Working Paper 99-68
Statistics and Econometrics Series 23
October 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

ASYMPTOTIC INFERENCE FOR NONSTATIONARY FRACTIONALLY INTEGRATED PROCESSES

Juan J. Dolado and Francesc Marmol *

Abstract

This paper studies the asymptotic of nonstationary fractionally integrated (NFI) multivariate processes with memory parameter $d > 1/2$. We provide conditions to establish a functional central limit theorem and weak convergence of stochastic integrals for NFI processes under the assumptions of these results are given. More specifically, we obtain the rates of convergence and limiting distributions of the OLS estimators of cointegrating vectors in triangular representations. Further, we extend Sims, Stock and Watson's (1990) analysis on estimation and hypothesis testing in vector autoregressions with integrated processes and deterministic components to the more general fractional framework. We show how their main conclusions remain valid when dealing with NFI processes. That is, whenever a block of coefficients can be written as coefficients on zero mean I(O) 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 nonstandard limiting distributions.

Key Words

Fractionally integrated processes, functional central limit theorem; stochastic integration; vector autoregressions; Granger causality; lag selection; OLS estimation.

*Dolado, Department of Economics, Universidad Carlos III de Madrid, Spain, e-mail: dolado@eco.uc3m.es; Marmol, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain, e-mail: fmarmol@est-econ.uc3m.es; The authors thank Javier Hidalgo, Domenico Marinucci, Daniel Peña, Peter M. Robinson, Carlos Velasco, Santiago Velilla and participants at CEF99, Boston College, for helpful comments and suggestions on earlier drafts of this paper.

ASYMPTOTIC INFERENCE FOR NONSTATIONARY FRACTIONALLY INTEGRATED PROCESSES

by

JUAN J. DOLADO

Department of Economics, Universidad Carlos III *de Madrid*

and

FRANCESC MARMOL*

Department of Statistics and Econometrics, Universidad Carlos III *de Madrid*

July 1999

ABSTRACT

This paper studies the asymptotics of nonstationary fractionally integrated (*NFI*) multivariate processes with memory parameter $d > 1/2$. We provide conditions to establish a functional central limit theorem and weak convergence of stochastic integrals for *NFI* processes under the assumption that the innovations are $I(0)$ linear processes. Several applications of these results are given. More specifically, we obtain the rates of convergence and limiting distributions of the *OLS* estimators of cointegrating vectors in triangular representations. Further, we extend Sims, Stock and Watson's (1990) analysis on estimation and hypothesis testing in vector autoregressions with integrated processes and deterministic components to the more general fractional framework. We show how their main conclusions remain valid when dealing with *NFI* processes. That is, 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 nonstandard limiting distributions.

KEYWORDS : Fractionally integrated processes ; functional central limit theorem ; stochastic integration ; vector autoregressions ; Granger causality ; lag selection ; OLS estimation.

---

* The authors thank Javier Hidalgo, Domenico Marinucci, Daniel Peña, Peter M. Robinson, Carlos Velasco, Santiago Velilla and participants at CEF99, Boston College, for helpful comments and suggestions on earlier drafts of this paper.

1. INTRODUCTION

The aim of this paper is twofold. On the one hand, we provide conditions to establish a functional central limit theorem (*FCLT*) and weak convergence of stochastic integrals for nonstationary fractionally integrated (*NFI*) multivariate processes with memory parameter $d \in \partial = \{d \in \Re | d > \frac{1}{2}, d \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots\}$. On the other hand, these theoretical results are applied to the analysis of inference in linear time series models with some fractional unit roots.

Statistical analysis of time series data exhibiting high power at very low frequencies, in general, and fractionally integrated processes (Granger and Joyeux, 1980, Hosking, 1981), in particular, have attracted the attention of both macroeconomists and econometricians and the field is unusually active with both theoretical and empirical research moving forward together. See, e.g., Robinson (1994a) and Baillie (1996) for recent overviews of the subject. In this sense, most of the recent empirical studies seem to locate the memory parameter of these fractionally integrated processes in the nonstationary $(d > \frac{1}{2})$ range.

As is well-known, when dealing with nonstationary data, the asymptotic analysis requires different techniques from those used in the standard stationary case. Indeed, in the nonstationary case, the asymptotic analysis involves the combined use of a multivariate *FCLT* along with the Continuous Mapping Theorem (*CMT*) to establish the weak convergence of those processes to functionals of Brownian motions or related Gaussian processes. Invariance principles for multivariate integrated processes have been derived, among others, by Phillips and Durlauf (1986), Chan and Wei (1988) and more recently by de Jong and Davidson (1999).

As a first contribution, we provide in this paper a *FCLT* for multivariate *NFI* processes under the assumption that the innovations are fairly general $I(0)$ linear processes. Our invariance principle nests as particular cases most of the previously derived *FCLT* for integrated processes. At this stage it is important to remark that ours is not the only attempt to obtain a multivariate *FCLT* for *NFI* processes. Indeed, Davidson and de Jong (1999) and Marinucci and Robinson (1998) have recently derived two invariance principles for fractionally integrated processes which may look similar to ours. However, our *FCLT* differs from the ones obtained in those papers in two significant ways. On the one hand, the *FCLT* derived in Davidson and de Jong (1999) assumes that the innovations are *NED* processes and that the multivariate fractional process is *stationary*, i.e., $|d| < \frac{1}{2}$. Such a process, suitably normalized, converges weakly to a multivariate fractional Brownian motion with *stationary increments* (a *Type I* fractional Brownian motion in Marinucci and Robinson's, 1999, sense). Inference for *NFI* processes is then conducted using the *CMT* by means of partial sums of the underlying stationary fractionally integrated process as in the unit root literature. However, as pointed out by Marinucci and Robinson (1998), partial sum processes can be restrictive for many applications and more general forms of dependence may be considered, for instance the nonstationary fractional integration in Heyde and Yang's (1997) sense. By contrast, our *FCLT* is explicitly derived for *NFI* processes and includes the extended version of nonstationary long range dependence proposed by Heyde and Yang. We show that, after a suitable normalization, a *NFI* process converges weakly to a fractional Brownian motion with *nonstationary increments* and defined for all $d \in \partial$ (a so-called *Type II* fractional Brownian motion).

On the other hand, Marinucci and Robinson (1998) derive the *FCLT* assuming that the fractionally integrated process is nonstationary as we do, but they assume different initial conditions and moment requirements on the innovations driving the multivariate *NFI* processes. See Remark 5 below for further details. Therefore we consider that none of the invariance principles derived by those authors make our results redundant. On the contrary, our analysis nicely complements theirs since comparison of their invariance principles with ours highlights the existing trade-off between alternative assumptions and/or definitions of a fractionally integrated process and the requirements to achieve the desired results.

Another result in the paper relates the asymptotic properties of statistics which are functions of sample moments in which the data is a mixture of nonstationary and stationary processes, where the corresponding limits appear to be (Ito) stochastic integrals. Under the assumption that the nonstationary data is composed by integrated processes, this weak convergence has been studied, for example, by Chan and Wei (1988), Phillips (1988), Hansen (1992) and de Jong and Davidson (1999) under different conditions on the amount and form of dependence allowed in the innovations. As a second contribution, herein we prove the stochastic integral convergence for *NFI* processes with respect to weakly dependent integrator processes. As before, this question has also been handled by Davidson and de Jong (1999). However, their use of a *Type I* fractional Brownian motion makes their results different from ours.

Finally, as a third contribution, we have considered a generalization of the results in Sims et al. (1990) (hereinafter denoted SSW) regarding inference in *VAR* models with integer unit roots, to the more general case where variables include *NFI* processes. In particular, we analyze under which conditions their conclusions regarding the asymptotic distribution of causality tests, lag length selection tests and estimation of cointegrating vectors remain valid in this more general framework. We show how their main conclusions remain valid in this more general framework, namely, 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 nonstandard limiting distributions.

The rest of the paper is structured as follows. In Section 2 we derive the multivariate *FCLT* for *NFI* processes, whereas in Section 3 we are concerned with issues on stochastic integration of *NFI* processes with respect to fairly general *I*(0) linear processes. In Section 4 we use the previous theoretical results to extend SSW set-up by allowing for *VAR* models with *NFI* processes. Next, some applications of this extension are provided in Section 5. In particular, we prove that the Granger causality test statistic will have a limiting chi-square distribution whenever the *NFI* processes in the *VAR* system are $CI(d,d)$, $d \in \partial$, while the test statistic for lag length selection does not have, in general, a limiting chi-square distribution in our set-up, where $d \in \partial$ and the only stationary terms considered are *I*(0), except in the particular case where d is an integer number, i.e., the situation considered by SSW. Lastly, we prove that the *OLS* estimators of cointegrating vectors in triangular representations are $O_p\left(T^{-d}\right)$ if $d \geq 1$ and $O_p\left(T^{1-2d}\right)$ if $\frac{1}{2} < d < 1$. Indeed, when the regressors are strongly exogenous with respect to the corresponding parameters of the cointegrating vector, then *OLS* is $O_p\left(T^{-d}\right)$ for all $d \in \partial$. Furthermore, in this particular case, the limiting *OLS* distribution is a mixture of normals from which standard inference can be implemented. Finally, some conclusions are summarized in Section 6. Proofs of all the theoretical results are collected in the Appendix. The notation throughout the paper is as follows: $[Tr]$ denotes the greatest lesser integer part of Tr, where $r \in [0,1]$, "$\Rightarrow$" denotes weak convergence of the associated probability measures, "$\xrightarrow{p}$" denotes convergence in probability and "$\equiv$" denotes equality in distribution.

2. A MULTIVARIATE FCLT FOR FRACTIONALLY INTEGRATED PROCESSES

Let $x_t = \left(x_{1t},\ldots,x_{nt}\right)'$ be an n-dimensional vector of nonstationary fractionally integrated (hereafter denoted *NFI*) processes with Wold representation given by

(2.1) $\Delta^d x_t = \varepsilon_t, \quad t = 1,2,\ldots,$

where the fractional difference operator Δ^d (with $\Delta = 1 - B$ where B is the lag operator) can be expressed in terms of a Maclaurin expansion as

(2.2) $$\Delta^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)B^k}{\Gamma(k+1)\Gamma(-d)} = \sum_{k=0}^{\infty} \pi_k B^k, \qquad \pi_k = \frac{k-1-d}{k}\pi_{k-1}, \qquad \pi_0 = 1,$$

with $\Gamma(\cdot)$ denoting the gamma or generalized binomial function and where ε_t is a square integrable martingale difference sequence (mds) with constant conditional variance Σ, i.e.,

$E\left(\varepsilon_t \middle| \varepsilon_{t-1},\ldots,\varepsilon_1\right)=0$ and $E\left(\varepsilon_t\varepsilon_t' \middle| \varepsilon_{t-1},\ldots,\varepsilon_1\right)=\Sigma$. Further, define $\varepsilon_t=\Sigma^{1/2}\eta_t$ so that η_t is an n-dimensional mds with constant conditional variance the identity matrix of order n, I_n. With respect to the memory parameter d, we assume that it belongs to the set $\partial=\left\{d\in\Re \middle| d>\tfrac{1}{2}, d\neq\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2},\cdots\right\}$. This is done without loss of generality in the sense that the set $\left\{\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2},\cdots\right\}$ has Lebesgue measure zero. See, however, Remark 4. As a technical assumption, we further assume that $\left\{\eta_s\right\}_{-\infty}^{0}=0$. Such a restriction will not be assumed to hold when dealing with stationary process.

Instead of using x_t, let $\xi_t(d)=\Delta^{-d}\eta_t$ and $\tilde{\xi}_t(d)=\Delta^{-d}\varepsilon_t$ denote the relevant vectors of *NFI* processes, where the slight change of notation emphasizes the memory of the process and the distributional properties of the corresponding innovations. Note that, in the integer case where $d=\{1,2,3,\ldots\}$, the recursive partial sums of η_t are defined as $\xi_t(d)=\sum_{s=1}^{t}\xi_s(d-1)$, with $\xi_t(1)=\xi_t=\sum_{s=1}^{t}\eta_s$. When d is a real number, $\xi_t(d)=\Delta^{-d}\eta_t$ becomes the continuous equivalence to the above $(d-1)$-fold discrete version. On the other hand, the shocks η_t are assumed to satisfy the following assumption.

ASSUMPTION A : $max_i \sup_t E\left|\eta_{it}\right|^{g+\zeta}<\infty,\quad g=max\left\{2,\dfrac{1}{d-\frac{1}{2}}\right\},\quad \zeta>0$.

Under this assumption, we obtain the following multivariate invariance principle for *NFI* processes. The univariate version was first proved by Akonom and Gourieroux (1988).

THEOREM 1 : *Under Assumption A with* $d\in\partial$, *asymptotically, as the sample size* $T\to\infty$,

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

(2.4) $\quad T^{1/2-d}\tilde{\xi}_{[Tr]}(d)\Rightarrow V(d,r)$,

where

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

$V(d,r)=\Sigma^{1/2}W(d,r)$ *and* $W(r)$ *is a standard n-dimensional Brownian motion,* $W(r)\equiv BM(I_n)$, *such that* $W(1,r)=W(r), r\in[0,1]$.

REMARK 1. The n-dimensional process $W(d,r)$ would be the generalization of $W^m(r)$, $m=\{1,2,3,\ldots\}$, the $(m-1)$-fold integral recursively defined by $W^m(r) = \int_0^r W^{m-1}(s)ds$ with $W^1(r) \equiv W(r)$. Formally, $W(d,r)$ is not the usual (multivariate) fractional Brownian motion as defined by Mandelbrot and Van Ness (1968), but rather a Holmgren-Riemann-Liouville fractional integral (*cf.*, Lévy, 1953), where the parameter d is allowed to take values in the nonstationary $d > \frac{1}{2}$ range.

In effect, and assuming $n=1$ for simplicity, $W(d,r)$ is a Gaussian process with almost surely continuous sample paths and nonstationary (and non independent) increments which differs from the more standard version of the fractional Brownian motion, as originally introduced by Mandelbrot and Van Ness (1968), namely,

$$\widetilde{W}(d,r) = k\left\{\int_{-\infty}^{0}\left\{(r-s)^d - (-s)^d\right\}dW(s) + \int_{0}^{r}(r-s)^d\,dW(s)\right\},$$

where k denotes a constant term. It can be proved (see, e.g., Samorodnitsky and Taqqu, 1994) that $\widetilde{W}(d,r)$ is a Gaussian process with almost surely continuous sample paths and stationary increments, only defined for values of d in the stationary $\left(-\frac{1}{2},\frac{1}{2}\right)$ range. Moreover, it can be proved that for $r \geq 0$, $\widetilde{W}(d,r)$ is composed of two independent components, where one of them is a scaled $W(d,r)$. Therefore, $W(d,r)$ and $\widetilde{W}(d,r)$ have different covariance structures. In terms of Marinucci and Robinson (1999), $\widetilde{W}(d,r)$ and $W(d,r)$ correspond to *Type I* and *Type II* fractional Brownian motions, respectively. Indeed, the invariance principle derived by Davidson and de Jong (1999), extending previous results by Davydov (1970), Taqqu (1975) and Chan and Terrin (1995), is based on $\widetilde{W}(d,r)$, and it differs, correspondingly, from that obtained in Theorem 1. See Marinucci and Robinson (1999) for further insights.

REMARK 2. When dealing with *NFI* processes, one cannot simply use Donker's Theorem jointly with the Continuous Mapping Theorem (*CMT*) to obtain expressions (2.3) and (2.4), since the *CMT* can only be used for $d \geq 1$. In effect, and assuming again for simplicity that $n=1$, then the mapping defined on $C[0,1]$ by $W(r) \to \int_0^r (r-s)^{d-1} W(r)dr$ will be uniformly continuous if $\int_0^r (r-s)^{d-2} ds < \infty$, i.e., if $d \geq 1$, and therefore, the *CMT* can be directly invoked in this case. However, in the remaining cases, when $\frac{1}{2} < d < 1$, it is not possible to use the latter result, and hence it is necessary to prove the weak convergence in an explicit way as we do in Theorem 1. See Akonom and Gourieroux (1988) for further details.

REMARK 3. Theorem 1 can be readily extended to allow for more general situations, such as different memory parameters $(d_1, d_2, \ldots, d_n)$, much in the same way as in Marinucci and Robinson (1998) and Davidson and de Jong (1999). We have not considered such an extension in this paper since we are primarily concerned with the cointegrated case where a necessary condition for cointegration with $(d_1, d_2, \ldots, d_n) \in \partial$ is that $d_1 = d_2 = \cdots = d_n = d$.

REMARK 4. Along this paper we confine the analysis to memory parameters lying in the set $\partial = \{d \in \Re | d > \frac{1}{2}, d \neq \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots\}$. This is done without loss of generality in the sense that the set $\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots\}$ has Lebesgue measure zero. However, as recently proved by Liu (1998), this restriction *is not* without loss of generality with respect to the assumed initial conditions. In effect, if we impose that $\{\eta_s\}_{-\infty}^{0} = 0$ so that $\xi_t(d) = \sum_{j=1}^{t} d_{t-j}\eta_j$, with $d_0 = 1, d_j \sim j^{d-1}$, then (2.3) not only holds for $d \in \partial$ but also for $d \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots\}$. See Marinucci and Robinson (1998, Theorem 1). Such a restriction on the initial conditions of $\xi_t(d)$ is necessary because the d_j's are not square summable for $d > \frac{1}{2}$. However, without imposing that the MA representation for the nonstationary long memory process $\xi_t(d)$ takes value 0 at time 0, Liu (1998, Theorem 2.3) shows that the convergence rate at the points $d \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \cdots\}$ has an additional component, namely, $\log T$, when compared to the convergence rates for the other values of d in ∂. Hence, in light of these results, and since we must impose that $\{\eta_s\}_{-\infty}^{0} = 0$ for the sum $\xi_t(d) = \sum_{j=1}^{t} d_{t-j}\eta_j$ to be finite, we decided to restrict our analysis to the case where $d \in \partial$. By doing this, it can be verified that $\text{var}\left(\sum_{t=1}^{p} \xi_t(d)\right) \sim cp^{2d+1}$ as $p \to \infty$, $0 < c < \infty$ and thus $\xi_t(d)$ is nonstationary long range dependent in Heyde and Yang's (1997) sense.

So far, we have assumed the simple Wold representation (2.1) in order to assess the requirements of associated moments and because it is the most useful representation to develop the set-up introduced in Section 4. In practical applications, however, the assumption that the *NFI* processes of interest have mds sequences can be too restrictive, and much more generality can be achieved by allowing the underlying innovations to follow linear processes. For this, let us now assume that the relevant data generating process is given by

(2.5) $\Delta^d x_t = u_t$,

where $u_t = C(L)\varepsilon_t = C(L)\Sigma^{1/2}\eta_t = \Psi(L)\eta_t$ and $\Psi(L) = \sum_{j=0}^{\infty}\Psi_j L^j$ is a 1-summable one-sided linear filter, i.e., $\sum_{j=0}^{\infty} j\left|\Psi_j\right| < \infty$, where $|A| = max_{i,j}\left|A_{ij}\right|$ for a matrix A. For this set-up, let $\hat{\xi}_t(d) = \Delta^{-d}u_t$, and assume the following condition.

ASSUMPTION B : $max_i \sup_t E\left|\eta_{it}\right|^{g+\zeta} < \infty, \zeta > 0$, *where*

(i) $g = 2$ *if* $d > \frac{3}{2}$,

(ii) $g = 4$ *if* $\frac{3}{4} \leq d < \frac{3}{2}$, *and*

(iii) $g = \dfrac{8(1-d)}{2d-1}$ *if* $\frac{1}{2} < d < \frac{3}{4}$.

Then, we obtain the following invariance principle for multivariate *NFI* processes with innovations driven by linear processes, where the long-run covariance of the innovation vector process will be denoted by $\Omega = \Psi(1)\Psi(1)' = C(1)\Sigma C(1)'$.

THEOREM 2. *Under Assumption B with* $d \in \partial$, *asymptotically, as the sample size* $T \to \infty$,

(2.6) $T^{1/2-d}\hat{\xi}_{[Tr]}(d) \Rightarrow B(d,r)$,

where

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

and $B(r) \equiv BM(\Omega)$.

The most important message of Theorem 2 is that, for $d > \frac{3}{2}$, only finite second moments of η_t, together with some restrictions on the dependence embodied in the lag polynomial *C(L)*, are needed for the multivariate invariance principle with *NFI* processes to hold. This result contrasts sharply with previous assumptions reported in the literature of *I(d)* processes, which require uniformly bounded finite fourth moments for all $d = 1, 2, \ldots$. See, for instance, SSW (Condition 1) and Stock and Watson (1993, Assumption C).

REMARK 5. In independent work, Marinucci and Robinson (1998, Theorem 1) obtain expression (2.6) under Assumption A but assuming that $u_t = 0$ for $t \leq 0$ and η_t~iid. By contrast, and following most of the relevant econometric literature (*vid.*, Stock, 1987, SSW and Stock and Watson, 1993, *inter alia*) we have assumed throughout the paper that $\eta_t = 0$ for $t \leq 0$ and η_t~mds. Of course, those different assumptions on the initial conditions of the underlying processes have negligible asymptotic effects. Moreover, it is not difficult to show that both sets of initial conditions give rise to the same sequence of solutions for the $\{x_t\}$ process. Thus, Marinucci and Robinson's result nicely complements our Theorem 2, highlighting the existing trade-off between the distributional assumptions on the error terms and moment requirements: if η_t is assumed to be iid, then these authors prove that Assumption A is all that is needed to obtain expression (2.6). By contrast, if η_t is assumed to be a mds, as we do here, then the stronger Assumption B is required.

3. STOCHASTIC INTEGRATION

The combination of Assumption *B* and the summability conditions imposed on *C*(*L*), together with *CMT*, are the only tools which are needed to ensure the convergence of most of the relevant random matrices which appear in estimation and testing of long-run relationships among *NFI* processes.

However, the above theory does not cover the very important case of the weak convergence of sample covariance matrices $\sum_{t=1}^{T} x_t u_t'$ to matrix stochastic integrals of the form $\int_0^1 B(d,r)dB(r)' + \aleph$, where $\aleph$ is a constant matrix of bias terms, since in this case the convergence cannot be obtained from a routine application of the *CMT* and a multivariate invariance principle. As in the previous section, an explicit proof of the weak convergence of the sample matrices $\sum_{t=1}^{T} x_t \eta_t'$ and $\sum_{t=1}^{T} x_t u_t'$ is required. For this, assume the following condition to hold.

ASSUMPTION C : $E\left(\eta_t'\eta_t \middle| \eta_{t-1}, \eta_{t-2}, \ldots\right) \leq c$ *a.s. for some constant* $c > 0$.

THEOREM 3. *Under Assumptions A and C, then for all* $d \in \partial$, *as* $T \to \infty$,

$$(3.1) \quad T^{-d}\sum_{t=1}^{T}\xi_{t-1}(d)\eta_t' \Rightarrow \int_0^1 W(d,r)dW(r)'$$

$$(3.2) \quad T^{-d}\sum_{t=1}^{T}\tilde{\xi}_{t-1}(d)\varepsilon_t' \Rightarrow \int_0^1 V(d,r)dV(r)'.$$

Under Assumptions B and C, then, as $T \to \infty$,

$$(3.3) \quad T^{-d}\sum_{t=1}^{T}\hat{\xi}_{t-1}(d)u_t' \Rightarrow \int_0^1 B(d,r)dB(r)', \text{ when } d > 1,$$

$$(3.4) \quad T^{-1}\sum_{t=1}^{T}\hat{\xi}_{t-1}u_t' \Rightarrow \int_0^1 B(r)dB(r)' + \Lambda, \text{ when } d = 1,$$

and

$$(3.5) \quad T^{-1}\sum_{t=1}^{T}\hat{\xi}_{t-1}(d)u_t' \xrightarrow{p} \Lambda(d), \text{ when } \tfrac{1}{2} < d < 1,$$

where $\Lambda \equiv \sum_{k=1}^{\infty} E(u_0 u_k')$ *and* $\Lambda(d) \equiv \sum_{k=1}^{\infty} E(\hat{\xi}_0(d-1)u_k')$.

From Theorem 3 we learn that $\sum_{t=1}^{T} x_t\eta_t'$ and $\sum_{t=1}^{T} x_t\varepsilon_t'$ are $O_p(T^d)$ for all $d \in \partial$. This is only true for the sample matrix $\sum_{t=1}^{T} x_t u_t'$ whenever $d \geq 1$. Otherwise, when $\frac{1}{2} < d < 1$, it is $O_p(T)$, degenerating in probability to a nonstochastic limit. Notice also that, in the particular case where $d = 1$, $\sum_{t=1}^{T} x_t u_t'$ has a bias (in mean) matrix, Λ.

On the other hand, given expressions (3.1)-(3.5), the next result follows directly.

COROLLARY 1. *Under the same assumptions as in Theorem 3, then, as* $T \to \infty$,

$$(3.6) \quad T^{-d}\sum_{t=1}^{T}\xi_t(d)\eta_t' \Rightarrow \int_0^1 W(d,r)dW(r)', \text{ for all } d \in \partial,$$

$$(3.7) \quad T^{-d}\sum_{t=1}^{T}\tilde{\xi}_t(d)\varepsilon_t' \Rightarrow \int_0^1 V(d,r)dV(r)', \text{ for all } d \in \partial,$$

$$(3.8) \quad T^{-d}\sum_{t=1}^{T}\hat{\xi}_t(d)u_t' \Rightarrow \int_0^1 B(d,r)dB(r)', \text{ when } d > 1,$$

$$(3.9) \quad T^{-1}\sum_{t=1}^{T}\hat{\xi}_t u_t' \Rightarrow \int_0^1 B(r)dB(r)' + \aleph, \text{ when } d = 1,$$

and

$$(3.10) \quad T^{-1}\sum_{t=1}^{T}\tilde{\xi}_t(d)u_t' \xrightarrow{p} \aleph(d), \text{ when } \tfrac{1}{2}<d<1,$$

where $\aleph \equiv \sum_{k=0}^{\infty} E(u_0 u_k')$ *and* $\aleph(d) \equiv \sum_{k=0}^{\infty} E(\tilde{\xi}_0(d-1)u_k')$.

Finally, to conclude this section, consider the following partition of (2.5), $x_t = (x_{1t}' \quad x_{2t}')'$, where x_{1t} and x_{2t} have dimensions n_1 and n_2 $(n_1 + n_2 = n)$, respectively. Further, divide $u_t, \Omega, \Lambda, \tilde{\xi}_t(d)$ and $B(r)$ conformably with x_{1t} and x_{2t}. Then, a very important particular case arises when the innovations u_{1t} and u_{2s} are independent for all *t*, *s*, so that $\Omega_{21} = \Lambda_{21} = 0$. In this case we obtain the following result, stated as a corollary.

COROLLARY 2. *Under the same assumptions as in Theorem 3 and for all* $d \in \partial$, *if* $\Omega_{21} = \Lambda_{21} = 0$, *then, as* $T \to \infty$,

$$(3.11) \quad T^{-d}\sum_{t=1}^{T}\tilde{\xi}_{2,t-1}(d)u_{1t}' \Rightarrow \int_0^1 B_2(d,r)dB_1(r)'$$

and

$$(3.12) \quad T^{-d}\sum_{t=1}^{T}\tilde{\xi}_{2t}(d)u_{1t}' \Rightarrow \int_0^1 B_2(d,r)dB_1(r)',$$

where $B_1(r)$ *and* $B_2(d,r)$ *are stochastically independent processes. Furthermore, in this particular case, the limiting distribution* $\int_0^1 B_2(d,r)dB_1(r)'$ *is mixed normal.*

4. ESTIMATION OF COINTEGRATING VECTORS IN FRACTIONAL SYSTEMS

As an application of the previous general results, consider the *VAR* model[1]

$$(4.1) \quad Y_t = \alpha + \sum_{i=1}^{p}\Phi_i Y_{t-i} + \varepsilon_t,$$

where Y_t is an $n \times 1$ vector. Assume that the determinant of the autoregressive polynomial $|I_n - \Phi_1 z - \Phi_1 z^2 - \cdots - \Phi_p z^p|$ has all its roots on or outside the unit circle, and that the maximum order of integration of any element of Y_t is $d \in \partial$. Moreover, assume that there are no cross equation restrictions, so that the efficient linear estimators correspond to the equation-by-equation *OLS* estimators. This assumption is only made for simplicity. The generalization to more general frameworks, such as those considered, for instance, by Park and Phillips (1988, 1989) and SSW, can be obtained by a direct application of the techniques and analyses discussed here.

Previous to this analysis, however, we will include the following useful lemma, of independent interest, which extends Lemma 1 in SSW and Lemma A.2 in Stock and Watson (1993) to the more general fractional set-up considered in this paper.

LEMMA 1. *Assume that Assumptions B and C hold and that the lag operators* $F(L)=\sum_{j=0}^{\infty}F_jL^j$ *and* $G(L)=\sum_{j=0}^{\infty}G_jL^j$ *are* 1*-summable and* $F(1)$ *nonsingular. Then the following converge jointly*:

(4.2) $$T^{-(m+d+1/2)}\sum_{t=1}^{T}t^m\xi_t(d)'\Rightarrow\int_0^1 r^m W(d,r)'dr,\quad m\geq 0, d\in\partial,$$

(4.3) $$T^{-(m+d)}\sum_{t=1}^{T}\xi_t(m)\xi_t(d)'\Rightarrow\int_0^1 W(m,r)W(d,r)'dr,\quad m,d\in\partial,$$

(4.4) $$T^{-(m+1)}\sum_{t=1}^{T}t^m\to\frac{1}{(m+1)},\quad m\geq 0,$$

(4.5) $$T^{-(m+1/2)}\sum_{t=1}^{T}t^m\left[F(L)\eta_{t+1}\right]'\Rightarrow\int_0^1 r^m dW(r)'F(1)',\quad m\geq 0,$$

(4.6) $$T^{-1/2}\sum_{t=1}^{T}F(L)\eta_t\Rightarrow F(1)\int_0^1 W(r)dr,$$

(4.7) $$T^{-1}\sum_{t=1}^{T}\left[F(L)\eta_t\right]\left[G(L)\eta_t\right]'\xrightarrow{p}\sum_{i=1}^{\infty}F_iG_i',$$

(4.8) $$T^{-1}\sum_{t=1}^{T}\xi_t\left(F(L)\eta_t\right)'\Rightarrow F(1)'+\int_0^1 W(r)dW(r)'F(1)',\quad if\ d=1,$$

(4.9) $$T^{-d}\sum_{t=1}^{T}\xi_t(d)\left(F(L)\eta_t\right)'\Rightarrow\int_0^1 W(d,r)dW(r)'F(1)',\quad if\ d>1,$$

and

(4.10) $$T^{-1}\sum_{t=1}^{T}\xi_t(d)\left(F(L)\eta_t\right)'\xrightarrow{p}F(1)^{-1}\aleph(d),\quad if\ \tfrac{1}{2}<d<1.$$

[1] A different *VAR* model with $I(1)$ regressors and stationary fractionally integrated errors has been recently considered by Jeganathan (1999).

Next, following Watson (1994), consider the ith equation of model (4.1),

$$(4.11)\quad y_{it} = X_t'\beta + \varepsilon_{it},$$

where y_{it} is the ith element of Y_t, $X_t = \left(1, Y_{t-1}', Y_{t-2}', \ldots, Y_{t-p}'\right)'$ is the $(np+1)$ vector of regressors, β denotes the corresponding vector of regression coefficients, and $\varepsilon_{it} = \omega'\eta_t = \eta_t'\omega$ is the ith element of ε_t and ω' is the ith row of the covariance matrix Σ.

Now, following SSW, obtaining the asymptotic behavior of the *OLS* estimator $\hat{\beta}$ of β is greatly facilitated by transforming the regressors in a way that isolates the various stochastic and deterministic trends. In particular, the regressors are transformed as $Z_t = DX_t$, where the nonsingular square matrix D is chosen in such a way that Z_t has a simple representation in terms of the fundamental stochastic and nonstochastic 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 $D'\hat{\gamma} = \hat{\beta}$. The regressors Z_t are related to the deterministic and stochastic trends given in Lemma 1 by the transformation $Z_t = F(L)v_t$, where $F(L)$ is a lower triangular matrix and the variates v_t are referred to as the canonical regressors associated with Y_t. See SSW for a general procedure for transforming regressors from an integrated *VAR* into canonical form.

The advantage of this transformation is that it isolates the terms of different orders of probability. Notice, however, that, contrary to the case where d takes only integer values, the vector v_t would be an infinite dimensional random vector in the fractional case and for any $d \in \partial$. Therefore, among the infinite available possibilities, we have chosen the following one :

$$(4.12)\quad v_t = \begin{pmatrix} \eta_t \\ 1 \\ \xi_t(d) \\ \xi_t \\ t \end{pmatrix},$$

where $\xi_t(d)$ stands for a *NFI* process of order $\frac{1}{2} < d < 1$, thus allowing for a *VAR* model where its elements Y_t are individually *I*(0), *I*(1) or *I*(*d*) with $\frac{1}{2} < d < 1$ (i.e., nonstationary but mean reverting) processes, possibly around a linear time trend. Many authors have stressed the empirical relevance of that class of processes. See, e.g., Baillie (1996). The main conclusions of our analysis remain invariant to other possible configurations of v_t.

The corresponding regressors Z_t can now be related to the deterministic and stochastic trends given in Lemma 1 by the following transformation

$$(4.13)\qquad \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{3t} \\ z_{4t} \\ z_{5t} \end{pmatrix} = \begin{pmatrix} F_{11}(L) & 0 & 0 & 0 & 0 \\ 0 & F_{22} & 0 & 0 & 0 \\ F_{31}(L) & F_{32} & F_{33} & 0 & 0 \\ F_{41}(L) & F_{42} & F_{43} & F_{44} & 0 \\ F_{51}(L) & F_{52} & F_{53} & F_{54} & F_{55} \end{pmatrix} \begin{pmatrix} \eta_t \\ 1 \\ \xi_t(d) \\ \xi_t \\ t \end{pmatrix}.$$

Further, assume that z_{1t} contains k_1 elements, z_3 contains k_3 elements, z_4 contains $B_1(r)$ elements and define the scaling matrix

$$\Psi_T = diag\left\{T^{1/2}I_{k_1}, T^{1/2}, T^d I_{k_3}, TI_{k_4}, T^{3/2}\right\}.$$

Then, using Lemma 1, we are now able to obtain the following result concerning the limiting behavior of the moment matrices based on the transformed regressors Z_t.

LEMMA 2 . *Under the same assumptions as in Lemma* 1, *the following converge jointly* :

(*a*) $\Psi_T^{-1}\left(\sum_{t=1}^{T} Z_t Z_t^{'}\right)\Psi_T^{-1} \Rightarrow \Pi$, *where, partitioning* Π *conformably with* Z_t,

$$\Pi_{11} = \sum_{j=0}^{\infty} F_{11,j}F_{11,j}^{'},$$

$$\Pi_{22} = F_{22}^2,$$

$$\Pi_{33} = F_{33}\int_0^1 W(d,r)W(d,r)'\,drF_{33}^{'},$$

$$\Pi_{44} = F_{44}\int_0^1 W(r)W(r)'drF_{44}^{'},$$

$$\Pi_{55} = \frac{F_{55}^2}{3},$$

$$\Pi_{1j} = \Pi_{j1}^{'} = 0, \quad j = 2,3,4,5,$$

$$\Pi_{23} = \Pi_{32}^{'} = F_{22}\int_0^1 W(d,r)'\,drF_{33}^{'},$$

$$\Pi_{24} = \Pi_{42}^{'} = F_{22}\int_0^1 W(r)'\,drF_{44}^{'},$$

$$\Pi_{25} = \Pi_{52} = \frac{F_{22}F_{55}}{2},$$

$$\Pi_{34} = \Pi_{43}^{'} = F_{33}\int_0^1 W(d,r)W(r)'drF_{44}^{'},$$

$$\Pi_{35} = \Pi_{53}^{'} = F_{33}\int_0^1 rW(d,r)drF_{55},$$

$$\Pi_{45} = \Pi_{54}^{'} = F_{44}\int_0^1 rW(r)drF_{55}.$$

(*b*) $\Psi_T^{-1}\sum_{t=1}^{T} Z_t\eta_t^{'}\omega \Rightarrow \mathrm{A}$, *where, partitioning* A *conformably with* Z_t,

$$A_1 = N\{0,(\omega'\omega)\Pi_{11}\},$$

$$A_2 = F_{22}\int_0^1 dW(r)'\omega,$$

$$A_3 = F_{33}\int_0^1 W(d,r)dW(r)'\omega,$$

$$A_4 = F_{44}\int_0^1 W(r)dW(r)'\omega,$$

$$A_5 = F_{55}\int_0^1 rdW(r)'\omega.$$

Application of Lemma 2, in turn, makes the asymptotic analysis of the *OLS* estimators $\hat{\gamma}$ and $\hat{\beta}$ straightforward, as stated in the following theorem.

THEOREM 4. *Under the same assumptions as in Lemma* 1, *asymptotically,*

(4.14) $\Psi_T(\hat{\gamma}-\gamma) \Rightarrow \Pi^{-1}A,$

and

(4.15) $\Psi_T(D')^{-1}(\hat{\beta}-\beta) \Rightarrow \Pi^{-1}A.$

From this theorem, the following comments apply. First, when the model is correctly specified, in the sense that the errors are mds, then $\hat{\gamma}$ and $\hat{\beta}$ are consistent when there are deterministic time trends and an arbitrary number of unit roots. The individual coefficients converge to their theoretical counterparts at different rates. Secondly, when some transformed regressors are dominated by stochastic trends, their joint distribution will be non-normal. When there are no Z_t regressors dominated by stochastic trends, $\hat{\gamma}$ (and thus $\hat{\beta}$) has an asymptotically normal joint distribution. Thirdly, the block diagonality of Π implies that $T^{1/2}(\hat{\gamma}_1 - \gamma_1) \xrightarrow{\ell} N(0, \omega'\omega\Pi_{11}^{-1})$. 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 $T^{1/2}(\hat{\gamma}_1 - \gamma_1)$ is asymptotically independent of the other estimated coefficients. By contrast, all of the other coefficients will have non-normal limiting distributions, in general.

All the above results are well known from SSW and Watson (1994). Herein we have shown how they extend to the fractional framework. Thus, as in the case of an integer degree of integration, they provide a very useful sufficient condition for estimating coefficients with asymptotically normal limiting distributions. 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.

Moreover, if we consider Wald statistics for linear hypothesis of the form

$$H_0:\ R\beta = r \quad \text{vs.} \quad H_1:\ R\beta \neq r,$$

where R denotes a full column rank matrix of q restrictions,

$$W = \frac{\left(R\hat{\beta} - r\right)'\left[R\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} R'\right]^{-1}\left(R\hat{\beta} - r\right)}{\hat{\sigma}_i^2},$$

where $\sigma_i^2 = \text{var}(\varepsilon_{it})$, then proceeding as in SSW (pp. 124-127) or Watson (1994, pp. 2858-2860), 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. This is so since W has in this case a limiting chi-square distribution. Otherwise, in general, the statistics will have nonstandard limiting distributions.

5. SOME APPLICATIONS

Lemma 2 and Theorem 4 together provide the basis for extending SSW's analysis to the fractional case. In this section we will investigate how some of their results carry over the more general fractional framework. For this, we shall follow Watson's (1994) notation as close as possible in order to compare our findings with the existing results for the $I(1)$ case.

5.1. *Testing for Granger causality*

Following Watson (1994), consider the bivariate *VAR* model

$$(5.1) \quad y_{1t} = \alpha_1 + \sum_{i=1}^{p} \phi_{11,i} y_{1,t-i} + \sum_{i=1}^{p} \phi_{12,i} y_{2,t-i} + \varepsilon_{1t} ,$$

$$(5.2) \quad y_{2t} = \alpha_2 + \sum_{i=1}^{p} \phi_{21,i} y_{1,t-i} + \sum_{i=1}^{p} \phi_{22,i} y_{2,t-i} + \varepsilon_{2t} ,$$

with the restriction that y_{2t} does not Granger-cause y_{1t} corresponding to the null hypothesis H_0: $\phi_{12,1} = \phi_{12,2} = \cdots = \phi_{12,p} = 0$.

When y_{1t} and y_{2t} are both *I*(0) processes, it is well known that the resulting Wald, LR or LM test for this hypothesis have a limiting chi-square distribution. On the other hand, when y_{1t} and y_{2t} are integrated, SSW, Toda and Phillips (1993a, 1993b) and Watson (1994) prove 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 *I*(1) but cointegrated, 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 above results, it is not difficult to show that the same comments apply in the fractional case, namely, when y_{1t} and y_{2t} are *NFI* processes with $d \in \partial$. So, the Granger-causality test statistic will have a limiting chi-square distribution whenever they are cointegrated, i.e., when there is an *I*(0) linear combination of the variables, say $c_t = y_{2t} - \lambda y_{1t}$, so that (5.1) can be rewritten as

$$(5.3) \quad y_{1t} = \tilde{\alpha}_1 + \sum_{i=1}^{p} \tilde{\phi}_{11,i} y_{1,t-i} + \sum_{i=1}^{p} \phi_{12,i} \left(c_{t-i} - \mu_c\right) + \varepsilon_{1t} ,$$

where μ_c is the mean of c_t, $\tilde{\alpha}_1 = \alpha_1 + \sum_{i=1}^{p} \phi_{12,i} \mu_c$ and $\tilde{\phi}_{11,i} = \phi_{11,i} + \phi_{12,i} \lambda, \quad i = 1, \ldots, p$. Now, since the Granger-causality restriction in the transformed regression corresponds to the restriction that the terms $c_{t-i} - \mu_c$ do not enter the regression and these are zero mean *I*(0) regressors in a regression that includes a constant, 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 c_t is not an *I*(0) process but a fractionally integrated process of order $d_c < d$, then the resulting test statistic will not, in general, have a limiting chi-square distribution.

5.2. Testing lag length restrictions

Second, another important issue in specifying *VAR*'s is the determination of the correct lag length. For this, consider the $VAR(p+s)$ model,

$$(5.4) \quad Y_t = \alpha + \sum_{i=1}^{p+s} \Phi_i Y_{t-i} + \varepsilon_t ,$$

and the null hypothesis that the true model is a $VAR(p)$ process, i.e., $H_0 : \Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$.

When $p \geq 1$, SSW and Watson (1994) prove that if ΔY_t is *I*(0) with mean μ, then the usual Wald (and LR and LM) test statistic for H_0 has an asymptotic chi-square distribution under the null in the integer case.

In effect, in this case, model (2.37) can be rewrite as

$$(5.5) \quad Y_t = \widetilde{\alpha} + \mathrm{H} Y_{t-1} + \sum_{i=1}^{p+s-1} \mathrm{M}_i (\Delta Y_{t-i} - \mu) + \varepsilon_t ,$$

where $\mathrm{H} = \sum_{i=1}^{p+s} \Phi_i$, $M_i = -\sum_{j=i+1}^{p+s} \Phi_j$ and $\widetilde{\alpha} = \alpha + \sum_{i=1}^{p+s-1} M_i \mu$, so that the restrictions $\Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$ in the original model are equivalent to $M_p = M_{p+1} = \cdots = M_{p+s-1} = 0$ in the transformed model. Since these coefficients are zero mean *I*(0) regressors in regression equations which include a constant term, the test statistics will have the usual large sample chi-square distribution.

On the other hand, this conclusion can be readily extended to the integer case where the maximal order of integration of Y_t is $d = 1, 2, \ldots$, just by iterating the above procedure. In effect, if, for instance, *I*(2)-ness is the maximal order of integration to consider so that $\Delta^2 Y_t$ is *I*(0) with mean μ, then it is direct to prove that model (5.4) can be now transformed as

$$(5.6) \quad Y_t = \widetilde{\widetilde{\alpha}} + \mathrm{H} Y_{t-1} + M \Delta Y_{t-1} + \sum_{i=1}^{p+s-2} N_i (\Delta^2 Y_{t-i} - \mu) + \varepsilon_t ,$$

where $\widetilde{\widetilde{\alpha}} = \alpha + \sum_{i=1}^{p+s-2} N_i \mu$, $M = \sum_{i=1}^{p+s-1} M_i$, $N_i = -\sum_{j=i+1}^{p+s-1} M_j$ and the restrictions $\Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$ in the original model become now equivalent to $N_{p-1} = N_p = \cdots = N_{p+s-2} = 0$ in (5.6). Again, these coefficients are zero mean *I*(0) regressors in regression equations that contain a constant term, so the test statistics will have a limiting chi-square distribution.

By contrast, in the fractional case, when the vector Y_t is such that $\Delta^d Y_t \sim I(0)$ and $E(\Delta^d Y_t) = \mu$, $d \in \partial$, $d \neq 1, 2, \ldots$, since the fractional difference operator Δ^d has an expansion with infinite terms, given by expression (2.2), so that there does not exist a finite state space representation for this kind of processes (see, *e.g.*, Chan and Palma, 1998), then the restrictions $\Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$ in the original model cannot be transformed into a set of restrictions on the coefficients of zero mean $I(0)$ regressors in a transformed model containing a constant term. Consequently, the test statistic for lag length selection will not have in general a limiting chi-square distribution in our set-up, where $d \in \partial$ and the only stationary terms considered are $I(0)$.

5.3. Estimating cointegrating vectors by OLS

Finally, as a third application, assume that the scalar random processes y_{1t} and y_{2t} are generated by the triangular representation

(5.7) $y_{1t} = \beta y_{2t} + u_{1t}$,

(5.8) $\Delta^d y_{2t} = u_{2t}, \quad d \in \partial$,

with the innovations $u_t = (u_{1t}, u_{2t})' = C(L)\varepsilon_t = \Psi(L)\eta_t$ being a linear process satisfying Assumptions *B* and *C*. Then, using Theorem 2, Corollary 1 and the *CMT*, it follows that, when $d = 1$, the limiting distribution of the *OLS* estimator of β in model (5.7)-(5.8) is given by

$$(5.9) \quad T\left(\hat{\beta} - \beta\right) \Rightarrow \frac{\int_0^1 B_2(r)dB_1(r) + \aleph_{21}}{\int_0^1 B_2^2(r)dr},$$

using the obvious notation, so that $\hat{\beta}$ is a super consistent estimator of β which suffers from mean and median biases. Moreover, from Phillips and Park (1988) and Corollary 2, it follows that the limiting *OLS* distribution is no longer mixed normal. Result (5.9) was first shown by Stock (1987).

Likewise, when $d > 1$, Theorem 2, Corollary 1 and the *CMT* yield

$$(5.10) \quad T^d\left(\hat{\beta} - \beta\right) \Rightarrow \frac{\int_0^1 B_2(d,r)dB_1(r)}{\int_0^1 B_2^2(d,r)dr},$$

and thus, in this case, $\hat{\beta}$ is super consistent, median biased and (from Corollary 2) not mixed normal. Yet, notice that the drift term $\aleph_{21}$ is no longer present and therefore the mean bias disappears.

On the other hand, Theorem 2, Corollary 1 and the *CMT* taken together imply that when $\frac{1}{2} < d < 1$, then

$$(5.11) \quad T^{2d-1}\left(\hat{\beta} - \beta\right) \Rightarrow \frac{\aleph_{21}(d)}{\int_0^1 B_2^2(d,r)dr},$$

so that $\hat{\beta}$ is mean biased and has a limiting distribution that it is not a mixture of normals. Further, in this case $\hat{\beta}$ is super consistent only if $d > \frac{3}{4}$. Otherwise, the *OLS* estimator converges to the theoretical counterpart at a rate slower than the standard $T^{1/2}$ rate.

6. CONCLUDING REMARKS

In this paper we have been concerned with the asymptotics of *NFI* multivariate processes. We have provided conditions to establish both a multivariate invariance principle and weak convergence of stochastic integrals for *NFI* processes under the assumption that the innovations are $I(0)$ linear processes. Several applications of these results are given. In particular, we extend SSW analysis on estimation and hypothesis testing in vector autoregressions with integrated processes and deterministic components to the more general fractional framework. We show how their main conclusions remain valid when dealing with *NFI* processes. That is, 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 nonstandard limiting distributions.

As in the integer case, $d = 1, 2, \ldots$, the statistical procedures analyzed here require at least partial knowledge of which variables cointegrate and of the memory parameters of the individual series. In the fractional case, estimation of the memory parameters $d_1, d_2, \ldots, d_n$ and testing for cointegration can be based on Robinson's (1995a,b) results. See Robinson and Marinucci (1998) for some empirical applications.

On the other hand, we have assumed in this paper that all of the fractionally integrated processes considered in a vector autoregression are nonstationary with $I(0)$ innovations. This is clearly a limitation of our study. For instance, when searching for the correct lag length in model (5.4), if we define the integer number $q = 1, 2, \ldots$, so that $d = q + \delta$ with $|\delta| < \frac{1}{2}$, then, it is not difficult to prove that after q iterations, one can rewrite the restrictions $\Phi_{p+1} = \Phi_{p+2} = \cdots = \Phi_{p+s} = 0$ in (5.4) in terms of the coefficients of zero mean $I(\delta)$

regressors in a transformed model containing a constant term. The relevant question in this case is whether the corresponding test statistic has a standard limiting distribution as in the $I(0)$ case.

The answer to such a question, in turn, amounts to deriving the weak convergence of sample matrices of the form $\sum_{t=1}^{T} x_{t-1} z_t'$ with z_t denoting a multivariate stationary fractionally integrated process. This is not a straightforward problem, however, since the corresponding limiting stochastic integral cannot be defined as an Ito integral with respect to $\widetilde{W}(d,r)$ as in Section 3, because fractional Brownian motion is not a semimartingale. In this sense, Chan and Terrin (1995) have proposed a definition of stochastic integration with respect to a fractional Brownian motion in terms of the the so-called harmonizable representation of a *Type I* fractional Brownian motion (*cf.*, Samorodnitsky and Taqqu, 1994), proving the weak convergence of $\sum_{t=1}^{T} x_{t-1} z_t'$ to their specific definition and imposing rather strong assumptions on the range of possible values of the memory parameters of x_t and z_t. See also Marinucci (1998, Section 3). Therefore, we consider the weak convergence of $\sum_{t=1}^{T} x_{t-1} z_t'$ to be still an open question in need of further research.

APPENDIX: MATHEMATICAL PROOFS

PROOF OF THEOREM 1. The process $\xi_{[Tr]}(d), r \in [0,1]$ have associated sample paths which are elements of $D[0,1]^n = D[0,1] \times \cdots \times D[0,1]$, the n-dimensional product metric space of all real valued vector functions on $[0,1]$ that are right continuous at each element of $[0,1]$ and possess finite left limits. Endow each component space $D[0,1]$ with the Skorokhod topology (see Billingsley, 1968, chapter 3). The weak convergence appearing in Theorem 1 is associated with this particular topology.

The proof of the theorem now will be based on Prohorov's Theorem (*cf.*, Billingsley, 1968, pages 35-40), so that, in order to prove weak convergence, we need to verify if the family of joint probability measures on the product space $D[0,1]^N$ are tight as well as the convergence of its finite dimensional distributions of $T^{1/2-d} \xi_{[Tr]}(d)$. However, since $D[0,1]^n$ is separable and complete under the Skorokhod metric, tightness will follow if and only if the marginal probability measures of the components spaces are tight.

Consider now the arbitrary linear combination $\chi_T(r) = \wp' \left(T^{1/2-d} \xi_{[Tr]}(d) \right)$, for arbitrary $\wp \in \Re^n$, $\wp' \wp = 1$. Under Assumption A and with $C[0,1]$ denoting the space of continuous functions defined on the unit interval, weak convergence of this sequence to a univariate Holmgren-Riemann-Lioville fractional integral on $C[0,1]$, say $Z(d,r)$, can be proved as in Silveira (1991). Hence, writing $Z(d,r)=\wp' W(d,r)$, we deduce that $\chi_T(r)\Rightarrow \wp' W(d,r)$ for arbitrary $\wp$. Consequently, by the Cramèr-Wold device, it follows that the finite dimensional distributions of $T^{1/2-d}\xi_{[Tr]}(d)$ converge weakly to those of the vector process $W(d,r)$ and expression (2.3) of the theorem is proved.

On the other hand, expression (2.4) follows in a direct manner from the *CMT* and by noticing that $\varepsilon_t=\Sigma^{1/2}\eta_t$ so that $\tilde{\xi}_{[Tr]}(d)=\Delta^{-d}\varepsilon_{[Tr]}=\Delta^{-d}\Sigma^{1/2}\eta_{[Tr]}=\Sigma^{1/2}\xi_{[Tr]}(d)$. ■

PROOF OF THEOREM 2. Since $\Psi(L)=\sum_{j=0}^{\infty}\Psi_j L^j$ is 1-summable, then, it follows from Lemma 2.1 in Phillips and Solo (1992) that $\Psi(L)=\Psi(1)-(1-L)\hat{\Psi}(L)$, where $\hat{\Psi}(L)=\sum_{i=0}^{\infty}\hat{\Psi}_i L^i$, $\hat{\Psi}_i=\sum_{i=j+1}^{\infty}\Psi_i$, $\sum_{i=0}^{\infty}|\hat{\Psi}_i|<\infty$ and $\Psi(1)<\infty$. Consequently,

$$T^{1/2-d}\tilde{\xi}_{[Tr]}(d)=T^{1/2-d}\Delta^{-d}u_{[Tr]}=T^{1/2-d}\Delta^{-d}\left\{\Psi(1)-\Delta\hat{\Psi}(L)\right\}\eta_{[Rr]}$$
$$=\Psi(1)T^{1/2-d}\xi_{[Tr]}(d)-\hat{\Psi}(L)T^{1/2-d}\xi_{[Tr]}(d-1),$$

and, since using Theorem 1 and the *CMT* we know that

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

in order to prove the theorem, we only need to demonstrate that

$$\hat{\Psi}(L)T^{1/2-d}\xi_{[Tr]}(d-1)=o_p(1).$$

Now, if $d>\frac{3}{2}$, then $d-1>\frac{1}{2}$, and $\xi_t(d-1)$ is a *NFI* process for which Theorem 1 applies so that $T^{3/2-d}\xi_{[Tr]}(d-1)\Rightarrow W(d-1,r)$. This fact, together with the absolute summability of $\hat{\Psi}(L)$ implies that

$$\hat{\Psi}(L)T^{1/2-d}\xi_{[Tr]}(d-1)=T^{-1}\left\{\hat{\Psi}(L)T^{3/2-d}\xi_{[Tr]}(d-1)\right\}=T^{-1}O_p(1)=O_p(T^{-1})=o_p(1).$$

Notice that, in order to derive this result, we only need Assumption A to hold, so that when $d>\frac{3}{2}$ we only require that $max_i \sup_t E|\eta_{it}|^{2+\zeta}<\infty$, $\zeta>0$ for expression (2.6) to hold.

On the other hand, when $\frac{1}{2}<d<\frac{3}{2}$, i.e., when $-\frac{1}{2}<d-1<\frac{3}{2}$, $\xi_t(d-1)$ is a stationary fractionally integrated (denoted *SFI*) process. Letting $\delta=d-1$ so that $|\delta|<\frac{1}{2}$ and given the absolute summability of $\hat{\Psi}(L)$, the theorem will be proved if we show that $T^{-1/2-\delta}\xi_{[Tr]}(d-1)=o_p(1)$. For this, we have that for arbitrary $\lambda>0$ and for $i=1,2,\ldots,N$,

$$P\left\{\sup_{r\in[0,1]}\left|T^{-1/2-\delta}\xi_{i,[Tr]}(d-1)\right|>\lambda\right\}\leq TP\left\{\left|T^{-1/2-\delta}\xi_{i,t}(d-1)\right|>\lambda\right\}$$

$$\leq T^{1-1/2g-g\delta}\frac{E\left|\xi_{i,t}(d-1)\right|^{g}}{\lambda^{g}},\quad g>0\text{, by Markov's inequality,}$$

$$=\frac{1}{\lambda^{g}}T^{-1}T^{2-(1/2+\delta)g}E\left|\xi_{i,t}(d-1)\right|^{g}.$$

Hence, when $-\frac{1}{4}\leq\delta<\frac{1}{2}$, i.e., when $\frac{3}{4}\leq d<\frac{3}{2}$, if $g=4$, then

$$P\left\{\sup_{r\in[0,1]}\left|T^{-1/2-\delta}\xi_{i,[Tr]}(d-1)\right|>\lambda\right\}\leq\frac{1}{\lambda^{4}}T^{-1-4\delta}E\left|\xi_{i,t}(d-1)\right|^{4}\xrightarrow[T\to\infty]{}0.$$

However, when $-\frac{1}{2}<\delta<-\frac{1}{4}$, i.e., when $\frac{1}{2}<d<\frac{3}{4}$, we need the existence of the moments of $\xi_{i,t}(d-1)$ at least of order $-8\delta(1+2\delta)^{-1}$, since in this case,

$$1-\frac{g}{2}-g\delta=1+\frac{4\delta}{1+2\delta}+\frac{8\delta^{2}}{1+2\delta}=8\delta^{2}+6\delta+1,$$

which is a quadratic equation with roots $\delta_1=-0.25$ and $\delta_2=-0.5$. Hence, if $g>-8\delta(1+2\delta)^{-1}$ in the interval $-\frac{1}{2}<\delta<-\frac{1}{4}$, then

$$P\left\{\sup_{r\in[0,1]}\left|T^{-1/2-\delta}\xi_{i,[Tr]}(d-1)\right|>\lambda\right\}\xrightarrow[T\to\infty]{}0,$$

as required. Finally, since $\xi_{i,t}(d-1)$ is a stationary fractionally integrated process of order $d-1$,

$$\xi_{i,t}(d-1)=\sum_{j=0}^{\infty}c_{j}(d-1)\eta_{i,t-j},$$

with $c_j(d-1)\approx\Gamma(d-1)^{-1}j^{d-2}$, the existence of $E\left|\xi_{i,t}(d-1)\right|^{g}$ is guaranteed by the existence of the corresponding moments of $E\left|\eta_{i,t}\right|^{g}$. ■

PROOF OF THEOREM 3. Since $\{\eta_s\}_{-\infty}^{0}=0$, then

$$\xi_{i,t}(d)=\sum_{k=0}^{\infty}c_{k}(d)\eta_{i,t-k}=\sum_{k=0}^{t-1}c_{k}(d)\eta_{i,t-k}=\sum_{k=1}^{t}c_{t-k}(d)\eta_{i,k},\quad i,j=1,\ldots,n,$$

with $c_k(d)\approx k^{d-1}/\Gamma(d)$. Let $U_T(r)\equiv\sum_{k=1}^{[Tr]}c_{[Tr]-k}(d)\eta_{i,k}$ and $V_T(r)=\sum_{k=1}^{[Tr]}\eta_{j,k}$. Notice that, since $W_i(d,r)$, $W_j(r)\in C[0,1]$ a.s., the convergence in the Skorokhod topology is equivalent (*cf.*, Billingsley, 1968, page 112) to uniform convergence, where $W_i(d,r)$ and $W_j(r)$ stand for the *i*th and *j*th components of the *n*-dimensional vectors $W(d,r)$ and $W(r)$, respectively, $i, j = 1,\ldots,n$.. Thus, by the Skorokhod representation theorem, there are a probability space Θ and random elements $U^T, V^T \in D[0,1]$ such that

(A.1) $\left\| \left\{ U^T, V^T \right\} - \left\{ W_i^d, W_j \right\} \right\|_\infty \to 0$ a.s.

and

(A.2) $\left\{ U^T, V^T \right\} =_\ell \left\{ T^{1/2-d} U_T, T^{-1/2} V_T \right\}$,

where $=_\ell$ denotes equality in distribution and $\|\cdot\|_\infty$ the essential sup norm. Moreover, by (A1) and Egoroff's theorem, given $\varepsilon > 0$, there is an event $\Theta_\varepsilon \subset \Theta$ such that $P(\Theta_\varepsilon) \geq 1 - \varepsilon$ and $\sup\left[\left\| \left\{ U^T(\omega), V^T(\omega) \right\} - \left\{ W_i^d(\omega), W_j(\omega) \right\} \right\|_\infty : \omega \in \Theta_\varepsilon \right] = \tau_T \to 0$, where the sequence of constants τ_T is the uniform distance between $\left\{ U^T, V^T \right\}$ and $\left\{ U_T, V_T \right\}$ except on a set of arbitrarily small probability. Therefore, under Assumption A Theorem 1 holds, yielding

$$\left\{ T^{1/2-d} U_T, T^{-1/2} V_T \right\} \Rightarrow \left\{ W_i(d,r), W_j(r) \right\},$$

Consequently, since η_t is a square integrable martingale difference, expressions (3.1) and (3.2) follow by Theorem 2.1 in Hansen (1992), using a similar argument to that leading to Theorem 1 (see also Chan and Wei, 1988, Remark in page 377).

As regards expressions (3.3)-(3.5), following Hall and Heyde (1980), define $\gamma_t = \sum_{k=0}^{\infty} \left(E_t u_{t+k} - E_{t-1} u_{t+k} \right)$ and $\upsilon_t = \sum_{k=1}^{\infty} E_t u_{t+k}$, so that $u_t = \gamma_t + \upsilon_{t-1} - \upsilon_t$ and $E_{t-1}\gamma_t = 0$, where $E_t Y$ stands for $E(Y|\Im_t)$ for any random process *Y*. Notice that $\left\{ \gamma_t, \Im_t \right\}$ is a mds, where γ_t is ergodic, stationary and square integrable with covariance matrix $\Omega = C(1)\Sigma C(1)'$, and υ_t is also ergodic, stationary and square integrable.

Thus,

(A.3) $T^{-d} \sum_{t=1}^{T} \hat{\xi}_{t-1}(d) u_t' = T^{-d} \sum_{t=1}^{T} \hat{\xi}_{t-1}(d) \gamma_t' + \Lambda_T$,

with

(A.4) $\Lambda_T = T^{-d} \sum_{t=1}^{T} \hat{\xi}_{t-1}(d) \upsilon_{t-1}' - T^{-d} \sum_{t=1}^{T} \hat{\xi}_{t-1}(d) \upsilon_t'$.

Now, given that γ_t is a mds, it follows from (3.2) that, for all $d \in \partial$,

(A.5) $T^{-d} \sum_{t=1}^{T} \hat{\xi}_{t-1}(d) \gamma_t' \Rightarrow \int_0^1 B(d,r) dB(r)'$.

Therefore, to prove expressions (3.3)-(3.5), it only remains to show the convergence of Λ_T. To do this, notice that Λ_T can also be rewritten as

(A.6) $$\Lambda_T = T^{-d}\sum_{t=1}^{T}\left(\widetilde{\xi}_{t-1}(d)-\widetilde{\xi}_{t-2}(d)\right)\upsilon_{t-1}^{'} - T^{-d}\widetilde{\xi}_{t-1}(d)\upsilon_t^{'} .$$

Now, using previous results, it follows that for all $d \in \partial$,

(A.7) $$T^{-d}\widetilde{\xi}_{t-1}(d)\upsilon_t^{'} = \left(T^{1/2-d}\widetilde{\xi}_{t-1}(d)\right)\left(T^{-1/2}\upsilon_t^{'}\right) = O_p(1)o_p(1) = o_p(1).$$

Hence,

(A.8) $$\Lambda_T = T^{-d}\sum_{t=1}^{T}\left(\widetilde{\xi}_{t-1}(d)-\widetilde{\xi}_{t-2}(d)\right)\upsilon_{t-1}^{'} + o_p(1).$$

When $d = 1$, Phillips (1988) showed that

(A.9) $$\Lambda_T \xrightarrow{p} \sum_{k=1}^{\infty} E\left(u_0 u_k^{'}\right).$$

Moreover, in the case where $d = 2$, Hansen (1992) showed that $\Lambda_T \xrightarrow{p} 0$. Consider now the case where $d > \frac{3}{2}$. In this case, noticing that $d = j + \delta$, $j = 2,3,\ldots$, we have that $\Delta\widetilde{\xi}_t(d)$ is a *NFI* process for which

$$\Lambda_T = T^{-d}\sum_{t=1}^{T}\Delta\widetilde{\xi}_{t-1}(d)\upsilon_{t-1}^{'} + o_p(1) = T^{-1}\left(T^{1-d}\sum_{t=1}^{T}\Delta\widetilde{\xi}_{t-1}(d)\upsilon_{t-1}^{'}\right) + o_p(1) = o_p(1),$$

and then, $\Lambda_T \xrightarrow{p} 0$. On the other hand, when $\frac{1}{2} < d < \frac{3}{2}$, using the decomposition $d = 1 + \delta$, we have that $\Delta^d x_t = u_t \Leftrightarrow \Delta x_t = \Xi_t, \Delta^\delta \Xi_t = u_t$, so that Ξ_t is a *SFI* process of order δ, and then, $\widetilde{\xi}_t(d) - \widetilde{\xi}_{t-1}(d) = \Xi_t$. Therefore,

(A.10) $$\Lambda_T = T^{-d}\sum_{t=1}^{T}\Xi_{t-1}\upsilon_{t-1}^{'} + o_p(1),$$

so that, in the case where $1 < d < \frac{3}{2}$, using the Cauchy-Schwartz inequality it is straightforward to show that

(A.11) $$\Lambda_T = T^{-\delta}\left(T^{-1}\sum_{t=1}^{T}\Xi_{t-1}\upsilon_{t-1}^{'}\right) = o_p(1).$$

Finally, when $\frac{1}{2} < d < 1$, i.e., when $-\frac{1}{2} < \delta < 0$, from the weak law of large numbers we have

(A.12) $$T^{-1}\sum_{t=1}^{T}\Xi_{t-1}\upsilon_{t-1}^{'} \xrightarrow{p} E\left(\Xi_{t-1}\upsilon_{t-1}^{'}\right) = E\left(\Xi_{t-1}\sum_{k=1}^{\infty}E_{t-1}u_{t-1+k}^{'}\right) = \sum_{k=1}^{\infty}E\left(\Xi_0 u_k^{'}\right),$$

that is well-defined since Ξ_t and u_t are both linear processes with absolutely summable coefficients.

To prove the last statement, let us focus for simplicity the attention on the univariate case and assume that the stationary process Ξ_t has nonzero initial conditions. Then, we obtain

$$\sum_{k=1}^{\infty} E(\Xi_0 u_k) = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} c_j(\delta)\gamma_u(k+j) = \sum_{j=0}^{\infty} c_j(\delta)\sum_{k=1}^{\infty}\gamma_u(k+j) = \sum_{j=0}^{\infty} c_j(\delta) z_j ,$$

where $\gamma_u(l)$ denotes the autocovariance of u_t at lag l, and where the interchange of the summation operators is justified by the absolute summability of the $\{c_l(\delta)\}$ and $\{\gamma_u(l)\}$ sequences. Therefore, since $z_j = \text{var}(\eta_1)\sum_{k=1}^{\infty}\sum_{i=0}^{\infty}\Psi_i\Psi_{i+k+j}$, it follows that

$$\sum_{j=0}^{\infty}\left|c_j(\delta)z_j\right| \le M\left(\sum_{m=0}^{\infty}\left|\Psi_m\right|\right)^2 \sum_{j=0}^{\infty}\left|c_j(\delta)\right|$$

(M being a constant term)

$$= M\left(\sum_{m=0}^{\infty}\left|\Psi_m\right|\right)^2 \left\{1+\sum_{j=1}^{N-1}\left|c_j(\delta)\right| + \sum_{j=N}^{\infty}\left|c_j(\delta)\right|\right\}$$

(N being a sufficiently large real number)

$$\approx M\left(\sum_{m=0}^{\infty}\left|\Psi_m\right|\right)^2 \left\{1+\sum_{j=1}^{N-1}\left|c_j(\delta)\right| + \int_N^{\infty} x^{\delta-1}dx\right\}$$

(using Sheppard's lemma). Since $\int_N^{\infty} x^{\delta-1}dx$ converges, it follows from the Cauchy integral criterion that $\sum_{j=0}^{\infty}\left|c_j(\delta)\right| < \infty$, and hence, $\sum_{j=0}^{\infty}\left|c_j(\delta)z_j\right| < \infty$. Lastly, it is well known that the binomial coefficients $\{c_j(\delta)\}_{j=0}^{\infty}$ sum up to zero (see, e.g., Knopp, 1964, page 440). Since those coefficients are all of them negative for $j = 1, 2, \ldots,$ it follows that $\sum_{j=0}^{\infty}\left|c_j(\delta)\right| \neq 0$. ■

PROOF OF COROLLARY 1. Straightforward using the results obtained in Theorem 3 jointly with the *CMT* and the fact that

$$\sum_{t=1}^{T}\hat{\xi}_t(d)u_t' = \sum_{t=1}^{T}\hat{\xi}_{t-1}(d)u_t' + \sum_{t=1}^{T}\left[\Delta\hat{\xi}_{t-1}(d)\right]u_t' . \quad ■$$

PROOF OF COROLLARY 2. When $\Omega_{21} = \Lambda_{21} = 0$ it turns out that the Gaussian vector processes $B_1(r)$ and $B_2(r)$ are independent of each other, and then so are $B_1(r)$ and

$B_2(d,r)$ since $B_2(d,r)$ is a uniformly continuous functional of $B_2(r)$. Therefore, expressions (3.3), (3.4), (3.8) and (3.9) become equal to (3.11) and (3.12) for $d \geq 1, d \in \partial$.

On the other hand, when $\frac{1}{2} < d < 1$, notice that $\Lambda_{21} = 0$ implies $\Lambda_{21}(d) = 0$ (uniformly in *d*). Consequently, expression (A.25) becomes

$$\text{(A.26)} \quad T^{-1}\sum_{t=1}^{T} \Xi_{t-1} \upsilon_{t-1}^{'} \xrightarrow{p} \sum_{k=1}^{\infty} E\left(\Xi_0 u_k^{'}\right) = 0,$$

and in this case a standard *CLT* for linear processes with square summable coefficients can be invoked, yielding

$$\text{(A.27)} \quad T^{-d}\sum_{t=1}^{T} \hat{\xi}_{2,t-1}(d) u_{1t}^{'} = T^{-d}\sum_{t=1}^{T} \hat{\xi}_{2t-1}(d) \gamma_{1t}^{'} + \Lambda_{21,T} = T^{-d}\sum_{t=1}^{T} \hat{\xi}_{2t-1}(d) \gamma_{1t}^{'} + o_p(1),$$

and expressions (3.11) and (3.12) now follow, when $\frac{1}{2} < d < 1$, by the use of (A.18).

Finally, given that $B_2(d,r)$ is a uniformly continuous functional of $B_2(r)$, it follows from Park and Phillips (1988, Lemma 2.1) that

$$\text{(A.28)} \quad \left. \int_0^1 B_2(d,r) dB_1(r)' \right|_{\sigma(B_2(r))} \equiv N\left\{0, \Omega_{11} \otimes \int_0^1 B_2(d,r) B_2(d,r)' dr \right\},$$

where $\cdot|_{\sigma(B_2(r))}$ denotes the conditional distribution with respect to the σ-algebra generated by $B_2(r)$. ■

PROOF OF LEMMA 1. Proofs of equations (4.2) and (4.3) follow in a straightforward manner by using Theorem 1 and the *CMT*, while expressions (4.4)-(4.8) are proved by SSW, Lemma 1 and Stock and Watson (1993), Lemma A.2. In order to prove the remaining cases, let us notice the 1-summability of *F(L)*, which allows us to write

$$\text{(A.29)} \quad \sum_{t=1}^{T} \xi_t(d) \left(F(L)\eta_t\right)' = \sum_{t=1}^{T} \xi_t(d) \eta_t^{'} F(1)' + \sum_{t=1}^{T} \xi_t(d) \left(F^*(L)\Delta\eta_t\right)',$$

where $F^*(L)$ is absolutely summable with coefficients $F_j^* = -\sum_{I=J+1}^{\infty} F_j$ (*c.f.*, Phillips and Solo, 1992, Lemma 2.1).

From (3.6) and the *CMT*, the first term in the right hand side of (A29) converges weakly to

$$\text{(A.30)} \quad T^{-d}\sum_{t=1}^{T} \xi_t(d) \eta_t^{'} F(1)' \Rightarrow \int_0^1 W(d,r) dW(r)' F(1)'.$$

With respect to the second term in the right hand side of (A29), after some manipulation, one obtains

(A.31) $\sum_{t=1}^{T} \xi_t(d)\left(F^*(L)\Delta\eta_t\right)' = \sum_{t=1}^{T} \xi_t(d-1)\left(F^*(L)\eta_T\right)' - \sum_{t=1}^{T} \xi_t(d-1)\left(F^*(L)\eta_{t-1}\right)'$.

When $d > \frac{3}{2}$, then

(A.32) $T^{-d}\sum_{t=1}^{T} \xi_t(d-1)\left(F^*(L)\eta_T\right)' = \left[T^{-(d-1/25)}\sum_{t=1}^{T} \xi_t(d-1)\right]\left[T^{-1/2}F^*(L)\eta_T\right]'$

which vanishes by (4.2) and the absolute summability of the lag operator $F^*(L)$. On the other hand, using Markov's inequality and assuming for notational convenience that ξ_t and $F^*(L)\eta_{t-1}$ are scalars, yields

(A.33) $E\left|T^{-d}\sum_{t=1}^{T} \xi_t(d-1)\left(F^*(L)\eta_{t-1}\right)\right| \le T^{-3/2}\sum_{t=1}^{T} E\left(\left|T^{3/2-d}\xi_t(d-1)\right|\left|F^*(L)\eta_{t-1}\right|\right)$

$$\le T^{-3/2}\sum_{t=1}^{T}\left[E\left(T^{3/2-d}\xi_t(d-1)\right)^2\right]^{1/2}\left[E\left(F^*(L)\eta_{t-1}\right)^2\right]^{1/2}$$

$$\le T^{-1}\sum_{t=1}^{T}\left[E\left(T^{3/2-d}\xi_t(d-1)\right)^2\right]^{1/2} T^{-1/2}\left(\sum_{j=0}^{\infty}\left|F_j^*\right|\right)$$

$$\Rightarrow O_p\left(T^{-1/2}\right)\int_0^1\left[E\left(W(d-1,r)^2\right)\right]^{1/2} dr = o_p(1).$$

When $1 < d < \frac{3}{2}$, then $0 < d-1 = \delta < \frac{1}{2}$, expression (A.31) is now the sum of two $O_p(1)$ terms by the weak law of large numbers, implying that

(A.34) $T^{-d}\sum_{t=1}^{T} \xi_t(d)\left(F^*(L)\Delta\eta_t\right)' = O_p\left(T^{-\delta}\right) = o_p(1)$.

Now, by collecting expressions (A.29)-(A.34), we get

(A.35) $T^{-d}\sum_{t=1}^{T} \xi_t(d)\left(F(L)\eta_t\right)' \Rightarrow \int_0^1 W(d,r)dW(r)'F(1)'$

whenever $d > 1$.

Lastly, in the case where $\frac{1}{2} < d < 1$, first notice from (3.10) and Phillips and Solo's Lemma 2.1 that

(A.36) $T^{-1}\sum_{t=1}^{T}\left\{F(1) + \Delta F^*(L)\right\}\xi_t(d)\eta_t' F(L)' \xrightarrow{p} \aleph(d)$,

which together with the fact that $\sum_{t=1}^{T} F^*(L)\Delta\xi_t(d)\eta_t' F(L)' = O_p\left(T^{1/2}\right)$ imply

(A.37) $T^{-1}F(1)\sum_{t=1}^{T} \xi_t(d)\left(F(L)\eta_t\right)' \xrightarrow{p} \aleph(d)$. ■

PROOF OF LEMMA 2. The limits of the blocks $\Pi_{ii}, i = 1,2,4,5$, $\Pi_{1j}, j = 2,4,5$, Π_{24}, Π_{25}, Π_{45} and A_m, $m = 1,2,4,5$, follow from Lemma 2 in SSW, whereas the limits of the blocks Π_{i3}, $i = 1,2,3$, Π_{34}, Π_{35} and A_3 follow in a direct way from (4.13), Lemma 1 and the *CMT*. ■

PROOF OF THEOREM 4. Straightforward using Lemma 2, the *CMT* and the relationships $D'\hat{\gamma} = \hat{\beta}$ and $\Psi_T(\hat{\gamma} - \gamma) = \left\{\Psi_T^{-1}\sum_{t=1}^{T} Z_t Z_t' \Psi_T^{-1}\right\}^{-1}\left\{\Psi_T^{-1}\sum_{t=1}^{T} Z_t \eta_t' \omega\right\}$. ■

REFERENCES

[1] AKONOM, J. and C. GOURIEROUX (1988), "A Functional Central Limit Theorem for Fractional Processes", CEPREMAP Working Paper 8801.

[2] BAILLIE, R. (1996), "Long Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics* 73, 5-59.

[3] BILLINGSLEY, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

[4] CHAN, N.H. and W. PALMA (1998), "State Space Modeling of Long-Memory Processes", *Annals of Statistics* 26, 719-740.

[5] CHAN, N.H. and N. TERRIN (1995), "Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Root Autoregressions", *Annals of Statistics* 23, 1662-1683.

[6] CHAN, N.H. and C.Z. WEI (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics* 16, 367-401.

[7] DAVIDSON J. and R.M. de JONG (1999): "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals II: Fractionally Integrated Processes", Unpublished manuscript, Michigan State University.

[8] DAVYDOV, Y.A. (1970), "The Invariance Principle for Stationary Processes", *Theory of Probability and its Applications* 15, 487-498.

[9] DE JONG, R.M. and J. DAVIDSON (1999), "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes", Unpublished manuscript, Michigan State University.

[10] GRANGER, C.W.J. and R. JOYEUX (1980), "An Introduction to Long Memory Time Series Models and Fractional Differencing", *Journal of Time Series Analysis* 1, 15-39.

[11] HALL, P. and C.C. HEYDE (1980), *Martingale Limit Theory and its Applications*, New York: Academic Press.

[12] HANSEN, B. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes", *Econometric Theory* 8, 489-500.

[13] HEYDE, C.C. and Y. YANG (1997), "On Defining Long-Range Dependence", *Journal of Applied Probability* 34, 939-944.

[14] HOSKING, J.R.M. (1981), "Fractional differencing", *Biometrika* 68, 165-176.

[15] JEGANATHAN, P. (1999), "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors", forthcoming in *Econometric Theory.*

[16] KNOPP, K. (1964), *Theorie und Anwendung der unendlichen Reihen*, 5th edn., Berlin: Springer-Verlag.

[17] LÉVY, P. (1953), "Random Functions: General Theory with Special Reference to Laplacian Random Functions", *University of California Publications in Statistics* 1, 331-390.

[18] LIU, M. (1998), "Asymptotics of Nonstationary Fractional Integrated Series", *Econometric Theory* 14, 641-662.

[19] MANDELBROT, B.B. and J.W. VAN NESS (1968), "Fractional Brownian Motions, Fractional Noises and Applications", *SIAM Review* 10, 422-437.

[20] MARINUCCI, D. (1998), "Band Spectrum Regression for Cointegrated Time Series with Long Memory Innovations", *STICERD Discussion Paper* #353, LSE.

[21] MARINUCCI, D. and P.M. ROBINSON (1998): "Weak Convergence of Multivariate Fractional Processes", *STICERD Discussion Paper* #352, LSE.

[22] MARINUCCI, D. and P.M. ROBINSON (1999): "Alternative Forms of Fractional Brownian Motion", forthcoming in *Journal of Statistical Planning and Inference.*

[23] PARK, J.Y. and P.C.B. PHILLIPS (1988), "Statistical Inference in Regressions with Integrated Processes: Part I", *Econometric Theory* 4, 468-497.

[24] PARK, J.Y. and P.C.B. PHILLIPS (1989), "Statistical Inference in Regressions with Integrated Processes: Part II", *Econometric Theory* 5, 95-131.

[25] PHILLIPS, P.C.B. (1988), "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations", *Econometric Theory* 4, 528-533.

[26] PHILLIPS, PC.B. and S.N. DURLAUF (1986), "Multiple Time Series Regression with Integrated Processes", *Review of Economic Studies* 53, 473-496.

[27] PHILLIPS, P.C.B. and J.Y. PARK (1988), "Asymptotic Equivalence of OLS and GLS in Regression with Integrated Regressors", *Journal of the American Statistical Association* 83, 111-115.

[28] PHILLIPS, P.C.B. and V. SOLO (1992), "Asymptotics for Linear Processes", *Annals of Statistics* 20, 971-1001.

[29] ROBINSON, P.M. (1994a), "Time Series with Strong Dependence", in C.A. Sims (ed.), *Advances in Econometrics. Sixth World Congress*, vol. I, Cambridge: Cambridge University Press.

[30] ROBINSON, P.M. (1994b), "Efficient Tests of Nonstationary Hypothesis", *Journal of the American Statistical Association* 89, 1420-1437.

[31] ROBINSON, P.M. (1995a), "Log-Periodogram Regression of Time Series with Long Range Dependence", *Annals of Statistics* 23, 1048-1072.

[32] ROBINSON, P.M. (1995b), "Gaussian Semiparametric Estimation of Time Series with Long Range Dependence", *Annals of Statistics* 23, 1630-1661.

[33] ROBINSON, P.M. and D. MARINUCCI (1998), "Semiparametric Frequency Domain Analysis of Fractional Cointegration", *STICERD Discussion Paper* #348, LSE.

[34] SAMORODNISTSKY, G. and M. S. TAQQU (1994): *Stable Non-Gaussian Random Processes*, New York: Chapman and Hall.

[35] SILVERIA, G. (1991): *Contributions to Strong Approximations in Time Series with Applications in Nonparametric Statistics and Functional Central Limit Theorems*, Unpublished Ph.D. thesis, University of London.

[36] SIMS, C.A., STOCK, J.H. and M.W. WATSON (1990), "Inference in Linear Time Series with Some Unit Roots", *Econometrica* 58, 113-144.

[37] STOCK, J. H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica* 55, 1035-1056.

[38] STOCK, J.H. and M.W. WATSON (1993), "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems", *Econometrica* 61, 783-820.

[39] TAQQU, M.S. (1975), "Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process", *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 50, 53-83.

[40] TODA, H.Y. and P.C.B. PHILLIPS (1993a), "Vector Autoregressions and Causality", *Econometrica* 62, 1367-1394.

[41] TODA, H.Y. and P.C.B. PHILLIPS (1993b), "Vector Autoregressions and Causality. A Theoretical Overview and Simulation Study", *Econometric Reviews* 12, 321-364.

[42] WATSON, M.W. (1994), "Vector Autoregressions and Cointegration", in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, Amsterdam: Elsevier.