

Asymptotic inference results for multivariate long-memory processes

JUAN J. DOLADO AND FRANCESC MARMOL



*Department of Economics, Universidad Carlos III de Madrid, Madrid 126, 28903, Getafe
(Madrid), Spain*

E-mail : dolado@eco.uc3m.es; francesc.marmol@uc3m.es

Received: October 2002

Summary In this paper, we extend the well-known Sims, Stock and Watson (SSW) (Sims et al. 1990; *Econometrica* 56, 113–44), analysis on estimation and testing in vector autoregressive process (VARs) with integer unit roots and deterministic components to a more general set-up where non-stationary fractionally integrated (NFI) processes are considered. In particular, we focus on partial VAR models where the conditioning variables are NFI since this is the only finite-lag VAR model compatible with such processes. We show how SSW's conclusions remain valid. This means that whenever a block of coefficients in the partial VAR can be written as coefficients on zero-mean $I(0)$ regressors in models including a constant term, they will have a joint asymptotic normal distribution. Monte Carlo simulations and an empirical application of our theoretical results are also provided.

Keywords: *Vector fractionally integrated processes, Fractional cointegration, Granger causality, Permanent income hypothesis.*

1. INTRODUCTION

Fractionally integrated (FI) processes have recently drawn a great deal of attention among macroeconomists and econometricians with both theoretical and empirical research growing fast and in parallel. Loosely speaking, a series x_t is said to be *FI of order d* , in short $I(d)$, if $\Delta^d x_t$ is a weakly stationary or short memory $I(0)$ process, with $\Delta^d = \sum_{k=0}^{\infty} \pi_k (-d) L^k$, $\pi_k (-d) = \frac{k-1-d}{k} \pi_{k-1}(-d)$, $\pi_0(-d) = 1$, where d is a real number. The degree of integration or *memory parameter*, d , determines the key dynamic or memory properties of the series. In particular, an $I(d)$ process is stationary if $d < \frac{1}{2}$. When $\frac{1}{2} < d < 1$, despite being non-stationary, the process is non-persistent, i.e. any random shock has only a transitory influence on the series. Finally, when $d \geq 1$, the process is both non-stationary and persistent, i.e. any random shock has a permanent effect on the future path of the series. Consequently, FI processes cover a wide range of dynamic behaviour which is ruled out *a priori* if d is restricted to take integer values as in the literature on unit roots.

Our goal in this paper is to analyse the properties of tests of hypotheses based on Ordinary Least Squares (OLS) estimators of the parameters in reduced-form autoregressive distributive lag (ADL) models whose variables include non-stationary FI (NFI) processes with *a priori* known order of integration. These models, in which a set of endogenous variables are regressed on their

own lags and lags of a set of conditioning variables, are often used in applied work to test relevant economic hypotheses. A popular application is the test of the Rational Expectations Permanent-Income Hypothesis of consumption which typically relies upon regressing (logged) consumption on its own lags and lags of (logged) disposable income, to further test for the joint significance of the latter set of coefficients.

In particular, we analyse whether the well-known guidelines proposed by Sims *et al.* (1990) (SSW) and Stock and Watson (1993) for a vector autoregressive process (VAR) with integer unit roots ($d = 1, 2, \dots$) continue to be valid when variables are extended to be NFI processes. In this respect, it should be stressed at the outset that dealing with NFI processes precludes the analysis of full-system VARs with a finite number of lags in the levels of the series (as in SSW) since finite-order VAR models with *i.i.d.* innovations cannot generate FI series. This is so because the application of the filter Δ^d to any vector of series gives rise to an infinite number of parameters in its AR representation whenever d is a real number. Thus, the analysis in what follows is restricted to partial VARs where the conditioning variates are NFI processes. This set-up automatically implies that, if the roots of the ADL lag polynomial are outside the unit circle, there will be a *cointegrating relationship* among the variables in the system. In particular, the single-equation ADL model, which constitutes the focus of most of this paper, can be interpreted as a specific representation of a partial VAR where only one of the variables is modelled.

More concretely, we study reduced-form ADL models whose elements are individually $I(0)$ and $I(d)$ with $\frac{1}{2} < d < \frac{3}{2}$ processes, possibly around a linear time trend, since this range of values is the most relevant one when modelling economic time series (see, e.g. Baillie 1996). We show that SSWs main conclusions, drawn for the case of integer unit roots, remain valid in our more general fractional set-up. This implies that whenever a block of coefficients can be written as coefficients on zero-mean $I(0)$ regressors in a model that includes a constant term, they will have a joint asymptotic normal distribution. Therefore, linear restrictions on this set of parameters can be tested using standard asymptotic chi-square distribution theory. Otherwise, the associated test statistics will have non-standard limiting distributions. Moreover, we show that the OLS estimator of the cointegrating vector is consistent with the individual coefficients converging to their theoretical counterparts at different rates, depending on the memory parameters.

The rest of this paper is structured as follows. In Section 2, we rely upon a multivariate functional central limit theorem (FCLT) for NFI processes due to Marinucci and Robinson (2000), based on the so-called *Type-II* definition of fractional Brownian motion, to derive some new results concerning weak convergence of stochastic integrals having fractional integrands and weakly stationary integrators. Next, in Section 3, we use these convergence results to extend the SSW set-up allowing for partial VAR models with NFI processes. We introduce the relevant partial VAR model and then provide the asymptotic representation of the OLS estimator by transforming the regressors in the way suggested by SSW, namely, by isolating in different blocks the various stochastic and deterministic trends. In Section 4, we illustrate how our asymptotic results perform in finite samples by means of a small Monte Carlo study. Section 5, in turn, provides an empirical application where some of the available evidence regarding the random-walk property of consumption and its long-run relationship with disposable income is revisited in light of the results of this paper. Finally, Section 6 concludes. Proofs of the main results are gathered in the Appendix.

2. PRELIMINARY THEORY

2.1. Definitions of NFI processes

Let $\{\eta_t\}$ be an n -dimensional *i.i.d.* sequence of random vectors with $E(\eta_1) = 0$ whose covariance matrix is the identity matrix of order n , I_n . Further, let $a_t = \Upsilon(L)\eta_t$, $\Upsilon(L) = \sum_{j=0}^{\infty} \Upsilon_j L^j$ (L denotes the lag operator), and let $x_t = (x_{1t}, \dots, x_{nt})'$ be an n -dimensional vector of NFI processes with Wold representation given by

$$x_t = \Delta(L)^{-1} a_t 1 \{t > 0\}, \quad t = 1, 2, \dots, \quad (1)$$

where $\Delta(L) = \text{diag}\{\Delta^{d_1}, \Delta^{d_2}, \dots, \Delta^{d_n}\}$, $\Delta^{d_i} = (1-L)^{d_i} = \sum_{k=0}^{\infty} \pi_{ki}(-d_i)L^k$, $\pi_{ki}(-d_i) = \frac{k-1-d_i}{k} \pi_{k-1}(-d_i)$, $\pi_{0i}(-d_i) = 1$, $\frac{1}{2} < d_i < \frac{3}{2}$, $i = 1, 2, \dots, n$, and $1\{\cdot\}$ is the indicator function. Note that the assumption about the initial condition, i.e. $a_t = 0$ for $t \leq 0$, is necessary to cater for cases where $d \geq \frac{1}{2}$, because otherwise the right-hand side of (1) do not converge in mean square and hence x_t is not well defined. Therefore (1) implies $x_t = 0$, $t \leq 0$, and it can be verified that x_t is non-stationary long-range dependent in the extended sense of Heyde and Yang (1997) (*cf.* Robinson and Marinucci 2001, Lemma 3.4).

As pointed out by Marinucci and Robinson (2001), an alternative definition for an NFI process with $\frac{1}{2} < d < \frac{3}{2}$ could be

$$x_t = \Delta^{-1} \zeta_t 1 \{t > 0\}, \quad t = 1, 2, \dots, \quad (1')$$

such that $\Delta = 1 - L$ and ζ_t is a stationary $I(d-1)$, $-\frac{1}{2} < d-1 < \frac{1}{2}$. The distinction between definitions (1) and (1') has been discussed at length in Marinucci and Robinson (1999) where it has been shown that, for $\frac{1}{2} < d < \frac{3}{2}$, first differences of x_t using either definition are asymptotically equivalent in the mean square sense (see also Lemmae 1 and 2 in Dolado *et al.* 2002). However, an appropriately scaled partial sum of x_t defined according to (1') converges in distribution to the so-called *Type-I* fractional Brownian motion (fBM) whereas, if defined according to (1), it converges to a *Type-II* fBM (*cf.* Levy (1953) and Mandelbrot and Van Ness (1968)). As argued by Marinucci and Robinson (2001), in common with the literature on integer unit-roots which defines an $I(d)$ process, $d = 1, 2, \dots$, as $\Delta^d x_t$ being $I(0)$, definition (1) is adopted in the sequel as the more natural one when considering the range $\frac{1}{2} < d < \frac{3}{2}$.

Next, in dealing with the more general case where $d > \frac{1}{2}$, the following assumptions are introduced.

Assumption A $\sum_{j=0}^{\infty} j \|\Upsilon_j\| < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.

Assumption B $\Upsilon(1)$ is non-singular.

Assumption C $E\|\eta_1\|^q < \infty$, $q > \max\{4, \frac{2}{2d-1}\}$, $d = \min_{1 \leq i \leq n} d_i$.

Assumption A is satisfied, among others, by all (asymptotically) stationary and invertible ARMA processes. Assumption B ensures that the asymptotic limit process will have non-degenerate finite dimensional distributions. Finally, Assumption C states that a larger amount of persistence, i.e. a larger d , implies weaker moment conditions, at least for $\frac{1}{2} < d < \frac{3}{4}$.

Define the normalising matrix function $\mathfrak{F}_T = \text{diag}\{T^{1/2-d_1}, \dots, T^{1/2-d_n}\}$, where T is the sample size, and let $x_T(r) = \mathfrak{F}_T x_{\lfloor Tr \rfloor}$, for $0 \leq r \leq 1$, where $\lfloor \cdot \rfloor$ is integer part, so that $x_T(r) \in D[0, 1]^n$, the space of \mathfrak{R}^n -valued vector functions on $[0, 1]$ whose components are continuous on the right and with finite left limit. Then, under Assumptions A–C, Marinucci and Robinson (2000)

have proved that, as $T \rightarrow \infty$,

$$x_T(r) \Rightarrow B(d_x, r), \quad (2)$$

where ' \Rightarrow ' signifies convergence in the Skorohod J_1 topology of $D[0, 1]^n$, $d_x = (d_1, \dots, d_n)$ and $B(d_x, r)$ is a multivariate fBM for $r \geq 0$, defined as

$$B(d_x, r) = (0, \dots, 0)' \quad \text{a.s., } r = 0, \quad (3)$$

$$B(d_x, r) = \int_0^r G(r, s) dB(s), \quad r > 0, \quad (4)$$

such that $B(r)$ is an n -dimensional Brownian motion with covariance matrix $\Omega = \Upsilon(1) \Upsilon(1)'$, and the $(n \times n)$ matrix $G(r, s)$ has (i, j) th element $\frac{1}{\Gamma(d_i)}(r-s)^{d_i-1}$, $i, j = 1, \dots, n$, for $0 \leq s \leq r$, and zero otherwise. $\Gamma(\cdot)$ stands for the gamma or generalised binomial function.

Formally, $B(d_x, r)$ is a Gaussian process with almost surely continuous sample paths and non-stationary (and non-independent) increments. As discussed above, $B(d_x, r)$ corresponds to a *Type-II* fBM in terms of Marinucci and Robinson's (1999) terminology.

In the empirically relevant case where $d_1 = d_2 = \dots = d_n = d$, (1) becomes

$$\begin{aligned} x_t &= \Delta^{-d} a_t 1\{t > 0\} = \Upsilon(1) \Delta^{-d} \eta_t 1\{t > 0\} + \Upsilon^*(L) \Delta^{1-d} \eta_t 1\{t > 0\} \\ &= \Upsilon(1) \xi_t(d) + \Upsilon^*(L) \Delta^{1-d} \eta_t 1\{t > 0\}, \end{aligned} \quad (5)$$

where use has been made of the well-known multivariate Beveridge–Nelson decomposition (cf. Phillips and Solo 1992), with $\xi_t(d) = \sum_{k=0}^{t-1} \pi_k(d) \eta_{t-k}$, $\xi_t(1) = \xi_t$, and $\Upsilon^*(L) = \sum_{j=0}^{\infty} \Upsilon_j^* \eta_{t-j}$ is absolutely summable with coefficients $\Upsilon_j^* = \sum_{i=j+1}^{\infty} \Upsilon_i$. Since $T^{1/2-d} \Upsilon^*(L) \Delta^{1-d} \eta_{[Tr]}$ is $o_p(1)$ uniformly in r , (2)–(5) imply

$$T^{1/2-d} x_{[Tr]} \Rightarrow B(d, r) \equiv \Upsilon(1)W(d, r), \quad (6)$$

and

$$T^{1/2-d} \xi_{[Tr]}(d) \Rightarrow W(d, r), \quad (7)$$

where a standardised *Type-II* fBM is defined as $W(d, r) = \frac{1}{\Gamma(d)} \int_0^r (r-s)^{d-1} dW(s)$, $r > 0$, and $W(r)$ is an n -dimensional standard Brownian motion.

2.2. Weak convergence of sample covariance matrices

The use of the above multivariate functional central limit theorem, together with the continuous mapping theorem (CMT), provides the necessary tools to ensure the convergence of most of the relevant random matrices appearing in the estimation and testing of long-run relationships among NFI processes considered in this paper. It should be noted that invariance principles for *Type-I* fBM have been previously established, among others, by Davydov (1970) and Taqqu (1975), and extended by Chan and Terrin (1995), Csörgo and Mielniczuk (1995) and Davidson and de Jong (2000). Similarly, Akonom and Gourieroux (1987) and Marinucci and Robinson (1999) have derived similar results using *Type-II* fBM. However, the available theory for NFI processes as defined in (1) does not cover the very important case of the weak convergence of the sample covariance matrices $\sum_{t=1}^T x_t a_t'$ since convergence cannot be obtained in this case from a routine

application of the CMT and a multivariate invariance principle. As will be seen below, expressions of the form $\sum_{t=1}^T x_t a_t'$ arise as key elements in the theory of cointegrating regressions with weakly dependent error terms, where the numerator of the deviations of regression coefficients from their true parameter values involves sample covariances of NFI series (regressors, x_t) and weakly stationary series (disturbances, a_t).

Nonetheless, there are related results concerning the weak convergence of $\sum_{t=1}^T x_t a_t'$ in a different set-up to ours. In the integer case, with $d = 1, 2, \dots$, results have been provided, *inter alia*, by Strasser (1986), Chan and Wei (1988), Phillips (1988), Hansen (1992) and de Jong and Davidson (2000) under different assumptions on the innovation sequence. Moreover, the fractional case has been studied by Fox and Taqqu (1987), Chan and Terrin (1995) and Marinucci (2000), whose results are cast in the frequency domain, in terms of the Wiener–Itô calculus of the spectral representation of such a covariance. As regards the results cast in the time domain, weak convergence of $\sum_{t=1}^T x_t a_t'$ for NFI processes and weakly dependent innovations, has been derived by Davidson and de Jong (2000) but using the *Type-I* definition of fBM given in (1'). Thus, our first contribution here is to establish weak convergence of such a covariance term using definition (1) and, therefore, the asymptotic theory based on *Type-II* fBMs.

From (5), the relevant cross sample moment turns out to be $\sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]'$. Asymptotic results for the $I(1)$ case have been previously provided by SSW (Lemma 1) who prove that, under Assumptions A–C, as $T \rightarrow \infty$,

$$T^{-1} \sum_{t=1}^T \xi_t [\Upsilon(L) \eta_t]' \Rightarrow \Upsilon(1)' + \int_0^1 W(r) dW(r)' \Upsilon(1)' \quad (8)$$

In the following lemma we extend their result by providing the limiting distribution of $\sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]'$ for $d > \frac{1}{2}$. Notice that we establish a more general result than that needed for the range $\frac{1}{2} < d < \frac{3}{2}$ which is the one considered below when dealing with partial VARs.

Lemma 1 *Under definition (1) of an NFI process and assumptions A–C, as $T \rightarrow \infty$,*

$$T^{-d} \sum_{t=1}^T \xi_{t-1}(d) [\Upsilon(L) \eta_t]' \Rightarrow \int_0^1 W(d, r) dW(r)' \Upsilon(1)', \quad \text{if } d > 1, \quad (9)$$

$$T^{-1} \sum_{t=1}^T \xi_{t-1}(d) [\Upsilon(L) \eta_t]' \xrightarrow{p} \sum_{j=0}^{\infty} \pi_j(d-1) \Upsilon_j^{*'} \quad \text{if } \frac{1}{2} < d < 1, \quad (10)$$

$$T^{-d} \sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]' \Rightarrow \int_0^1 W(d, r) dW(r)' \Upsilon(1)', \quad \text{if } d > 1, \quad (11)$$

and

$$T^{-1} \sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]' \xrightarrow{p} \sum_{j=0}^{\infty} \pi_j(d-1) \tilde{\Upsilon}_j', \quad \text{if } \frac{1}{2} < d < 1, \quad (12)$$

where $\pi_j(d-1)$ stands for the j -th element in the binomial expansion of Δ^{1-d} , $\tilde{\Upsilon}_j = \sum_{i=j}^{\infty} \Upsilon_i$ and \xrightarrow{p} denotes convergence in probability.

Lemma 1 states that, after proper normalisation, $\sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]'$ converges weakly to a stochastic integral with respect to a Brownian motion for all $d \geq 1$, plus a drift term in the particular unit root ($d = 1$) case. By contrast, in the non-persistent case, with $\frac{1}{2} < d < 1$, as long as $\sum_{j=0}^{\infty} \pi_j(d-1) \tilde{\Upsilon}_j' \neq 0$, it holds that $\sum_{t=1}^T \xi_t(d) [\Upsilon(L) \eta_t]'$ is $O_p(T)$ for all $\frac{1}{2} < d < 1$, converging in probability to finite constants.

3. ESTIMATION OF COINTEGRATING VECTORS IN FRACTIONAL SYSTEMS

In order to apply the previous asymptotic results to scenarios of potential economic interest, we assume that the relevant DGP has the following (reduced form) triangular representation

$$y_{1t} = \mu + \theta' y_{2t-1} + u_{1t}, \quad (13)$$

$$\Delta_*(L) y_{2t} = u_{2t}, \quad (14)$$

where $x_t = (y_{1t}, y_{2t}')'$ is an n -dimensional vector of variables ($n = m + 1$) such that y_{1t} is a scalar NFI process of order d_1 , y_{2t} is an m -dimensional vector of NFI time series with $\Delta_*(L) = \text{diag}\{\Delta^{d_2}, \dots, \Delta^{d_m}\}$, $d_1 \geq d_2 \geq \dots \geq d_m$, $d_i \in (\frac{1}{2}, \frac{3}{2})$, $i = 1, 2, \dots, n$, and $\theta = (\theta_1, \dots, \theta_m)'$. A DGP like (13) and (14) appears in various set-ups relevant for economics. For instance, when $m = 1$, a well-known application in finance is to check for efficient markets by regressing spot returns (y_{1t}) on the lag of the forward premium (y_{2t-1}) in order to test whether $\theta = 1$.

However, before discussing in detail the previous DGP, it is convenient to digress briefly on the role of timing when NFI processes are present in the model. In conventional models of cointegration, where x_t is a vector of $I(1)$ processes and u_t is a vector of $I(0)$ disturbances, it is well known that the cointegrating relation is preserved under mixed time translations of the regressors, namely, the cointegrating relationship derived from (13) involving contemporaneous values of all variables, instead of $(y_{1t} - \theta' y_{2t-1})$, remains $I(0)$ since Δy_{2t} is $I(0)$. However, this is no longer the case when x_t is a vector of *NFI* processes. In effect, assuming for simplicity that $d_1 = d_2 = \dots = d_n = d$, the structural form of (13) and (14) becomes

$$y_{1t} = \mu + \theta' y_{2t} + v_{1t}, \quad (15)$$

$$\Delta^d y_{2t} = u_{2t}, \quad \frac{1}{2} < d < \frac{3}{2}, \quad (16)$$

$$v_{1t} = u_{1t} - \theta' \Delta y_{2t}, \quad (17)$$

where Δy_{2t} is $I(d-1)$. Thus, for instance, if $1 < d < \frac{3}{2}$, v_{1t} is a fractionally integrated process of order $d-1$. However, if $\frac{1}{2} < d < 1$, then v_{1t} will be $I(0)$ since u_{1t} is an $I(0)$ process and it will dominate the behaviour of Δy_{2t} which is $I(d-1)$ with $-\frac{1}{2} < d-1 < 0$. Hence, in the NFI set-up, the transient dynamics of the system are affected by the transformation from reduced form to structural form and conversely. Consequently, synchronisation issues may well be important in cointegration with fractionally integrated processes, in contrast to conventional models with $I(1)$ variables. Moreover, v_{1t} carries information about θ , so that synchronisation does also play a role in the efficient estimation of long-run relationships among NFI processes.

Once this issue has been clarified, we proceed next to adapt our framework to SSW's set-up. To do so, we partition u_t conformably with x_t , so that $u_t = (u_{1t}, u_{2t}')'$, and assume that u_t is

generated by

$$u_t = \sum_{j=1}^{p-1} A_j u_{t-j} + \varepsilon_t, \quad (18)$$

where ε_t is an n -dimensional *i.i.d.* sequence of random vectors with $E(\varepsilon_1) = 0$ and covariance matrix $E(\varepsilon_1 \varepsilon_1') = \Sigma > 0$, so that $\varepsilon_t = \Sigma^{1/2} \eta_t$. Assuming that the determinant of the autoregressive polynomial $|I_n - A_1 z - A_2 z^2 - \dots - A_p z^{p-1}|$ has all its roots outside the unit circle, where

$$A_j = \begin{pmatrix} a_{11,j} & 0 \\ a_{21,j} & a_{22,j} \\ (1) & (m) \end{pmatrix} \quad \begin{matrix} (1) \\ (m) \end{matrix},$$

$j = 1, 2, \dots, p-1$, then $u_t \sim AR(p-1)$ process. Therefore, (13) and (14) can be rewritten as

$$\begin{aligned} y_{1t} &= \alpha + (a_{11,1}L + a_{11,2}L^2 + \dots + a_{11,p-1}L^{p-1}) \\ y_{2t} &+ -\theta'(L + a_{11,1}L^2 + \dots + a_{11,p-1}L^p) y_{2t} + \varepsilon_{1t}, \end{aligned} \quad (19)$$

with $\alpha = \mu(1 - a_{11,1} - a_{11,2} - \dots - a_{11,p-1})$ and such that $\varepsilon_{1t} = \omega' \eta_t = \eta_t' \omega$ is the first element of ε_t , where $\omega' = (\omega_1, \omega_2, \dots, \omega_n)$ is the first row of the covariance matrix Σ . Note that the assumption that $a_{12,j} = 0$, for all $j = 1, 2, \dots, p-1$, is needed to ensure that no terms in $\Delta_*(L) y_{2t}$ appear in (19) in order to mimic the linear-in-levels reduced-form ADL model considered by SSW in the context of $I(d)$, $d = 1, 2, \dots$, variables.

For convenience, let us rewrite the reduced form of (19) in an unrestricted way as

$$\begin{aligned} y_{1t} &= \alpha + \sum_{i=1}^p \phi_{11,i} y_{1,t-i} + \sum_{i=1}^p \Phi'_{12,i} y_{2,t-i} + \varepsilon_{1t} \\ &= X_t' \beta + \varepsilon_{1t}, \end{aligned} \quad (20)$$

where $X_t = (1, x'_{t-1}, x'_{t-2}, \dots, x'_{t-p})'$ is the $(np+1)$ vector of regressors and β denotes the corresponding vector of regression coefficients, so that a finite lag length dynamic model among NFI processes is a valid representation. To obtain the asymptotic behaviour of the OLS estimator of β , $\hat{\beta}$, it is convenient to transform the regressors in the way suggested by SSW, i.e. isolating the various stochastic and deterministic components in different blocks. In particular, the regressors can be transformed as $Z_t = DX_t$, where the non-singular square matrix D is chosen in such a way that the g -dimensional ($g = np+1$) vector Z_t has a simple representation in terms of the fundamental stochastic and non-stochastic components. Notice that $X_t' D'(D')^{-1} \beta = Z_t' \gamma$, with $\gamma = (D')^{-1} \beta$, so that the OLS estimators of the original and transformed models are related by the mapping $D' \hat{\gamma} = \hat{\beta}$.

In the fractional case, the regressors Z_t are related to a vector of so-called *canonical regressors*, v_t , of dimension g , by means of the moving-average transformation $Z_t = C(L) v_t$, where

$$v_t = (\eta_t', 1, \xi_t'(d_f), \xi_t'(d_{f-1}), \dots, \xi_t'(d_i), \dots, \xi_t'(d_1), t)', \quad (21)$$

so that v_t has dimension $h = 2 + n(f + 1)$, $f = g - 3$. By letting $C(L)$ be a lower triangular matrix such that $(z_{1t}, z_{2t}, z_{3t}, \dots, z_{g-1,t}, z_{gt})'$ equals

$$\begin{pmatrix} C_{11}(L) & 0 & 0 & \cdots & 0 & 0 \\ 0 & C_{22} & 0 & \cdots & 0 & 0 \\ C_{31}(L) & C_{32} & C_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{g-1,1}(L) & C_{g-1,2} & C_{g-1,3} & \cdots & C_{g-1,h-1} & 0 \\ C_{g1}(L) & C_{g2} & C_{g3} & \cdots & C_{g,h-1} & C_{gh} \end{pmatrix} \begin{pmatrix} \eta_t \\ 1 \\ \xi_t(d_f) \\ \vdots \\ \xi_t(d_1) \\ t \end{pmatrix}, \quad (22)$$

we allow for a model where its elements are individually $I(0)$ or NFI with $d \in (\frac{1}{2}, \frac{3}{2})$, possibly around a linear time trend. In (22) it is assumed that: (i) z_{1t} contains k_1 elements with $\sum_{j=0}^{\infty} C_{11j}C'_{11j}$ being non-singular, (ii) z_{jt} contains k_j elements, $j = 3, \dots, g - 1$, and (iii) C_{jj} has full row rank k_j (possibly equal to zero) for $j = 3, \dots, g - 1$. Finally, let us further assume that $\frac{3}{2} > d_1 \geq d_2 \geq \dots \geq d_i \geq 1$ and $1 > d_{i+1} \geq \dots \geq d_f > \frac{1}{2}$ and define the scaling matrix

$$\Psi_T = \text{diag} \{ T^{1/2}I_{k_1}, T^{1/2}, TI_{k_3}, \dots, TI_{k_{f+2}}, T^{d_i}I_{k_{f+3-i}}, \dots, T^{d_1}I_{k_{g-1}}, T^{3/2} \}.$$

In this fashion, application of Lemma 1 leads to the following result concerning the asymptotic analysis of the OLS estimators $\hat{\gamma}$ and $\hat{\beta}$.

Lemma 2 Under Assumptions B and C with $C_{j1}(L)$, $j = 1, \dots, g$, I -summable, as $T \rightarrow \infty$,

(a)

$$\Psi_T^{-1} \left(\sum_{t=1}^T Z_t Z_t' \right) \Psi_T^{-1} \Rightarrow \Pi, \quad (23)$$

where, partitioning Π conformably with Z_t ,

$$\Pi_{11} = \sum_{j=0}^{\infty} C_{11j}C'_{11j}, \Pi_{22} = C_{22}^2, \Pi_{1j} = \Pi'_{j1} = 0, j = 2, \dots, g,$$

$$\Pi_{jk} = \Pi'_{kj} = \begin{cases} 0, & j, k = 3, \dots, f + 2 - i, \\ C_{jj} \int_0^1 W(d_{f+3-j}, r) W(d_{f+3-k}, r)' dr C'_{kk}, \\ & j, k = f + 3 - i, \dots, g - 1, \end{cases}$$

$$\Pi_{gg} = \frac{1}{3}C_{gh}^2, \Pi_{2g} = \Pi_{g2} = \frac{1}{2}C_{22}C_{gh},$$

$$\Pi_{2j} = \Pi'_{j2} = \begin{cases} 0, & j = 3, \dots, f + 2 - i, \\ C_{22} \int_0^1 W(d_{f+3-j}, r)' dr C'_{jj}, & j = f + 3 - i, \dots, g - 1, \end{cases}$$

$$\Pi_{jg} = \Pi'_{gj} = \begin{cases} 0, & j = 3, \dots, f + 2 - i, \\ C_{gh} C_{jj} \int_0^1 r W(d_{f+3-j}, r) dr, & j = f + 3 - i, \dots, g - 1, \end{cases}$$

$$\Pi_{jk} = \Pi'_{kj} = 0, j = 3, \dots, f + 2 - i, k = f + 3 - i, \dots, g - 1.$$

(b)

$$\Psi_T^{-1} \sum_{t=1}^T Z_t \eta_t' \omega \Rightarrow A, \quad (24)$$

where, partitioning A conformably with Z_t ,

$$\begin{aligned} A_1 &= N\{0, (\omega' \omega) \Pi_{11}\}, \\ A_2 &= C_{22} \int_0^1 dW(r)' \omega, \\ A_j &= \begin{cases} \left(C_{jj} + \sum_{k=1}^{\infty} C_{j1,k} \right) \omega, & j = 3, \dots, (f+2-i), \\ C_{jj} \int_0^1 W(d_{f+3-j}, r) dW(r)' \omega, & j = (f+3-i), \dots, (g-1), \end{cases} \\ A_g &= C_{gh} \int_0^1 r dW(r)' \omega. \end{aligned}$$

(c)

$$\Psi_T (\widehat{\gamma} - \gamma) \Rightarrow \Pi^{-1} A, \quad (25)$$

and

$$\Psi_T (D')^{-1} (\widehat{\beta} - \beta) \Rightarrow \Pi^{-1} A. \quad (26)$$

From Lemma 2, the following remarks apply. *Firstly*, $\widehat{\gamma}$ (and $\widehat{\beta}$) are consistent estimators when there are deterministic time trends and an arbitrary number of NFI processes, such that the individual coefficients converge to their theoretical counterparts at different rates. *Secondly*, when some transformed regressors are dominated by stochastic trends, their joint limiting distribution will be non-normal. However, when there are no Z_t regressors dominated by stochastic trends, $\widehat{\gamma}$ (and $\widehat{\beta}$) has an asymptotically normal joint distribution. *Thirdly*, the block diagonality of Π implies that $T^{1/2}(\widehat{\gamma}_1 - \gamma_1) \Rightarrow N\{0, (\omega' \omega) \Pi_{11}^{-1}\}$. Moreover, Theorem 2.2 in Chan and Wei (1988) applies in our context, implying that A_1 is independent of A_j for $j > 1$, so that $T^{1/2}(\widehat{\gamma}_1 - \gamma_1)$ is asymptotically independent of the remaining estimated coefficients.

Thus, as in the case of an integer d , the results above provide a very useful sufficient condition for estimating coefficients with asymptotically normal limiting distributions. Like in the original SSW set-up, all that is needed is that a block of coefficients can be written as coefficients on zero mean $I(0)$ regressors in a model that includes a constant term. As in SSW (pp. 124–127), it is straightforward to obtain their same general result stating that restrictions involving subsets of coefficients that can be written as coefficients on zero-mean $I(0)$ regressors in regressions that include constant terms, can be tested using standard asymptotic distribution theory. Otherwise, in general, the test statistics will have non-standard limiting distributions.

4. MONTE CARLO EVIDENCE

This section reports the results from a small Monte Carlo study carried out to examine how the asymptotic theory derived above performs in finite samples. The following two DGPs have been considered in order to illustrate the timing-synchronisation issue discussed above.

DGP 1: We have generated two NFI processes (y_{1t}, y_{2t}) according to definition (1) with the same memory parameter, d , such that

$$y_{1t} = y_{2t} + v_{1t}, \quad (27)$$

$$\Delta^d y_{2t} = v_{2t} \quad (28)$$

$$\Delta^b v_{1t} = \varepsilon_{1t}, v_{2t} = \varepsilon_{2t}, (\varepsilon_{1t}, \varepsilon_{2t})' \sim N(0, I_2), \quad (29)$$

where b and d ($b < d$) are assumed to be known and take the values $b = \{0, 0.1, 0.4\}$ and $d = \{0.8, 1, 1.4\}$. Equations (27)–(29) taken together imply that $y_{1t} = y_{2,t-1} + e_t$ where $e_t = v_{1t} + \Delta^{1-d} v_{2t}$. Thus, e_t is the sum of an $I(b)$ and an $I(d-1)$ process. Hence, whenever $d-1 < b$, e_t will be $I(b)$, whereas it will be $I(d-1)$ when $b < d-1$. Hereby, e_t is $I(0)$ for the following parameter configurations: $b = 0$ and $d = 0.8, 1$. Similarly, it will be $I(0.1)$ for $b = 0.1$ and $d = 1$. Finally, it will be $I(0.4)$ for: $b = 0, 0.1$ and $d = 1.4$, and $b = 0.4$ and any of the four values of d . Accordingly, DGP 1 allows us to generate cases where the variables are cointegrated $CI(d, d)$ and others where they are $CI(d, c)$ with $d-c = b$ for $b > 0$ and $d-1 < b$, or $d-c = d-1$ for $d-1 > b$. Note that, although the theoretical results derived in the previous section are restricted to the $CI(d, d)$ case, we have extended the simulations to include $CI(d, c)$ processes in order to check how robust are the results to small deviations from the ideal case where $d = c$ or $b = 0$.

DGP 2: In this case, y_{1t} and y_{2t} are generated as $I(d)$ processes such that

$$y_{1t} = y_{2,t-1} + v_{1t}, \quad (30)$$

$$\Delta^d y_{2t} = v_{2t}, \quad (31)$$

where v_{1t} and v_{2t} have the same properties as in DGP1 and the rest of assumptions hold as well. Thus, in this case, $y_{1t} = y_{2,t-1} + e_t$ where now $e_t (= v_{1t} = \Delta^{-b} \varepsilon_{1t} 1_{\{t > 0\}})$ is always $I(b)$, for any value of d .

For both DGPs, the sample size considered in the simulations is $T = 100$, and the number of replications is $N = 2000$. In line with equation (19), the underlying regression model considered here to carry out inference under each of the DGPs is

$$y_{1t} - y_{2,t-1} = \mu + \beta y_{1,t-1} + \theta y_{2,t-1} + e_t, \quad (32)$$

where $\theta = (\pi_1 - 1)$. Then, the following set of hypotheses is considered

CASE 1:

$$H_0 : \theta = 0, H_0' : \beta = 0, \quad (33)$$

CASE 2:

$$H_0'' : \beta = 0 \text{ and } \theta = 0. \quad (34)$$

The interpretation of each test is as follows. In CASE 1, we tests for parameter restrictions giving rise to a relationship of the form $y_{1t} = y_{2,t-1} + e_t$, based on t -ratios of individual coefficients (β and θ) in a regression with $CI(d, d)$ or $CI(d, c)$ variables. Notice that, in accord with DGP 1, $\beta = 0$ and $\theta = 0$, and that the equilibrium relationship ($y_{1t} - y_{2t}$) can be $I(0)$, $I(0.1)$ or $I(0.4)$ depending on the parameter configurations described above. By contrast, under DGP 2, $(y_{1t} - y_{2t})$ is $I(0)$ when $b = 0$, and $I(b)$ for $b > 0$, independently of the value of d . Since the regression model (32) can be reparameterized so that θ and β become the coefficients on $(y_{1,t-1} - y_{2,t-1})$, that is

$$y_{1t} - y_{2,t-1} = \mu + (\beta + \theta)y_{1,t-1} - \theta(y_{1,t-1} - y_{2,t-1}) + e_t,$$

or

$$y_{1t} - y_{2,t-1} = \mu + \beta(y_{1,t-1} - y_{2,t-1}) + (\beta + \theta)y_{2,t-1} + e_t,$$

then the corresponding t -ratios of $H_0 : \theta = 0$ and $H'_0 : \beta = 0$ will have a standard limiting distribution when $(y_{1t} - y_{2t})$ is $I(0)$ and a non-standard distribution when $(y_{1t} - y_{2t})$ is $I(0.1)$ or $I(0.4)$.

The test considered in CASE 2 turns out to be a joint F -test for a long-run relationship between y_1 and y_2 with cointegrating slope ($= \frac{1+\theta}{1-\beta} = \frac{\pi_1}{1-\beta}$) equal to unity. Following a similar argument as before, it is clear that while β and θ can be individually rewritten as coefficients on zero-mean $I(0)$ variables when $b = 0$ and $d - 1 < b$ for DGP 1, and $b = 0$ for DGP 2, the test of $H''_0 : \theta = 0$ and $\beta = 0$ will always have a non-standard distribution since $(\beta + \theta)$ is associated with $y_{1,t-1}$ or $y_{2,t-1}$, which under both DGPs are $I(d)$, $\frac{1}{2} < d < \frac{3}{2}$.

Table 1 reports both the percentage of rejections (for 5% nominal size) and the empirical quantiles of the finite sample distributions of the t -test of $H_0 : \theta = 0$ in CASE 1. The results for $H'_0 : \beta = 0$ are omitted since the conclusions drawn for this case are the same. As can be observed, the results obtained by SSW (for $d = 1$) extend to the more general case where $d > 1/2$ and $b = 0$ for DGP 2. However, they only hold for DGP 1 when e_t is $I(0)$. Notwithstanding, the results when e_t is $I(0.1)$ are not too different (given the standard error of the simulation ($= \sqrt{(0.05 * 0.95)/2000} \approx 0.005$). By contrast, when e_t is $I(0.4)$, the size distortions are very large. Table 2, in turn, presents the percentage of rejections (for 5% nominal size) in an F -test for $H''_0 : \beta = \theta = 0$ in CASE 2. In all cases, the empirical size is significantly larger than the nominal size. Thus, overall, the experimental evidence presented above is supportive of our theoretical results.

5. AN EMPIRICAL ILLUSTRATION

In this section, we apply some of the previous results to one of the best known empirical applications of cointegration, namely the relationship between (logged) consumption (c) and (logged) disposable income (i^d), which has been the subject of many studies over the last two decades; cf Muellbauer and Lattimore (1995) for an excellent review of the relevant literature. In particular we focus on the econometric implications of the *permanent income hypothesis* (PIH) of consumption, as derived by Stock and West (1988). The PIH implies that c_t is an $I(1)$ variable, more precisely a random walk whose error term is the innovation of permanent income. Further, if i_t^d is $I(1)$ as well, the saving rate $s_t \approx i_t^d - c_t$ will be $I(0)$, namely i_t^d and c_t will be cointegrated $CI(1, 1)$. Accordingly, one of the standard tests for PIH in the literature relies upon testing for the absence of Granger causality of i_t^d on c_t when lagged consumption is allowed for in the

Table 1. Tests on individual restrictions.

| $d \backslash$ quantiles | 0.025 | 0.05 | 0.50 | 0.90 | 0.95 | 0.975 | Size(5%) |
|-----------------------------------|--------|--------|--------|-------|-------|-------|----------|
| <i>(a) b = 0, DGP 1, CASE 1</i> | | | | | | | |
| 0.8 | -2.094 | -1.696 | -0.142 | 1.193 | 1.586 | 1.914 | 5.6 |
| 1.0 | -2.059 | -1.683 | -0.102 | 1.258 | 1.661 | 1.932 | 5.3 |
| 1.4 | -2.654 | -2.432 | -0.761 | 0.632 | 1.025 | 1.203 | 22.8 |
| <i>(b) b = 0.1, DGP 1, CASE 1</i> | | | | | | | |
| 0.8 | -2.232 | -1.882 | -0.453 | 1.034 | 1.366 | 1.685 | 9.2 |
| 1.0 | -2.183 | -1.762 | -0.212 | 1.114 | 1.496 | 1.813 | 6.9 |
| 1.4 | -2.721 | -2.462 | -0.820 | 0.617 | 1.019 | 1.186 | 23.2 |
| <i>(c) b = 0.4, DGP 1, CASE 1</i> | | | | | | | |
| 0.8 | -2.812 | -2.534 | -0.853 | 0.649 | 1.123 | 1.227 | 22.0 |
| 1.0 | -2.798 | -2.502 | -0.796 | 0.627 | 1.152 | 1.342 | 18.0 |
| 1.4 | -2.808 | -2.593 | -0.813 | 0.597 | 1.125 | 1.263 | 20.6 |
| <i>(d) b = 0, DGP 2, CASE 1</i> | | | | | | | |
| 0.8 | -2.008 | -1.677 | -0.082 | 1.262 | 1.591 | 1.930 | 5.4 |
| 1.0 | -1.987 | -1.665 | -0.071 | 1.269 | 1.603 | 1.938 | 5.2 |
| 1.4 | -2.045 | -1.703 | -0.112 | 1.245 | 1.578 | 1.904 | 5.7 |
| <i>(a) b = 0.1, DGP 2, CASE 1</i> | | | | | | | |
| 0.8 | -2.202 | -1.765 | -0.154 | 1.203 | 1.558 | 1.873 | 6.7 |
| 1.0 | -2.121 | -1.723 | -0.164 | 1.197 | 1.576 | 1.889 | 5.9 |
| 1.4 | -2.132 | -1.709 | -0.172 | 1.208 | 1.558 | 1.823 | 6.9 |
| <i>(a) b = 0.4, DGP 2, CASE 1</i> | | | | | | | |
| 0.8 | -2.834 | -2.632 | -0.796 | 0.632 | 1.232 | 1.286 | 19.0 |
| 1.0 | -2.693 | -2.567 | -0.758 | 0.644 | 1.223 | 1.314 | 19.3 |
| 1.4 | -2.815 | -2.587 | -0.823 | 0.612 | 1.143 | 1.264 | 20.7 |

Note: t -test on $H_0 : \theta = 0$; $T = 100$.

regression. Using SSW's guidelines, t -ratios on the individual coefficients or F -statistics on the joint significance of the π_i 's in the ADL model

$$c_t = \mu + \beta c_{t-1} + \pi_1 i_{t-1}^d + \dots + \pi_m i_{t-m}^d + e_t, \quad (35)$$

will follow standard asymptotic distributions since (35) can always be reparameterised in such a way that $\pi_i (i = 1, \dots, l)$ becomes the coefficient of $(c_{t-1} - i_{t-1}^d)$, which is $I(0)$ since it is the sum of $(c_{t-1} - i_{t-1}^d)$ and $(i_{t-1}^d - i_{t-1}^d)$ which are $I(0)$ under the previous assumptions.

Indeed, using the unrestricted ADL model considered in (20) with $m = 1$

$$y_{1t} = \alpha + \sum_{i=1}^p \phi_{11,i} y_{1,t-i} + \sum_{i=1}^p \phi_{12,i} y_{2,t-i} + \varepsilon_{1t}, \quad (36)$$

the restriction that y_{2t} does not Granger-cause y_{1t} corresponds to the null hypothesis

$$H_0 : \phi_{12,1} = \phi_{12,2} = \dots = \phi_{12,p} = 0. \quad (37)$$

Table 2. Tests on joint restrictions.

| $d \backslash$ quantiles | Empirical size |
|-----------------------------------|----------------|
| <i>(a) b = 0, DGP 1, CASE 2</i> | |
| 0.8 | 14.8 |
| 1.0 | 6.6 |
| 1.4 | 18.9 |
| <i>(b) b = 0.1, DGP 1, CASE 2</i> | |
| 0.8 | 8.8 |
| 1.0 | 7.4 |
| 1.4 | 18.7 |
| <i>(c) b = 0.4, DGP 1, CASE 2</i> | |
| 0.8 | 23.2 |
| 1.0 | 18.6 |
| 1.4 | 22.3 |
| <i>(d) b = 0, DGP 2, CASE 2</i> | |
| 0.8 | 8.5 |
| 1.1 | 7.7 |
| 1.4 | 17.3 |
| <i>(e) b = 0.1, DGP 2, CASE 2</i> | |
| 0.8 | 8.3 |
| 1.0 | 9.5 |
| 1.4 | 13.5 |
| <i>(f) b = 0.4, DGP 2, CASE 2</i> | |
| 0.8 | 16.8 |
| 1.0 | 15.3 |
| 1.4 | 23.2 |

Note: F-test on $H_0' : \beta = 0 \text{ \& } \theta = 0; T = 100$. Nominal size 5%.

In general, when y_{1t} and y_{2t} are integrated, SSW and Toda and Phillips (1993) had proved that the distribution of the test statistic depends on the location of unit roots in the system. If y_{1t} is $I(1)$ and y_{2t} is $I(0)$, or if both y_{1t} and y_{2t} are cointegrated $I(1)$ processes, then the test statistic has a limiting chi-square distribution. Otherwise, i.e. if y_{1t} and y_{2t} are not cointegrated $I(1)$ processes, then the Granger-causality test statistic will not be in general asymptotically chi-square.

From the results in the previous sections we have shown that the same properties apply when y_{1t} and y_{2t} are NFI processes. Thus, the Granger-causality test statistic will have a limiting chi-square distribution whenever there is an $I(0)$ linear combination of the variables, say $v_t = y_{2t} - \lambda y_{1t}$, so that (36) can be rewritten as

$$y_{1t} = \tilde{\alpha} + \sum_{i=1}^p \tilde{\phi}_{11,i} y_{1,t-i} + \sum_{i=1}^p \phi_{12,i} (v_{t-i} - \mu_c) + \varepsilon_{1t}, \quad (38)$$

where μ_c is the mean of v_t , $\tilde{\alpha} = \alpha + \sum_{i=1}^p \phi_{12,i} \mu_c$ and $\tilde{\phi}_{11,i} = \phi_{11,i} + \phi_{12,i} \lambda$, $i = 1, \dots, p$. Now, since the Granger-causality restriction in the transformed regression corresponds to the restriction

that the terms $v_{t-i} - \mu_c$ do not enter the regression, being these terms zero-mean $I(0)$ regressors in a regression that includes a constant term, then the resulting test statistics will have a limiting χ_p^2 distribution. On the contrary, when either the NFI processes y_{1t} and y_{2t} are not cointegrated or v_t is not an $I(0)$ process, in general, the resulting test statistics will not have a limiting chi-square distribution and the appropriate asymptotic distributions need to be tabulated. In this vein, Hidalgo (2000) has recently introduced a non-parametric Granger-causality test for covariance stationary linear processes under, possibly, the presence of long-range dependence. The test is consistent and has power against a sequence of local alternatives converging to the null at the parametric rate $T^{-1/2}$. This test has asymptotic relative efficiency greater than zero when compared to other parametric tests based on a correct specification of the model.

In the sequel, we will test whether the implications of the PIH are verified in the more general fractional set-up where c_t and i_t^d are considered to be $I(d)$ variates and s_t is $I(\delta)$, with $\frac{1}{2} < d < \frac{3}{2}$ and $0 \leq \delta \leq d$. We use real (1992 dollars) quarterly (seasonally adjusted) data of c_t (non-durable goods plus services) and i_t^d for the *U.S.* economy, from 1947: 3 to 1998: 4. Data come from the Bureau of Economic Analysis and is displayed in Figure 1. We also considered the sample period 1953: 1–1986: 4 as in Campbell and Mankiw (1989), obtaining similar results which are not reported for the sake of brevity.

In Tables 3–5 we present the *ACF* and *PACF* of the levels and first differences of the series. As regards c_t and i_t^d , we observe in Tables 3 and 4 that the autocorrelations decline very slowly, whereas the partial autocorrelations decline to zero after the first one. On the other hand, the autocorrelations of the *change* in both series are small, specially in i_t^d . With respect to s_t , Table 5 shows that the autocorrelations decay at a faster rate, but not so fast as would be expected for a weakly stationary series. In fact, the first autocorrelation of Δs_t is negative, albeit small in absolute value.

Table 6 offers the results of applying the well-known ADF test statistic to c_t and i_t^d . As can be observed, we cannot reject the null hypothesis of unit root at any conventional significance level, whereas it is rejected for s_t . Thus, on the basis of the evidence reported in Tables 3–6, one is tempted to conclude that c_t and i_t^d are $CI(1, 1)$. Consequently, the Wald statistic for zero coefficients on the lags of i_t^d in model (35) will have a limiting chi-square distribution.

An OLS regression of c_t on a constant, c_{t-1} , and four lags of i_t^d yields

$$\begin{aligned}
 c_t = & \underset{(0.006)}{0.0167} + \underset{(0.0237)}{1.0198}c_{t-1} + \underset{(0.0364)}{0.1066}i_{t-1}^d - \underset{(0.0476)}{0.0558}i_{t-2}^d \\
 & - \underset{(0.0464)}{0.0809}i_{t-3}^d + \underset{(0.033)}{0.0096}i_{t-4}^d + res
 \end{aligned} \tag{39}$$

where the Schwartz information criterion has been used to select the lag length (OLS standard errors in parenthesis). In this model, the F -statistic for zero coefficients on the disposable income variables is 4.28. As the (approximate) corresponding critical value is about 2.40, we reject the joint null hypothesis $H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0$ (p -value equal to 0.002), providing evidence against the PIH.

These findings, however, have been obtained under the assumption that the relevant series are $CI(1, 1)$. Let us now consider the more general fractional set-up envisaged in this paper. First, we have estimated the memory parameters of c_t , i_t^d and s_t by means of the exact maximum likelihood (ML) procedure using the package *ARFIMA 1.0* in *Ox* (see Doornik and Ooms, 1999). As the method is only well defined for stationary processes, we have fitted four *ARFIMA*(p, δ, q) models

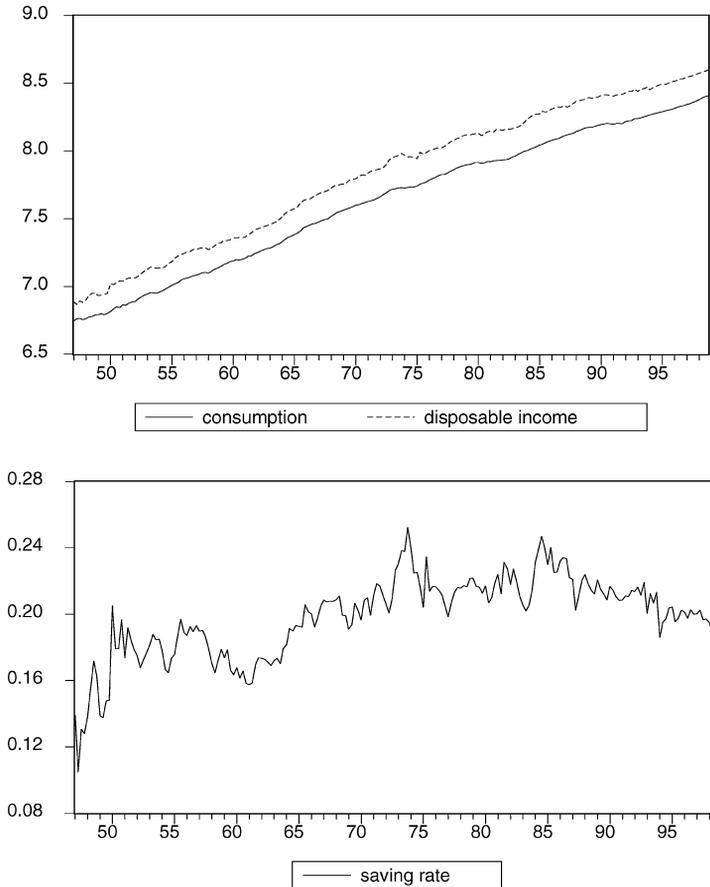


Figure 1. Consumption, disposable income and saving-rate series. Sample period:1947:1–1998:4.

($p, q = 0, 1$) to Δc_t , Δi_t^d and Δs_t , and hence $\hat{d} = \hat{\delta} + 1$. For the three series, the Akaike and Schwartz information criteria selected the $ARFIMA(0, \delta, 0)$ specification as the appropriate one.

The estimates of δ for Δc_t and Δi_t^d are $\hat{\delta}_{\Delta c_t} = 0.17(0.053)$ and $\hat{\delta}_{\Delta i_t^d} = 0.08(0.054)$, respectively, with standard errors given in parentheses. Therefore, whereas we reject at the 5% significance level that c_t is $I(1)$, we cannot reject that i_t^d is $I(1)$. These results are broadly in agreement with the previous ones (see Tables 3 and 4), namely, that i_t^d seems to follow a random walk process whereas c_t is slightly more non-stationary. As $\delta_{\Delta c_t} = \delta_{\Delta i_t^d}$ is necessary for cointegration in our fractional set-up, we also test for $H_0 : \delta_{\Delta c_t} - \delta_{\Delta i_t^d} = 0$ using the Wald test proposed by Marinucci and Robinson (2001, p. 234) with estimates of d based this time on the log-periodogram regression as in Geweke and Porter-Hudak (1983), which yielded very similar estimates with a bandwidth parameter of 25. The test is asymptotically distributed as $\chi^2(1)$, and

Table 3. Autocorrelations of consumption.

| | Levels | First differences |
|--------|--------|-------------------|
| ACF | | |
| First | 0.986 | 0.218 |
| Second | 0.973 | 0.147 |
| Third | 0.959 | 0.169 |
| Fourth | 0.945 | 0.026 |
| Fifth | 0.931 | -0.027 |
| PACF | | |
| First | 0.986 | 0.218 |
| Second | -0.004 | 0.104 |
| Third | -0.010 | 0.125 |
| Fourth | -0.015 | -0.047 |
| Fifth | -0.004 | -0.060 |

Note: ACF = Autocorrelation Function; PACF = Partial Autocorrelation Function.

Table 4. Autocorrelations of disposable income.

| | Levels | First differences |
|--------|--------|-------------------|
| ACF | | |
| First | 0.986 | -0.005 |
| Second | 0.972 | 0.065 |
| Third | 0.958 | 0.034 |
| Fourth | 0.944 | -0.137 |
| Fifth | 0.930 | -0.106 |
| PACF | | |
| First | 0.986 | -0.005 |
| Second | -0.023 | 0.065 |
| Third | 0.004 | 0.034 |
| Fourth | -0.021 | -0.142 |
| Fifth | 0.003 | -0.115 |

Note: See Note to Table 3.

yields a value of 1.83 below 3.84, i.e. the 5% critical value. Thus the null hypothesis of equal δ 's is not rejected.

On the other hand, the estimated $\delta_{\Delta s_t}$ for Δs_t is -0.29 (0.049), i.e. the memory parameter of s_t was found to be 0.71. To provide additional evidence on the null hypothesis that the memory parameter of the saving rate series is $d = 1$ versus the alternative of $d < 1$, we make use of the so-called Augmented Fractional Dickey-Fuller (AFDF) test proposed by Dolado *et al.* (2002), which relies upon the statistical significance of the coefficient ϕ in the following auxiliary regression

$$\Delta s_t = \mu + \phi \Delta^{d_s} s_{t-1} + \sum_{i=1}^k \xi_i \Delta s_{t-i} + e_{kt}, \quad (40)$$

Table 5. Autocorrelations of saving rate.

| | Levels | First differences |
|--------|--------|-------------------|
| ACF | | |
| First | 0.906 | -0.244 |
| Second | 0.834 | 0.020 |
| Third | 0.772 | 0.034 |
| Fourth | 0.708 | -0.137 |
| Fifth | 0.686 | -0.106 |
| PACF | | |
| First | 0.906 | -0.244 |
| Second | 0.072 | -0.042 |
| Third | 0.004 | -0.001 |
| Fourth | -0.241 | -0.255 |
| Fifth | -0.015 | -0.156 |

Note: See Note to Table 3.

Table 6. ADF test statistic.

| | |
|-------------------|---------|
| Consumption | -2.3040 |
| Disposable income | -2.008 |
| Saving rate | -3.829* |

Note: Lag-length selection using SIC. The asterisk (*) denotes significance at the 0.01 level.

where $k = o(T^{1/3})$ and d_a stands for the proposed value of d under the alternative hypothesis. Dolado *et al.* (2002) prove that, provided d_a is estimated using a \sqrt{T} -consistent estimation procedure, the ADF test based on the t -ratio on ϕ asymptotically follows a normal distribution. In our case, $t_\phi = -4.5$ so that we reject the null hypothesis that the saving rate follows an $I(1)$ process at any conventional significance level. Moreover, we also reject the $I(0)$ null hypothesis at the 1% significance level on the grounds of the corresponding ML t -ratio.

Further evidence on the fractional nature of s_t is heuristically provided in Figure 2, the interpretation of which is as follows. With $I(0)$ series, the log-periodogram should be scattered near the origin around a constant. If $d > 0$ the points in the figure should be scattered near the origin around a negative slope. By contrary, if we overdifferentiate so that $d < 0$, such a slope should be positive near the origin. As the slope of the *levels* is negative and the slope of the *changes* seems positive, Figure 2 reinforces our belief that s_t follows a (mean-reverting) fractional process along the sample period.

Therefore, according with the results obtained in Section 4, for the specific sample period considered above, the F -statistic for zero coefficients on the disposable income variables in model (35) will *not* have a standard limiting distribution. Hence, the p -value reported just below (39) is not correct. One possible route to compute the correct critical values of the F -statistic would be to use bootstrap methods as in Davidson (2002), yet this computation goes beyond the scope of this paper.

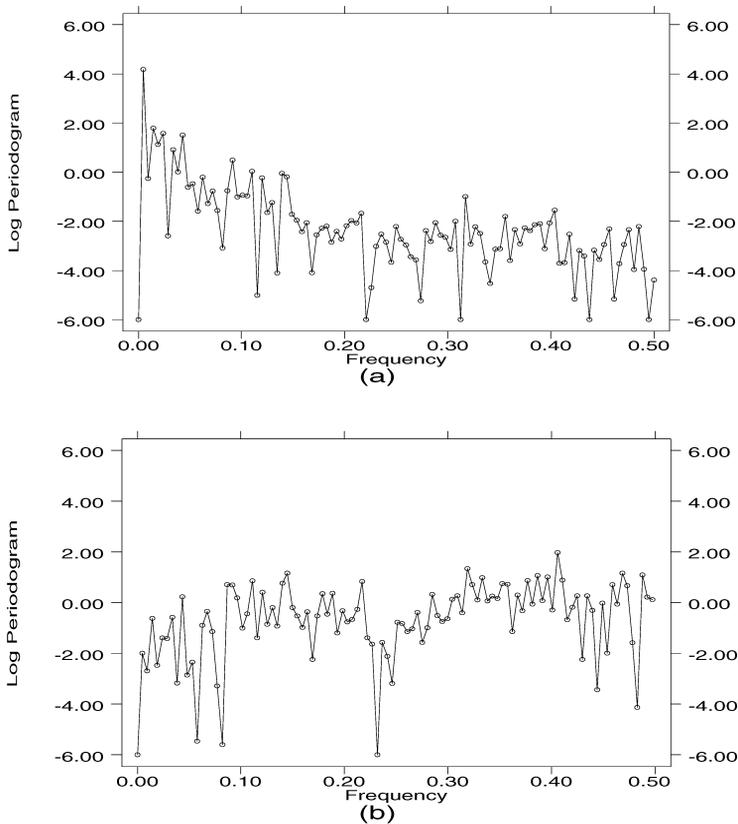


Figure 2. Log periodogram of the saving-rate series. Part (a) shows the log periodogram of the levels of the series. Part (b) shows the log periodogram of the first difference of the series. In both cases the periodogram has been evaluated at natural frequencies.

6. CONCLUSIONS

In this paper we extend the SSW analysis on estimation and hypothesis testing in VAR models with integrated processes and deterministic components to the more general fractional framework. In particular, we have considered partial VARs where the conditioning variables are NFI processes since this is the only finite-lag VAR model compatible with such processes. We show that SSW's conclusions remain valid in this more general case. This means that whenever a block of coefficients can be written as coefficients on zero-mean $I(0)$ regressors in a model that includes a constant term, they will have a joint asymptotic normal distribution, so that the corresponding restrictions can be tested using standard asymptotic chi-square distribution theory. Otherwise, in general, the associated statistics will have non-standard limiting distributions. As in the integer

case, $d = 1, 2, \dots$, notice that the statistical procedures analysed here require at least partial knowledge of which variables are cointegrated and of the memory parameters of the individual series.

In this respect, two restrictive assumptions made in this paper merit further research. On the one hand, we have taken for granted that the memory parameters of the relevant NFI processes were known. However, in practice one may need to estimate the memory parameters of the processes at the same time that one tests for the null hypothesis. Fortunately, there currently exist a few rigorous statistical procedures that can be used to extend our results to the less restrictive case where d is taken to be unknown (cf. Marinucci and Robinson (2001) and Robinson and Hualde (2003)).

On the other hand, it has been assumed that stationary variables in the partial VAR model are all $I(0)$. This could be again a clear limitation of our analysis and would be desirable to extend our analysis to cover cases with stationary fractionally integrated error terms. Addressing such an issue amounts to deriving weak convergence results for stochastic integrals having multivariate fractional integrands and integrators. In this respect, recent results by Davidson (2003), considering fractional integrands and integrators, with a *Type-I* definition of fBM could shed light for future research on the corresponding convergence results with the *Type-II* definition used in this paper.

ACKNOWLEDGEMENTS

We are grateful to R. J. Smith and two anonymous referees for very useful comments that led to a considerable improvement of the paper. We also thank C. Alonso-Borrego, J. Davidson, J. Hidalgo, M. Jensen, D. Marinucci, P. M. Robinson, C. Velasco and participants at ESEM01, Laussane, for helpful comments and suggestions. Finally, we are especially grateful to Laura Mayoral for her help in designing the Monte Carlo simulations. This research has been partially financed with DGICYT grants PB98-0026 and SEC2001-0890. The second author thanks Instituto Flores de Lemus for financial support.

REFERENCES

- Akonom, J. and C. Gouieroux (1987). A functional central limit theorem for fractional processes. Technical Report no 8801, *CEPREMAP*, Paris.
- Baillie, R. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.
- Campbell, J. Y. and N. G. Mankiw (1989). Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual* 4, 185–216.
- Chan, N. H. and N. Terrin (1995). Inference for unstable long-memory processes with applications to fractional unit root autoregressions. *Annals of Statistics* 23, 1662–83.
- Chan, N. H. and C. Z. Wei (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367–401.
- Csörgö, S. and J. Mielniczuk (1995). Distant long-range dependent sums and regression estimation. *Stochastic Processes and Their Applications* 59, 143–55.
- Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics* 110, 187–212.

- Davidson, J. (2003). Convergence to stochastic integrals with fractionally integrated integrator processes: Theory and applications to cointegrating regression. Mimeo, Cardiff University.
- Davidson, J. and R. M. De Jong (2000). The functional central limit theorem and weak convergence to stochastic integrals II. *Econometric Theory* 16, 643–66.
- Davydov, Y. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications* 15, 487–98.
- de Jong, R. M. and J. Davidson (2000). The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory* 16, 621–42.
- Dolado, J. J., Gonzalo, J. and L. Mayoral (2002). A fractional Dickey-Fuller test for unit roots. *Econometrica* 70, 1963–2006.
- Doornik, J. A. and M. Ooms (1999). A package for estimating, forecasting and simulating ARFIMA models: Arfima package 1.0 for Ox. Preprint, Erasmus University.
- Fox, R. and M. S. Taquq (1987). Central limit theorem for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields* 74, 213–40.
- Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–38.
- Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications*. New York: Academic Press.
- Hansen, B. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- Heyde, C. C. and Y. Yang (1997). On defining long-range dependence. *Journal of Applied Probability* 34, 939–44.
- Hidalgo, J. (2000). Nonparametric test for causality with long-range dependence. *Econometrica* 68, 1465–90.
- Levy, P. (1953). Random functions: General theory with special reference in laplacian random functions. *University of California Publication Statistics* 1, 331–90.
- Liu, M. (1998). Asymptotics of nonstationary fractional integrated series. *Econometric Theory* 14, 641–62.
- Mandelbrot, B. B. and J. W. Van Ness (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review* 10, 422–37.
- Marinucci, D. (2000). Spectral regression for cointegrated time series with long-memory innovations. *Journal of Time Series Analysis* 21, 685–705.
- Marinucci, D. and P. M. Robinson (1999). Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference* 80, 111–22.
- Marinucci, D. and P. M. Robinson (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103–120.
- Marinucci, D. and P. M. Robinson (2001). Semiparametric fractional cointegration analysis. *Journal of Econometrics* 105, 225–47.
- Muellbauer, J. and R. Lattimore (1995). The consumption function: A Theoretical and Empirical Overview. In Pesaran, M. H. and M. Wickens (Eds.), *Handbook of Applied Econometrics*, Chapter 5, Oxford: Blackwell.
- Phillips, P. C. B. (1988). Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–33.
- Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
- Robinson, P. M. and D. Marinucci (2001). Narrow-band analysis of nonstationary processes. *Annals of Statistics* 29, 947–86.
- Robinson, P. M. and J. Hualde (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* 71, 1727–76.
- Sims, C. A., J. H., Stock and M. W. Watson (1990). Inference in linear time series with some unit roots. *Econometrica* 58, 113–44.

- Stock, J. H. and M. W. Watson (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
- Stock, J. H. and K. D. West (1988). Integrated regressors and tests of the permanent income hypothesis. *Journal of Monetary Economics* 21, 85–95.
- Strasser, H. (1986). Martingale difference arrays and stochastic integrals. *Probability Theory and Related Fields* 72, 83–98.
- Taqqu, M. (1975). Weak convergence to fractional brownian motion and to the rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 40, 203–38.
- Toda, H. Y. and P. C. B. Phillips (1993). Vector autoregressions and causality. *Econometrica* 62, 1367–94.

APPENDIX

Proof of Lemma 1. Define $\gamma_t = \sum_{k=0}^{\infty} (E_t u_{t+k} - E_{t-1} u_{t+k}) = \Upsilon(1) \eta_t$ and $\kappa_t = \sum_{k=1}^{\infty} E_t u_{t+k} = \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} \Upsilon_i \eta_{t-j} = \sum_{j=0}^{\infty} \Upsilon_j^* \eta_{t-j} = \Upsilon^*(L) \eta_t$, $\Upsilon_j^* = \sum_{i=j+1}^{\infty} \Upsilon_i$, so that $u_t = \gamma_t + \kappa_{t-1} - \kappa_t$, where $E_t u_{t+k} = E(u_{t+k} | F_t)$, κ_t is strictly stationary and square integrable (cf Phillips and Solo 1992) and F_t is the smallest sigma-field containing the past history of the innovation sequence. This decomposition allows us to write

$$\sum_{t=1}^T \xi_{t-1}(d) (\Upsilon(L) \eta_t)' = \sum_{t=1}^T \xi_{t-1}(d) \eta_t' \Upsilon(1)' - \sum_{t=1}^T \xi_{t-1}(d) (\Upsilon^*(L) \Delta \eta_t)'. \quad (\text{A.1})$$

Let $\wp_T(r) = T^{-1/2} \wp_{|T|T}$, $\wp_t = \sum_{i=1}^t \eta_i$. Under the assumptions of the lemma, $(T^{1/2-d} \xi_t(d), \wp_T(r)) \Rightarrow (W(d, r), W(r))$, and since $\{\eta_t, F_t\}$ is a square integrable martingale difference sequence, it follows from Hansen (1992, Theorem 2.1) that

$$T^{-d} \sum_{t=1}^T \xi_{t-1}(d) \eta_t' \Upsilon(1)' \Rightarrow \int_0^1 W(d, r) dW(r)' \Upsilon(1)'. \quad (\text{A.2})$$

Hence, to prove expressions (9) and (10) it only remains to show the convergence of $\Lambda_T = T^{-d} \sum_{t=1}^T \xi_{t-1}(d) (\Upsilon^*(L) \Delta \eta_t)' = T^{-d} \sum_{t=1}^T \xi_{t-1}(d-1) \kappa_{t-1}' - T^{-d} \xi_{t-1}(d) \kappa_t' = \Lambda_T^1 - \Lambda_T^2$, say. Since the limit process is non-random, it is sufficient to consider the case where Λ_T is scalar.

As regards the term Λ_T^1 , notice that for all $d > \frac{1}{2}$,

$$\sup_{t \leq T} T^{-d} |\xi_{t-1}(d) \kappa_t| \leq \sup_{t \leq T} |T^{1/2-d} \xi_{t-1}(d)| T^{-1/2} \sup_{t \leq T} |\kappa_t| \xrightarrow{p} 0 \quad (\text{A.3})$$

because $\sup_{t \leq T} |\xi_t(d)| = O_p(T^{d-1/2})$ and $T^{-1/2} \sup_{t \leq T} |\kappa_t| \xrightarrow{p} 0$. See Hall and Heyde (1980, pp. 142–43).

Next, with regard to the term Λ_T^2 , we obtain the following results for different values of d in decreasing order.

Firstly, for all $d > \frac{3}{2}$, we have that $E|T^{-d} \sum_{t=1}^T \xi_{t-1}(d-1) \kappa_{t-1}|$ is upper bounded by

$$\begin{aligned} & T^{-3/2} \sum_{t=1}^T E(|T^{3/2-d} \xi_{t-1}(d-1)| |\Upsilon^*(L) \eta_{t-1}|) \\ & \leq T^{-3/2} \sum_{t=1}^T \left(E(T^{3/2-d} \xi_{t-1}(d-1))^2 \right)^{1/2} \left(E(\Upsilon^*(L) \eta_{t-1})^2 \right)^{1/2} \\ & \leq T^{-1} \sum_{t=1}^T \left(E(T^{3/2-d} \xi_{t-1}(d-1))^2 \right)^{1/2} T^{-1/2} \left(E(\eta_1^2) \sum_{j=0}^{\infty} |\Upsilon_j^*| \right) \rightarrow 0, \end{aligned}$$

using Markov's inequality, where the last line follows from

$$T^{-1} \sum_{t=1}^T \left(E \left(T^{3/2-d} \xi_{t-1} (d-1) \right)^2 \right)^{1/2} \Rightarrow \int_0^1 E \left(W(d-1, r)^2 \right)^{1/2} dr < \infty$$

and from $\Upsilon^*(L)$ being absolutely summable.

Secondly, when $d = \frac{3}{2}$, $\xi_t(d-1)$ is an NFI process of order $\frac{1}{2}$ for which the invariance principle (7) does not apply. Nonetheless, in this case Liu (1998, Theorem 2.2) proved that, under the assumptions of the lemma, $(\log^{-1/2} T) \xi_{\lfloor Tr \rfloor}(d-1) \Rightarrow KrW(1)$ where K denotes a positive constant independent of r . Thus, using again Markov's inequality, it turns out that when $d = \frac{3}{2}$, $E|T^{-3/2} \sum_{t=1}^T \xi_{t-1}(\frac{1}{2})\kappa_{t-1}|$ is upper bounded by

$$\begin{aligned} & (\log^{1/2} T) T^{-3/2} \sum_{t=1}^T E \left(\left| \log^{-1/2} T \xi_{t-1} \left(\frac{1}{2} \right) \right| |\Upsilon^*(L)\eta_{t-1}| \right) \\ & \leq T^{-1} \sum_{t=1}^T \left(E \left(\log^{-1/2} T \xi_{t-1} \left(\frac{1}{2} \right) \right)^2 \right)^{1/2} (\log^{1/2} T) T^{-1/2} \left(E \left(\eta_1^2 \sum_{j=0}^{\infty} |\Upsilon_j^*| \right) \right) \rightarrow 0, \end{aligned}$$

since $T^{-1} \sum_{t=1}^T (E(\log^{-1/2} T \xi_{t-1}(\frac{1}{2}))^2)^{1/2} \Rightarrow \int_0^1 E(K^2 r^2 W^2(1))^{1/2} dr < \infty$.

Thirdly, for the range $\frac{1}{2} < d < \frac{3}{2}$, $\xi_t(d-1)$ becomes a stationary fractionally integrated process of order $\delta := d-1$, say, with $-\frac{1}{2} < \delta < \frac{1}{2}$. We proceed in this range by first considering the case where $1 < d < \frac{3}{2}$. Since $0 < \delta < \frac{1}{2}$, $E|T^{-d} \sum_{t=1}^T \xi_{t-1}(\delta)\kappa_{t-1}|$ is upper bounded by

$$T^{-d} \sum_{t=1}^T (E(\xi_{t-1}(\delta))^2)^{1/2} \left(E \left(\eta_1^2 \sum_{j=0}^{\infty} |\Upsilon_j^*| \right) \right) = O(T^{-\delta}). \quad (\text{A.4})$$

Consider now the other relevant range where $\frac{1}{2} < d < 1$, and hence $-\frac{1}{2} < \delta < 0$. Define $\Xi_t \equiv \xi_t(\delta)\kappa_t - E(\xi_t(\delta)\kappa_t)$. Given that $\Xi_t = (\sum_{j=0}^{\infty} \pi_j(\delta) \eta_{t-j})(\sum_{i=0}^{\infty} \Upsilon_i^* \eta_{t-i}) - E(\xi_t(\delta)\kappa_t) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \pi_j(\delta) \Upsilon_i^* [\eta_{t-j}\eta_{t-i} - E(\eta_{t-j}\eta_{t-i})]$, with $\pi_j(\delta)$ standing for the j -th element in the binomial expansion of $\Delta^{-\delta}$, it turns out that

$$\begin{aligned} & E|E(\Xi_t | F_{t-m})| \\ & = E \left| \sum_{j=m}^{\infty} \sum_{i=m}^{\infty} \pi_j(\delta) \Upsilon_i^* [\eta_{t-j}\eta_{t-i} - E(\eta_{t-j}\eta_{t-i})] \right| \\ & \leq E \left(\sum_{j=m}^{\infty} \sum_{i=m}^{\infty} |\pi_j(\delta) \Upsilon_i^*| |[\eta_{t-j}\eta_{t-i} - E(\eta_{t-j}\eta_{t-i})]| \right) \\ & \leq C_1 \sum_{j=m}^{\infty} \sum_{i=m}^{\infty} |\pi_j(\delta) \Upsilon_i^*| = C_1 \sum_{j=m}^{\infty} |\pi_j(\delta)| \sum_{i=m}^{\infty} |\Upsilon_i^*|, \end{aligned}$$

where C_1 is a positive constant, $m > 0$. Define $\zeta_m := \sum_{j=m}^{\infty} |\pi_j(\delta)| \sum_{i=m}^{\infty} |\Upsilon_i^*|$. As $\pi_j(\delta)$ and Υ_j^* are absolutely summable sequences, $\lim_{m \rightarrow \infty} \zeta_m = 0$. It follows that Ξ_t is an L^1 -mixingale with respect to F_t . Indeed,

$$E|\Xi_t| \leq C_1 \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\pi_j(\delta) \Upsilon_i^*| = C_2 < \infty,$$

whence, for $k > 0$,

$$\begin{aligned} E(|\Xi_t| 1\{\Xi_t > k\}) &\leq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\pi_j(\delta) \Upsilon_i^*| \{E[|[\eta_{t-j}\eta_{t-i} - E(\eta_{t-j}\eta_{t-i})]|^r]\}^{1/r} \\ &\times \{k^{-1} E|\Xi_t|\}^{(r-1)/r} \leq C_3 \left(\frac{C_2}{k}\right)^{(r-1)/r} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\pi_j(\delta) \Upsilon_i^*|, \end{aligned} \quad (\text{A.5})$$

where C_2, C_3 denote positive constants, $r > 1$, and the first inequality follows using Hölder's inequality. As $\sum_{j=0}^{\infty} \sum_{i=0}^{\infty} |\pi_j(\delta) \Upsilon_i^*|$ is finite, (A.5) can be made as small as desired by choosing k sufficiently large. Thus, Ξ_t is uniformly integrable.

As $\xi_t(\delta)\kappa_t - E(\xi_t(\delta)\kappa_t)$ is an uniformly integrable L^1 -mixingale, it follows from Hansen (1992, Corollary to Theorem 3.3) that

$$\sup_{t \leq T} \left| T^{-1} \sum_{i=1}^t (\xi_i(\delta)\kappa_i - E(\xi_i(\delta)\kappa_i)) \right| \xrightarrow{p} 0, \quad (\text{A.6})$$

noting that

$$T^{-1} E \left(\sum_{i=1}^T \xi_{i-1}(\delta) (\Upsilon^*(L)\eta_{i-1})' \right) = \sum_{j=0}^{\infty} \pi_j(\delta) \Upsilon_j^*. \quad (\text{A.7})$$

Expressions (9) and (10) finally follow by collecting the previous results.

As regards (11) and (12), they follow from both (9) and (10) and the use of the identity $\xi_t(d) = \xi_{t-1}(d) + \xi_t(\delta)$ so that

$$\sum_{i=1}^T \xi_i(d) (\Upsilon(L)\eta_i)' = \sum_{i=1}^T \xi_{i-1}(d) (\Upsilon(L)\eta_i)' + \sum_{i=1}^T \xi_i(\delta) (\Upsilon(L)\eta_i)', \quad (\text{A.8})$$

and by noting that, as in (A.7),

$$T^{-1} \sum_{i=1}^T \xi_i(\delta) (\Upsilon(L)\eta_i)' \xrightarrow{p} \sum_{j=0}^{\infty} \pi_j(\delta) \Upsilon_j'. \quad (\text{A.9})$$

Notice also that in the particular unit root ($d = 1$) case, (A.7) and (A.9) become $\Upsilon_0^{*'} = \sum_{j=1}^{\infty} \Upsilon_j'$ and Υ_0' , respectively, yielding (8). \square

Proof of Lemma 2. Straightforward using Lemma 1, Lemma 2 in SSW and the relationships $D'\hat{\gamma} = \hat{\beta}$ and $\Psi_T(\hat{\gamma} - \gamma) = \{\Psi_T^{-1} \sum_{i=1}^T Z_i Z_i' \Psi_T^{-1}\}^{-1} \{\Psi_T^{-1} \sum_{i=1}^T Z_i \eta_i' \omega_i'\}$. \square