x_t + u_t, \quad u_t = \varphi u_{t-1} + \varepsilon_t, \quad |\varphi| < 1,$

(27) $\Delta x_t = \varepsilon_{2t},$
Under this set-up, Gonzalo (1994) proves that

\[(29) \quad T(\hat{\beta}_{\text{OLS}} - \beta) \xrightarrow{d} \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \left( \frac{\sigma_1}{1 - \varphi} \right)(1 - \vartheta^2)^{1/2} \int_0^1 B_2(r)dB_2(r) + \left( \frac{1}{1 - \varphi} \right) \vartheta \sigma \right\}
\]

\[\left\{ \int_0^1 B_2(r)dB_2(r) + \left( \frac{1}{1 - \varphi} \right) \sigma \right\},\]

\[(30) \quad T(\hat{\beta}_{\text{NLS}} - \beta) \xrightarrow{d} \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \left( \frac{\sigma_1}{1 - \varphi} \right)(1 - \vartheta^2)^{1/2} \int_0^1 B_2(r)dB_2(r) + \left( \frac{1}{1 - \varphi} \right) \right\}
\]

\[\left\{ \beta + \vartheta \frac{\sigma_1}{\sigma_2} \right\} \int_0^1 B_2(r)dB_2(r) \},\]

\[(31) \quad T(\hat{\beta}_{\text{MLECM}} - \beta) \xrightarrow{d} \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \left( \frac{\sigma_1}{1 - \varphi} \right)(1 - \vartheta^2)^{1/2} \int_0^1 B_2(r)dB_2(r) + \left( \frac{1}{1 - \varphi} \right) \right\}
\]

\[\theta \frac{\sigma_1}{\sigma_2} \int_0^1 B_2(r)dB_2(r) \},\]

where \( W_i(r) \) is a standard Brownian motion independent of \( B_2(r) \), and where \( \hat{\beta}_{\text{NLS}} \) and \( \hat{\beta}_{\text{MLECM}} \) stand for the nonlinear least squares and the maximum likelihood in a fully specified error correction model estimators of \( \beta \), respectively. Also, from (24) it is not difficult to show that

\[(32) \quad T(\hat{\beta}_{\text{MLECM}} - \beta) \xrightarrow{d} \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \left\{ \left( \frac{\sigma_1}{1 - \varphi} \right)(1 - \vartheta^2)^{1/2} \int_0^1 B_2(r)dB_2(r) + \left( \frac{1}{1 - \varphi} \right) \right\}
\]

\[\theta \frac{\sigma_1}{\sigma_2} \int_0^1 B_2(r)dB_2(r) \}.
\]

Now, from expressions (28)-(31) the following comments can be readily deduced.

First, with respect to the limiting distribution of the \( OLS \) estimator, notice that it involves three different parts. The first one is a mixture of normals, the second one is a unit root term, \( \left\{ \int_0^1 B_2^2(r)dr \right\}^{-1} \int_0^1 B_2(r)dB_2(r) \), making the distribution nonsymmetrical 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 \( \omega_{21} \) and \( \Delta_{21} \), respectively.

Second, comparing the asymptotic distributions of \( OLS \) and \( NLS \), we can see that \( \Delta_{21} \) 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 to perform better than NLS in finite samples. This is the case if $\beta$ is large and $\theta$ is close to zero.

Third, the presence of the nuisance parameter $\omega_{21}$ 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 for looking at expressions (19) and (24), comparing the asymptotic distributions of PIV and OLS, it clearly appears preferable the former for any value of the parameter space. On the other hand, comparing the asymptotic distributions of PIV and NLS, it shows up that NLS can be less asymmetric than PIV if, for instance, $\beta$ is negative and $\theta$ positive.

Fourth, the limiting distribution of MLECM is a median-unbiased mixture of normals. In fact, it is asymptotically efficient as proved by Phillips (1991b) and Saikkonen (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 (31).

5. ASYMPTOTIC EFFICIENCY OF PIV

From the previous section we obtained a lack of optimality of PIV for all $d \in \partial$. Moreover, with respect to OLS we found advantages in using PIV only in the unit root case. Since the optimality of the FM-OLS estimator for $d > 1$ has been recently dealt with by Dolado and Marmol (1998), in the rest of the paper we will focus the attention on the DGP (5)-(7) in the particular $d = 1$ and $d_z = 0$ case.

To overcome the optimality problems of PIV in the unit root case, herein we suggest a semi-parametric correction of this estimator as that originally proposed by Park and Phillips (1988, 1989) and Phillips and Hansen (1990). This procedure, known as fully modified, proceeds as follows. First of all, notice that $\hat{\beta}_{piv}$ is not optimal because $\omega_{21} \neq 0$. Therefore, as a first step, define the kernel estimators

\begin{align*}
(33) \quad \hat{\omega}_{21} &= \sum_{j=-k}^{k} \ell \left( \frac{j}{k} \right) T^{-1} \sum \left( \Delta x_{t-j} \hat{u}_t \right), \\
(34) \quad \hat{\omega}_{22} &= \sum_{j=-k}^{k} \ell \left( \frac{j}{k} \right) T^{-1} \sum \left( \Delta x_{t-j} \Delta x_t \right),
\end{align*}

where $\Sigma'$ signifies summation over $1 \leq t, t - j \leq T$ and $\hat{u}_t$ denotes the PIV residuals from equation (5).
The kernel function $\ell(\cdot): \mathbb{R} \to [-1, 1]$ 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 $\Omega$.

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

Dolado and Marmol (1998) show that, in the OLS case, if $k, T \to \infty$ but $kT^{-1/2} \to 0$, then $\hat{\omega}_{2i} \sim \omega_{2i}$, $i = 1, 2$ for all $d \in \partial$. The same remains true in the PIV case using similar arguments. Hence, in terms of $O_k$, this implies that $k = O_k(T^k)$ for some $k \in (0, 1)$. 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 $k - cT^{2/3}$ (cf., Andrews, 1991), with $c$ a constant that applies when minimizing the asymptotic mean squared error of kernel estimates such as (33) and (34).

Now define the endogeneity bias-corrected disturbance

$$u_{it}^* = u_{it} - \frac{\omega_{12}}{\omega_{22}} u_{2t},$$

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

$$(36) \quad \hat{u}_{it}^* = \hat{u}_{it} - \frac{\hat{\omega}_{12}}{\hat{\omega}_{22}} \hat{\Delta} x_i.$$

Subtracting $\hat{\omega}_{12}\hat{\omega}_{22}^{-1}\hat{\Delta} x_i$ from both sides of (5) finally yields

$$(5') \quad \hat{y}_t^* = \beta x_i + \hat{u}_{it}^*,$$

where $\hat{y}_t^* = y_t - \hat{\omega}_{12}\hat{\omega}_{22}^{-1}\hat{\Delta} x_i$. Our proposed estimator, called fully modified pseudo instrumental variable (FM-PIV), will have then the expression

$$(36) \quad \hat{\beta}_{PIV}^* = \frac{\sum_{t=r+1}^{T} \hat{y}_t^* x_{t-r}}{\sum_{t=r+1}^{T} x_t x_{t-r}}.$$
THEOREM 4. Given the DGP (5)-(7) with $d = 1$ and $d_z = 0$, then, under Assumption 1 with $k = O_e(T^k)$ for some $k \in (0, \frac{1}{2})$ and $\tau = o(T^{1/4})$, asymptotically, as $T \to \infty$ and $\tau \to \infty$,

(37) \( \hat{\beta}_{piv} \overset{p}{\to} \beta \)

and

(38) \( T(\hat{\beta}_{piv} - \beta) \Rightarrow \frac{\int_0^1 B_2(r)dB_{12}(r)}{\int_0^1 B_2^2(r)dr} \),

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

Notice that the truncation lag $\tau$ should increase with the sample size $T$ at the same rate as the Newey and West’s (1987) correction. On the other hand from equation (37) it follows that $FM-PIV$ is a consistent estimator of $\beta$. Moreover, from equation (38), it is also median-unbiased, a mixture of normals and asymptotically efficient (Phillips, 1991b, Saikkonen, 1991). In particular, and returning to the data generating process (26)-(28), Gonzalo (1994) proves that in this case \( \omega_{12}^{1/2} = \sigma_1(1-\theta^2)^{1/2}/(1-\phi) \), and since \( B_{12}(r) = \omega_{12}^{1/2}W_1(r) \), it follows from equations (31) and (38) 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 $\beta$,

(39) \( t_\beta = \frac{\hat{\beta}_{piv} - \beta}{(\hat{\omega}_{12})^{1/2}\left(\sum_{i=1}^T x_i^2\right)^{-1/2}} \).

Then, by using Park and Phillips’ (1988) Lemma 5.1, under the null hypothesis $H_0: \beta = \beta_0$ and under the same assumptions as in Theorem 4, it is straightforward to prove that, asymptotically, $t_\beta \Rightarrow N(0,1)$. More in 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.

6. CONCLUDING REMARKS

In this paper we have studied the properties of an alternative characterization of the concept of linear cointegration that could be extended to the nonlinear case along the lines of Aparicio and Escribano (1999) since it is not model dependent. Such a characterization exploited the relationship between the autocorrelation and the cross-correlation 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 a delayed replica of the regressor variable in the cointegrating equation.

In the unit root case and after a semi-parametric correction of the endogeneity bias of the type originally proposed by Park and Phillips (1988, 1989) and Phillips and Hansen (1990), we derived an estimator, called FM-PIV, with a median unbiased mixture of normals that is asymptotically efficient and from which standard inference can be conducted.

For practical purposes, FM-PIV estimators require the specification of the kernel function, \( \ell(\cdot) \), the bandwidth parameter, \( k \), and the truncation lag, \( \tau \). 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, for instance, 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 \( k \sim cT^{\frac{1}{3}} \), the method proposed by Andrews (1991), possibly after prewhitening with a first-order vector autoregression prior to kernel estimation, as suggested by Andrews and Monahan (1992), seems to be a good choice. For instance, Haug (1998) employs a quadratic spectral kernel with the associated automatic, data-dependent, plug-in bandwidth estimator to construct the kernel estimators of the fully-modified procedure.
Third, as regards the choice of $\tau$, the hint from Theorem 4 is that the truncation lag $\tau$ of the instrumental variable should increase with the sample size $T$ but being $o(T^{1/4})$. If the truncation lag is chosen too small, the tests could be biased. Instead, if the truncation point is too large, there could be a loss of efficiency. As usual, it appears that one the best option is to employ data dependent rules which incorporate the sample information, for instance,

$$\tau_{opt} = \arg\min_{\tau} \left\{\log \left[ \frac{1}{T-\tau} \sum_{t=\tau+1}^{T} (y_t - \hat{\beta}_{PIV} x_{t-\tau})^2 + \frac{C_T \tau}{T} \right] \right\},$$

where $C_T$ denotes a function of $T$ such that $C_T > 0$ and $T^{-1}C_T \xrightarrow{\tau \to \infty} 0$ (see, e.g., Andrews, 1991) or to use some alternative standard order selection criteria.

Finally, the practical implementation of the $FM-PIV$ estimator could be done, for instance, by using the GAUSS subroutines provided in the computer software COINT (Ouliaris and Phillips, 1993).

**MATHEMATICAL APPENDIX**

**PROOF OF THEOREM 1.** From the definition of fractional cointegration, it follows that when $x_t, y_t$ are cointegrated, there exist a nonzero finite real number, $\beta$, such that

(A.1) \[ \lim_{\tau \to \infty} \rho_{yx}(\tau, t) = \beta \lim_{\tau \to \infty} \rho_x(\tau, t) + \lim_{\tau \to \infty} \rho_{xx}(\tau, t) \]

or

(A.2) \[ \lim_{\tau \to \infty} \frac{\rho_{yx}(\tau, t)}{\rho_x(\tau, t)} = \frac{\beta + \lim_{\tau \to \infty} \rho_{xx}(\tau, t)}{\rho_x(\tau, t)} = \beta + \lim_{\tau \to \infty} \frac{E(z_t x_{t-\tau})}{E(x, x_{t-\tau})}. \]

Thus, in order to prove the theorem, it suffices to show that the ratio of the right-hand side of (A.2) vanishes as $\tau$ goes to infinity. For this, from the Wold representations of $x_t$ and $z_t$,

(A.3) \[ \Delta^d z_t = u_{1t}, \]

(A.4) \[ \Delta^d x_t = u_{2t}, \]

where, under Assumption 1, $u_t = (u_{1t}, u_{2t})'$ is a linear process, we obtain

(A.5) \[ u_t = \sum_{j=0}^{m} \Pi_j e_{t-j}, e_t = 0 \text{ for } t \leq 0, \]

or

\[ 20 \]
\[(A.6) \quad u_t = \sum_{j=0}^{t-1} \Pi_j e_{t-j}, \]

where we have imposed the initial conditions \( e_t = 0 \) for \( t \leq 0 \), to obtain well-defined nonstationary processes. Now defining \( Y_t = (z_t, x_t)' \) we have that

\[(A.7) \quad Y_t = \begin{pmatrix} \Delta^{-d} u_{t_1} \\ \Delta^{-d} u_{t_2} \end{pmatrix} = \sum_{j=0}^{\infty} \begin{pmatrix} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} & 0 \\ 0 & \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \end{pmatrix} \begin{pmatrix} u_{t-j} \\ u_{t_2-t-j} \end{pmatrix} = \sum_{j=0}^{\infty} H_j u_{t-j}. \]

Using (A.6) and (A.7), after a little algebra we obtain

\[(A.8) \quad Y_t = \sum_{j=0}^{t-1} M_j e_{t-j}, \]

where

\[(A.9) \quad M_j = \sum_{i=0}^{j} H_{j-i} \Pi_i, \]

is the general term of the convolution of the matrix sequences \( \{H_j\} \), \( \{\Pi_j\} \).

Consequently,

\[(A.10) \quad E(Y_t Y_{t-t}) = E\left( \sum_{j=0}^{t-1} M_j e_{t-j} \right) \left( \sum_{i=0}^{t-1} M_i e_{t-i} \right) = \sum_{j=0}^{t-1} M_{t-j} M_j. \]

Therefore, given that, for \( \tau \) large, the convolution (A.9) is dominated by the diagonal matrix \( H_j \) and using Sheppard's lemma, expression (A.10) becomes asymptotically equivalent to

\[(A.11) \quad \sum_{j=0}^{t-1} \text{diag}\{j^{d-1}, j^{d-1}\} A \text{diag}\{j^{d-1}, j^{d-1}\}, \]

with \( A \) denoting a positive matrix of constant elements. Expression (A.11), in turn, for \( \tau \) large is of the same order as

\[(A.12) \quad \begin{pmatrix} g_{zz} \tau^{2d-1} & g_{zx} \tau^{d+z-1} \\ g_{zx} \tau^{d+z-1} & g_{xx} \tau^{2d-1} \end{pmatrix}, \]

for suitable positive constants \( g_{ij}, i, j = z, x \).

Putting together (A.10) and (A.12) yield

\[(A.13) \quad \frac{\rho_{zz}(\tau,t)}{\rho_z(\tau,t)} = \frac{E(z_{t-t})}{E(z_{t})} \propto \tau^{d-\tau-d}, \quad \tau > 0, \]

\[\text{for } g > 0.\]
for $\tau$ sufficiently large. Consequently, given that, under the cointegration hypothesis, $d_z < d$, (A.13) converges to zero as $\tau$ goes to infinity, proving expression (9). The converse statement follows using the same argument.

Finally, notice from (A.2) that, under the spurious hypothesis, i.e., when $d_z = d$ and $x_t, y_t$ are independent of each other for all $t, t'$ so that $\beta = 0$, then

$$\rho_{xy}(\tau, t) = E(y_t x_{t-\tau}) = E(y_t)E(x_{t-\tau}) = 0,$$

for all $\tau$ sufficiently large, so that expression (10) follows in a trivial way. \qed

**Proof of Theorem 3.** Without any loss of generality, we will prove the theorem for $d \in \left(\frac{1}{2}, \frac{3}{2}\right)$, the extension to $d \in \partial$ being straightforward by noting that any $d \in \partial$ can always been decomposed as $d = q + d^*$, with $q = 0, 1, 2, \ldots$, and $d^* \in \left(\frac{1}{2}, \frac{3}{2}\right)$, so that a fractionally integrated process of order $d$ can be regarded as the $q$-fold of a fractionally integrated process of order $d^*$.

Assume then that $d \in \left(\frac{1}{2}, \frac{3}{2}\right)$. To prove the theorem, we proceed by parts. Hence, we will first analyze the denominator of

$$\hat{\beta}_{pvy} = \frac{\sum_{t=r+1}^{T} u_t x_{t-r}}{\sum_{t=r+1}^{T} x_{t-r}}.
\text{(A.16)}$$

We have

$$\sum_{t=r+1}^{T} x_{t-r} = \frac{1}{2} \sum_{t=r+1}^{T} x_t^2 + \frac{1}{2} \sum_{t=r+1}^{T} x_{t-r}^2 - \frac{1}{2} \sum_{t=r+1}^{T} (x_t - x_{t-r})^2.
\text{ (A.17)}$$

On the one hand,

$$\frac{1}{2} \sum_{t=r+1}^{T} x_t^2 = \frac{1}{2} \left( \sum_{t=1}^{T} x_t^2 - \sum_{t=1}^{[T]} x_t^2 \right),$$

where $\alpha = r/T$ and $[\cdot]$ denotes the integer part operator. Then, from Dolado and Marmol (1998, Theorem 1) and the continuous mapping theorem (CMT), asymptotically, when $T \to \infty$, $\tau$ fix, we obtain

$$\frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_t^2 \Rightarrow \frac{1}{2} \int_{0}^{1} B_2^{d}(d, r) dr - \frac{1}{2} \int_{0}^{\alpha} B_2^{d}(d, r) dr,$$

and as $\tau \to \infty$ with $\alpha = \tau/T \to 0$, yields
(A.18) $\frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_t^2 \Rightarrow \frac{1}{2} \int_0^1 B_2^2(d,r)dr$.

On the other hand, noting that $\sum_{t=r+1}^{T} x_t^2 = \sum_{t=r+1}^{T} x_t^2 - \sum_{t=T-r+1}^{T} x_t^2$, under Assumption 1, it follows from Dolado and Marmol (1998, Theorem 1) and the CMT that

(A.19) $\frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_t^2 \Rightarrow \frac{1}{2} \int_0^1 B_2^2(d,r)dr$.

Rewrite now $x_t$ as $\Delta x_t = \eta_{2t}$ with $\Delta^d \eta_{2t} = u_{2t}$, $|\delta| < \frac{1}{2}$, so that $x_t = x_{t-r} + e_{2t}(r)$, where $e_{2t}(r) = \sum_{j=0}^{r-1} \eta_{2t-j}$. Then,

$$\frac{1}{2} \sum_{t=r+1}^{T} (x_t - x_{t-r})^2 = \frac{1}{2} \sum_{t=r+1}^{T} e_{2t}^2(r),$$

but since

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

it follows that

(A.20) $T^{-1} \sum_{t=r+1}^{T} e_{2t}^2(r) \xrightarrow{p} \frac{1}{2} \omega_{2d}(d)$,

where $\omega_{2d}(d)$ denotes the long-run variance of $\eta_{2t}$. Now, using (A18)-(A.20) and the CMT, yields

(A.21) $T^{-2d} \sum_{t=r+1}^{T} x_t x_{t-r} = \frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_t^2 + \frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_{t-r}^2$

$$- \frac{1}{2} \left( \frac{T}{T} \right)^{2d-1} T^{-1} \sum_{t=r+1}^{T} (x_t - x_{t-r})^2 =$$

$$= \frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_t^2 + \frac{1}{2} T^{-2d} \sum_{t=r+1}^{T} x_{t-r}^2 + o_p(1) \Rightarrow \frac{1}{2} \int_0^1 B_2^2(d,r)dr.$$

Consider now the numerator of (A.16),

(A.22) $\sum_{t=r+1}^{T} u_{1t} x_{t-r} = \sum_{t=r+1}^{T} u_{1t} x_t - \sum_{t=r+1}^{T} e_{2t}(r) u_{1t}$.

With regard the term $\sum_{t=r+1}^{T} u_{1t} x_t$, it follows under Assumption 1 from Dolado and Marmol (1998, Theorem 1) and the CMT that, as $\alpha, T \to \infty$,
(A.23) \( T^{-d} \sum_{t=r+1}^{T} x_i u_{it} \Rightarrow \int_0^1 B_1(d,r)dB_1(r) \) if \( d > 1 \),

(A.24) \( T^{-1} \sum_{t=r+1}^{T} x_i u_{it} \Rightarrow \int_0^1 B_1(r)dB_1(r) \) if \( d = 1 \)

and

(A.25) \( T^{-1} \sum_{t=r+1}^{T} x_i u_{it} \xrightarrow{p} 0 \) if \( \frac{1}{2} < d < 1 \).

As regards the second term in the right hand side of equation (A.22), notice that

\[
\sum_{t=r+1}^{T} e_{2t}(r)u_{it} = \sum_{t=r+1}^{T} \eta_{2t}u_{it} + \sum_{t=r+1}^{T-1} \sum_{i=1}^{r-1} \eta_{2t,i}u_{it},
\]

and then, when \( T \to \infty \) and \( \tau \) fix, using the ergodic theorem we have that

\[
T^{-1} \sum_{t=r+1}^{T} e_{2t}(r)u_{it} \xrightarrow{p} \sum_{k=0}^{\infty} E(\eta_{2,0}u_{ik}),
\]

and thus, when \( \tau \to \infty \),

(A.26) \( T^{-1} \sum_{t=r+1}^{T} e_{2t}(r)u_{it} \xrightarrow{p} \sum_{k=0}^{\infty} E(\eta_{2,0}u_{ik}) = \Delta_{21}(d). \)

Theorem 3 now follows from collecting all the previous results and apply the CMT.

PROOF OF THEOREM 4. Given the expression of the FM-PIV estimator,

(A.27) \( \hat{\beta}_{PIV} = \frac{\sum_{t=r+1}^{T} \hat{\mu}_{it}^* x_{t-r}}{\sum_{t=r+1}^{T} x_i x_{t-r}} \),

we have

(A.28) \( \sum_{t=r+1}^{T} \hat{\mu}_{it}^* x_{t-r} = \sum_{t=r+1}^{T} u_{it} x_{t-r} - \hat{\omega}_{1z} \hat{\varphi}_{22}^{-1} \sum_{t=r+1}^{T} u_{2t} x_{t-r} \).

With respect to the first term in the right hand side of (A.28), we know from Theorem 3 that

(A.29) \( T^{-1} \sum_{t=r+1}^{T} u_{it} x_{t-r} \Rightarrow \int_0^1 B_1(r)dB_1(r) \).

As regards the second term in the right hand side of (A.28), proceeding as in the proof of Theorem 3 yields
so that, using the same arguments as in the preceding theorem, we get

\[(A.30) \quad T^{-1} \sum_{t=r+1}^T u_{2t} x_{t-r} = T^{-1} \sum_{t=r+1}^T u_{2t} x_t - T^{-1} \sum_{t=r+1}^T e_{2t}(\tau)u_{2t}, \]

provided that \( a \to 0 \) as \( r \to \infty \) and \( T \to \infty \).

Consequently, from (A.28)-(A.30) and the consistency of \( \omega_{12}, \omega_{22} \), one obtains

\[(A.31) \quad T^{-1} \sum_{t=r+1}^T \hat{u}_{2t} x_{t-r} \Rightarrow \int_0^1 B_2(r)dB_1(r) - \omega_{12} \omega_{22}^{-1} \int_0^1 B_2(r)dB_2(r) = \int_0^1 B_1(r)dB_{12}(r), \]

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

Lastly, as regards the denominator in (A.27), notice from the corresponding part of the proof of Theorem 3 that for expression (A.21) to hold, we can apply Newey and West’s (1987) results to the term \( \sum_{t=r+1}^T e_{2t}^2(\tau) \) such that equation (A.20) also holds for \( \tau = o(T^{1/4}) \). The theorem finally follows from (A.21), (A.31) and the CMT.

REFERENCES


