SEMIPARAMETRIC FREQUENCY DOMAIN ANALYSIS OF FRACTIONALLY INTEGRATED AND COINTEGRATED TIME SERIES

1998

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PH.D.) AT THE UNIVERSITY OF LONDON BY DOMENICO MARINUCCI, REGISTERED AT THE LONDON SCHOOL OF ECONOMICS

ABSTRACT

The concept of cointegration has principally been developed under the assumption that the raw data vector z_t is $I(1)$ and the cointegrating residual e_t is $I(0)$; we call this framework the $CI(1)$ case. The purpose of this thesis is to consider more general fractional circumstances, where z_t is stationary with long memory and e_t is stationary with less memory, or where z_t is nonstationary while e_t is less nonstationary or stationary, possibly with long memory. First we establish weak convergence to what we term "type II fractional Brownian motion" for a wide class of nonstationary fractionally integrated processes, then we go on to investigate the behaviour of the discretely averaged periodogram for processes that are not second order stationary. These results are exploited for the analysis of a procedure originally proposed by Robinson (1994a), which we call Frequency Domain Least Squares (FDLS). FDLS yield estimates of the cointegrating vector that are consistent for stationary and nonstationary z_t, asymptotically equivalent to OLS in some circumstances, and superior in many others, including the standard $CI(1)$ case; a semiparametric methodology for fractional cointegration analysis is applied to data sets on eleven US macroeconomic variables. Finally, we investigate an alternative definition of fractional cointegration, for which we introduce a continuously averaged version of FDLS, obtaining consistent estimates in both the stationary and the nonstationary case. Asymptotic distributions and Monte Carlo evidence on finite sample performance are also provided.

ACKNOWLEDGEMENTS

I am immensely grateful to my supervisor, Prof.P.M.Robinson, for the strong support during my Ph.D. programme, for the key contributions to several parts of the thesis, and for providing me an unrivalled example of rigour and research commitment.

I thank my friends and colleagues Carlos, Luis, Paolo, and especially Claudio, for many useful discussions, and Prof.F.H.Hahn for encouraging me to pursue graduate studies at the LSE.

A Luisa

CONTENTS

LIST OF TABLES

Chapter 1

COINTEGRATION AND FRACTIONAL INTEGRATION: AN OVERVIEW

1.1 INTRODUCTION

Cointegration analysis has developed as a major theme of time series econometrics since the article of Engle and Granger (1987), much applied interest prompting considerable methodological and theoretical development during the past decade. The reference framework can be summarized as follows: for a zero-mean, covariance stationary $p \times 1$ vector sequence $\eta_t = (\eta_{1t}, ..., \eta_{pt})'$, $t = 0, 1, ...$, introduce the spectral density matrix $f(\lambda) = \{f_{ab}(\lambda)\}$, $a, b = 1, ..., p$, which satisfies

$$\gamma_{ab}(\tau) = \int_{-\pi}^{\pi} f_{ab}(\lambda) e^{i\lambda\tau} d\lambda \ , \text{ for } \gamma_{ab}(\tau) = E\eta_{a0}\eta_{b\tau} \ . \tag{1.1}$$

We term the series η_{at} short range dependent or $I(0)$ if $0 < f_{aa}(0) < \infty$; in case all the components of η_t are short range dependent we can introduce the convenient notation $\Omega = 2\pi f(0)$. We define z_{at} (fractionally) integrated of order d, written $z_{at} \sim I(d_a)$, for d_a

real, if

$$z_{at} = (1-L)^{-d_a}\eta^*_{at}, \; \eta^*_{at} = \begin{cases} \eta_{at}, & t > 0 \\ 0, & t \leq 0 \end{cases}, \tag{1.2}$$

where L is the lag operator and formally

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j \;, \; \Gamma(\alpha) = \int_0^{\infty} s^{\alpha-1} e^{-s} ds \;. \tag{1.3}$$

The standard paradigm for cointegration analysis envisages a vector of economic variables z_t which are all $I(1)$; these variables are cointegrated if there exists a linear combination $e_t = \alpha' z_t$ which is $I(0)$, the prime denoting transposition. Hereafter we term this framework the $CI(1)$ case.

Several reasons motivate the interest aroused by cointegration analysis among theoretical and applied researchers. First, the idea of cointegration as a characterization of long run equilibria is appealing from the point of view of economic theory; numerous stationary relationships can be conjectured among nonstationary economic variables, such as stock prices and dividends, consumption and income, wages and prices, short- and long-run interest rates, monetary aggregates and nominal GNP, exchange rates and prices, GNP and public debt and many others. Second, it has been argued that the notion of equilibrium implicit in cointegrating relationships is more meaningful for a non-experimental science as economics than others, which typically find their roots in the natural sciences (cf. Banerjee et al..(1993)). Third, cointegration analysis has reconciled different methodologies for the analysis of time series data, i.e. the "statistical" $ARIMA$ approach pioneered by Box and Jenkins (1976), who advocated preliminary differencing of the raw data in order to make classical inference procedure applicable, and the "econometric" simultaneous equation approach typically applied to levels to mantain information on long run equilibrium relationships. Granger's (1987) representation theorem proves that deviations from long-run relationships can be introduced as explanatory variables for short run dynamics (modelled on differenced data), so that attitudes from both methodologies are included by default in cointegration analysis. Fourth, empirical studies starting from the seminal paper by Nelson and Plosser (1982) have highlighted the practical relevance of stochastic trends and unit roots nonstationarity in very many observed time series. Finally, a number of papers mainly by Phillips and his coauthors (Phillips and Durlauf (1986), Park and Phillips (1988), Phillips and Hansen (1990), Phillips (1991a,b)) demonstrated that statistical inference procedures can be developed for $I(1)$ processes, and indeed under appropriate assumptions have even more attractive properties than in the stationary framework.

The considerable theoretical and empirical success of the cointegration paradigm motivates an investigation into its possible generalizations and/or its robustness against departures from the $CI(1)$ case. The $I(1)$ and $I(0)$ classes are in fact highly specialized forms of, respectively, nonstationary and stationary processes when nested in the $I(d)$ class (1.2), for real-valued d. A definition of cointegration for $I(d)$ variables is as follows:

Definition 1.1 For a $p \times 1$ vector z_t whose a-th element $z_{at} \sim I(d_a)$, $d_a > 0$, $a = 1, ..., p$, we say $z_t \sim FCI(d_1, d_2, ..., d_p; d_e))$ if there exists a $p \times 1$ non-null vector α (called the *cointegrating vector*) such that $\alpha' z_t = e_t \sim I(d_e)$, where $\min_{1 \leq a \leq p} d_a > d_e \geq 0$.

Although the possibility of fractional (i.e. non-integer) d_a or d_e was already mentioned in Engle and Granger (1987), that paper and the bulk of the subsequent theoretical and empirical literature on cointegration focused on the $CI(1)$ case we introduced earlier. There are, however, several arguments from both economics and statistics to support the investigation of the more general circumstances described in Definition 1.1. Unit roots in multivariate time series can sometimes be viewed as a consequence of economic theory; however, as argued by Sims (1988), even these cases rest on very strong hypotheses, or provide only asymptotic justification. The power of unit roots tests has often been criticised, and the bulk of these have been directed against $I(0)$ alternatives, and data that seems consistent with the $I(1)$ hypothesis might well also be consistent with $I(d)$

processes, at least for a certain interval of d values others than unity. Many different procedures have been proposed for the estimation of the parameters governing the cointegrating relationships; these methods are all specifically designed for the $CI(1)$ case, and they typically lack any sound statistical basis once one departs from the unit root assumption. In the $CI(1)$ case, cointegration is typically a true/false proposition, and researchers cannot qualify in any quantitative way the "strength" of such equilibrium relationships, nor indeed assess the speed of convergence to long run equilibria: both notions have however a clear theoretical and empirical relevance.

The purpose of this thesis is to analyze the generalization of the notion of cointegration beyond the unit root framework. We shall consider multivariate fractionally integrated processes which satisfy Definition 1.1 without restricting d_e and d_a to be integers, and in this sense are fractionally cointegrated. This allows for many possibilities not covered by the $CI(1)$ paradigm: z_t can be stationary with long memory and e_t stationary with less memory, or z_t can be nonstationary and e_t less nonstationary or stationary, possibly with long memory; the value of d_e can be used as a measure of the amount of memory left in the cointegrating residual, which can be functionally related to the rate of convergence to the long run equilibrium (Diebold and Rudebusch (1989)); the value of $d_a - d_e$ can be linked to the strength of the cointegrating relationship and to the asymptotic behaviour of feasible estimates for the cointegrating relationship α; more generally, cointegration can be reconsidered in a broader sense than usual as a relationship that yields residuals which, while not necessarily short range dependent nor stationary, have less memory than the original series. This enlarged framework also calls for new probabilistic foundations and statistical inference procedures.

The plan of the thesis is as follows. The remaining part of this chapter reviews rapidly part of the econometric and statistical literature on cointegrated time series and fractionally integrated processes. The contributions in these areas has been so vast in the last decade that we shall make no attempt to be complete or representative. In Section 2 we shall concentrate on the asymptotics for some estimates of the parameters that govern the cointegrating relationships; we favour those papers that adopt a semiparametric or nonparametric specification for the disturbances, and the procedures that can be granted a frequency-domain interpretation, even when this perspective is not adopted in the original contributions: this motivated some major exclusions, such as the highly popular Johansen's (1988,1991) Full Information Maximum Likelihood procedure which is deeply rooted within a time-domain, parametric autoregressive specification. In Section 3 we shall be concerned with the asymptotic behaviour of parametric and semiparametric estimates of the "memory" parameter d, while a more rapid consideration is given to the asymptotic behaviour of quadratic forms in variables having long range dependence and to regression with fractionally integrated regressors and/or residuals. Chapter 2 discusses two alternative definitions of fractional Brownian motion, and shows how these definitions have occasionally led to some confusion in the econometric literature. We then generalize and correct unpublished results of Akonom and Gourieroux (1987) and Silveira (1991) to demonstrate weak convergence of a class of nonstationary fractionally integrated processes to what we term "type II fractional Brownian motion". Chapter 3 investigates the behaviour of a well-known frequency-domain statistic, the discretely averaged periodogram, in somewhat unfamiliar circumstances, namely for processes that are not second order stationary, and hence have not a proper spectral density: here the main concern is on approximation of sample moments by narrow-band averages. These results are then exploited in Chapter 4, where, as an estimate of the cointegrating vector α, we analyze for the stationary and the nonstationary case a procedure originally considered in Robinson (1994a), which we call Frequency Domain Least Squares (FDLS). This procedure is shown to yield consistent estimates for any positive value of d_a, to be asymptotically equivalent to OLS in some circumstances, and superior in many others, including the standard $CI(1)$ case. Monte Carlo evidence of finite sample performance is also provided, while a semiparametric methodology for fractional cointegration analysis is applied to data considered by Engle and Granger (1987) and Campbell and Shiller (1987). Finally, in Chapter 5 we explore a definition of fractional cointegration which cannot be nested into (1.2)/(1.3); for this framework we propose a semiparametric procedure, denoted Weighted Covariance Estimate (WCE), which is not equivalent to FDLS. Consistency is shown to hold under both stationary and nonstationary circumstances; the asymptotic distribution is derived in the nonstationary case building upon a contribution by Chan and Terrin (1995). Encouraging Monte Carlo evidence of finite sample performance is also included.

Throughout this thesis we shall adopt the following notation: $(A : B)$ will indicate the $(p_1 + p_2) \times q$ matrix obtained by stacking the $p_1 \times q$ matrix A over the $p_2 \times q$ matrix B, $|A|$, $rank(A)$ and $tr(A)$ respectively the determinant, the rank and the trace of the (square) matrix A, $\|.\|$ the Euclidean norm of a matrix such that $\|A\| = \sqrt{tr(A'A)}$, I_p the p-rowed identity matrix; we write $A > 0$ to signify that A is positive definite. "[.]" will signify the integer part of a real number, "$\otimes$" will indicate the Kronecker product, "vec" the vec operator stacking the colums of a matrix one over the other; " $\equiv$ " will indicate equality in distribution and " $\triangleq$ " equality by definition; C will stand for a generic, finite and positive constant, which need not be the same all time it is used.

1.2 ASYMPTOTICS FOR COINTEGRATED TIME SERIES

We define Brownian motion $B(r; \sigma^2)$, $r \in R$, $\sigma^2 > 0$, to be a real-valued Gaussian process with independent increments, $B(0; \sigma^2) = 0$, a.s., and

$$\begin{aligned} EB(r;\sigma^2) &= 0, r \in R, && (1.4) \\ EB(r_1;\sigma^2)B(r_2;\sigma^2) &= \sigma^2 \min(r_1, r_2),\ r_1, r_2 \geq 0\ . && (1.5) \end{aligned}$$

Functional central limit theorems (weak invariance principles) entail weak convergence of random variables to Brownian motion in a suitable metric space; authoritative references are Billingsley (1968) and Pollard (1984). Denote by $(\mathcal{X}, \partial)$ a complete separable metric space with metric ∂ and probability measures μ_i, $i \geq 0$, on the Borel sets of $\mathcal{X}$. We say that μ_n converges weakly to μ_0 in $(\mathcal{X}, \partial)$ if for every bounded continuous function h in $\mathcal{X}$, $\lim_{n\to\infty} \int h d\mu_n = \int h d\mu_0$. We can construct a probability space $\{\mathcal{S}, \Im, P\}$ with random elements ξ_n, $n \geq 0$, of $\mathcal{X}$ having distributions μ_n respectively. If μ_n converges weakly to μ_0 then we write $\xi_n \Rightarrow \xi_0$. Two metric spaces which feature considerably in the theory of weak convergence are $C[0,1]$, the space of continuous functions on $[0,1]$ endowed with the uniform topology, and $D[0,1]$, the space of functions on $[0,1]$ that are continuous on the right with finite left limits, endowed with the Skorohod (1956) J_1 topology.

Now consider the process z_t defined by (1.2), where the a-th component of z_t satisfies $z_{at} \sim I(1)$, $a = 1, 2, ..., p$, i.e.

$$z_0 = \underline{0} \ , \tag{1.6}$$

$$z_t = \eta_1 + ... + \eta_t, \quad t \geq 1 \ , \tag{1.7}$$

for $\underline{0}= (0, ..., 0)'$. Introduce the vector random elements

$$z_n(r) = n^{-1/2} z_{[nr]} \ ; \tag{1.8}$$

we have

$$z_n(r) \in D[0,1]^p = D[0,1] \times ... \times D[0,1] \ , \tag{1.9}$$

the space of R^p-valued vector functions whose components are continuous on the right and with finite left limit. For the product space, Phillips and Durlauf (1986) suggest the metric

$$d^p(g(.), h(.)) = \max_{a=1,...,p} \{d(g_a(.), h_a(.))\} \ , \ g(.), h(.) \in D[0,1]^p \ , \tag{1.10}$$

with $g_a(.)$ and $h_a(.)$ the a-th components of g and h and $d(g_a(.), h_a(.))$ defined by

$$d(g_a(.), h_a(.)) = \inf_{\psi \in \Psi} \left\{ \varepsilon : \|\psi\|_B \leq \varepsilon, \ \sup_{r \in [0,1]} |g_a(\psi(r)) - h_a(r)| \leq \varepsilon \right\} \tag{1.11}$$

where Ψ is the class of all homeomorfisms from $[0,1]$ onto itself with $\psi(0) = 0$ and

$$\|\psi\|_B = \sup_{r_2 \neq r_1} |\log \frac{\psi(r_2) - \psi(r_1)}{r_2 - r_1}| \ , \ r_2, r_1 \in [0,1] \ , \tag{1.12}$$

(see also Billingsley (1968)). This choice ensures that d^p induces on $D[0,1]^p$ a σ-field $\mathcal{D}^p$ which is the Cartesian product of the σ-fields generated by the open sets on the component spaces. Also, $D[0,1]^p$ is a complete separable metric space, like $D[0,1]$.

Prohorov's (1956) theorem shows that in a complete and separable space a family of probability measure is relatively compact if and only if it is tight; Billingsley (1968, p.41) ensures that probability measures in a product space are tight if and only if the marginal probability measures are tight on the component space; Phillips and Durlauf (1986) show that the finite dimensional sets obtained from the inverse of the projection mapping form a determining class on $\mathcal{D}^p$, the class of Borel sets on $D[0,1]^p$. Hence weak convergence in $D[0,1]^p$ can be established by the same procedure discussed by Billingsley (1968) for the univariate case, i.e. establishing convergence of the finite-dimensional distributions and tightness of the sequence of probability measures, for which a sufficient condition (see Billingsley, 1968, p.128) is that for some $c > 0$, $\alpha > 1$, $a = 1, ..., p$

$$E\,|z_{an}(r) - z_{an}(r_1)|^c\,|z_{an}(r_2) - z_{an}(r)|^c \leq C\,|r_2 - r_1|^\alpha \ , \tag{1.13}$$

for all r, r_1, r_2 such that $0 \leq r_1 \leq r \leq r_2 \leq 1$. More precisely, define multivariate Brownian motion $B(r;\Omega)$, $r \in R$, $\Omega > 0$, to be a vector-valued Gaussian process with independent increments, $B(0;\Omega) = \underline{0}$, a.s., and

$$EB(r;\Omega) = \underline{0}\ ,\ r \in R, \quad EB(r_1;\Omega)B'(r_2;\Omega) = \Omega \min(r_1, r_2),\ r_1, r_2 \geq 0\ . \tag{1.14}$$

We shall write $B(r)$ for $B(r;\Omega)$ when $\Omega = I_p$. Now introduce the uniform and strong mixing coefficients φ_m, α_m defined as

$$\varphi_m = \sup_t \sup_{\substack{B \in \Im_{t+m}^{\infty} \\ A \in \Im_{-\infty}^{t}}} |P(A|B) - P(A)| \tag{1.15}$$

$$\alpha_m = \sup_t \sup_{\substack{B \in \Im_{t+m}^{\infty} \\ A \in \Im_{-\infty}^{t}}} |P(A \cap B) - P(A)P(B)| \tag{1.16}$$

with $\Im_{-\infty}^{t} = \sigma(\eta_t, \eta_{t-1}, ...\eta_{t-j},\ j \geq 0)$ the σ-field representing the "information" up to time t, and $\Im_{t+m}^{\infty} = \sigma(\eta_{t+m}, ..., \eta_{t+m+j},\ j \geq 0)$. From Phillips and Durlauf (1986) we learn

Theorem 1.1 *(Phillips and Durlauf (1986))* Let η_t be a zero-mean $p \times 1$ vector sequence satisfying

(a) $E\|\eta_t\|^{\gamma} < \infty$, $2 \leq \gamma < \infty$

(b) $\sum_{m=1}^{\infty} \varphi_m^{1-1/\gamma} < \infty$ or $\gamma > 2$ and $\sum_{m=1}^{\infty} \alpha_m^{1-2/\gamma} < \infty$

Then under (1.6)/(1.7), if $\operatorname{rank}(\Omega) = p$, we have

$$z_n(r) \Rightarrow B(r;\Omega) \text{ as } n \to \infty\ . \tag{1.17}$$

For instance for η_t covariance stationary and $r = 1$, (1.17) entails $n^{-1/2}\sum_{t=1}^{n} \eta_t \Rightarrow N(0, 2\pi f(0))$. Therefore large-sample inference on time series models, often requires consistent estimation of $f(0)$. Such estimates are typically elaborated using techniques from the vast spectral density estimation literature (for instance Press and Tukey (1956), Eicker (1967), Brillinger (1981), Robinson (1991), (1997)). For a sequence of column vectors u_t, $t = 1, ..., n$, define the discrete Fourier transform

$$w_u(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} (u_t - Eu_t) e^{it\lambda} , \tag{1.18}$$

where mean correction need not be entertained if we focus on the Fourier frequencies λ_j, $j = 1, 2, ...$. If v_t, $t = 1, ..., n$, is also a sequence of column vectors, define the (cross-) periodogram matrix

$$I_{uv}(\lambda) = w_u(\lambda) w_v^*(\lambda) , \tag{1.19}$$

where * indicates transposition combined with complex conjugation. Further define the (real part of the) averaged periodogram

$$\widehat{F}_{uv}(k, \ell) = \frac{2\pi}{n} \sum_{j=k}^{\ell} \mathrm{Re}\, \{I_{uv}(\lambda_j)\} , \; 1 \leq k \leq \ell \leq n - 1 . \tag{1.20}$$

In case $\widehat{F}_{uv}(k, \ell)$ is a vector we shall denote its a-th element $\widehat{F}_{uv}^{(a)}(k, \ell)$, and in case it is a matrix we shall denote its (a,b)-th element $\widehat{F}_{uv}^{(ab)}(k, \ell)$; analogously we will use $I_{uv}^{(a)}(\lambda)$ and $I_{uv}^{(ab)}(\lambda)$ to denote respectively the a-th element of the vector $I_{uv}\lambda)$ and the (a,b)-th element of the matrix $I_{uv}(\lambda)$. We shall drop the subscripts and the superscripts whenever this is made possible by the context. Also, for $k = 1$ we will occasionally use the alternative notation $\widehat{F}(1, \ell) = \widehat{F}(\lambda_\ell)$; moreover, we shall focus when convenient on the continuously averaged version $\widetilde{F}(\lambda_\ell) = \int_0^{\lambda_\ell} I(\lambda) d\lambda$. Under short range dependence and regularity conditions, a consistent estimate of $f(0)$ is provided by the smoothed statistics

$$\widehat{f}(0) = \sum_{j=1}^{m} K_m(\lambda_j) I(\lambda_j) , \; m < n , \tag{1.21}$$

and

$$\widetilde{f}(0) = \int_{-\pi}^{\pi} K_m(\lambda) I(\lambda) d\lambda , \tag{1.22}$$

(1.21) and (1.22) popularly labelled as discretely and continuously averaged periodogram estimates, respectively. The kernel $K_m(.)$ typically satisfies $K_m(-\lambda) = K_m(\lambda)$, $-\pi < \lambda < \pi$, and $\int_{-\pi}^{\pi} K_m(\lambda) d\lambda = 1$. Recent contributions have also allowed for more general conditions than covariance stationarity, as motivated by the econometric environment, for instance non-trending heteroscedasticity in Newey and West (1987), Andrews (1991), Andrews and Monahan (1992), and Hansen (1992a). Robinson (1997) discusses conditions under which smoothing and short range dependence are not necessary for consistency to occur under nonparametric autocorrelation.

Consider now the multiple regression problem:

$$\begin{aligned} y_t &= x_t' B + e_t \ , && (1.23) \\ x_t &= x_{t-1} + u_t \ , \ x_0 = 0, \ t = 1, 2, ..., && (1.24) \end{aligned}$$

where B is a $p_2 \times p_1$ matrix of unknown coefficients, $p_1 + p_2 = p$, $(y_t : x_t) = z_t$ and $(e_t : u_t) = \eta_t$ is a $p \times 1$ vector satisfying the assumptions of Theorem 1.1. This model is general enough to cover a great variety of serial dependence and simultaneity assumptions; in particular, it could be regarded as the reduced form of a dynamic simultaneous equations system where all variables are jointly endogenous and $E x_t e_t' \neq 0$.

By Parseval's identity

$$\widehat{F}_{xx}(1, n-1) = \frac{1}{n} \sum_{t=1}^{n-1} (x_t - \overline{x})(x_t - \overline{x})' \ , \ \widehat{F}_{xy}(1, n-1) = \frac{1}{n} \sum_{t=1}^{n-1} (x_t - \overline{x})(y_t - \overline{y})' \ , \quad (1.25)$$

where $\overline{x} = n^{-1} \sum_{t=1}^{n} x_t$, $\overline{y} = n^{-1} \sum_{t=1}^{n} y_t$; hence the least squares regression estimates for B are given by

$$\widehat{B} = (F_{xx}(1, n-1))^{-1} F_{xy}(1, n-1) \ ,$$

provided $F_{xx}(1, n-1)$ has full rank.

We decompose $\Omega = \Sigma + \Lambda + \Lambda'$, where $\Lambda = \sum_{\tau=1}^{\infty} \Gamma(\tau)$, $\Sigma = \Gamma(0)$, $\Gamma(\tau) = E \eta_0 \eta_\tau'$; also we partition

$$\Omega = \begin{bmatrix} \{\Omega\}_{11} & \{\Omega\}_{12} \\ \{\Omega\}_{21} & \{\Omega\}_{22} \end{bmatrix}, \tag{1.26}$$

where $\{A\}_{ab}$ is the a, b-th block component of a generic matrix A (to achieve notational convenience, whenever the context does not allow for ambiguity we shall use the shorthand A_{ab}). Phillips and Durlauf (1986) present the following result:

Theorem 1.2 *(Phillips and Durlauf (1986))* If η_t satisfies the conditions of Theorem 1.1, then as $n \to \infty$

$$n(\widehat{B} - B) \Rightarrow \left(\left\{\int_0^1 B(r;\Omega)B'(r;\Omega)dr\right\}_{22}\right)^{-1} \left\{\int_0^1 B(r;\Omega)dB'(r;\Omega)dr + \Sigma + \Lambda\right\}_{21}. \tag{1.27}$$

The proof of (1.27) entails showing that

$$\frac{1}{n}\widehat{F}_{xx}(1, n-1) \Rightarrow \left\{\int_0^1 B(r;\Omega)B'(r;\Omega)dr\right\}_{22}, \tag{1.28}$$

and

$$\widehat{F}_{xe}(1, n-1) \Rightarrow \left\{\int_0^1 B(r;\Omega)dB'(r;\Omega)dr + \Sigma + \Lambda\right\}_{21}. \tag{1.29}$$

The continuous mapping theorem allows to link these two results to yield the asymptotic behaviour of $\widetilde{B}$, provided the following condition is established:

$$\left|\left\{\int_0^1 B(r;\Omega)B'(r;\Omega)dr\right\}_{22}\right| > 0, \text{ a.s. }. \tag{1.30}$$

Now (1.28) follows from Theorem 1.1, the continuous mapping theorem and easy manipulations. The proof of (1.30), which is not trivial, is delayed to Phillips and Hansen (1990), Lemma A.3. The proof of (1.29) requires a much more advanced mathematical treatment than (1.28) - a similar dichotomy we shall face again in Chapter 4. A first attempt to establish (1.29) is in Phillips (1988a), but the argument is flawed by an error, as noted by Hansen (1992b). Hence this issue deserves a more careful treatment.

The asymptotic behaviour of

$$\widehat{F}_{xe}(1, n-1) = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{t} u_s e_t' \tag{1.31}$$

when $\{(e_t : u_t), \Im_{-\infty}^{t}\}$ is a multivariate martingale difference sequence was first investigated by Chan and Wei (1988); a highly comprehensive treatment within the martingale assumption was subsequently achieved by Kurtz and Protter (1990). Define

$$\frac{1}{n} \sum_{t=1}^{[nq]} x_t e_t' = \int_0^q x_n(r) de_n'(r), \;\; x_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} x_t \; , \; e_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} e_t \; . \tag{1.32}$$

We have the following

Theorem 1.3 *(Chan and Wei (1988))* Assume (1.23)/(1.24) holds, and for $\eta_t = (e_t : u_t)$ let $\{\eta_t, \Im_t\}$ be a martingale difference sequence such that $E(\eta_t \eta_t' | \Im_t) = \Sigma$, a.s. . Then as $n \to \infty$ we have

$$x_n(r) \Rightarrow B(r; \Sigma_{22}) \; , \; \frac{1}{n} \sum_{t=1}^{n} x_{t-1} e_t' \Rightarrow \int_0^1 B(r; \Sigma_{22}) dB'(r; \Sigma_{11}) \; . \tag{1.33}$$

Theorem 1.4 *(Kurtz and Protter (1990))* Let (1.32) hold, where $\{(e_n(r), x_n(r)), \Im_{[nr]}\}$ is a martingale array with

$$(e_n(r), x_n(r)) \Rightarrow (B(r; \Sigma_{11}), B(r; \Sigma_{22})) \; . \tag{1.34}$$

Let

$$\sup_n \frac{1}{n} \sum_{t=1}^{n} \|e_t\|^2 < \infty \; , \text{ a.s. } . \tag{1.35}$$

Then we have, as $n \to \infty$

$$\left(e_n(r), x_n(r), \int_0^q x_n(r)de_n'(r)\right) \Rightarrow \left(B(r;\Sigma_{11}), B(r;\Sigma_{22}), \int_0^q B(r;\Sigma_{22})dB'(r;\Sigma_{11})\right) . \tag{1.36}$$

The results in Chan and Wei (1988) and in Kurtz and Protter (1990) as they stand are of limited applicability in time series econometrics because the martingale assumption rules out any form of serial dependence. However, these results are instrumental for the analysis of more general cases through martingale approximations theorems. An example of this approach is provided by Phillips (1988b); a suitable version of his results is the following:

Theorem 1.5 *(Phillips (1988b))* Let (1.23)/(1.24) hold, with $\eta_t = (e_t : u_t)$ and

$$\eta_t = A(L)\varepsilon_t,\ A(L) = \sum_{j=0}^{\infty} A_j L^j , \tag{1.37}$$

$$\sum_{j=0}^{\infty} \|A_j\| < \infty,\ \sum_{k=1}^{\infty}\left[\|\sum_{j=k}^{\infty} A_j\|\right] < \infty , \tag{1.38}$$

where the sequence of random vectors (ε_t) is $i.i.d.(0, \Sigma_\varepsilon)$ with $rank(\Sigma_\varepsilon) = p$. Then

$$\frac{1}{n}\sum_{t=1}^{n} x_{t-1}e_t' \Rightarrow \int_0^1 B(r;\Omega_{22})dB'(r;\Omega_{11}) + \Lambda_{21} , \tag{1.39}$$

where $\Omega = A(1)\Sigma_\varepsilon A(1)'$.

A subsequent contribution by Hansen (1992b) is aimed at obtaining weak convergence of $n^{-1}\sum_t^n x_{t-1}e_t'$ to $\int BdB'$ relaxing the linearity and strict stationarity assumptions in Phillips (1988b), again exploiting a martingale approximation argument.

Theorem 1.6 *(Hansen (1992b))* Let η_t be a zero mean, strong mixing sequence with mixing coefficients $\alpha_m = o(m^{-\delta\gamma/(\delta-\gamma)+\varepsilon})$ for some $\varepsilon > 0$, $\delta > \gamma > 2$. Assume that (1.6)/(1.7) hold and $\sup_{t\geq 1} E\|\eta_t\|^{2\delta} < \infty$, where

$$\lim_{n\to\infty} \frac{1}{n} E x_n x_n' = \Omega < \infty \ , \ \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{t} E\eta_s \eta_t' = \Lambda \ , \ rank(\Omega) = p \ . \tag{1.40}$$

Then for $0 < q \leq 1$

$$\frac{1}{n} \sum_{t=1}^{[nq]} x_{t-1} e_t' \Rightarrow \int_0^q B(r;\Omega_{22}) dB'(r;\Omega_{11}) dr + q\Lambda_{21} \text{ as } n \to \infty \ . \tag{1.41}$$

We note that Theorem 1.5 and Theorem 1.6 do not imply each other, because a strong mixing sequence need not be stationary, while on the other hand a strictly stationary linear process need not be strong mixing (cf. for instance Ibragimov and Rozanov, (1977), Andrews (1984), Pham and Tran (1985)). Recently, a more general result than Theorems 1.5-1.6 has been given by Davidson and DeJong (1997), but we shall not deal with such extension in this thesis.

Gathering together the results from Theorems 1.1,1.2, and 1.5, (1.27) is established; we can rewrite the right hand side as $\Xi^{-1}\varkappa$, where

$$\varkappa = \int_0^1 B(r;\Omega_{22}) dB'(r;\Omega_{11}) dr + \Sigma_{21} + \Lambda_{21} \ , \tag{1.42}$$

$$\Xi = \int_0^1 B(r;\Omega_{22}) B'(r;\Omega_{22}) dr \ . \tag{1.43}$$

Some features of (1.27) are as follows. Convergence occurs at rate n and hence faster than the usual square root rate which applies to the stationary case; on the other hand, the limiting distribution is non-Gaussian and depends on nuisance parameters that inhibit statistical inference. For instance, if we consider the classical linear hypotheses of form

$$H_0 : Qvec(B) = q \ , \ H_1 : Qvec(B) \neq q \ , \tag{1.44}$$

where Q and q are respectively $g \times p_1 p_2$, $g \times 1$ matrices respectively, and Q has full rank g, the Wald statistic takes the form

$$\tau_W = (Qvec(\widehat{B}) - q)'[Q(2\pi \widehat{f}(0) \otimes (X'X)^{-1})Q']^{-1}(Qvec(\widehat{B}) - q) \,, \tag{1.45}$$

whose asymptotic distribution is very complicated under the conditions of Theorems 1.1,1.2, 1.5 and cannot be tabulated without some form of a priori knowledge of the covariance structure of the innovations. Furthermore, the distribution (1.27) is affected by a bias term $\{\Sigma + \Lambda\}_{21}$ which is second order in the sense that it does not prevent (super-) consistency but it does jeopardize noticeably the behaviour of OLS in finite sample (see for instance the simulation results in Banerjee et al.. (1993)).

In order to investigate in greater detail the structure of the asymptotic distribution (1.27), we introduce

$$B_{1.2}(r; \Omega_{1.2}) = B_1(r; \Omega_{11}) - \Omega_{12}\Omega_{22}^{-1} B_2(r; \Omega_{22}) \,, \tag{1.46}$$

where we denote by $\Omega_{1.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ the covariance matrix of $B_{1.2}(r)$, the latter interpreted as the component of $B_1(r; \Omega_{11})$ which is orthogonal (and hence by Gaussianity independent) from $B_2(r; \Omega_{22})$. We can rewrite (1.27) as

$$\begin{aligned} n(\widehat{B} - B) &\Rightarrow \Xi^{-1}(\varkappa_{1.2} + \Omega_{22}^{-1}\Omega_{21}\varkappa_2 + \Sigma_{21} + \Lambda_{21}) & (1.47) \\ \varkappa_{1.2} &= \int_0^1 B_2(r; \Omega_{22}) dB'_{1.2}(r; \Omega_{1.2}) \,, \; \varkappa_2 = \int_0^1 B_2(r; \Omega_{22}) dB'_2(r; \Omega_{22}) \,. & (1.48) \end{aligned}$$

The asymptotic distribution of $\widehat{F}_{xe}(1, n-1)$ is then partitioned into three components: (i) a stochastic integral $(\varkappa_{1.2})$ where the integrand and the integrator are independent processes; (ii) the "unit root distribution", i.e., a stochastic integral driven by identical integrand and integrator, multiplied by a constant matrix which can be interpreted as the least squares coefficient in a sort of "long run regression" between e_t and u_t; (iii) a bias term induced by the long-run correlation between e_t and u_t. Note that if the data generating process driving e_t and u_t is such that the two innovations are uncorrelated at all leads and lags, the terms (*ii*) and (*iii*) vanish; namely, if Ω has a block diagonal structure, with Λ_{21} and Σ_{21} equal to the $p_2 \times p_1$ null matrix, we have that

$$n(\widehat{B} - B) \Rightarrow \Xi^{-1} \varkappa_{1.2} \ . \tag{1.49}$$

Let us consider more carefully the asymptotic distribution of (1.49).

Theorem 1.7 *(Park and Phillips (1988))* The following equality in distribution holds:

$$vec(\varkappa_{1.2}' \Xi^{-1/2}) \equiv N(\underline{0}, \Omega_{1.2} \otimes I_{p_2}) \ . \tag{1.50}$$

Theorem 1.7 implies that the asymptotic distribution of OLS when the innovations in e_t and u_t are uncorrelated at all leads and lags is given by

$$n(vec(\widehat{B}) - vec(B)) \Rightarrow \int_{G>0} N(\underline{0}, K_{p_1 p_2} \left\{\Omega_{11} \otimes G^{-1}\right\} K_{p_1 p_2}') dP(G) \ , \tag{1.51}$$

where $K_{p_1 p_2}$ is the commutation matrix of order $p_1 p_2$ (Magnus and Neudecker, (1991)), $P(.)$ is the probability measure associated with G and

$$G = \int_0^1 B_2(r; \Omega_{22}) B_2'(r; \Omega_{22}) dr \ . \tag{1.52}$$

Hence the asymptotic distribution is a mixture of normals which can be visualized as the outcome of a two stage random experiment, such that in the first part we draw the random covariance matrix G, while in the second we operate another random drawing from a multivariate Gaussian distribution with a covariance matrix that depends on G.

Thus the asymptotic distribution belongs to the so called LAMN family introduced by Jeganathan (1980), LAMN standing for Locally Asymptotic Mixed Normal. Hence under these circumstances OLS estimates share all the asymptotical properties of this class, namely (i) they are symmetrically distributed around zero; (ii) the nuisance parameters involve only scale effects and can be easily eliminated for the purpose of inference; (iii) an optimal theory of inference applies (see LeCam 1986); (iv) hypothesis testing can be conducted within the usual asymptotic chi-squared paradigm.

Theorem 1.7 motivates most of the procedures proposed for the analysis of cointegrated time series, the underlying thread being the implementation of parametric, semiparametric or nonparametric procedures to eliminate the "extra terms" in (1.47). A first example of this approach is provided by Phillips and Hansen (1990). Let us concentrate on the case of a single cointegrating relationship, so that (1.23)/(1.24) becomes

$$y_t = x_t'\beta + e_t \ , \tag{1.53}$$

$$x_t = x_{t-1} + u_t, \ x_0 = 0 \ , \ t = 1, 2, \dots \ , \tag{1.54}$$

where β is a $(p-1) \times 1$ non-null vector ($\alpha = (1 : \beta')$ in the sense of Definition 1.1). Consider the $p-$dimensional process $(B_{1.2}(r) : B_2(r))$, with variance covariance matrix given by

$$\begin{pmatrix} \omega_{11.2} & \underline{0}' \\ \underline{0} & \Omega_{22} \end{pmatrix} , \tag{1.55}$$

where $\omega_{11.2} = \omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. Now let $e_{t.u} = e_t - u_t'\Omega_{22}^{-1}\Omega_{21}$, and note that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} e_{t.u}^2 = \omega_{11.2} \ , \ \text{a.s.} \ , \tag{1.56}$$

$$\frac{1}{n} \sum_{t=1}^{n} x_t e_{t.u} \Rightarrow \int_0^1 B_2(r; \Omega_{22}) dB_{1.2}(r; \Omega_{1.2}) + \delta^+ \ , \tag{1.57}$$

$$\delta^+ = \Sigma_{21} + \Lambda_{21} - (\Sigma_{22} + \Lambda_{22})\Omega_{22}^{-1}\Omega_{21} \ . \tag{1.58}$$

Heuristically, taking $y_t^+ = y_t + e_{t.u} - e_t$ as the endogenous variable eliminates the "unit root distribution" from (1.47).

Concerning δ^+, feasible estimates of Σ_{21} and Σ_{22} are naturally provided by the correspondig sample covariances; for Λ_{21} and Λ_{22}, Phillips and Hansen (1990) suggest to adopt kernel estimates as in Newey and West (1987) and Andrews (1991). As discussed before, such consistent estimates of Ω_{2a}, $a = 1,2$ are due to the literature on kernel spectral density estimation, for instance Press and Tukey (1956), Eicker (1967), Brillinger (1979). This duality cannot, however, be exploited in the estimation of Λ_{2a}, $a = 1,2$, because this quantity derives from a "one-sided" summation of autocovariances and hence cannot be granted a spectral interpretation. We shall get back on the connection between long-run covariance matrices and the spectral density matrix at frequency zero in Chapter 5.

A feasible three-step procedure is then the following: (i) estimate the OLS regression coefficients and residuals as $\widehat{\beta} = (\widehat{F}_{xx}(1,n-1))^{-1}(\widehat{F}_{xe}(1,n-1))$, $\widehat{e} = y_t - x_t'\widehat{\beta}$; (ii) obtain consistent estimates of δ^+, e_t and $e_{t.u}$ from $\widehat{e}_t$ and $u_t = x_t - x_{t-1}$. (iii) implement the fully-modified procedure as

$$\widehat{\beta}^+ = (\widehat{F}_{xx}(1,n-1))^{-1}(\widehat{F}_{xy^+}(1,n-1) - n\widehat{\delta}^+) \ , \ \widehat{y}_t^+ = y_t + \widehat{e}_{t.u} - \widehat{e}_t \ . \tag{1.59}$$

The previous heuristics is formalized in the following

Theorem 1.8 *(Phillips and Hansen (1990))* Under the conditions of Theorems 1.1 and 1.5, (1.52) and (1.59), we have as $n \to \infty$

$$n(\widehat{\beta}^+ - \beta) \Rightarrow \int_{G>0} N(0, \omega_{11.2}G^{-1})dP(G) \ ; \tag{1.60}$$

also, if we consider the set of linear hypotheses

$$H_0 : Q\beta = q \ , \ H_1 : Q\beta \neq q \ , \tag{1.61}$$

for fixed $g \times (p-1)$ matrix Q and $g \times 1$ vector $r \neq 0$ we have

$$\tau_W = (Q\widehat{\beta}^+ - r)'(\widehat{\omega}_{11.2}Q(X'X)^{-1}Q')^{-1}(Q\widehat{\beta}^+ - r) \Rightarrow \chi^2_g \ , \tag{1.62}$$

where χ^2_g is the standard chi-squared distribution with g degress of freedom.

An important feature of Theorem 1.8 is that the Wald statistic (1.62) is asymptotically distributed as a chi-squared, so that hypothesis testing can be performed within the standard paradigm.

The case of multiple cointegrating relationships is investigated by Phillips (1991a). We rewrite (1.23)/(1.24) in an Error-Correction Mechanism (ECM) format as

$$\Delta z_t = -\Gamma \left[I, -B' \right] z_{t-1} + v_t \ , \tag{1.63}$$

where

$$\Gamma = \begin{bmatrix} I_{p_1} \\ 0 \end{bmatrix} , \ v_t = \begin{bmatrix} I_{p_1} & B' \\ 0 & I_{p_2} \end{bmatrix} \eta_t \ , \ \eta_t = (e_t : u_t) \ . \tag{1.64}$$

(1.63) is not the Error Correction representation advocated by Granger (1987), the short-run dynamics being entirely confined to the error term v_t. We assume that v_t is generated by the linear model

$$v_t = \sum_{j=0}^{\infty} A_j(\theta)\varepsilon_{t-j} \ , \ \varepsilon_t \equiv i.i.d.(0, \Sigma(\theta)) \ , \tag{1.65}$$

where θ is a $q \times 1$ vector of parameters, $A_0 = I_p$ and

$$\sum_{j=0}^{\infty} j^{1/2} A_j(\theta) < \infty \ , \ \forall \theta \ . \tag{1.66}$$

The frequency-domain approximation to the Gaussian log-likelihood (the so-called Whittle likelihood) is given by

$$\mathcal{L}_n^W(B,\theta) = \ln |\Sigma(\theta)| + n^{-1} \sum_{j=1}^{n-1} tr \left\{ f_{vv}(\lambda_j;\theta)^{-1} I_{vv}(\lambda_j) \right\} \ . \tag{1.67}$$

Phillips (1991a) consider the estimates

$$\widetilde{B}_{ML}, \widetilde{\theta}_{ML} = \arg\min_{B,\theta} \mathcal{L}_n^W(B,\theta) \ , \tag{1.68}$$

where the minimization is carried over a compact parameter space Θ. We have the following

Theorem 1.9 *(Phillips (1991a))* Let (1.63), (1.65) and (1.66) hold. Assume also the regularity conditions of Dunsmuir (1979) are satisfied. We have

$$n(vec(\widetilde{B}_{ML}) - vec(B)) \Rightarrow \int_{G>0} N(0, K_{p_1p_2}\left\{\Omega_{11.2} \otimes G^{-1}\right\} K'_{p1p_2}) dP(G) \ . \tag{1.69}$$

Under the conditions of Theorem 1.9 a standard theory of inference applies and classical testing procedures lie entirely within the usual chi-squared paradigm, despite the convergence of the Hessian to a random limit rather than a constant matrix. The key to achieve an asymptotic distribution within the LAMN family is again the elimination of the "extra terms" in $\varkappa$. This goal was pursued through semiparametric corrections in Phillips and Hansen (1990); in this case, instead, the same objective is pursued through the specification of an appropriate ECM mechanism, such that all the *a priori* information on the numbers of unit roots of the series is properly incorporated in the system and an efficient MLE procedure can be implemented.

In Theorem 1.9, parametric correlation of a known form is assumed for the residual v_t. This assumption is relaxed by Phillips (1991b), where the idea is to adapt to the cointegration case efficient techniques for the analysis of a multiple system of linear equations in the presence of autocorrelation of unknown form; these techniques were introduced by Hannan (1963) and extended in Hannan (1971), Robinson (1972), Hannan and Robinson (1973), Robinson (1991). The starting point is again the ECM representation (1.63), which in the case of a single cointegrating relationship we rewrite as

$$\Delta z_t = -\gamma\alpha' z_{t-1} + v_t \ , \ \gamma' = (1,0), \ \alpha' = (1, -\beta') \ , \tag{1.70}$$

$$v_t = \begin{bmatrix} 1 & \beta' \\ 0 & I_{p-1} \end{bmatrix} \eta_t \ , \ \eta_t = (e_t : u_t) \ . \tag{1.71}$$

As in Theorems 1.1-1.6, we assume that suitable regularity conditions hold such that as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \eta_t \Rightarrow B(r; \Omega) \ , \tag{1.72}$$

$$\frac{1}{n} \sum_{t=1}^{[nq]} x_t e_t' \Rightarrow \int_0^q B(r; \Omega_{22}) dB'(r; \Omega_{11}) + q(\Sigma_{21} + \Lambda_{21}) \ . \tag{1.73}$$

Moreover, we shall use the following abbreviated notation for the discrete Fourier transform of generic lagged or first differenced sequences:

$$w_{Lu}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} u_{t-1} e^{i\lambda t} \ , \tag{1.74}$$

$$w_{\Delta u}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} (u_t - u_{t-1}) e^{i\lambda t} \ . \tag{1.75}$$

Consider now the estimate

$$\widehat{\alpha}_{SR} = \arg\min_{\alpha} \sum_{j=1}^{n} tr[[w_{\Delta z}(\lambda_j) - \gamma\alpha' w_{Lz}(\lambda_j)][w_{\Delta z}(\lambda_j) - \gamma\alpha' w_{Lz}(\lambda_j)]^* \mathcal{W}^{-1}(\lambda_j)] \ , \tag{1.76}$$

where $\mathcal{W}(.)$ is a given positive definite Hermitian matrix. The minimization of Hermitian form (1.76) would indeed be equivalent to the maximization of the Whittle's spectral likelihood function if the $w_v(.)$ were independent complex normal random vectors with covariance matrix given by $\mathcal{W}(.)$. In the more general case where Gaussianity does not hold, these estimates will no longer be fully efficient; however, it can be shown that under a great variety of weak dependence assumptions we have asymptotically (Brillinger (1981)), as $n \to \infty$ and $\lambda_j \to \lambda$,

$$w_v(\lambda_j) \Rightarrow N^c(0, f_{vv}(\lambda)), \quad \lambda \neq 0, \pi \ , \tag{1.77}$$

where $N^c(.,.)$ indicates the complex normal distribution. Hence, under the assumption that $f_{vv}(\lambda) > 0$ for $-\pi \leq \lambda \leq \pi$, a natural and feasible choice for $\mathcal{W}(.)$ is provided by $\mathcal{W}(\lambda) = \widetilde{f}_{vv}(\lambda)^{-1}$, where the estimate $\widetilde{f}_{vv}(.)$, based on residuals from a preliminary least squares regression, must be such that

$$\sup_{0 \leq \lambda < \pi} |\widetilde{f}_{vv}(\lambda) - f_{vv}(\lambda)| \ = o_p(1) \ . \tag{1.78}$$

More rigorously, the minimization problem (1.76) can be shown to approximate a time domain generalized least squares criterion, and it is solved by $\widehat{\alpha}_{SR} = (1, -\widehat{\beta}_{SR})$, where

$$\widehat{\beta}_{SR} = \left(\sum_{j=1}^{n-1} (\widetilde{f}_{vv}(\lambda_j))^{-1} w_x(\lambda_j) w_x(\lambda_j)^*\right)^{-1} \left(\sum_{j=1}^{n-1} (\widetilde{f}_{vv}(\lambda_j))^{-1} w_x(\lambda_j) w_y(\lambda_j)^*\right) \ . \tag{1.79}$$

This is not the expression considered by Phillips (1991b), who analyzes instead

$$\widetilde{\beta}_{SR} = \left(\sum_{j=1}^{n-1} (\widetilde{f}_{vv}(\lambda_j))^{-1} \widetilde{f}_{xx}(\lambda_j)\right)^{-1} \left(\sum_{j=1}^{n-1} (\widetilde{f}_{vv}(\lambda_j))^{-1} \widetilde{f}_{xy}(\lambda_j)\right) \ , \tag{1.80}$$

where

$$\widetilde{f}_{xx}(\lambda_j) \ = \ \sum_{\tau=-M}^{M} k(\frac{\tau}{M}) c_{xx}(\tau), \ c_{xx}(\tau) = \frac{1}{n} \sum_{t=1}^{n-|\tau|} x_t x'_{t+\tau} \ , \tag{1.81}$$

$$\widetilde{f}_{xy}(\lambda_j) \ = \ \sum_{\tau=-M}^{M} k(\frac{\tau}{M}) c_{xy}(\tau), \ c_{xy}(\tau) = \frac{1}{n} \sum_{t=1}^{n-|\tau|} x_t y_{t+\tau} \ , \tag{1.82}$$

and $k(u)$ is a lag window such that $k(0) = 1$ and $k(u) = 0$ when $|u| > 1$. (1.79) and (1.80) are not asymptotically equivalent in the case of nonstationary cointegrated variables (cf

Chapter 4); (1.80) can be interpreted as a form of weighted covariance spectral estimate, (1.79) being a form of smoothed periodogram estimate: the relationship between $K_m(.)$ and $k(.)$ will be discussed in Chapter 5. Phillips (1991b) considers also the estimate

$$\widetilde{\beta}_{(0)} = \left((\widetilde{f}_{vv}(0))^{-1}\widetilde{f}_{xx}(0)\right)^{-1}\left((\widetilde{f}_{vv}(0))^{-1}\widetilde{f}_{xy}(0)\right) . \tag{1.83}$$

Let us now analyze the asymptotic behaviour of (1.80)/(1.83).

Theorem 1.10 *(Phillips (1991b))* Under (1.72), (1.73) and $1/M + M^2/n \to 0$ as $n \to \infty$ we have

$$n(\widetilde{\beta}_{SR} - \beta) \Rightarrow \Xi^{-1}\varkappa_{1.2} , \tag{1.84}$$

$$n(\widetilde{\beta}_{(0)} - \beta) \Rightarrow \Xi^{-1}\varkappa_{1.2} . \tag{1.85}$$

The proof of Theorem 1.10 is based mainly on the Skorohod representation theorem, which at crucial steps allows us to interchange almost sure convergence and convergence in probability. A fundamental ingredient of the construction is the fact that the number $(2M + 1)$ of covariances considered in $\widetilde{f}_{xx}(.)$ and $\widetilde{f}_{xy}(.)$ is $o(n^{1/2})$, a condition which can be easily assumed to hold for weighted covariance procedures (see for instance Hannan (1970)) but which fails in general for the weighted periodogram estimates.

Theorem 1.10 implies that the nonparametric treatment of the vector of innovations $(e_t : u_t)$ implicit in the spectral regression procedure achieves the asymptotic optimality typical of the LAMN class, as in Theorem 1.8, 1.9. Also, (1.85) shows that narrow-band and full band procedures can be equivalent under nonstationarity, the heuristic rationale being that the "signal" concentrates eventually on the lowest frequencies for (co)-integrated variables; in other nonstationary circumstances narrow-band procedures can show some superiority, as discussed in Chapter 4.

1.3 ASYMPTOTICS FOR FRACTIONALLY INTEGRATED TIME SERIES

The analysis of the previous section was characterized by the sharp distinction between processes that are integrated and nonstationary (i.e. $I(d)$, for $d = 1, 2, ...$), and processes that are stationary with summable autocovariances, which we defined to be short range dependent or $I(0)$. We anticipated that the gap between these two classes can be smoothed by a broader definition of integration, namely (1.2). For $p = 1$, $-\frac{1}{2} < d < \frac{1}{2}$, z_t defined there is asymptotically a covariance stationary and invertible scalar process, and we might replace the right hand side of (1.2) by η_t to obtain a covariance stationary process, ζ_t (say); under these circumstances, other characterizations are possible for fractional integration: more precisely, in terms of the behaviour of the spectral density, we have the class of processes

$$f(\lambda) \approx C\lambda^{-2d} \text{ as } \lambda \to 0^+ , \tag{1.86}$$

where " $\approx$ " indicates that the ratio of the left- and right-hand sides tends to one. Under mild additional conditions, (Yong (1974)), (1.86) is equivalent to

$$\gamma(\tau) \approx C\tau^{2d-1} \text{ as } \tau \to \infty . \tag{1.87}$$

Processes satisfying (1.86)/(1.87) are called long-memory or long range dependent for $0 < d < \frac{1}{2}$ and negative dependent for $-\frac{1}{2} < d < 0$; these processes featured in the probabilistic literature much before the present interest in unit roots, cf. for instance Kolmogorov (1940,1941). Major interest was aroused however by Mandelbrot and Van Ness (1968) and subsequent papers, who considered applications to hydrology and economics of fractional Gaussian noise, i.e. the covariance stationary Gaussian process with spectral density

$$f(\lambda) = C(1 - \cos(\lambda)) \sum_{j=-\infty}^{\infty} \frac{1}{|\lambda + 2\pi j|^{2d+2}} , \tag{1.88}$$

(see Sinai (1976)). In the econometric and statistical literature, the introduction of fractional processes is due to Adenstedt (1974), Robinson (1978), Granger (1980), Granger and Joyeux (1980), and Hosking (1981), and in particular a widely considered process has been the fractional ARIMA, with spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{i\lambda}|^{-2d} \frac{|\vartheta(e^{i\lambda})|^2}{|\varphi(e^{i\lambda})|^2} , \tag{1.89}$$

where $\varphi(.)$ and $\vartheta(.)$ have their roots outside the unit circle; again the literature herefter has been so vast that we shall make no attempt to do it justice. In the sequel we shall concentrate mainly on statistical inference for the memory parameter in case $0 < d < \frac{1}{2}$.

The problem of estimating d was addressed already in the early eighties by Mohr (1981), Janacek (1982), Geweke and Porter-Hudak (1983) and many others, and before them by Mandelbrot and Wallis (1969); for none of these estimates, which are all of a semiparametric nature, was a rigorous asymptotic justification provided. We shall delay an analysis of the asymptotic behaviour of semiparametric procedures, and consider first fully parametric estimates; instrumental for the analysis of the asymptotic properties of many such estimates is a careful study of the behaviour of quadratic forms in random variables having long range dependence. A first step in this direction is provided by the following results.

Theorem 1.11 *(Fox and Taqqu (1985))* Let ζ_t be a zero mean covariance stationary Gaussian process such that $E\zeta_t^2 = 1$ and (1.87) holds . Let a_s, $-\infty < s < \infty$ be a sequence of constants satysfying $a_s = a_{-s}$ and $\sum_{s=-\infty}^{\infty} |a_s| < \infty$. Then if $\frac{1}{4} < d < \frac{1}{2}$, as

$n \to \infty$

$$Z_n(r) \triangleq \frac{1}{n^{2d}}(\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]} a_{t-s}\zeta_t\zeta_s - E\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]} a_{t-s}\zeta_t\zeta_s) \Rightarrow (\sum_{s=-\infty}^{\infty} a_s)R(r) , \tag{1.90}$$

where $R(r)$ is the Rosenblatt process, which is represented by the following multiple Ito integral

$$R(r) = \frac{1}{2\Gamma(1-2d)\cos((1-2d)\pi/2)} \int_{R^2}'' \frac{e^{i(s_1+s_2)r}-1}{i(s_1+s_2)} |s_1|^{d-1}|s_2|^{d-1} dB(s_1)dB(s_2) , \tag{1.91}$$

where $\int''$ signifies that the integral excludes the diagonals $s_1 = \pm s_2$.

Theorem 1.12 *(Fox and Taqqu (1985))* Let ζ_t and a_s satisfy the assumptions of Theorem 1.11 with $0 < d < \frac{1}{4}$; then as $n \to \infty$, $0 \leq r \leq 1$

$$Z_n(r) \Rightarrow \sigma_1^2 B(r) , \tag{1.92}$$

where

$$\sigma_1^2 = 2 \sum_{k=-\infty}^{\infty} \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} a_{j_1}a_{j_2}\gamma(k)\gamma(k+j_1-j_2) . \tag{1.93}$$

The proof of Theorem 1.12 is based on the representation of the cumulants of order p of the normal random variable $Z_n(1)$ as

$$cum_p(Z_n(1)) = 2^{p-1}(p-1)!tr(Q(n)A(n)) , \tag{1.94}$$

where $Q(n)$ and $A(n)$ are Toeplitz matrices such that

$$Q(n) = \{\gamma(i-j)\}_{i,j=1,\ldots n} , \quad A(n) = \{a_{i-j}\}_{i,j=1,\ldots,n} . \tag{1.95}$$

The argument is then completed showing that all cumulants of order greater than two of

$Z_n(.)$ are $o(1)$ and then appealing to the Frechet-Shohat "moment convergence theorem", and unique determination of the Gaussian distribution by its moments; this result is achieved by a (lengthy and demanding) argument based on power counting theorems from mathematical physics.

The essential insight we receive from Theorems 1.11 and 1.12 is that the limit of appropriately normalized Gaussian quadratic forms will typically be Gaussian for $d < \frac{1}{4}$, non-Gaussian for $d > \frac{1}{4}$. Of course this dichotomy severely inhibits statistical inference. Note however that Theorem 1.11 implies the convergence in probability to zero of $Z_n(r)$ when $\sum_{s=-\infty}^{\infty} a_s = 0$; it is therefore possible to conjecture that a Gaussian limiting behaviour can be achieved if a smaller normalization (in terms of the sequence a_s) is adopted. Define $g(\lambda) = \sum_{s=-\infty}^{\infty} a_s e^{-i\lambda s}$; the previous conjecture is confirmed by the following result

Theorem 1.13 *(Fox and Taqqu (1987))* Let a_s and ζ_t satisfy the assumptions of Theorem 1.11, and in addition let $f(\lambda)$ and $g(\lambda)$ be almost everywhere continuous and bounded in $[\delta, \pi]$, for all $\delta > 0$. Assume also that there exist $\alpha_1, \alpha_2 < 1$ such that $\alpha_1 + \alpha_2 < \frac{1}{2}$ and for each $\varepsilon > 0$

$$f(\lambda) = O(\lambda^{-\alpha_1-\varepsilon}) \text{ as } \lambda \to 0^+ , \qquad (1.96)$$

$$g(\lambda) = O(\lambda^{-\alpha_2-\varepsilon}) \text{ as } \lambda \to 0^+ . \qquad (1.97)$$

Then

$$Z_n(1) \Rightarrow N(0, \sigma_2^2) , \qquad (1.98)$$

where

$$\sigma_2^2 = 16\pi^3 \int_{-\pi}^{\pi} \{f(\lambda)g(\lambda)\}^2 d\lambda . \qquad (1.99)$$

Theorem 1.13 is the fundamental tool that allows efficient estimates of the parameters of a long memory process to be implemented via maximum likelihood procedures. Let us recall from Section 2 that the Whittle approximation to the likelihood function for univariate Gaussian processes is defined as

$$\mathcal{L}_n^W(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f(\lambda;\theta) + \frac{I(\lambda)}{f(\lambda;\theta)} \right\} d\lambda \ , \tag{1.100}$$

where θ is a finite-dimensional vector of parameters; assume that $d = d(\theta)$ is also determined by θ. Consider the quasi-maximum likelihood estimates

$$\widetilde{\theta}_n^W = \arg\min_{\theta \in \Theta} \mathcal{L}_n^W(\theta) \ , \tag{1.101}$$

where the parameter space Θ is assumed to be compact and regularity conditions hold for ζ_t, a Gaussian process whose behaviour is entirely determined by θ. The properties of the vector of estimates $\widetilde{\theta}_n^W$ are given in the following

Theorem 1.14 *(Fox and Taqqu (1986))* Let ζ_t satisfy the assumptions of Theorem 1.11 or 1.12 with spectral density $f(\lambda, \theta_0)$ satisfying suitable regularity conditions on its first and second derivatives. Then, as $n \to \infty$

$$\lim_{n \to \infty} \widetilde{\theta}_n^W = \theta_0 \text{ a.s.} \ , \tag{1.102}$$

and

$$\sqrt{n}(\widetilde{\theta}_n^W - \theta_0) \Rightarrow N(0, 4\pi W^{-1}(\theta_0)) \ , \tag{1.103}$$

where $W(\theta)$ is a $k \times k$ Toeplitz matrix of with elements a, b-th element

$$\int_{-\pi}^{\pi} f(\lambda, \theta) \frac{\partial^2}{\partial \theta_a \partial \theta_b} f^{-1}(\lambda, \theta) d\lambda \ . \tag{1.104}$$

The proof exploit standard arguments for implicitly-defined extremum estimates, together with a central limit theorem as Theorem 1.13 for the quadratic form

$$\int_{-\pi}^{\pi} (I(\lambda) - EI(\lambda))g(\lambda)d\lambda \ , \tag{1.105}$$

where in this case $g(\lambda) = \partial f^{-1}(\lambda, \theta)/\partial\theta_a$. Note that the result is made possible by the special nature of the function $g(\lambda)$, because as a consequence of Theorem 1.11 the limit of (1.105) can be non-Gaussian in general for other choices of $g(.)$.

Further improvements over Fox and Taqqu's result are provided by Dahlhaus (1989). This author shows under Gaussianity of ζ_t and mild additional restrictions that (i) the same asymptotics as for $\widetilde{\theta}_n^W$ holds for $\widetilde{\theta}_n$, i.e. the exact maximum likelihood estimates obtained as

$$\widetilde{\theta}_n = \arg\min_{\theta\in\Theta} \mathcal{L}_n \ , \ \mathcal{L}_n = \frac{1}{2n} \log|Q(n)| + \frac{1}{2n}(\zeta_1, ...\zeta_n)' Q^{-1}(n)(\zeta_1, ..., \zeta_n) \ , \tag{1.106}$$

where $Q(n)$ is given by (1.95), and (ii) $\widetilde{\theta}_n$ and $\widetilde{\theta}_n^W$ are efficient in the sense of Fisher, i.e. they attain the Cramer-Rao lower bound. A major generalization of Fox and Taqqu (1987) results is provided by Giraitis and Surgailis (1990). These authors are concerned with the behaviour of $Z_n(1)$, cf. (1.90), for a_s a sequence of constants as in Theorems 1.11-1.13, and ζ_t a linear stationary process defined as

$$\zeta_t = \sum_{j=0}^{\infty} \psi_{t-j}\varepsilon_j, \ \varepsilon_j \equiv i.i.d.(0, \sigma^2) \ , \ \sum_{j=0}^{\infty} \psi_j^2 < \infty \ . \tag{1.107}$$

Note that by Wold representation theorem any Gaussian stationary sequence satisfies the above representation, so that the assumptions of Fox and Taqqu (1987) are nested as a special case. For $g(\lambda)$ defined as in (1.105), $Q(n)$ as in (1.95), Giraitis and Surgailis (1990) prove the following

Theorem 1.15 *(Giraitis and Surgailis (1990))* Assume there exist constants α_1 and α_2, with $\alpha_1 < 1$, $\alpha_2 < 1$, $\alpha_1 + \alpha_2 < 1$, such that

$$f(\lambda) = O(\lambda^{-\alpha_1}),\ g(\lambda) = O(\lambda^{-\alpha_2}),\ \lambda \in [0, \pi]\ ; \tag{1.108}$$

assume also

$$\lim_{n\to\infty} \frac{tr(Q(n)A(n))}{n} \to 8\pi^3 \int_{-\pi}^{\pi} \{f(\lambda)g(\lambda)\}^2 d\lambda < \infty\ ; \tag{1.109}$$

then as $n \to \infty$,

$$Z_n(1) \Rightarrow N(0, \sigma_1^2)\ , \tag{1.110}$$

with

$$\sigma_1^2 = 16\pi^3 \int_{-\pi}^{\pi} (f(\lambda)g(\lambda))^2 d\lambda + cum_4(2\pi \int_{-\pi}^{\pi} f(\lambda)g(\lambda)d\lambda)^2\ . \tag{1.111}$$

In the same paper Giraitis and Surgailis (1990) derive also asymptotic normality of Whittle's estimates under the conditions of Theorem 1.15. In a related work, linearity is assumed also by Beran and Terrin (1994); these authors advocate a slightly modified version of Whittle's estimate which has some computational advantage for very long time series as those arising in communication engineering. On the other hand in the probabilistic literature the results of Theorem 1.15 are generalized to partial sums of Appell polynomials by Terrin and Taqqu (1991) and to continuous stochastic processes by Ginovian (1994).

In view of Theorems 1.11-1.15, we can say in short that the problem of estimation and inference is solved for fully parametric stationary long memory series. However, we note that these estimates will typically be inconsistent if the parametric model is misspecified, for instance if the short run dynamics is represented as a stationary $ARMA$ model and both the number of lags included are incorrect. Since fractional integration is a long run phenomenon, in that it relates to the behaviour of the series only at the smallest frequencies, it could be preferable to adopt a semiparametric approach and concentrate only on a proper subset of the available information. In other words, rather than specifying a full parametric model for the sequence ζ_t, we might impose conditions only on a degenerating band of low frequencies, such as (1.86), which is fulfilled for instance by fractional Gaussian noise (1.88) and stationary fractional $ARIMA$ models (1.89). A rigorous attempt to estimate d under (1.86) is provided by Robinson (1994a). We note firstly that for $\lambda = 0$ the spectral density has a singularity and therefore of course it cannot be estimated. On the other hand, by covariance stationarity the quantity

$$F(\lambda_m) = \int_0^{\lambda_m} f(s)ds \tag{1.112}$$

will be finite, and indeed under (1.86) $F(\lambda_m) \approx C\lambda_m^{1-2d}$; $F(\lambda_m)$ is a population analogue of the averaged periodogram $\widehat{F}(\lambda_m)$, cf. (1.20). Now introduce the bandwidth condition

$$\frac{1}{m} + \frac{m}{n} \to 0 \ , \ \text{as } n \to \infty \ . \tag{1.113}$$

Robinson (1994a) proves (a generalized version of) the following

Theorem 1.16 *(Robinson (1994a))* Assume that (1.86) and (1.113) hold, and

$$z_t = \mu + \sum_{j=0}^{\infty} \psi_j \xi_{t-j} \ , \ \sum_{j=0}^{\infty} \psi_j^2 < \infty \ , \tag{1.114}$$

where $E\xi_t^2 = \sigma^2$, while ξ_t and $\xi_t^2 - \sigma^2$ are martingale differences sequences with respect to the filtration $\Im_t = \sigma(\xi_t, \xi_{t-1}, ...)$. Then

$$\frac{\widehat{F}(\lambda_m)}{F(\lambda_m)} \to_p 1 \text{ as } n \to \infty \ . \tag{1.115}$$

Now note that for any $q > 0$, $F(q\lambda)/F(\lambda) \approx q^{2(1/2-d)}$. This suggests the Averaged

Periodogram Estimate (APE)

$$\widehat{d}_{mq} = \frac{1}{2} - \frac{\log\left\{\widehat{F}(q\lambda_m)/\widehat{F}(\lambda_m)\right\}}{2\log(q)} \ . \tag{1.116}$$

A consequence of Theorem 1.16 is the following

Theorem 1.17 *(Robinson (1994a))* Under the conditions of Theorem 1.16, as $n \to \infty$

$$\widehat{d}_{mq} \to_p d \ . \tag{1.117}$$

The issue of optimal bandwidth for the APE and for semiparametric estimates in general is discussed respectively in Robinson (1994c) and Giraitis, Robinson and Samarov (1995). Although the conditions for the consistency of $\widehat{d}_{mq}$ are very mild, the APE has been of limited use in practice, mainly because its limiting asymptotic distribution is very complicated (Lobato and Robinson (1996)). However, the relevance of Theorem 1.16 (which has been generalized to the multivariate case by Lobato (1997)) goes far beyond Theorem 1.17, as we shall see in the next chapters.

A semiparametric procedure which has met considerable success in practical applications is the log-periodogram regression proposed originally by Geweke and Porter-Hudak (1983). The idea is to estimate d from the least squares regression

$$\log I(\lambda_k) = C - 2d \log \lambda_k + v_k \ , \tag{1.118}$$

where by an analogy with the weak dependent case the v_k ($\approx \log(I(\lambda_k)/f(\lambda_k))$ were conjectured by Geweke and Porter-Hudak (1983) to be asymptotically *i.i.d.* for $k/n \to 0$ as $n \to \infty$. This conjecture is proven false by Robinson (1995a), where the validity of a generalized and modified version of (1.118) is studied, extending the result to vector valued time series, under the assumption of Gaussianity. Let ζ_t be a $p \times 1$ stationary process such that the following conditions holds on the spectral density matrix:

$$f_{aa}(\lambda) \approx C_a \lambda^{-2d_a} \text{ as } \lambda \to 0^+ \ , \ a = 1, ..., p \ . \tag{1.119}$$

Consider the OLS estimates of $\underline{C}= (C_1, ..., C_p)$ and $\underline{d}= (d_1, ..., d_p)$, given by

$$\begin{bmatrix} \widetilde{C} \\ \widetilde{d} \end{bmatrix} = vec\left\{H'W(W'W)^{-1}\right\} \ , \tag{1.120}$$

for

$$\begin{aligned} W &= (W_{l+1}, ..., W_m)' \ , \ H = (H_1, ..., H_p) \ , & (1.121) \\ W_k &= (1, -2\log\lambda_k)' \ , \ H_a = (\log I_{aa}(\lambda_{l+1}), ..., \log I_{aa}(\lambda_m)) \ . & (1.122) \end{aligned}$$

where l is a trimming number such that $1/l + l/m \to 0$ as $n \to \infty$, so that the very lowest frequencies are excluded from the regression. We have

Theorem 1.18 *(Robinson (1995a)* Under (1.86), Gaussianity of ζ_t and suitable regularity conditions on $f(\lambda)$ and l, m, n, we have as $n \to \infty$

$$2\sqrt{m}(\widetilde{d} - \underline{d}) \Rightarrow N(0, \mathcal{V}) \ , \tag{1.123}$$

where $\mathcal{V}$ is consistently estimated by the least squares residuals, i.e

$$\widetilde{\mathcal{V}} \to_p \mathcal{V} \ , \ \widetilde{\mathcal{V}} = \frac{1}{m-l} \sum_{k=l+1}^{m} \widetilde{V}_k \widetilde{V}_k' \ , \ \widetilde{V}_k = H_k - \widetilde{C} + 2\widetilde{d} \log \lambda_k \ .$$

A consequence of Theorem 1.18 is that a standard theory of inference will hold for classical testing procedures that have an asymptotic justification; for instance, for a suitable constraint matrix Q the null hypothesis $H_0 : Q\underline{d} = 0$ can be tested through the usual Wald statistic

$$\tau_W = \tilde{d}'Q' \left\{(0:Q)'((W'W)^{-1} \otimes \tilde{\mathcal{V}}))(0:Q')\right\}^{-1} Q\tilde{d} \Rightarrow \chi^2_g \,, \tag{1.124}$$

where g is the number of restrictions, i.e. the number of rows of the (full row rank) matrix Q; in particular, for fractional cointegration analysis it could be relevant to test the hypothesis that some or all of the components of the vector process z_t share the same d parameter (Chapter 4). In general, assume we have some a priori knowledge, such as $\underline{d}= R\theta$, for a known matrix R of appropriate dimensions; this information can be exploited together with a preliminary estimate of $\mathcal{V}$ to set up a GLS type estimate of the form

$$\tilde{\theta} = \left[\left\{\begin{array}{cc} I_p & 0 \\ 0 & R' \end{array}\right\}(W'W \otimes \tilde{\mathcal{V}}^{-1})\left\{\begin{array}{cc} I_p & 0 \\ 0 & R \end{array}\right\}\right]^{-1}\left\{\begin{array}{cc} I_p & 0 \\ 0 & R' \end{array}\right\} vec(\tilde{\mathcal{V}}^{-1}H'W) \,. \tag{1.125}$$

We have

Theorem 1.19 *(Robinson (1995a))* Under (1.86),(1.113), Gaussianity of ζ_t, and suitable regularity conditions on $f(\lambda)$ and l, m, n, as $n \to \infty$

$$2\sqrt{m}(\tilde{d}_{GLS} - \underline{d}) \Rightarrow N(0, R(R'\mathcal{V}^{-1}R)^{-1}R') \,, \tag{1.126}$$

where $\tilde{d}_{GLS} = R\tilde{\theta}$ and the covariance matrix is consistenty estimated by $R(R'\tilde{\mathcal{V}}^{-1}R)^{-1}R'$.

Since $\mathcal{V} - R(R'\mathcal{V}^{-1}R)^{-1}R'$ is a positive definite matrix, the GLS procedure achieves in this framework efficiency improvements of a standard typology.

The assumption of Gaussianity for the ζ_t is restrictive, but it is extremely convenient because in these estimates the periodogram enters the procedure in a highly nonlinear way. Velasco (1997a) obtained consistency for the log-periodogram regression under more general conditions, in particular relaxing Gaussianity to boundedness and integrability of the characteristic function of ζ_t, which implies ζ_t has a bounded continuous density. In order to obtain the asymptotic normality of the estimates, however, this author introduces a condition very close to Gaussianity, namely the existence of finite moments of all orders.

Consider now a sequence ζ_t satisfying condition (1.86), and introduce the *Gaussian semiparametric estimate*

$$\widehat{d} = \arg\min_{d\in\Theta} \mathcal{L}_m(d),\ \mathcal{L}_m(d) = \log(\frac{1}{m}\sum_{j=1}^{m}\frac{I(\lambda_j)}{\lambda_j^{-2d}}) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\ . \tag{1.127}$$

for Θ a compact subset of $(-\frac{1}{2},\frac{1}{2})$. If we were to assume the model $f(\lambda) = C\lambda^{-2d}$ over all frequencies in $(0,\pi)$ the estimate (1.127) with $m = n/2$ (or $m = (n-1/2)$ for n odd) would be a discretized form of the Whittle's estimate justified by Fox and Taqqu (1986) and Giraitis and Surgailis (1990). Robinson (1995b) considers instead the behaviour of (1.127) under (1.113), for ζ_t a linear process with innovations ξ_t assumed to be martingale differences with finite fourth moments. We have

Theorem 1.20 *(Robinson 1995b)* Under (1.86), (1.113), $-\frac{1}{2} < d < \frac{1}{2}$ and suitable regularity conditions on $f(\lambda)$ and m, for $\widehat{d}$ given by (1.127) we have, as $n \to \infty$

$$\sqrt{m}(\widehat{d} - d) \Rightarrow N(0,\frac{1}{4})\ . \tag{1.128}$$

The multivariate extension of Theorem 1.20 is provided by Lobato (1995).

Semiparametric estimates for $d < \frac{1}{2}$ are considered among the others also by Hall, Koul and Turlach (1996), Taqqu, Teverovsky and Willinger (1995), Teverovsky and Taqqu (1996), Giraitis, Robinson and Surgailis (1996), Hidalgo and Yajima (1997); Hidalgo (1997) is concerned with the estimation of the pole λ_0 for covariance stationary variables z_t with spectral density $f(\lambda)$ such that $f(\lambda) \approx C|\lambda - \lambda_0|^{-2d}$ as $\lambda \to \lambda_0$, with λ_0 unknown.

A major extension of Theorems 1.19 and 1.20 is given by Velasco (1996,1997b), building in part upon some ideas from Hurvich and Ray (1995). The literature reviewed to this stage was concerned with the covariance stationary case $d < \frac{1}{2}$; a definition of fractional integration for d which takes values on the entire real line is given in (1.2): this definition is not unique, however, and we might define as fractionally integrated of order d_I the process

$$x_t = \sum_{s=1}^{t} \zeta_s \ , \ t = 1, 2, ... \tag{1.129}$$

where $\zeta_s \sim I(d_I - 1)$ is a stationary long memory process, $\frac{1}{2} < d_I < \frac{3}{2}$. This is the definition adopted, for instance, by Hurvich and Ray (1995); the relationship between x_t and z_t as defined by (1.2) is investigated in detail in Chapter 2, but we anticipate that the two processes are in no sense equivalent. The discrete Fourier transform of x_t can be rewritten as

$$w_x(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \sum_{s=1}^{t} \zeta_s e^{-i\lambda_j s} \ , \tag{1.130}$$

which after some manipulations entails

$$EI_{xx}(\lambda_j) = \int_{-\pi}^{\pi} f_x(\lambda) K_n(\lambda - \lambda_j) d\lambda \ , \tag{1.131}$$

where $K_n(.)$ denotes the Fejer's kernel

$$K_n(\lambda) = \frac{1}{2\pi n} \left| \frac{\sin(n\lambda/2)}{\sin(\lambda/2)} \right|^2 , \tag{1.132}$$

and formally $f_x(\lambda) \triangleq |1 - e^{i\lambda}|^{-2} f_\zeta(\lambda)$, with

$$f_x(\lambda) \approx C\lambda^{-2d_I} \text{ as } \lambda \to 0^+ \ . \tag{1.133}$$

Although $f_x(\lambda)$ is not a proper spectral density, expression (1.133) suggests that the rationale for the log-periodogram and the Gaussian semiparametric procedure can still be valid for processes integrated of order $d_I > \frac{1}{2}$, at least for a certain interval of d_I values. This conjecture is proven correct by Velasco (1996, 1997b). This author firstly analyzes the log-periodogram estimates, and shows under regularity conditions that they can still be consistent for $d_I < 1$, and consistent and asymptotically normal for $d_I < \frac{3}{2}$ when a data taper is used (Velasco (1996)). Consistency and asymptotic normality of Gaussian semiparametric estimates of a positive d_I which can take values on the entire real line is demonstrated in Velasco (1997b), allowing for the existence of deterministic trends in x_t.

As a first attempt to narrow the gap between the literature on cointegration considered in Section 2 and the literature on long memory and fractionally integrated processes considered in this section, we shall now briefly concentrate on a few papers that concern regression models with fractionally integrated residuals and/or regressors. The properties of least squares for the regression model

$$y_t = \beta' w_t + e_t \ , \tag{1.134}$$

where e_t is stationary long memory and w_t is a vector of deterministic functions of time were firstly considered by Yajima (1988); OLS estimates are shown to be strongly consistent under broad conditions but no longer efficient with respect to GLS, so that Grenander's theorem (1954) does not hold (the rationale being that one of its condition fails, namely boundedness of the spectrum at the origin). In a later paper, Yajima (1991) shows that OLS and GLS are asymptotically normally distributed under conditions on the cumulants of all orders of e_t; the special case $w_t = \exp(i\lambda t)$ had been considered before by the same author (Yajima (1989)), providing a central limit theorem for discrete Fourier transform of long memory stationary variables *at fixed frequencies* λ, whereas Dahlhaus (1995) also investigates the behaviour of a form of generalized least squares with nonstochastic regressors and stationary long memory residuals.

Robinson and Hidalgo (1997) considered model (1.134) under the assumption that w_t and e_t are independent stationary long memory sequences. They consider the class of estimates

$$\widehat{\beta}_{\phi} = (\int_{-\pi}^{\pi} I_{ww}(\lambda)\phi(\lambda)d\lambda)^{-1} \int_{-\pi}^{\pi} I_{wy}(\lambda)\phi(\lambda)d\lambda \,, \tag{1.135}$$

which is OLS for $\phi(\lambda) = 1$, $-\pi \leq \lambda \leq \pi$; GLS for $\phi(\lambda) = f_e^{-1}(\lambda)$; and feasible GLS for $\phi(\lambda) = \widehat{f}_e^{-1}(\lambda)$, $\widehat{f}_e^{-1}(\lambda)$ being any consistent estimate of $f_e^{-1}(\lambda)$. Under these circumstances OLS need no longer be $\sqrt{n}$-consistent, while feasible GLS under suitable conditions converge at rate $\sqrt{n}$ to an asymptotic normal distribution.

As discussed in Chapter 2, model (1.134) when $w_t = y_{t-1}$ and $\beta = 1$ is considered by Sowell (1990), Chan and Terrin (1995), and Lubian (1996), with the purpose to investigate the robustness of unit root asymptotics to departures from the assumption that the innovations are weak dependent. The case of a nonstationary or "close" to nonstationarity w_t is also considered by Jeganathan (1996), as an application of ideas from Jeganathan (1980, 1988) and LeCam (1986). Dolado and Marmol (1996) attempted to extend the FM-OLS procedure to the case when the regressors w_t and the "cointegrating residuals" e_t are fractionally integrated, in the sense of (1.2) for $d > \frac{3}{2}$ and in the sense of (1.129) for $\frac{1}{2} < d < \frac{3}{2}$. The proof they provide is incorrect, however, and the possibility of a practical implementation of their ideas remains an open question. Recent surveys of further developments in the study of long memory processes in time series econometrics are in Robinson (1994b) and Baillie (1996).

Chapter 2

WEAK CONVERGENCE TO FRACTIONAL BROWNIAN MOTION[1]

2.1 INTRODUCTION

The second section of Chapter 1 discusses functional central limit theorems entailing convergence to Brownian motion of a normalized z_t in a suitable metric space, presenting also applications of these results to limit distribution theory for statistics that arise when investigating time series that have unit roots. In the univariate case, we can summarize this literature as follows. Suppose that the scalar sequence η_t is covariance stationary and has (without loss of generality) mean zero, and lag-τ autocovariance $\gamma(\tau) = \int_{-\pi}^{\pi} f(\lambda) \exp(i\lambda\tau) d\lambda$. Under the "short range dependence" assumption

$$0 < f(0) < \infty , \tag{2.1}$$

[1]**The content of this chapter is based on the papers "Alternative Definitions of Fractional Brownian Motion" and "A Functional Central Limit Theorem for Multivariate Fractional Processes"; both these papers are the outcome of joint work with P.M.Robinson.**

we have

$$n^{-\frac{1}{2}}\left(\sum_{t=1}^{[nr]}\eta_t\right) \Rightarrow B(r;2\pi f(0)), \quad \text{as } n\to\infty, \quad 0\le r\le 1\ , \tag{2.2}$$

under a variety of conditions, for example with η_t a linear process (e.g. Hannan (1979), Phillips and Solo (1992)), various mixing or functions-of-mixing processes (e.g. McLeish (1977), Herrndorf (1984)), Wooldridge and White (1988)), or with vector valued η_t (e.g. Phillips and Durlauf (1986)). We have also seen that in many cases of interest application of functional limit theorems of the form (2.2), and the continuous mapping theorem (see Billingsley (1968)), typically leads to limit distributions that are nonstandard functionals of Brownian motion.

The property (2.1) can be viewed as a mild form of short range dependence condition (while it is also possible to focus on behaviour at alternative frequencies λ); for instance, under Gaussianity it is known that (2.1) is implied if η_t is strong mixing with rate $\alpha_m = m^{-\gamma}$, any $\gamma > 0$ (Ibragimov and Rozanov, (1977)). Some of the work establishing (2.2) has allowed for forms of nonstationarity requiring $f(0)$ to have a broader interpretation, but nevertheless (2.1) still conveys a sense of weak dependence. While many standard time series models for η_t, including stationary and invertible mixed autoregressive moving averages, satisfy (2.1), recently there has been considerable interest in ones which do not, and exhibit instead long memory or long range dependence properties in the sense of Chapter 1, (1.86)/(1.87). Our purpose here is to bridge the gap between the material of Section 2 and 3 of the previous chapter, investigating weak convergence under fractional circumstances. Functional central limit theorems for partial sums of long range dependent innovations typically lead to an interest in forms of *fractional* Brownian motion; two alternative definitions of the latter process are discussed in Section 2 and Section 3, where we also show how these definitions occasionally led to some confusion in the econometric literature. We go on in Section 4 to establish a functional central limit theorem for a wide class of multivariate fractionally integrated processes; most proofs are collected in the Appendix.

2.2 "TYPE I" FRACTIONAL BROWNIAN MOTION

Mandelbrot and Van Ness (1968) introduced fractional Brownian motion $B(r;\sigma^2,d_I)$ which we present in the following form (see also Taqqu (1979), Samorodnitsky and Taqqu (1994)), for $\frac{1}{2} < d_I < \frac{3}{2}$:

$$B(r;\sigma^2,d_I) = \frac{1}{U(d_I)}\int_R \left[\{(r-s)_+\}^{d_I-1} - \{(-s)_+\}^{d_I-1}\right] dB(s;\sigma^2)\,, \quad r \in R\,, \tag{2.3}$$

with $B(0;\sigma^2,d_I) = 0$ a.s., and where $t_+ = \max(t,0)$,

$$U(d_I) = \left\{\frac{1}{2d_I-1} + T(d_I)\right\}^{\frac{1}{2}}, \quad T(d_I) = \int_0^\infty \left\{(1+s)^{d_I-1} - s^{d_I-1}\right\}^{\frac{1}{2}} ds\,. \tag{2.4}$$

We term $B(r;\sigma^2,d_I)$ "Type I" fractional Brownian motion.

For $d_I = 1$ (2.3) is interpreted as

$$B(r;\sigma^2,1) = \int_0^r dB(s;\sigma^2), \quad r \geq 0\,, \tag{2.5}$$

$$B(r;\sigma^2,1) = -\int_r^0 dB(s;\sigma^2), \quad r < 0\,, \tag{2.6}$$

so that standard Brownian motion $B(r;\sigma^2) = B(r;\sigma^2,1)$ is nested as a special case. For $d_I \neq 1$ can be formally interpreted as a fractional derivative or integral of $B(r;\sigma^2)$ in the sense of Weyl (1917). We can rewrite (2.3) as

$$\begin{aligned} B(r;\sigma^2,d_I) &= \frac{1}{U(d_I)}\int_{-\infty}^0 \left\{(r-s)^{d_I-1} - (-s)^{d_I-1}\right\} dB(s;\sigma^2) \\ &\quad + \frac{1}{U(d_I)}\int_0^r (r-s)^{d_I-1} dB(r;\sigma^2)\,, \end{aligned} \tag{2.7}$$

for $r \geq 0$ and

$$\begin{aligned} B(r;\sigma^2,d_I) &= \frac{1}{U(d_I)}\int_{-\infty}^{r}\left\{(r-s)^{d_I-1}-(-s)^{d_I-1}\right\}dB(s;\sigma^2)+ \\ &\quad -\int_{r}^{0}(-s)^{d_I-1}dB(s;\sigma^2)\,, \end{aligned} \tag{2.8}$$

for $r < 0$. It is then easily verified that $EB(r;\sigma^2,d_I)=0$, $r \in R$, and

$$EB(r;\sigma^2,d_I)B(r;\sigma^2,d_I)=\frac{1}{2}\left(|r_1|^{2d_I-1}+|r_2|^{2d_I-1}-|r_1-r_2|^{2d_I-1}\right), \quad r_1,r_2\in R\,. \tag{2.9}$$

Take for notational convenience $\sigma^2=1$, and identify $B(r;1,d_I)=B(r;d_I)$; the increment $B(r_2;d_I)-B(r_1;d_I)$, $r_2 \geq r_1$, has variance

$$E(B(r_2;d_I)-B(r_1;d_I))^2=|r_2-r_1|^{2d_I-1}\,. \tag{2.10}$$

Thus for integers $j=0,\pm1,\ldots$ the increments (which we term "type I fractional Gaussian noise")

$$b(j;d_I)=B(j+1;d_I)-B(j;d_I) \tag{2.11}$$

have zero mean, unit variance and autocovariance

$$Eb(j;d_I)b(k;d_I)=\frac{1}{2}\left(|j-k+1|^{2d_I-1}-2\,|j-k|^{2d_I-1}+|j-k-1|^{2d_I-1}\right)\,, \tag{2.12}$$

so they have standard Gaussian marginals and are stationary, with autocovariance function such that

$$Eb(j;d_I)b(j+\tau;d_I)\approx(2d_I-1)(d_I-1)\tau^{2d_I-3}, \qquad \text{as } \tau\to\infty\,. \tag{2.13}$$

Mandelbrot and Van Ness (1968) showed that $B(r;d_I)$ has almost all sample paths continuous, and is self similar with similarity parameter $H=d_I-\frac{1}{2}$; a process, $X(t)$, is said to be self-similar with similarity parameter H if the finite-dimensional joint distributions of $X(ct)$ are identical to those of $c^H X(t)$, for all $c > 0$. Verwaat (1985) showed that $B(r; d_I)$ can only be defined for $d_I > \frac{1}{2}$ in that for $d_I \leq \frac{1}{2}$ a self-similar process with stationary increments is almost surely identically zero, and also showed that for $\frac{1}{2} < d_I < \frac{3}{2}$ the paths of $B(r; d_I)$ have almost surely locally unbounded variation, in common with standard Brownian motion $B(r)$. Samorodnitsky and Taqqu (1994) indicate that $B(r; d_I)$ is not a unique representation of fractional Brownian motion, in that for any real α and β the process

$$\int_R \left[\alpha\left\{(r-s)_+^{d_I-1} - (-s)_+^{d_I-1}\right\} + \beta\left\{(r-s)_-^{d_I-1} - (-s)_-^{d_I-1}\right\}\right] dB(s) \ , \quad r \in R \ , \quad (2.14)$$

shares the same properties as $B(r; d_I)$, up to a multiplicative constant, where $t_- = -\min(t, 0)$; (2.14) provides a general expression for "moving average" representations of fractional Brownian motion. Samorodnitsky and Taqqu (1994) also discuss "harmonizable representations", which for real scalars α and β they present as

$$\int_{-\infty}^{\infty} \frac{e^{i\lambda r} - 1}{i\lambda} \left(\alpha(\lambda)_+^{-(d_I-1)} + \beta(\lambda)_-^{-(d_I-1)}\right) dM(\lambda) \ , \ r \in R \ , \qquad (2.15)$$

where $dM(\lambda)$ is a complex Gaussian random measure, such that

$$dM(\lambda) = d\overline{M(-\lambda)} \ , \ EdM(\lambda) = 0 \ , \ E|dM(\lambda)|^2 = d\lambda \ , \qquad (2.16)$$

the bar denoting complex conjugation, and

$$EdM(\lambda)^2 = EdM(\lambda)d\overline{M(\mu)} = 0 \ , \ \lambda \neq \mu \ . \qquad (2.17)$$

The representation (2.15) was introduced by Dobrushin (1979) and Dobrushin and Major (1979), while its equivalence (in the finite-dimensional distribution sense) with (2.14) was first proved by Taqqu (1979) .

We now consider how $B(r; d_I)$ describes the limiting behaviour of partial sums of long memory random variables ζ_t. The increment sequence $b(j; d_I)$ provides a clue as to the properties of ζ_t. Corresponding to (2.13), the stationary sequence $b(j; d_I)$ has spectral density $f(\lambda)$, $-\pi < \lambda \leq \pi$, satisfying

$$f(\lambda) \approx \left\{ \frac{(d_I - \frac{1}{2})\Gamma(2d_I - 1)\sin(\pi(d_I - \frac{1}{2}))}{\pi} \right\} \lambda^{2-2d_I} , \quad \text{as } \lambda \to 0^+ . \tag{2.18}$$

Thus it is seen that the $b(j; d_I)$ are long range dependent, in the sense of (1.86), after the identification $d = d_I - 1$. Correspondingly, the ζ_t whose partial sums are approximated by $B(r; d_I)$ typically have autocovariances $\gamma(\tau)$ that, up to a multiplicative constant, are approximated by the right side of (2.13) (and hence satisfy (1.87)), and/or have spectral density $f(\lambda)$ that is approximated, up to a multiplicative constant, by the right side of (2.18); also, from (2.13) it follows that $Var(\sum_1^n \zeta_t) \approx cn^{2d_I - 1}$ as $n \to \infty$. Thus we anticipate that under (2.18) and regularity conditions

$$(cn^{2d_I-1})^{-1/2} \left(\sum_{t=1}^{[nr]} \zeta_t \right) \Rightarrow B(r; d_I) \, , \quad 0 < c < \infty \, , \; 0 < r \leq 1 \, . \tag{2.19}$$

Davydov (1970) established (2.19) in the former case when ζ_t is a linear process with only square summable weights (which in itself permits long range dependence) and *i.i.d.* innovations ε_t. In particular, Davydov (1970) considers sequences such that

$$\zeta_t = \sum_{j=0}^{\infty} \psi_{t-j}\varepsilon_j \, , \; E|\varepsilon_j|^\gamma < \infty \, , \; \psi_j \approx j^{d_I - 2} \, , \tag{2.20}$$

$$\gamma > \max(4, \frac{-8(d_I - 1)}{2d_I - 1}) \, . \tag{2.21}$$

Gorodetskii (1977) extends Davydov's results establishing (2.19) under (2.20) and

$$\gamma > \max(2, \frac{1}{2d_I - 1}) \, . \tag{2.22}$$

Taqqu (1975) established (2.19) under a different type of condition on ζ_t. He assumed that $\zeta_t = G(v_t)$, where G is a possibly nonlinear function and v_t is a stationary Gaussian process with zero mean, unit variance and lag-τ autocovariance $\gamma(\tau) \approx c\tau^{2d_v - 1}$ as $\tau \to \infty$,

$0 < d_v < \frac{1}{2}$. Assuming $EG(v_t)^2 < \infty$, Taqqu introduced the Hermite rank of G: denoting by

$$H_j(x) = (-1)^j e^{\frac{1}{2}x^2} \left(\frac{d^j}{dx^j} \right) e^{-\frac{1}{2}x^2} \tag{2.23}$$

the j-th Hermite polynomial, and $J(j) = EG(v_t)H_j(v_t)$, then the Hermite rank of G is $m = \min_{j \geq 0}(j : J(j) \neq 0)$. Then (2.19) follows when $m = 1$; if $m \geq 2$, the series $G(v_t)$ is weakly dependent if $m > 1/(1-2d)$, otherwise the limit is non-Gaussian. The results of Davydov (1970), Gorodetskii (1977) and Taqqu (1975) are in fact more general than reported here because they allow for a slowly varying factor in $\gamma(\tau)$. Similar results have been given under various other conditions (e.g. Chan and Terrin, (1995), Csorgo and Mielniczuk, (1995)).

2.3 "TYPE II" FRACTIONAL BROWNIAN MOTION

Levy (1953), Mandelbrot and Van Ness (1968) mention an alternative definition of fractional Brownian motion, as a Holmgren-Riemann-Liouville fractional integral, which for $d > \frac{1}{2}$ we write as

$$W(r; \sigma^2, d) = (2d-1)^{\frac{1}{2}} \int_0^r (r-s)^{d-1} dB(s; \sigma^2), \quad r > 0 \ , \tag{2.24}$$

$$W(r; \sigma^2, d) = -(2d-1)^{\frac{1}{2}} \int_r^0 (s-r)^{d-1} dB(s; \sigma^2), \quad r < 0 \ , \tag{2.25}$$

with $W(r; \sigma^2, d) = 0$ a.s.; we call $W(r; \sigma^2, d)$ "Type II fractional Brownian motion". Clearly $W(r; \sigma^2, d)$ is again Gaussian with almost surely continuous sample paths, and for $d = 1$ (2.24) and (2.25) reduce to (2.5) and (2.6), thus nesting $B(r; \sigma^2)$ to the same extent as does $B(r; \sigma^2, d_I)$. As in Section 2 we take $\sigma^2 = 1$, and we modify the notation accordingly; we have

$$EW(r;d)=0\ ,\quad EW^2(r;d)=|r|^{2d-1}\ ,\quad r\in R\ , \tag{2.26}$$

so that (see (2.10)) the mean and variance of $W(r;d)$ have the same expression as those of $B(r;d_I)$. However, when $0<r_1<r_2$,

$$EW(r_1;d)W(r_2;d)=\frac{1}{2}\left(r_2^{2d-1}+r_1^{2d-1}-E\left(W(r_2;d)-W(r_1;d)\right)^2\right) \tag{2.27}$$

which does not agree with (2.9), because

$$\begin{aligned}E\left(W(r_2;d)-W(r_1;d)\right)^2 &= (2d-1)(\int_0^{r_1}\left\{(r_2-s)^{d-1}-(r_1-s)^{d-1}\right\}^2 ds+\int_{r_1}^{r_2}(r_2-s)^{2d-2}ds)\\ &= (2d-1)(r_2-r_1)^{2d-1}\times\\ &\quad (\frac{1}{2d-1}+\int_0^{r_1/(r_2-r_1)}\left\{(1+s)^{d-1}-s^{d-1}\right\}^2 ds)\ ,\end{aligned} \tag{2.28}$$

which is not the same as (2.10). Thus the increments of $W(r;d)$, even at equally-spaced intervals, are nonstationary, unlike those of $B(r;d_I)$, though

$$E\left(W(r_2;d)-W(r_1;d)\right)^2\approx(2d-1)U(d)^2(r_2-r_1)^{2d-1}\ \text{as}\ \frac{r_1}{r_2-r_1}\to\infty\ , \tag{2.29}$$

and

$$E\left(W(r_2;d)-W(r_1;d)\right)^2\approx(r_2-r_1)^{2d-1}\ \text{as}\ \frac{r_1}{r_2-r_1}\to 0\ , \tag{2.30}$$

the latter expression agreeing with (2.10) if we identify $d_I=d$. We can term "type II fractional Gaussian noise" the sequence of increments

$$w(j;d)=W(j+1;d)-W(j;d),\ j=1,2...\ . \tag{2.31}$$

The process $w(j;d)$ has mean zero and variance equal to $g(j)$, where

$$g(j) = 1 + (2d-1)\int_0^j \left\{(1+s)^{d-1} - s^{d-1}\right\}^2 ds \ , \ 1 \leq g(j) < (2d-1)U(d)^2 \ , \tag{2.32}$$

where $U(d)$ is defined in (2.4). The autocovariance between $w(j;d)$ and $w(k;d)$ is

$$Ew(j;d)w(k;d) = \frac{1}{2}(I - II - III + IV) \ , \ j > k+1 \ , \tag{2.33}$$

for

$$\begin{aligned}
I &= (j+1)^{2d-1} + (k+1)^{2d-1} - (j-k)^{2d-1} g(\frac{k+1}{j-k}) && (2.34)\\
II &= (j+1)^{2d-1} + k^{2d-1} - (j-k+1)^{2d-1} g(\frac{k}{j-k+1}) && (2.35)\\
III &= j^{2d-1} + (k+1)^{2d-1} - (j-k-1)^{2d-1} g(\frac{k+1}{j-k-1}) && (2.36)\\
IV &= j^{2d-1} + k^{2d-1} - (j-k)^{2d-1} g(\frac{k}{j-k}) \ , && (2.37)
\end{aligned}$$

with obvious modifications when $j = k+1$. It is then readily seen that as $\frac{k}{j-k} \to 0$

$$Ew(j;d)w(k;d) \approx \frac{1}{2}\left((j-k+1)^{2d-1} - 2(j-k)^{2d-1} + (j-k-1)^{2d-1}\right) \ , \tag{2.38}$$

which agrees with (2.12); in particular, we have, for $\tau, j \geq 0$ and $d = d_I$

$$Ew(0;d)w(\tau;d) = Eb(j;d_I)b(j+\tau;d_I) \ . \tag{2.39}$$

The greater dependence on the origin in $W(r;d)$, relative to $B(r;d_I)$, was offered as a criticism by Mandelbrot and Van Ness (1968). Another drawback with the joint distributions of $W(r;d)$ is that the self-similarity property only applies insofar as univariate marginal distributions are concerned. The possibility of providing type II fractional Brownian motion with a harmonizable representation in the sense of (2.15) is still an open field for research.

Three reasons can be advanced for interest in $W(r;d)$. First, it is defined on the same domain as $B(r)$. Second, while $B(r;d_I)$ is well defined only for $\frac{1}{2} < d_I < \frac{3}{2}$ (the integral $T(d_I)$ diverges when $d_I \geq \frac{3}{2}$), $W(r;d)$ is defined for all $d > \frac{1}{2}$. The value of this can be seen in connection with our third point, which indicates how $W(r;d)$ can describe the limit behaviour of certain nonstationary fractional processes. Define the process z_t as (1.2) in Chapter 1, i.e.

$$z_t = (1-L)^{-d}\eta_t^* \ , \quad t \geq 1 \ , \eta_t^* = \begin{cases} \eta_t, & t \geq 1 \\ 0 & t \leq 0 \end{cases} , \tag{2.40}$$

where the sequence η_t has zero mean and is covariance stationary with spectral density $f(\lambda)$ such that η_t is short range dependent, in the sense of (2.1). The convergence

$$\Gamma(d)\,(2d-1)^{\frac{1}{2}} n^{1/2-d} z_{[nr]} \triangleq \Gamma(d)\,(2d-1)^{\frac{1}{2}} z_n(r) \Rightarrow W(r; 2\pi f(0), d), \quad 0 \leq r \leq 1 \ , \tag{2.41}$$

has been discussed by Akonom and Gourieroux (1987) and Silveira (1991), the former assuming *i.i.d.* innovations, the latter considering more general absolutely regular processes. The convergence (2.41), combined with the continuous mapping theorem, is useful in characterizing the limit distribution of various statistics arising in inference on nonstationary, possibly fractionally integrated processes, especially in cointegration analysis of econometric time series (Chapter 4).

It is of some interest to remark that $W(r;d)$ is taken for granted as the proper definition of fractional Brownian motion in the bulk of the econometric time series literature, whereas the probabilistic literature focusses on $B(r;d_I)$. This dichotomy mirrors differing definitions of nonstationary fractionally integrated processes, as mentioned in Chapter 1. One definition, z_t, which led to $W(r;d)$, is given in (1.2)/(2.40). The other, which led to $B(r;d_I)$, was obtained in (1.129) where we set

$$x_t = \sum_{s=1}^{t} \zeta_s \ , \ t = 1, 2, \ldots \ , \tag{2.42}$$

where ζ_t is as before zero-mean, stationary long range dependent of order $d_I - 1$. Both x_t and z_t can be viewed as nonstationary fractionally integrated processes, but (2.42) allows only $d_I < \frac{3}{2}$. To compare (2.40) and (2.42) when $\frac{1}{2} < d < \frac{3}{2}$ rewrite (2.40) as

$$\begin{aligned} (1-L)z_t &= (1-L)^{d-1}\eta_t^* \\ &= \sum_{j=0}^{t-1} \psi_j \eta_{t-j} \, . \end{aligned} \tag{2.43}$$

Now in case x_t and z_t have the same order of integration ($d_I = d$) and the same short range dependent input, so that $\zeta_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$,(which satisfies (2.18), in view of (2.20)) we infer from (2.40) that

$$(1-L)(x_t - z_t) = \sum_{j=t}^{\infty} \psi_j \eta_{t-j}, \quad t \geq 1 \, . \tag{2.44}$$

In case the spectral density of η_t satisfies a stronger condition than short range dependence, namely $0 < f(\lambda) \leq C$, $\pi < \lambda \leq \pi$, (2.44) has variance

$$\begin{aligned} \int_{-\pi}^{\pi} \left| \sum_{j=t}^{\infty} \psi_j e^{ij\lambda} \right|^2 f(\lambda) d\lambda &\leq 2\pi C \sum_{j=t}^{\infty} \psi_j^2 \\ &= O\left(t^{2d-3}\right) \to 0 \end{aligned} \tag{2.45}$$

as $t \to \infty$, because $\psi_j = O(j^{d-2})$.

From (2.7)/(2.8) and (2.24) we may write down an identity between $B(r;d)$ and $W(r;d)$ for $r \geq 0$

$$\begin{aligned} B(r;d) &= \frac{1}{U(d)} \left\{ \frac{1}{\sqrt{(2d-1)}} W(r;d) + I(r;d) \right\} \\ &= \{1 + (2d-1)T(d)\}^{-1/2} W(r;d) + U(d)^{-1} I(r;d) \end{aligned} \tag{2.46}$$

where

$$I(r;d) = \int_{-\infty}^{0} \left\{ (r-s)^{d-1} - (-s)^{d-1} \right\} dB(s) \, . \tag{2.47}$$

Thus $B(r;d)$ is composed of two independent components, one of them a scaled $W(r;d)$.

Occasionally the different definitions of fractional Brownian motion on the one hand, and of fractionally integrated time series on the other, have led the definition (2.42) of fractionally integrated x_t to be incorrectly associated with $W(r;d)$. An important early theoretical econometric contribution in the literature is Sowell (1990), who considered the limiting distribution of the least square estimate of the coefficient of a first order autoregression in case the true coefficient is 1 and the innovations actually have long range or negative dependence. Sowell asserted under conditions assuring (2.18) and with $x_{[nr]}$ given by (2.42), that

$$\left(cn^{2d_I-1}\right)^{-1/2} x_{[nr]} \Rightarrow W(r;d_I),\ 0 \le r \le 1,\ \frac{1}{2} < d_I < \frac{3}{2}\ , \tag{2.48}$$

in contradiction to (2.19). Consequently Sowell's Theorem 3 requires correction by simply replacing $W(.,d)$ by $B(.;d)$. Related work of Dolado and Marmol (1996), Lubian (1996), Cappuccio and Lubian (1997), also appear to make use of (2.48) rather than (2.19) and can be corrected in a similar way. To be precise, Sowell considers (2.42) for $\zeta_t \sim I(d_I-1)$ and under the assumption that

$$\zeta_t = (1-L)^{-d_I-1}\eta_t\ ,\ \ \frac{1}{2} < d_I < \frac{3}{2}\ ,\ \eta_t \equiv i.i.d.\ ,\ E|\eta_t|^\gamma < \infty\ , \tag{2.49}$$

where γ satisfies (2.21). We can generalize Sowell's results as follows. Consider the case where (2.42) and (2.49) hold, but η_t rather than being *i.i.d.* is generated by the linear process

$$\eta_t = \vartheta(L)\xi_t,\ \vartheta(L) = \sum_{j=0}^{\infty}\vartheta_j L^j,\ \xi_t \sim i.i.d.(0,\sigma_\xi^2)\ , \tag{2.50}$$

$$\vartheta_j = O(j^{-\min(4,5-d_I)}),\ E|\xi_j|^\gamma < \infty,\ \gamma > \frac{1}{2d_I-1}\ . \tag{2.51}$$

We can rewrite

$$\zeta_t = (1-L)^{-d_I-1}\vartheta(L)\xi_t = \pi(L)\xi_t\ , \tag{2.52}$$

where $\pi(L) = \sum_j^\infty \pi_j L^j$ is such that $(1-L)^{-d_I-1}\vartheta(L) = \pi(L)$. From Lemma 2.4 in the Appendix we learn that $\pi_j \approx Cj^{d_I-2}$, and hence as a consequence of the results from Gorodetskii (1977), under (2.42) and (2.50)/(2.51) we obtain, as $n \to \infty$

$$n(\widehat{\varphi} - 1) \Rightarrow \frac{\frac{1}{2}B^2(1; d_I)}{\int_0^1 B^2(r; d_I)dr} \, , \; 1 < d_I < \frac{3}{2} \, , \tag{2.53}$$

$$n^{2d_I-1}(\widehat{\varphi} - 1) \Rightarrow -\frac{\frac{1}{2}\mathcal{C}}{\int_0^1 B^2(r; d_I)dr} \, , \; \frac{1}{2} < d_I < 1 \, , \tag{2.54}$$

where $\widehat{\varphi} = (\sum_{t=2}^n x_{t-1}^2)^{-1} \sum_{t=2}^n x_{t-1}x_t$ and

$$\mathcal{C} = \left(\lim_{n\to\infty} \frac{n^{2d_I-1}}{S_n^2}\right) \times \left(p \lim \frac{1}{n}\sum_{t=1}^n \zeta_t^2\right) \, , \; S_n^2 = Var(\sum_{t=1}^n \zeta_t) \, . \tag{2.55}$$

For $d_I > 1$ the fractional unit root distribution under these more general assumptions is the same as the result obtained by Sowell (1990). Also, by the ergodic theorem

$$p \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n \zeta_t^2 = \frac{\sigma_\xi^2}{2\pi}\int_{-\pi}^{\pi} |1 - e^{i\lambda}|^{-2d_I-2} |\vartheta(e^{i\lambda})|^2 d\lambda \text{ a.s. } . \tag{2.56}$$

Note that in a similar context, Chan and Terrin (1995) and Jeganathan (1996) make appropriate use of (2.19).

2.4 THE FUNCTIONAL CENTRAL LIMIT THEOREM

The purpose of this section is to establish a functional central limit theorem for multivariate fractionally integrated processes which will extend the works by Akonom and Gourieroux (1987) and Silveira (1991).

The main ingredient for the proof of (2.41) is the decomposition of $z_n(r)$ into two different elements, namely

$$z_n(r) = q_n(r) + r_n(r) , \tag{2.57}$$

$$\sup_{r\in[0,1]} |r_n(r)| = o_p(1) , \tag{2.58}$$

for $q_n(r) = (2d-1)^{1/2} \sum_{i=1}^{[nr]-1} (r - \frac{i}{n})^{d-1} w_i$, $w_i \equiv n.i.d.(0,1)$ and $r_n(r)$ a remainder term. Weak convergence of $q_n(r)$ to $W(r;d)$ entails proving convergence of the finite dimensional distributions and tightness of the sequence $q_n(r)$, as well-known; the argument provided by Akonom and Gourieroux (1987) to establish convergence of the finite dimensional distributions appeals directly to the definition of the stochastic integral $W(r;d)$ as the mean square limit of the partial sums $q_n(r)$. Here we disagree with their proof because $W(r;d)$ defined by (2.24) is not an Ito integral as a function of r, but only for fixed r as a function of s; more precisely, the integrand function $(r-s)^{d-1}$ in $W(r;d)$ is not adapted to the filtration $\Im_s = \sigma(B(\mu),\ \mu \leq s)$ as r varies in $[0,1]$. On the other hand, for (2.58) to obtain we need an approximation theorem for partial sums of short range dependent variables; Akonom and Gourieroux (1987) refer to results by Komlos, Major and Tusnady (1976) and Major (1976), which state that if $\{\nu_t\}_0^\infty$ is a sequence of *i.i.d.* random variables such that $E|\nu_1|^q < \infty$ for some $q > 2$, $E\nu_1 = 0$ and $E\nu_1^2 = 1$, we can construct a copy of $\{\nu_t\}_0^\infty$ (denoted $\{\widehat{\nu}_t\}_0^\infty$) and a sequence $\{w_t\}_0^\infty$ of *i.i.d.* random variables with $w_1 \equiv N(0,1)$ such that

$$\sup_{1\leq j\leq n} |\frac{\sum_{t=1}^{j} \widehat{\nu}_t - \sum_{t=1}^{j} w_t}{H(n)}| = o_p(1) \text{ a.s. } H(n) = n^{1/q} . \tag{2.59}$$

The results by Komlos, Major and Tusnady (1976) and Major (1976) are actually more general than reported here because they allow for a broader class of functions $H(n)$ than powers of n. The procedure we will pursue for the extension of (2.59) to the case of dependent innovations relies in part on ideas that have a systematic exposition in Phillips and Solo (1992) but can be traced back to Hannan (1970). We shall use the so-called

Beveridge-Nelson decomposition to rewrite partial sums of stationary linear sequences as partial sums of *i.i.d.* random vectors (to which multivariate results as in Einmahl (1989) can be applied) plus a stationary process, that we will bound conveniently under appropriate moment conditions. In the sequel, we indicate by $P \circ x$ the probability law of the random vector x; empty sums are taken to signify zero. The following assumptions introduce the generalization of (2.40) to the multivariate case

Assumption 2A1 For $d_a > \frac{1}{2}$, $a = 1, ..., p$, let

$$\Delta(L) = \sum_{j=0}^{\infty} \Delta_j L^j = diag\left\{(1-L)^{-d_1}, ..., (1-L)^{-d_p}\right\}, \ \Theta(L) = \sum_{j=0}^{\infty} \Theta_j L^j \ , \ \|\Theta(1)\| < \infty \ , \tag{2.60}$$

and

$$z_t = \Delta(L)\Theta(L)\eta_t^*, \ \eta_t^* = \begin{cases} \eta_t, & t \geq 0, \\ 0, & t < 0, \end{cases} \ , \ t = 0, \pm 1, ... \tag{2.61}$$

where

Assumption 2A2

$$\eta_t = A(L)\varepsilon_t \ , \ A(L) = \sum_{j=-\infty}^{\infty} A_j L^j \ , \ \|A(1)\| < \infty \ , \tag{2.62}$$

$$\sum_{j=0}^{\infty} \left\{ \sum_{k=j+1}^{\infty} \| A_k \|^2 + \sum_{k=j+1}^{\infty} \| A_{-k} \|^2 \right\} < \infty \, . \tag{2.63}$$

Assumption 2A3

$$\varepsilon_t \sim i.i.d. \ , \ E\varepsilon_1 = 0 \ , \ E\varepsilon_1\varepsilon_1' = \Sigma < \infty \ , \ E\|\varepsilon_1\|^q < \infty \ , \ q > 2 \ , \tag{2.64}$$

Assumption 2A4

$$rank(\Sigma) = rank(A(1)) = rank(\Theta(1)) = p \, . \tag{2.65}$$

We refer collectively to Assumptions 2A1-2A4 as Assumption 2A. For d_a a positive integer and $\Theta(.)$, $A(.)$ finite order matrix polynomials Assumption 2A covers for instance the class of Vector Autoregressive Integrated Moving Average processes ($VARIMA$), which extends to the multivariate case the well-known class of $ARIMA$ models; also, we note that for non-integer values of d_a, due to the truncation on the right hand side of (2.61), the class $\{z_t\}$ defined there is more general than

$$\widetilde{z}_t = \Delta(L)\widetilde{\eta}_t,\ \widetilde{\eta}_t = \begin{cases} \Theta(L)\eta_t, & t \geq 0, \\ 0, & t < 0, \end{cases},\ t = 0, \pm 1, \ldots . \tag{2.66}$$

The stationary linear specification adopted for $\{\eta_t\}$ in (2.62)/(2.64) is instrumental for Lemma 2.2 below; extensions to forms of asymptotic stationarity and stable heterogeneity can be allowed for, provided the approximation results of Lemma 2.2 are suitably generalized. (2.63) is a mild form of short range dependence condition, which is for instance implied by

$$\sum_{j=-\infty}^{\infty} j^{1/2}\|A_j\| < \infty\ , \tag{2.67}$$

compare Phillips and Solo (1992); (2.67) is implied, for instance, if the spectral density of η_t is differentiable at the origin. Condition (2.65) ensures that the asymptotic limit process will have nondegenerate finite dimensional distributions.

For Gaussian ε_t, the decomposition (2.57)/(2.58) is redundant and Assumption 2A suffices for weak convergence. We focus, however, on the more general case where the distribution of $\{\varepsilon_t\}$ is arbitrary; the following lemma follows from Theorems 1,2 and 4 in Einmahl (1989).

Lemma 2.1 *(Einmahl (1989))* Let $\{\varepsilon_t\} : \mathcal{S} \to R^p$ be a sequence of random vectors such that Assumptions 2A3 and 2B hold. Then we can construct a probability space $(\mathcal{S}_0, \Im_0, P_0)$ and two sequences of independent random vectors $\{\widehat{\varepsilon}_t\}$, $\{w_t\}$ with $P_0 \circ \widehat{\varepsilon}_t =$

$P \circ \varepsilon_t$, $P_0 \circ w_t = N(0, \Sigma)$, $n \in N$, such that

$$\| \sum_{t=1}^{n} \widehat{\varepsilon}_t - \sum_{t=1}^{n} w_t \| = o(n^{1/q}) \text{ a.s. } . \tag{2.68}$$

and

$$\max_{1 \le k \le n} \| \sum_{t=1}^{n} \widehat{\varepsilon}_t - \sum_{t=1}^{n} w_t \| = o_p(n^{1/q}) \; . \tag{2.69}$$

Results from Silveira (1991) suggest that under moment conditions stronger than Assumption 2A3 (2.68) can be extended for $q < 3$ to cover forms of dependence other than linearity, such as strong mixing; mixing conditions do not in general allow for more generality than linear models, however, (Pham and Tran, (1985)). We can now establish approximations for partial sums of multivariate linear sequences as follows.

Lemma 2.2 Under Assumptions 2A2-2A3 we can construct a probability space $(\mathcal{S}_0, \Im_0, P_0)$ and two sequences of independent random vectors $\{\widehat{\eta}_t\}$, $\{w_t\}$ with $P_0 \circ \widehat{\eta}_t = P \circ \eta_t$, $P_0 \circ w_t = N(0, \Sigma)$, $n \in N$, such that as $n \to \infty$, for $2 < s < q$

$$\begin{aligned} S(n) - V(n) &= o(n^{1/s}) \text{ a.s. } , & (2.70) \\ \sup_{j \le n} \frac{\|S(j) - V(j)\|}{n^{1/s}} &= o_p(1) \; , & (2.71) \end{aligned}$$

for $S(j) = \sum_{t=1}^{j} \widehat{\eta}_t$ and $V(j) = A(1) \sum_{t=1}^{j} w_t$.

Proof For scalar x, we have

$$A(x) = A(1) + (x - 1) \left\{ A^{+}(x) - A^{-}(x) \right\} \tag{2.72}$$

where

$$A^{+}(x) = \sum_{j=0}^{\infty} A_j^{+} x^j,\ A^{-}(x) = \sum_{j=0}^{\infty} A_j^{-} x^j, \tag{2.73}$$

$$A_j^{+} = \sum_{k=j+1}^{\infty} A_k,\ A_j^{-} = \sum_{k=j+1}^{\infty} A_{-k}\ . \tag{2.74}$$

Then

$$\sum_{t=1}^{n} \eta_t = A(1) \sum_{t=1}^{n} \varepsilon_t + \bar{\varepsilon}_{0n} \tag{2.75}$$

where $\bar{\varepsilon}_{0n} = \varepsilon_0^{+} - \varepsilon_0^{-} - \varepsilon_n^{+} + \varepsilon_n^{-}$, $\varepsilon_t^{+} = A^{+}(L)\varepsilon_t$, $\varepsilon_t^{-} = A^{-}(L)\varepsilon_t$. Also

$$E\|\bar{\varepsilon}_{0n}\|^q \le C\left\{E\|\varepsilon_0^{+}\|^q + E\|\varepsilon_0^{-}\|^q\right\}, \tag{2.76}$$

where

$$\begin{aligned} E\|\varepsilon_0^{+}\|^q &\le CE\left\{\sum_{j=0}^{\infty} \|A_j^{+}\|^2 \|\varepsilon_{-j}\|^2\right\}^{q/2} \\ &\le C\left\{\sum_{j=0}^{\infty} \left(E\|A_j^{+}\|^q \|\varepsilon_{-j}\|^q\right)^{2/q}\right\}^{q/2} \\ &\le C\left\{\sum_{j=0}^{\infty} \|A_j^{+}\|^2\right\}^{q/2} E\|\varepsilon_0\|^q < \infty \end{aligned} \tag{2.77}$$

using Burkholder's (1973) and Minkowski's inequalities. In the same way $E\|\varepsilon_0^{-}\|^q < \infty$. Thus $\bar{\varepsilon}_{0n} = o(n^{1/s})$ a.s. by Markov's inequality and the Borel-Cantelli lemma. The proof of (2.70) is completed by application of (2.68) to $A(1)\varepsilon_t$ and the identification $\widehat{\eta}_t = A(L)\widehat{\varepsilon}_t$, for $\widehat{\varepsilon}_t$ introduced in Lemma 2.1. To establish (2.71) write

$$S(j) - V(j) = A(1)\left\{\sum_{t=1}^{j} \widehat{\varepsilon}_t - \sum_{t=1}^{j} w_t\right\} + \bar{\varepsilon}_{0j}, \tag{2.78}$$

and hence $P(\sup_{j\le n}\|S(j)-V(j)\|>\lambda n^{1/s})$ is bounded by

$$P(\sup_{j\le n}\|A(1)(\sum_{t=1}^{j}\varepsilon_t-\sum_{t=1}^{j}w_t)\|>\frac{\lambda n^{1/s}}{2})+P(\sup_{j\le n}\|\bar{\varepsilon}_{0j}\|>\frac{\lambda n^{1/s}}{2}) \tag{2.79}$$

By 2A4, the first component of (2.79) is bounded by

$$P(\sup_{j\le n}\|(\sum_{t=1}^{j}\varepsilon_t-\sum_{t=1}^{j}w_t)\|>\delta\lambda n^{1/s})\ . \tag{2.80}$$

for some $0<\delta<\infty$. From (2.69) it follows that the w_t can be chosen as $n.i.d.(0,\Sigma)$ such that

$$\sup_{j\le n}\|(\sum_{t=1}^{j}\varepsilon_t-\sum_{t=1}^{j}w_t)\|=o_p(n^{1/s}) \tag{2.81}$$

because λ is arbitrary. The second component of (2.79) is bounded by

$$\begin{aligned}P(\sup_{j\le n}\|\varepsilon_{0j}\| &\ge \frac{\lambda n^{1/s}}{2})\le C\frac{E\sup_{j\le n}\|\varepsilon_{0j}\|^q}{(\lambda n^{1/s})^q}\\ &\le C\frac{nE(\|\varepsilon_0^+\|^q+\|\varepsilon_0^-\|^q)}{(\lambda n^{1/s})^q}\to 0 \end{aligned}\tag{2.82}$$

as $n\to\infty$ in view of previous evaluation. □

The only part of Lemma 2.2 which is used in the sequel is (2.71), but we have chosen to insert (2.70) to mirror the derivation of analogous results by Komlos, Major and Tusnady (1976) and Einmahl (1989). For a recent contribution on weak martingale approximations for sequences of linear processes see also Truong-Van (1995).

Assumption 2B For q defined by Assumption 2B we have

$$q>\max(2,\frac{2}{2d_*-1})\ ,\ d_*=\min_{1\le a\le p}(d_a)\ . \tag{2.83}$$

Assumption 2C For $\Theta(L)$ defined by Assumption 2A and $\vartheta_{ab,k}$ the a,b-th element of Θ_k , we have

$$|\vartheta_{ab,k}| \leq Ck^{-\delta_a} \ , \ a,b = 1,...,p \ , \ k = 1,2,... \ , \tag{2.84}$$

where $\delta_a = \max(4 - d_a, d_a)$.

In view of Assumption 2A3, Assumption 2B is vacuous for $d_* \geq 1$. In the literature on weak convergence there is typically an increasing relationship between the amount of "memory" one allows in the process and the order of the moment conditions imposed on the innovation sequence; for instance, in Herrndorf (1984) normalized partial sums of covariance stationary mixing sequences $\{\eta_t\}$ are considered, the argument to establish weak convergence requiring tighter moment conditions on η_t the smaller the mixing rate. On the other hand in (2.83) a larger amount of persistence, i.e. a larger d_*, entails weaker moment conditions, at least for $d < 1$. A heuristic explanation is as follows: while the mixing rate in classical central limit theorems does not affect the $\sqrt{n}$-normalization, here (Theorem 2.1) a lower value of d_* entails a smaller normalization, and hence tighter bounds on the remainder terms are needed (cf. also (2.21) and (2.22)). Assumption 2C imposes a mild upper bound on the asymptotic behaviour of the weight matrices Θ_k.

Define the normalizing matrix function $D(n; d_z)$ as

$$D(n; d_z) = diag\left\{ \left(\frac{n^{d_1 - 1/2}}{(2d_1 - 1)^{\frac{1}{2}}\Gamma(d_1)}\right)^{-1}, ..., \left(\frac{n^{d_p - 1/2}}{(2d_p - 1)^{\frac{1}{2}}\Gamma(d_p)}\right)^{-1} \right\} \ , \ d_z = (d_1, ..., d_p) \ , \tag{2.85}$$

and multivariate fractional Brownian motion for $r \geq 0$ as

$$W(r; d_z, \Omega) = \underline{0} \ , \text{ a.s.} \ , \ r = 0 \ , \tag{2.86}$$

$$W(r; d_z, \Omega) = \int_0^r G(r - s) dB(r; \Omega) \ , \ r > 0 \ , \tag{2.87}$$

where $\Omega = A(1)\Sigma A(1)'$ is a $p \times p$ full rank matrix by 2A4 and $G(.)$ has a, b-th element

$$G(r) = \left\{(2d_a - 1)^{\frac{1}{2}}\vartheta_{ab}(1)r^{d_a - 1}\right\}_{a,b}, \; a, b = 1, ..., p . \tag{2.88}$$

For $0 < r \leq 1$ we have $rank(G(r)) = rank(\Theta(1))$, for any d_z. Also, for $\{\widehat{\eta}_t\}$ introduced by Lemma 2.2 define a copy of $\{z_t\}$ as

$$\widehat{z}_t = \Delta(L)\Theta(L)\widehat{\eta}_t^*, \; \widehat{\eta}_t^* = \begin{cases} \widehat{\eta}_t, & t \geq 0, \\ 0, & t < 0, \end{cases}, \; t = 0, 1, \ldots . \tag{2.89}$$

Clearly the analysis of the weak convergence of a (suitably normalized) $\{\widehat{z}_t\}$ is sufficient to establish weak convergence for a (suitably normalized) $\{z_t\}$. Define for $0 \leq r \leq 1$

$$z_n(r) = D(n; d_z)\widehat{z}_{[nr]} , \tag{2.90}$$

and note that $z_n(r) \in D[0, 1]^p$ (Chapter 1). The proof of weak convergence in $D[0, 1]^p$ involves the same steps as for the univariate case (see e.g. Csörgö and Mielniczuk (1995)), namely convergence of the finite dimensional distributions and tightness of the components of $z_n(r)$.

Theorem 2.1 Under Assumptions 2A, 2B, and 2C, for $0 \leq r \leq 1$

$$z_n(r) \Rightarrow W(r; d_z, \Omega) \text{ as } n \to \infty , \tag{2.91}$$

where $\Rightarrow$ signifies convergence in the Skorohod J_1 topology of $D[0, 1]^p$.

Proof This proof draws from the argument of Akonom and Gourieroux (1987) for the univariate case, with modifications and corrections. For $S(j)$, $V(j)$ defined in Lemma 2.2 we can write

$$z_n(r) \;=\; D(n; d_z)\sum_{k=1}^{[nr]}\Pi_{[nr]-k}\left(S(k) - S(k-1)\right) ,$$

$$= \quad Q_{1n}(r) + Q_{2n}(r) + Q_{3n}(r) + Q_{4n}(r) + Q_{5n}(r) \ , \tag{2.92}$$

where the matrix sequence $\{\Pi_k\}$ is defined by $\Pi_k \triangleq \sum_{j=0}^{k} \Delta_j \Theta_{k-j}$, $k = 1, 2, ...,$ and

$$Q_{1n}(r) \quad = \quad \sum_{k=1}^{[nr]-1} G(r - k/n) n^{-1/2} \left[V(k) - V(k-1)\right] 1([nr] > 2) \ , \tag{2.93}$$

$$Q_{2n}(r) \quad = \quad D(n; d_z) \sum_{k=1}^{[nr]-1} \Pi_{[nr]-k} \left[(S(k) - S(k-1)) - (V(k) - V(k-1))\right] 1([nr] > 2) \tag{2.94}$$

$$Q_{3n}(r) \quad = \quad D(n; d_z) \sum_{k=1}^{[nr]} \left[\Pi_{[nr]-k} - \widetilde{G}(nr - k)\right] \left[V(k) - V(k-1)\right] 1([nr] > 2) \ , \tag{2.95}$$

$$Q_{4n}(r) \quad = \quad D(n; d_z) \left[S([nr]) - S([nr] - 1)\right] 1([nr] > 2) \ , \tag{2.96}$$

$$Q_{5n}(r) \quad = \quad D(n; d_z) \widehat{z}_{[nr]} 1([nr] \leq 2) \ , \tag{2.97}$$

where $1(.)$ is the indicator function and $\widetilde{G}(r) = \left\{\vartheta_{ab}(1) r^{d_a - 1}\right\}_{a,b}$, $a, b = 1, ..., p$, so that

$$D(n; d_z) \widetilde{G}(nr - k) = n^{-1/2} G(r - k/n) \ . \tag{2.98}$$

The result will follow if, as $n \to \infty$,

$$Q_{1n}(r) \quad \Rightarrow \quad W(r; d_z, \Omega), \ r \in [0, 1] \ , \tag{2.99}$$

$$\sup_{0 \leq r \leq 1} \|Q_{in}(r)\| \quad = \quad o_p(1) \ , \ i = 2, 3, 4, 5 \ . \tag{2.100}$$

Now (2.99) follows from Lemma 2.3, while (2.100) with $i = 2$ from Lemmas 2.4-2.6, with $i = 3$ from Lemma 2.7, and with $i = 4, 5$ from Lemma 2.8. □

Applications of Theorem 2.1 to asymptotic inference on frequency domain estimates for nonstationary time series are presented in Chapter 4. As noted by Akonom and Gourieroux (1987), for the special case where $d_* > 1$ a much simpler proof of Theorem 2.1 is made possible by Abel's formula of summation by parts and the continuous mapping theorem. Also, for the univariate case Silveira (1991) noted that the conditions on the moments of the innovation sequence $\{\varepsilon_t\}$ can be relaxed if we focus on the smoothed multivariate series

$$z^+_{[nr]} = \frac{1}{n}\sum_{t=1}^{[nr]} z_t \ , \ 0 \leq r \leq 1 \ . \tag{2.101}$$

Clearly for $r = 1$ we have $z^+_{[nr]} = \overline{z}_n$, i.e. the sample mean of $\{z_t\}$; the process $z^+_{[nr]}$ can be interpreted as representing the fluctuations of the partial means of z_t. We have the following

Corollary 2.1 Let Assumptions 2A and 2C hold. Then as $n \to \infty$, for $0 \leq r \leq 1$

$$nD(n; d_z^+) z^+_{[nr]} \Rightarrow W(r; d_z^+, \Omega) \ . \tag{2.102}$$

where $d_z^+ = (d_1 + 1, ..., d_p + 1)$.

Proof Rewrite

$$\begin{aligned} nD(n; d_z^+) z^+_{[nr]} &= D(n; d_z^+)(1 - L)^{-1} z_{[nr]} \\ &= D(n; d_z^+)\Delta^+(L)\Theta(L)\eta^*_{[nr]} \ , \end{aligned} \tag{2.103}$$

for $\Delta^+(L) = diag\left\{(1-L)^{-d_1-1}, ..., (1-L)^{-d_p-1}\right\}$ and $\Theta(L)$, $\eta^*_{[nr]}$ the same as in Assumption 2A. Hence $d_*^+ = \min_{1 \leq a \leq p}(d_a + 1) > \frac{3}{2}$, and then (2.102) follows immediately from Theorem 2.1. □

APPENDIX

Lemma 2.3 Under the assumptions of Theorem 2.1

$$Q_{1n}(r) \Rightarrow W(r;d_z,\Omega) \ , \ r \in [0,1] \ . \tag{2.104}$$

Proof Since $Q_{1n}(r)$ and $W(r;d,\Omega)$ are Gaussian, convergence of the finite dimensional distributions follows if we show equality of the first two moments. The fact that

$$\lim_{n\to\infty} EQ_{1n}(r) = EW(r;d_z,\Omega) = \underline{0} \ , \tag{2.105}$$

is obvious. Also, fix without loss of generality $r_2 \geq r_1$; from the well known identity $vec(ABC) = (C' \otimes A)vec(B)$, we have that

$$vec\left(EW(r_1;d_z,\Omega)W(r_2;d_z,\Omega)'\right) = \int_0^{r_1} \left(G(r_2-s)\otimes G(r_1-s)\right) vec(\Omega)ds \ . \tag{2.106}$$

Now note that the absolute value of the a,b-th component of $G(r-\frac{k}{n})$ is bounded by a costant if $d_a \geq 1$, and by the absolute value of the a,b-th component of $G(r-s)$ if $d_a < 1$, $\frac{k}{n} \leq s \leq \frac{k+1}{n}$; hence, by the dominated convergence theorem

$$\begin{aligned} vec\left(\lim_{n\to\infty} EQ_{1n}(r_1)Q_{1n}(r_2)'\right) &= \lim_{n\to\infty}\sum_{k=1}^{[nr_1]-1}\left(G(r_2-\frac{k}{n})\otimes G(r_1-\frac{k}{n})\right) vec(\Omega)\frac{1}{n} \\ &= \lim_{n\to\infty}\sum_{k=1}^{[nr_1]-1}\int_{k/n}^{(k+1)/n}\left(G(r_2-s)\otimes G(r_1-s)\right) vec(\Omega)ds \\ &= \int_0^{r_1}\left(G(r_2-s)\otimes G(r_1-s)\right) vec(\Omega)ds \ , \end{aligned} \tag{2.107}$$

so that (a) is established. As far as (b) is concerned, Akonom and Gourieroux (1987), p.13 show that a tightness criterion for Gaussian series is given by, for $a = 1,...,p$,

$0 \leq r_1 < r < r_2 \leq 1$

$$E|Q_{1n}^{(a)}(r) - Q_{1n}^{(a)}(r_1)|^2 E|Q_{1n}^{(a)}(r_2) - Q_{1n}^{(a)}(r)|^2 \leq C|r_2 - r_1|^{\gamma} \ , \ \gamma > 0 \ . \tag{2.108}$$

To prove (2.108), define for $0 \leq r \leq 1$, $a = 1, ..., p$,

$$R_{an}(r) = \sum_{k=1}^{[nr]} (r - \frac{k}{n})^{d_a - 1} \left(B(\frac{k}{n}) - B(\frac{k-1}{n}) \right) \ , \tag{2.109}$$

where $B(.)$ is as before univariate standard Brownian motion.

Consider first the case where $r_1 > 0$. The inequality (2.108) is trivial for any fixed n (n_0, say); we can take $n_0 = [\frac{2}{n}]$ and focus without loss of generality on $n > n_0$, so that $r_1 > \frac{2}{n}$ always holds. Note that, for a finite positive constant C depending in general on $\Omega, \Theta(1)$ and d_a but not on r_1, r_2, we have

$$\begin{aligned} & E\left\{Q_{1n}^{(a)}(r) - Q_{1n}^{(a)}(r_1)\right\}^2 E\left\{Q_{1n}^{(a)}(r_2) - Q_{1n}^{(a)}(r)\right\}^2 \\ \leq \ & CE\left\{R_{an}(r) - R_{an}(r_1)\right\}^2 E\left\{R_{an}(r_2) - R_{an}(r)\right\}^2 \ . \end{aligned} \tag{2.110}$$

From (2.109) we obtain easily, for $0 < \rho_1 < \rho_2 \leq 1$,

$$\begin{aligned} E\left\{R_{an}(\rho_2) - R_{an}(\rho_1)\right\}^2 &= \frac{1}{n} \sum_{k=1}^{[n\rho_1]} \left[(\rho_2 - \frac{k}{n})^{d_a-1} - (\rho_1 - \frac{k}{n})^{d_a-1} \right]^2 + \\ & \qquad \frac{1}{n} \sum_{k=1+[nr_1]}^{[ny_2]} (\rho_2 - \frac{k}{n})^{2d_a-2} \\ &= D_1(\rho_1, \rho_2) + D_2(\rho_1, \rho_2) \ . \end{aligned} \tag{2.111}$$

Now if $d_a > 2$, $D_1(\rho_1, \rho_2) \leq (d_a - 1)^2 (\rho_2 - \rho_1)^2$ by the mean value theorem and easy manipulations; if $1 < d_a \leq 2$, $D_1(\rho_1, \rho_2) \leq (\rho_2 - \rho_1)^{2d_a-2}$ from the inequality $|x + y|^{\theta} \leq |x|^{\theta} + |y|^{\theta}$, $1 \leq \theta \leq 2$; if $d_a = 1$, $D_1 = 0$. Finally, if $\frac{1}{2} < d_a < 1$, we note that for $\rho_2 > \rho_1$, $s \in (-\infty, \rho_1)$

$$f(s) = (\rho_1 - s)^{d_a-1} - (\rho_2 - s)^{d_a-1} \tag{2.112}$$

is non-decreasing, having derivative

$$(1-d_a)\left\{(\rho_1-s)^{d_a-2}-(\rho_2-s)^{d_a-2}\right\}>0\ . \tag{2.113}$$

Therefore

$$\begin{aligned} D_1(\rho_1,\rho_2) &\leq \int_0^{\rho_1} f(s)^2 ds \\ &= (\rho_2-\rho_1)^{2d_a-2}\int_0^{\rho_1}\left[(1+\frac{s}{\rho_2-\rho_1})^{d_a-1}-(\frac{s}{\rho_2-\rho_1})^{d_a-1}\right]^2 ds \\ &\leq (\rho_2-\rho_1)^{2d_a-1}\int_0^{\infty}\left[(1+v)^{d_a-1}-v^{d_a-1}\right]^2 dv = C(\rho_2-\rho_1)^{2d_a-1} \end{aligned} \tag{2.114}$$

because for $\frac{1}{2}<d_a<\frac{3}{2}$

$$\int_0^{\infty}\left[(1+v)^{d_a-1}-v^{d_a-1}\right]^2 dv<\infty\ , \tag{2.115}$$

as discussed in Samorodnitsky and Taqqu (1994). It follows that $D_1(\rho_1,\rho_2)\leq C(\rho_2-\rho_1)^{\gamma}$ for some $\gamma>0$. Let us now consider $D_2(\rho_1,\rho_2)$; we assume without loss of generality $[n\rho_2]>[n\rho_1]$. If $d_a\geq 1$, we have

$$\begin{aligned} D_2(\rho_1,\rho_2) &\leq \rho_2-\frac{[n\rho_1]+1}{n} \\ &\leq (\rho_2-\rho_1)+\frac{1}{n}\ . \end{aligned} \tag{2.116}$$

If instead $\frac{1}{2}<d_a<1$, we have that $(r-s)^{2d_a-2}$ is non-decreasing in s, for $s<r$. Hence

$$\begin{aligned} D_2(\rho_1,\rho_2) &\leq \int_{(1+[n\rho_1])/n}^{[n\rho_2]/n}(\rho_2-s)^{2d_a-2}ds \\ &\leq \int_{\rho_1}^{\rho_2}(\rho_2-s)^{2d_a-2}ds \\ &= \frac{1}{2d_a-1}(\rho_2-\rho_1)^{2d_a-1}\ , \end{aligned} \tag{2.117}$$

which. together with (2.116), gives

$$D_2(\rho_1, \rho_2) \le C(\rho_2 - \rho_1)^{\beta} + \frac{1}{n}, \text{ some } \beta > 0 \ . \tag{2.118}$$

for all $d_a > \frac{1}{2}$. Now we identify $\rho_2 = r$, $\rho_1 = r_1$ to bound the first element on the right hand side of (2.110), and $\rho_2 = r_2$, $\rho_1 = r$ to bound the second element on the right hand side of (2.110), so we consider together $D_1(r_1, r), D_2(r_1, r)$ and $D_1(r, r_2), D_2(r, r_2)$. For $r_2 - r_1 < \frac{1}{n}$ implies either $D_2(r_1, r) = 0$ or $D_2(r, r_2) = 0$; we assume $D_2(r_1, r) = 0$. Hence for $r_2 - r_1 > \frac{1}{n}$ we deduce from (2.116) and (2.117) that

$$\begin{aligned} E\{R_{an}(r) - R_{an}(r_1)\}^2 E\{R_{an}(r_2) - R_{an}(r)\}^2 &= CD_1(r_1, r)\left[D_1(r, r_2) + D_2(r, r_2)\right] \\ &\le C(r_2 - r_1)^{\gamma} \text{ some } \gamma > 0 \ . \end{aligned} \tag{2.119}$$

Otherwise, when $r_2 - r_1 > \frac{1}{n}$, we have

$$\begin{aligned} E\{R_{an}(r) - R_{an}(r_1)\}^2 E\{R_{an}(r_2) - R_{an}(r)\}^2 &\le C\max((r - r_1)^{\gamma}, (r_2 - r)^{\gamma}) + \frac{1}{n^2} \\ &\le C\max((r - r_1)^{\gamma}, (r_2 - r)^{\gamma}) + (r_2 - r_1)^2 \\ &\le C(r_2 - r_1)^{\gamma} \ , \ \gamma > 0 \ . \end{aligned} \tag{2.120}$$

The case $r_1 = 0$ can be dealt in a similar way; the result then follows from (2.110). □

Lemma 2.4 Let $d > -\frac{1}{2}$ and set

$$\delta = \begin{cases} 4, & d < 0 \ , \\ 4 - d, & 0 < d < 2 \ , \\ 2, & 2 < d \ , \end{cases} \ . \tag{2.121}$$

Consider the lag polynomial whose coefficients are determined by

$$\vartheta(x) = \sum_{k=0}^{\infty} \vartheta_k x^k, \ |\vartheta_k| \leq Ck^{-\delta}, \ |x| \leq 1, \tag{2.122}$$

and write

$$\pi_k = \sum_{j=0}^{k} \vartheta_{k-j} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}, \ k = 1, 2, \ldots. \tag{2.123}$$

Then as $k \to \infty$

$$|\pi_k - \frac{\vartheta(1)}{\Gamma(d)} k^{d-1}| \leq Ck^{d-2}. \tag{2.124}$$

Proof This lemma generalizes analogous results from Akonom and Gourieroux (1987), Silveira (1991) and Kokoszka and Taqqu (1995), on one hand allowing for algebraic rather than exponential decay of the coefficients $\{\vartheta_k\}$, on the other hand establishing a bound of order $O(j^{-\delta})$ rather than $O(j^{-1})$ for $d < \frac{1}{2}$. Note first that from Abramowitz and Stegun (1970), formula 6.1.47, there exist a constant C such that

$$|\frac{\Gamma(j+d)}{\Gamma(j+1)} - j^{d-1}| \leq Cj^{d-2}. \tag{2.125}$$

The left side of (2.124) is thus bounded by $\Gamma(d)^{-1}\{I + II + III + IV\}$, where

$$(I) = |\vartheta_k|\Gamma(d), \ (III) = \sum_{j=1}^{k} |\vartheta_{k-j}||j^{d-1} - k^{d-1}| \tag{2.126}$$

$$(II) = \sum_{j=1}^{k} |\vartheta_{k-j}||\frac{\Gamma(j+d)}{\Gamma(j+1)} - j^{d-1}|, \ (IV) = |\sum_{j=k}^{\infty} \vartheta_j| \frac{k^{d-1}}{\Gamma(d)}. \tag{2.127}$$

By (2.122), $I \leq Ck^{-\delta} \leq Ck^{d-2}$, and $IV \leq Ck^{d-\delta} \leq Ck^{d-2}$. By (2.125) $II \leq C\sum_{j=1}^{k} |\vartheta_{k-j}|j^{d-2}$. For $d \geq 2$ this is $O(k^{d-2})$ by summability of ϑ_j implied by (2.122). For $d < 2$ it is bounded by

$$\sum_{j=1}^{[k/2]} |\vartheta_{k-j}|j^{d-2} + \sum_{j=[k/2]}^{k} |\vartheta_{k-j}|j^{d-2} = \sum_{j=[k/2]}^{\infty} |\vartheta_j| + Ck^{d-2}\sum_{j=0}^{\infty} |\vartheta_j|, \tag{2.128}$$

and this is $O(k^{d-3} + k^{d-2}) = O(k^{d-2})$ by (2.122). For $d \geq 2$, by the mean value theorem

$III < Ck^{d-2}\sum_{j=1}^{k}|\vartheta_{k-j}|(k-j) \leq Ck^{d-2}$ by (2.122). Instead for $0 \leq d < 2$

$$\begin{aligned} III &\leq C\sum_{j=1}^{[k/2]}|\vartheta_{k-j}|(k-j) + Ck^{d-2}\sum_{j=[k/2]}^{k}|\vartheta_{k-j}|(k-j) \\ &\leq C(k^{-2}+k^{d-2}) \leq 2Ck^{d-2} \, . \end{aligned} \tag{2.129}$$

For $-\frac{1}{2} < d < 0$

$$III \leq \sum_{j=1}^{[k/2]}|\vartheta_{k-j}|(k^{1-d}-j^{1-d})(jk)^{d-1} + Ck^{d-2}\sum_{j=[k/2]}^{k}|\vartheta_{k-j}|(k-j) \, . \tag{2.130}$$

The second term is $O(k^{d-2})$, whereas the first is, by the mean value theorem, less than $Ck^{d-1}k^{-d}\sum_{j\geq[k/2]}j|\vartheta_j| \leq Ck^{-3}$. □

Lemma 2.5 Under the assumptions of Theorem 2.1, we have, for $\Pi_k = \{\pi_{ab,k}\}_{a,b}$, $a,b = 1,...,p$

$$\pi_{ab,k} - \frac{\vartheta_{ab}(1)}{\Gamma(d_i)}k^{d_a-1} = O(k^{d_a-2}) \ , \ \text{as } k \to \infty \, , \tag{2.131}$$

and for $0 \leq r \leq 1$

$$\|\Pi_{[nr]} - G(nr)\| = O(\sum_{i=1}^{p}[nr]^{d_a-2}) \, . \tag{2.132}$$

Proof Under Assumption 2C, (2.131) follows from a straightforward application of Lemma 2.4 to

$$\pi_{ab,k} = \sum_{j=0}^{k}\vartheta_{ab,k-j}\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} \quad , \ k = 1,2,... \, ; \tag{2.133}$$

for (2.132), rewrite

$$\|\Pi_{[nr]} - G(nr)\| \leq \|\Pi_{[nr]} - G([nr])\| + \|G([nr]) - G(nr)\| \, . \tag{2.134}$$

The first element on the right-hand side of (2.134) is $O(\sum_{a=1}^{p}[nr]^{d_a-2})$ by (2.131), while the second is $O(\sum_{a=1}^{p}[nr]^{d_a-2})$ by the mean value theorem, which entails that for $a = 1,...,p$

there exists a μ_a such that

$$(nr)^{d_a-1} = [nr]^{d_a-1} + (d_a-1)[nr+\mu_a]^{d_a-2}(nr-[nr]) \ , \ 0<\mu_a<1 \ . \ \square \tag{2.135}$$

Lemma 2.6 Under the assumptions of Theorem 2.1, we have

$$\sup_{0\le r\le 1} \|Q_{2n}(r)\| = o_p(1) \ . \tag{2.136}$$

Proof By Abel formula of summation by parts

$$\begin{aligned} Q_{2n}(r) &= D(n;d_z) \sum_{k=1}^{[nr]-1} \Pi_{[nr]-k} \left[(S(k)-S(k-1)) - V(k) - V(k-1)\right] \\ &= D(n;d_z) \sum_{k=1}^{[nr]-1} \left[\Pi_{[nr]-k} - \Pi_{[nr]-k-1}\right] [S(k)-V(k)] + \\ &\qquad D(n;d_z)\Pi_1 \left[S([nr]-1) - V([nr]-1)\right] \ . \end{aligned} \tag{2.137}$$

Define $\tilde{\Pi}_j = \Pi_j - \Pi_{j-1}$, where from (2.131) we obtain that $\tilde{\Pi}_j$ is such that

$$\|D(n;d_z)\tilde{\Pi}_j\| \le C \sum_{a=1}^{p} n^{1/2-d_a} j^{d_a-2} \ . \tag{2.138}$$

Thus

$$\begin{aligned} \|Q_{2n}(r)\| &\le C \sum_{j=1}^{[nr]-2} \left[\sum_{a=1}^{p} n^{1/2-d_a}([nr]-j)^{d_a-2} \|(S(j)-V(j))\|\right] + \\ &\qquad \sum_{i=1}^{p} n^{1/2-d_a} \|S([nr]-1) - V([nr]-1)\| \end{aligned} \tag{2.139}$$

and hence by Lemma 2.2

$$\sup_{0\le r\le 1} \|Q_{2n}(r)\| \ = \ o_p(1) n^{1/s} \sum_{a=1}^{p} \sum_{k=1}^{n-1} (n-k)^{d_a-2} n^{1/2-d_a}$$

$$= \quad o_p(1) n^{1/s} n^{-\min(1/2, d_* - 1/2)} = o_p(1) \,, \tag{2.140}$$

because by Assumption 2B we can choose $s > \max(2, 2/(2d_* - 1))$. □

Lemma 2.7 Under the assumptions of Theorem 2.1

$$\sup_{0 \le r \le 1} \|Q_{3n}(r)\| = o_p(1) \,. \tag{2.141}$$

Proof We have

$$\|Q_{3n}(r)\| \le \max_{j \le n} \|A(1) w_j\| \sup_{0 \le r \le 1} \sum_{k=1}^{[nr]-1} \|D(n; d_z)(\Pi_{[nr]-k} - \widetilde{G}(nr - k))\| \,. \tag{2.142}$$

As a consequence of Lemma 2.5,

$$\sup_{0 \le r \le 1} \|Q_{3n}(r)\| \le C \max_{j \le n} \|w_j\| \sup_{0 \le r \le 1} \sum_{a=1}^{p} n^{1/2 - d_a} \sum_{k=1}^{[nr]-1} ([nr] - k)^{d_{ia} - 2} \,, \tag{2.143}$$

which for $d_a > \frac{1}{2}$ is bounded by

$$c \sum_{a=1}^{p} n^{-\min(1/2, d_a - 1/2)} \max_{j \le n} \|w_j\| = C n^{-\delta} \max_{j \le n} \|w_j\|, \text{ some } \delta > 0 \,. \tag{2.144}$$

Denote by w_{aj} the a-th component of the vector process w_j , and recall that $w_{aj} \equiv n.i.d.(0, \sigma_a^2)$, for σ_a^2 the a-th element on the main diagonal of Σ ; now for any $\lambda > 0$, $i = 1, ..., p$

$$P\left\{n^{-\delta} \max_{j \le n} \|w_j\| > \lambda\right\} = O(n \sum_{a=1}^{p} P\left\{|w_{aj}| > \lambda n^{\delta}\right\} = o(n \sum_{a=1}^{p} e^{-(\lambda n^{\delta})/2\sigma_a^2}) = o(1) \,, \tag{2.145}$$

where the second step follows from the inequality

$$\int_{\mu}^{\infty} e^{-u^2/2} du < \frac{e^{-\mu^2/2}}{\mu} \,, \; \mu > 0 \,. \, \square \tag{2.146}$$

Lemma 2.8 Under the assumptions of Theorem 2.1

$$\sup_{0\le r\le 1} \|Q_{in}(r)\| = o_p(1)\ , \quad i = 4,5 \tag{2.147}$$

Proof For any $\lambda > 0$

$$\begin{aligned} P(\sup_{0\le r\le 1} \|Q_{4n}(r)\| &> \lambda) \le CP(\max_{k\le n} \|\widehat{\eta}_k\| > \lambda n^{d_*-1/2}) \\ &\le CnP(\|\widehat{\eta}_1\| \ge \lambda n^{d_*-1/2}) \\ &= O(n \times n^{q(1/2-d_*)}) = o(1)\ ; \end{aligned} \tag{2.148}$$

likewise

$$\begin{aligned} \sup_{0\le r\le 1} \|Q_{5n}(r)\| &= \max\{\|D(n;d_z)\widehat{\eta}_1\|, \|D(n;d_z)(\widehat{\eta}_1+\widehat{\eta}_2)\|\} \\ &= o_p(1)\ , \end{aligned} \tag{2.149}$$

because $E\|\widehat{\eta}_1+\widehat{\eta}_2\| = E\|\eta_1+\eta_2\| \le E\|\eta_1\| + E\|\eta_2\|$ and under Assumption 2A

$$E\|\eta_1\| = E\|\eta_2\| < \sum_{j=0}^{\infty} \|A_j\| E\|\varepsilon_{2-j}\| < \infty\ .\ \square \tag{2.150}$$

Chapter 3

THE AVERAGED PERIODOGRAM IN NONSTATIONARY ENVIRONMENTS[1]

3.1 INTRODUCTION

In Chapter 1 we introduced the discretely and continuously averaged periodogram matrices (also labelled as the integrated periodogram), which have long been known as fundamental statistics in time series analysis. The averaged periodogram of a stationary sequence has very much in common with the empirical distribution function of an *i.i.d.* sequence, and in fact this statistic is sometimes also called the empirical spectral distribution function (Anderson (1993), Kokoszka and Mikosch (1997)); this analogy has motivated many subsequent applications. Grenander and Rosenblatt (1984) constructed

[1] **The content of this chapter is based on the paper "Narrow Band Approximations of Sample Moments in the Presence of Deterministic and Stochastic Trends"; this paper is the outcome of joint work with Prof.P.M.Robinson.**

Kolmogorov-Smirnov type tests for the averaged periodogram of a finite variance process. Dahlhaus (1985) showed that the difference between the sample and process spectral distribution multiplied by the square root of the sample size converges weakly to a Gaussian process under several alternative conditions. Dahlhaus (1988) proved uniform central limit theorems for versions of the averaged periodogram indexed by set of functions (cf. Mikosch and Norvaisa (1995)). Anderson (1993) derived the limit distributions of test statistics of Kolmogorov-Smirnov and Cramer-Von Mises type for the averaged self-normalised periodogram. More recently nonstandard conditions, such as processes with infinite variance innovations, have also been considered. Kokoszka and Mikosch (1997) consider stationary linear sequences $\zeta_t = \sum_{j=0}^{\infty} \psi_j \xi_{t-j}$, where the sequence ψ_j need not be summable. In that paper, the self-normalised averaged periodogram for the strictly stationary linear process ζ_t is rewritten as

$$P_n(\lambda) = \int_{-\pi}^{\lambda} \frac{I_{\zeta\zeta}(s)}{|\sum_{j=0}^{\infty} \psi_j e^{-ijs}|^2} ds \ , \ -\pi < \lambda < \pi \ , \tag{3.1}$$

and the following approximation argument is shown to hold:

$$\begin{aligned} \int_{-\pi}^{\lambda} \frac{I_{\zeta\zeta}(s)}{|\sum_{j=0}^{\infty} \psi_j e^{-ijs}|^2} ds &\approx \int_{-\pi}^{\lambda} \frac{I_{\xi\xi}(s) |\sum_{j=0}^{\infty} \psi_j e^{-ijs}|^2}{|\sum_{j=0}^{\infty} \psi_j e^{-ijs}|^2} ds \\ &= \int_{-\pi}^{\lambda} I_{\xi\xi}(s) ds \\ &= (\lambda + \pi) c_{\xi\xi}(0) + 2 \sum_{\tau=1}^{n-1} \frac{\sin(\lambda\tau)}{\tau} c_{\xi\xi}(\tau) \ , \end{aligned} \tag{3.2}$$

where

$$c_{\xi\xi}(\tau) = n^{-1} \sum_{t=1}^{n-\tau} \xi_t \xi_{t+\tau} \ . \tag{3.3}$$

It is also shown that as $n \to \infty$ the process

$$\left(P_n(\lambda) - c_{\xi\xi}(0)\right)(\lambda + \pi) \ , \ -\pi < \lambda < \pi \tag{3.4}$$

weakly converges to a Brownian bridge process $\overset{\circ}{B}(r) \triangleq B(r) - rB(2\pi)$ defined on $[0, 2\pi]$ if $E\xi_t^2 < \infty$; otherwise the limit is non-Gaussian. Finally, applications of these results to Kolmogorov-Smirnov and Cramer-Von Mises tests of goodness of fit are considered (see also Kluppelberg and Mikosch (1998)).

The averaged periodogram is also the basis for the important Whittle estimate, which has been applied to the parameter estimation of long memory processes under Gaussianity (Fox and Taqqu (1986)) and under linearity (Giraitis and Surgailis (1990)), cf. Chapter 1. Kokoszka and Taqqu (1996) recently considered Whittle estimates for long memory processes with infinite variance, α-stable innovations.

The behaviour of the averaged periodogram on a degenerating band of low frequencies has also raised much interest, especially because of its applications in many semiparametric statistical procedures. In Chapter 1, we discussed the importance under short range dependence of $\widehat{F}(1, m)/\lambda_m$ as a consistent estimate for the spectral density matrix evaluated at the origin (Eicker (1967), Andrews (1992)); we also considered results for univariate and multivariate long memory time series (Robinson (1994a), Lobato (1997)), mentioning their application to the estimation of long memory parameters.

Under stationarity conditions, it seems therefore that there exists a large and distinguished literature on the asymptotic behaviour of the averaged periodogram matrix, covering also once unfamiliar circumstances as infinite variance innovations or processes with spectral singularities. Much less is known, on the other hand, on the asymptotic behaviour of $\widehat{F}(1, m)$ for variables that are not stationary and have moments that are not constant over time. To this area of research we point our attention in the next section, where we consider approximation of sample moments by narrow band periodogram averages in the presence of deterministic and stochastic trends. Stochastic trends are represented as moving averages of stationary innovations with algebraic weights (Theorems 3.1/3.2); this specification, as we shall discuss in the next chapter, covers multivariate fractionally integrated processes as introduced by Assumption 2A. We go on to cover smooth determinstic trends, for instance polynomial functions of time, in Theorem 3.3.

All proofs are collected in the Appendix.

3.2 NARROW-BAND APPROXIMATIONS OF SAMPLE MOMENTS

Consider a bivariate sequence $z_t = (z_{1t}, z_{2t})'$, $t = 1, 2, ...$, defined as

$$z_{at} = \sum_{j=1}^{t} \varphi_{a,t-j} \eta_{aj} \ , \ a = 1, 2 \ , \tag{3.5}$$

where we impose the following assumptions.

Assumption 3A (η_{1t}, η_{2t}), $t = 0, \pm 1, ...$, is a jointly covariance stationary process with zero mean and bounded spectral density matrix.

Assumption 3B For $a = 1, 2$, $0 \leq \gamma_2 \leq \gamma_1$, $\gamma_1 > \frac{1}{2}$, $\gamma_2 \neq \frac{1}{2}$, the sequences φ_{at} satisfy $\varphi_{at} = \varphi_t(\gamma_a)$, where for $t \geq 0$,

$$\begin{aligned} \varphi_t(\gamma) &= 1(t = 0) \ , \ \gamma = 0 \ , & (3.6) \\ &= O\left((1+t)^{\gamma - 1}\right) \ , \ \gamma > 0 \ , & (3.7) \\ &= 1 \ , \ \gamma = 1 \ , & (3.8) \\ \left|\varphi_t(\gamma) - \varphi_{t+1}(\gamma)\right| &= O\left(\frac{|\varphi_t(\gamma)|}{t}\right) \ , \ \gamma > 0 \ . & (3.9) \end{aligned}$$

As discussed in Chapter 4, Assumption 3B covers cases where z_{1t} is nonstationary whereas z_{2t} is either $I(0)$ $(\gamma_2 = 0)$, has asymptotically stationary long memory $(0 < \gamma_2 < \frac{1}{2})$, or is nonstationary $(\gamma_2 > \frac{1}{2})$.

We use here the abbreviated notation $\widehat{F}_{12}(.) = \widehat{F}_{z_1 z_2}(.)$, $I_{12}(.) = I_{z_1 z_2}(.)$, and we consider not only the statistic $\widehat{F}_{12}(1, m)$, but also $\widehat{F}_{12}(m+1, M)$, where $0 < m < M \leq$

$n/2$. The latter arises as follows. We have

$$\begin{aligned}\widehat{F}_{12}(1,m) &= \frac{\pi}{n}\sum_{j=1}^{m}\{I_{12}(\lambda_j)+I_{12}(\lambda_{n-j})\} \\ &= \frac{1}{2}\left\{\widehat{F}_{12}(1,m)+\widehat{F}_{12}(n-m,n-1)\right\} \\ &= \frac{1}{2}\widehat{F}_{12}(1,n-1)-\frac{1}{2}\widehat{F}_{12}(m+1,n-m-1)\,. \qquad (3.10)\end{aligned}$$

For n odd (3.10) is

$$\frac{1}{2}\widehat{F}_{12}(1,n-1)-\widehat{F}_{12}(m+1,(n-1)/2)\,, \qquad (3.11)$$

and for n even it is

$$\frac{1}{2}\widehat{F}_{12}(1,n-1)-\frac{1}{2}\widehat{F}_{12}(m+1,n/2)-\frac{1}{2}\widehat{F}_{12}(m+1,n/2-1)\,, \qquad (3.12)$$

where

$$\widehat{F}_{12}(1,n-1)=\frac{1}{n}\sum_{t=1}^{n}(z_{1t}-\overline{z}_1)(z_{2t}-\overline{z}_2)\ . \qquad (3.13)$$

The previous development follows Robinson (1994c). Because we are interested in approximations of an asymptotic nature, we introduce the following bandwidth condition

Assumption 3C

$$m\to\infty \text{ as } n\to\infty\ . \qquad (3.14)$$

Under Assumption 3C, we will deduce that, when $\gamma_1+\gamma_2>1$,

$$\widehat{F}_{12}(m+1,M)=o_p\left(n^{\gamma_1+\gamma_2-1}\right) \qquad (3.15)$$

by showing that both the mean and the standard deviation of the left side are $o\left(n^{\gamma_1+\gamma_2-1}\right)$. Thus if $\frac{1}{2}n^{1-\gamma_1-\gamma_2}\widehat{F}_{12}(1,n-1)$ has a nondegenerate limit distribution, $n^{1-\gamma_1-\gamma_2}\widehat{F}_{12}(1,m)$ shares it. For $\gamma_1+\gamma_2\le 1$, only the standard deviation of $\widehat{F}_{12}(m+1,M)$ is $o(n^{\gamma_1+\gamma_2-1})$

and it is necessary to estimate $E\widehat{F}_{12}(1, m)$, which differs non-negligibly from $E\widehat{F}_{12}(1, [(n-1)/2])$. We first consider the means.

Theorem 3.1 Under (3.5), Assumptions 3A, 3B and 3C, for $0 < m < M \leq n/2$

$$E\widehat{F}_{12}(m+1, M) = o\left(n^{\gamma_1+\gamma_2-1}\right) \ , \gamma_1 + \gamma_2 > 1 \ , \tag{3.16}$$

and

$$\begin{aligned} E\widehat{F}_{12}(1, m) &= O\left(n^{\gamma_1+\gamma_2-1}\right) \ , \gamma_1 + \gamma_2 > 1 \ , & (3.17) \\ &= O\left(\left(\frac{n}{m}\right)^{\gamma_1+\gamma_2-1}\right) \ , \gamma_1 + \gamma_2 < 1 \ . & (3.18) \end{aligned}$$

To consider the variances we impose the additional:

Assumption 3D (η_{1t}, η_{2t}) is fourth order stationary with bounded fourth-order cross-cumulant spectrum $f(\mu_1, \mu_2, \mu_3)$ satisfying

$$cum_4(j,k,l) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\mu_1, \mu_2, \mu_3) \exp(ij\mu_1 + ik\mu_2 + il\mu_3) \prod_{k=1}^{3} d\mu_k \ , \tag{3.19}$$

where $cum_4(j,k,l)$ is the fourth order cumulant of $\eta_{10}, \eta_{2j}, \eta_{1,j+k}, \eta_{2,j+k+l}$ for $j,k,l = 0, \pm 1, \ldots$.

Theorem 3.2 Under (3.5), Assumptions 3A, 3B, 3C and 3D, for $0 < m < M \leq n/2$, as $n \rightarrow \infty$

$$Var\left(\widehat{F}_{12}(m+1, M)\right) = o\left(n^{2(\gamma_1+\gamma_2-1)}\right) \tag{3.20}$$

and

$$Var\left(\widehat{F}_{12}(1, M)\right) = O\left(n^{2(\gamma_1+\gamma_2-1)}\right) \ . \tag{3.21}$$

In view of (3.13), $I_{12}(\lambda_j)$ distributes the sample covariance $\widehat{F}_{12}(1,n-1)$ across the Fourier frequencies λ_j, $j=1,...,n-1$. Theorems 3.1 and 3.2 suggest that $\widehat{F}_{12}(1,n-1)$ is dominated by the contributions from a possibly degenerating frequency band $(0,\lambda_m)$ when the collective memory in z_{1t}, z_{2t} is sufficiently strong $(\gamma_1+\gamma_2>1)$ while otherwise $\widehat{F}_{12}(1,m)-\frac{1}{2}\widehat{F}_{12}(1,n-1)$ is estimated by its mean, in view of (3.18).

For the final result of this chapter, we replace the stationary innovation sequence with uncorrelated errors that can exhibit non-trending heteroscedasticity; also, we allow for deterministic trends of a polynomial nature. More precisely, let us introduce the following

Assumption 3E Let $z_t=\chi_t+\omega_t$, where $\chi_t=(\chi_{1t},...,\chi_{pt})'$ is a p-dimensional deterministic function of t and ω_t is a real-valued, $p\times 1$ stochastic process such that

$$\begin{aligned}\omega_t &= \sum_{k=1}^{t}\Psi_{t-k}\varepsilon_k\ , &(3.22)\\ E\varepsilon_j\varepsilon_k' &= 0\ ,\ j\neq k\ ,\ E\|\varepsilon_k\|^2\leq C\ ,\ k\geq 1\ , &(3.23)\end{aligned}$$

with $\Psi_0=I_p$, $\Psi_{t-k}=\left\{\psi_{ab,t-k}\right\}$, $a,b=1,...,p$, and, for $\gamma_a>\frac{1}{2}$, $\delta_a>-\frac{1}{2}$,

$$\begin{aligned}|\psi_{ab,t}| &< C(1+t)^{\gamma_a-1}\ ,\ |\chi_{at}|<C(1+t)^{\delta_a}\ , &(3.24)\\ |\psi_{ab,t}-\psi_{ab,t+1}| &< C\frac{|\psi_{ab,t}|}{t}\ ,\ |\chi_{a,t}-\chi_{a,t+1}|<C\frac{|\chi_{a,t}|}{t}\ . &(3.25)\end{aligned}$$

Condition (3.23) is mild: the innovations ε_t need not be identically distributed nor satisfy martingale assumptions, also some stable heterogeneity is allowed for. Condition (3.24) sets an upper bound on the asymptotic behaviour of $\left\{\psi_{ab,k}\right\}$ and $\{\chi_{ak}\}$, which need not be decreasing sequences; (3.25) is a quasi-monotonic convergence condition (Yong (1974)). In view of (3.24)/(3.25), the sequences $\{\psi_{at}\}$ mirror the properties of φ_{at} as defined by

Assumption 3A; however, for simplicity, we consider here only the nonstationary case $\gamma_a > \frac{1}{2}$, $a = 1, ..., p$; also, because the innovations ε_t are now assumed to be uncorrelated, to allow for partial sums of short range dependent, possibly correlated processes the coefficients $\psi_{ab,t-k}$ need not be identically equal to unity for $\gamma_a = 1$. The deterministic component χ_{at} covers for instance the class

$$\chi_{at} = \ell(t)t^{\delta_a} \ , \ \delta_a > \frac{1}{2} \ , \tag{3.26}$$

where $\ell(.)$ is a bounded function varying slowly at infinity (Bingham et al., (1989)), i.e. a positive, measurable function such that $\ell(ct)/\ell(t) \approx 1$ as $t \to \infty$, for any positive c.

Theorem 3.3 Under Assumptions 3C and 3E, for $0 < m < M \leq n/2$, as $n \to \infty$ we have for $a, b = 1, ...p$

$$\widehat{F}^{ab}_{zz}(m+1, M) = o_p(n^{\gamma_a+\gamma_b-1} + n^{\delta_a+\delta_b}) \ , \tag{3.27}$$

for $\gamma_a, \gamma_b \neq 1$, $\delta_a, \delta_b \neq 0$, and

$$\widehat{F}^{ab}_{zz}(m+1, M) = o_p(n^{\gamma_a+\gamma_b-1+\varepsilon} + n^{\delta_a+\delta_b+\varepsilon}) \ , \tag{3.28}$$

any $\varepsilon > 0$, otherwise.

Theorem 3.3 provides the same sort of narrow-band approximation as Theorems 3.1 and 3.2, hence this result does not require any additional comment.

APPENDIX

To assist proof of Theorems 3.1, 3.2 and 3.3, we introduce the following Lemma.

Lemma 3.1 Let κ_t be a scalar deterministic sequence such that

$$|\kappa_t - \kappa_{t+1}| \leq C\frac{|\kappa_t|}{t}\ ,\ |\kappa_t| \leq C(1+t)^{\rho-1}\ ,\ t = 1, 2, \ldots\ .$$

Then

$$S_{uv}(\lambda) \triangleq \sum_{t=u}^{v} e^{it\lambda}\kappa_t \tag{3.29}$$

satisfies, for $0 \leq u < v$, $0 \leq |\lambda| \leq \pi$,

$$\begin{aligned} S_{uv}(\lambda) &= 1,\ u = 0\ , &(3.30)\\ &= 0,\ u > 0\ , &(3.31)\end{aligned}$$

and for $\rho > 0$

$$\begin{aligned} |S_{uv}(\lambda)| &\leq C\min\left(v^\rho,\ \frac{(u+1)^{\rho-1}}{|\lambda|},\ \frac{1}{|\lambda|^\rho}\right),\ 0 < \rho \leq 1\ , &(3.32)\\ |S_{uv}(\lambda)| &\leq C\frac{v^{\rho-1}}{|\lambda|}\ ,\ \rho > 1\ . &(3.33)\end{aligned}$$

Proof The proof for $\rho = 0$ is trivial, so we consider $\rho > 0$. Obviously $|S_{uv}(\lambda)| \leq Cv^\rho$. For $0 < \rho \leq 1$ we can write, for $u < s < v$,

$$S_{uv}(\lambda) = \sum_{t=u}^{s-1}\kappa_t e^{it\lambda} + \sum_{t=s}^{v-1}(\kappa_t - \kappa_{t+1})\sum_{v=s}^{t} e^{iv\lambda} + \kappa_v \sum_{t=s}^{v} e^{it\lambda} \tag{3.34}$$

by summation-by-parts. Thus because

$$|\sum_{v=s}^{t} e^{iv\lambda}| \leq \frac{C(t-s)}{1+(t-s)|\lambda|}, \ |\lambda| < \pi , \tag{3.35}$$

we have

$$|S_{uv}(\lambda)| \leq C\left((s-1)^{\rho} + \frac{s^{\rho-1}}{|\lambda|}\right) . \tag{3.36}$$

For $1/|\lambda| \leq Cv$ we may choose $s \approx |\lambda|^{-1}$ so that (3.36) is $O(|\lambda|^{-\rho})$. On the other hand we also have

$$S_{uv}(\lambda) = \sum_{t=u}^{v-1}(\kappa_t - \kappa_{t+1}) \sum_{s=u}^{t} e^{is\lambda} + \kappa_v \sum_{t=u}^{v} e^{it\lambda} , \tag{3.37}$$

to give $|S_{uv}(\lambda)| \leq C(u+1)^{\rho-1}/|\lambda|$. For $\rho > 1$, (3.37) gives instead

$$|S_{uv}(\lambda)| \leq Cv^{\rho-1}/|\lambda| . \tag{3.38}$$

□

Proof of Theorem 3.1 The discrete Fourier transform of z_{at} is, from (3.5),

$$\frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} z_{at} e^{it\lambda} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \varphi_{a,n-t}(\lambda) e^{it\lambda} \eta_{at} , \ a = 1, 2 , \tag{3.39}$$

where

$$\varphi_{at}(\lambda) \triangleq \sum_{s=0}^{t} \varphi_{as} e^{is\lambda} , \ a = 1, 2 . \tag{3.40}$$

Thus by Assumption 3A

$$EI_{12}(\lambda) = \frac{1}{2\pi n} \int_{-\pi}^{\pi} \Phi_1(\lambda, -\mu) \Phi_2(-\lambda, \mu) f_{12}(\mu) d\mu , \tag{3.41}$$

where $f_{ab}(\mu)$ is the cross spectral density of η_{at}, η_{bt} and

$$\Phi_a(\lambda, \mu) = \sum_{t=1}^{n} \varphi_{a,n-t}(\lambda) e^{it(\lambda+\mu)}, \ a = 1, 2 . \tag{3.42}$$

The modulus of (3.41) is bounded by

$$\frac{C}{n}\sup_{\mu}|f_{12}(\mu)|\left\{\int_{-\pi}^{\pi}|\Phi_1(\lambda,-\mu)|^2d\mu\int_{-\pi}^{\pi}|\Phi_2(-\lambda,\mu)|^2d\mu\right\}^{\frac{1}{2}}$$

$$\leq \frac{C}{n}\left\{\prod_{j=1}^{2}\sup_{\mu} f_{jj}(\mu)|\sum_{t=1}^{n}\varphi_{j,n-t}(\lambda)|^2\right\}^{\frac{1}{2}}, \tag{3.43}$$

from Assumption 3A. From Lemma 3.1, for $0<|\lambda|<\pi$, $t=1,...,n$ and $a=1,2$

$$\begin{aligned}|\varphi_{at}(\lambda)| &= O\left(\frac{n^{\gamma_a-1}}{|\lambda|}1(\gamma_a>1)+\frac{1}{|\lambda|^{\gamma_a}}1(\gamma_a\leq 1)\right)\\ &= O\left(\frac{n^{\max(\gamma_a-1,0)}}{|\lambda|^{\min(\gamma_a,1)}}\right)\end{aligned} \tag{3.44}$$

when $\gamma_a>0$. The latter bound also applies for $\gamma_a=0$, when it is $O(1)$. Thus (3.43) is, for $0<|\lambda|<\pi$,

$$O\left(\frac{n^{\max(\gamma_1-1,0)+\max(\gamma_2-1,0)}}{|\lambda|^{\min(\gamma_1,1)+\min(\gamma_2,1)}}\right). \tag{3.45}$$

For $\lambda=\lambda_j$, $j=1,...,M$, it is

$$O\left(\frac{n^{\gamma_1+\gamma_2}}{j^{\min(\gamma_1,1)+\min(\gamma_2,1)}}\right). \tag{3.46}$$

Hence, when $\gamma_1+\gamma_2>1$, by (3.14)

$$\begin{aligned}\left|E\widehat{F}_{12}(m+1,M)\right| &\leq \frac{2\pi}{n}\sum_{j=m+1}^{M}|EI_{12}(\lambda_j| \\ &\leq Cn^{\gamma_1+\gamma_2-1}\sum_{j=m}^{\infty}j^{-\min(\gamma_1,1)-\min(\gamma_2,1)}\\ &= o\left(n^{\gamma_1+\gamma_2-1}\right), \end{aligned} \tag{3.47}$$

$$\begin{aligned}\left|E\widehat{F}_{12}(1,m)\right| &\leq Cn^{\gamma_1+\gamma_2-1}\sum_{j=1}^{m}j^{-\min(\gamma_1,1)-\min(\gamma_2,1)}\\ &= O\left(n^{\gamma_1+\gamma_2-1}\right). \end{aligned} \tag{3.48}$$

Likewise, when $\gamma_1 + \gamma_2 < 1$,

$$
\begin{aligned}
\left|E\widehat{F}_{12}(1,M)\right| &\leq Cn^{\gamma_1+\gamma_2-1}\sum_{j=1}^{m}\frac{1}{j^{\gamma_1+\gamma_2}} \\
&= O\left(\left(\frac{n}{m}\right)^{\gamma_1+\gamma_2-1}\right) . \qquad (3.49)
\end{aligned}
$$

□

Proof of Theorem 3.2 We first assume that $\gamma_2 > 0$. From (3.5) and (3.41)

$$
I_{12}(\lambda) - EI_{12}(\lambda) = \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n}\varphi_{1,n-t}(\lambda)\varphi_{2,n-s}(-\lambda)e^{i(t-s)\lambda}\left\{\eta_{1t}\eta_{2s} - \gamma_{12}(s-t)\right\} , \quad (3.50)
$$

with $\gamma_{ab}(s-t) = E\eta_{at}\eta_{bs}$. The left hand side of (3.20) is thus bounded by the real part of

$$
\begin{aligned}
&\frac{1}{(2\pi)^2 n^4}\widetilde{\sum}\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1,n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j - i(r-q)\lambda_k} \\
&\qquad \times\left\{E\eta_{1t}\eta_{2s}\eta_{1r}\eta_{2q} - \gamma_{12}(s-t)\gamma_{12}(q-r)\right\} , \qquad (3.51)
\end{aligned}
$$

where

$$
\widetilde{\sum} = \sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{q=1}^{n} . \qquad (3.52)
$$

(3.51) can be written as $a_1 + a_2 + a_3$, where the three terms represent contributions from

$$
\gamma_{12}(q-t)\gamma_{12}(s-r),\ \gamma_{11}(r-t)\gamma_{22}(q-s),\ k(s-t, r-s, q-r) , \qquad (3.53)
$$

respectively, in the last line in (3.51). Now

$$
\begin{aligned}
a_1 &= \frac{1}{4\pi^2 n^4}\widetilde{\sum}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1,n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k) \\
&\qquad \times e^{i(t-s)\lambda_j - i(r-q)\lambda_k}e^{i(q-t)\lambda + i(s-r)\mu}f_{12}(\lambda)f_{12}(\mu)d\lambda d\mu \\
&= \frac{1}{4\pi^2 n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}\varphi_{1,n-t}(\lambda_j)\sum_{s=1}^{n}\varphi_{2,n-s}(-\lambda_j)e^{-is(\lambda_j-\mu)}
\end{aligned}
$$

$$\times \sum_{r=1}^{n} \varphi_{1.n-r}(-\lambda_j) e^{-ir(\lambda_k+\mu)} \sum_{q=1}^{n} \varphi_{2,n-q}(\lambda_k) e^{iq(\lambda_k+\lambda)} f_{12}(\lambda) f_{12}(\mu) d\lambda d\mu$$

$$= \frac{1}{4\pi^2 n^4} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{j=m+1}^{M} \sum_{k=m+1}^{M} \Phi_1(\lambda_j, -\lambda)\Phi_2(-\lambda_j, \mu)$$

$$\times \Phi_1(-\lambda_k, -\mu)\Phi_2(\lambda_k, \lambda) f_{12}(\lambda) f_{12}(\mu) d\lambda d\mu \, , \tag{3.54}$$

which is bounded in modulus by

$$\frac{C}{n^4} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{j=m+1}^{M} \Phi_1(\lambda_j, -\lambda)\Phi_2(-\lambda_j, \mu) \right|$$

$$\times \left| \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, -\mu)\Phi_2(\lambda_k, \lambda) \right| d\mu d\lambda$$

$$\leq \frac{C}{n^4} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{j=m+1}^{M} \Phi_1(\lambda_j, -\lambda)\Phi_2(-\lambda_j, \mu) \right|^2 d\mu d\lambda$$

$$= \frac{C}{n^4} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{j=m+1}^{M} \sum_{k=m+1}^{M} \Phi_1(\lambda_j, -\lambda)\Phi_2(-\lambda_j, \mu)$$

$$\times \Phi_2(-\lambda_k, \lambda)\Phi_2(\lambda_k, -\mu) d\mu d\lambda, \tag{3.55}$$

due to Assumption 3C and the fact that $\overline{\Phi_\ell(\lambda, \mu)} = \Phi_\ell(-\lambda, -\mu)$. Now for $\ell = 1, 2$

$$\int_{-\pi}^{\pi} \Phi_\ell(\lambda_j, -\lambda)\Phi_\ell(-\lambda_k, \lambda) d\lambda = \int_{-\pi}^{\pi} \sum_{t=1}^{n} \varphi_{\ell,n-t}(\lambda_j) e^{it(\lambda_j-\lambda)} \sum_{s=1}^{n} \varphi_{\ell,n-s}(-\lambda_k) e^{-is(\lambda_k-\lambda)} d\lambda$$

$$= 2\pi c_{jk,\ell} \, , \tag{3.56}$$

where

$$c_{jk,\ell} = \sum_{t=1}^{n} \varphi_{\ell,n-t}(\lambda_j)\varphi_{\ell,n-t}(-\lambda_k) e^{it(\lambda_j-\lambda_k)}, \; \ell = 1, 2 \, . \tag{3.57}$$

Thus (3.55) is

$$\frac{C}{n^4} \sum_{j=m+1}^{M} \sum_{k=m+1}^{M} c_{jk,1} c_{kj,2} \, . \tag{3.58}$$

Consider first the case $\gamma_1 + \gamma_2 > 1$. By (3.44) and elementary inequalities

$$|c_{jk,\ell}| \leq C \frac{n^{\max(\gamma_\ell,1)}}{(\lambda_j \lambda_k)^{\min(\gamma_\ell,1)}} \,, \tag{3.59}$$

so that (3.58) is bounded in modulus by

$$\begin{aligned} \frac{C}{n^2} \left(\sum_{j=m+1}^{M} \frac{n^{\max(\gamma_1-1,0)+\max(\gamma_2-1,0)}}{\lambda_j^{\min(\gamma_1,1)+\min(\gamma_2,1)}} \right)^2 &\leq Cn^{2(\gamma_1+\gamma_2-1)} \left(\sum_{j=m+1}^{M} j^{-\min(\gamma_1,1)-\min(\gamma_2,1)} \right)^2 \\ &\leq C \frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{\min(\gamma_1,1)+\min(\gamma_2,1)-1}} \\ &= o\left(n^{2(\gamma_1+\gamma_2-1)}\right) \,, \end{aligned} \tag{3.60}$$

using (3.14). Now consider the case $\gamma_1 + \gamma_2 \leq 1$ but $\frac{1}{2} < \gamma_1 < 1$ and $0 < \gamma_2 < \frac{1}{2}$ so that $\gamma_1 + \gamma_2 > \frac{1}{2}$. First we deduce from Lemma 3.1 the estimate

$$|c_{jk,1}| \leq \frac{Cn}{(\lambda_j \lambda_k)^{\gamma_1}} \,. \tag{3.61}$$

Define

$$\widetilde{\varphi}_{2t}(\lambda) = \varphi_{2,n-1}(\lambda) - \varphi_{2,n-t}(\lambda) = \sum_{s=n-t+1}^{n-1} \varphi_{2s} e^{is\lambda} \,,\ t = 2, ..., n \tag{3.62}$$

and $\widetilde{\varphi}_{21}(\lambda) = 0$, so that $c_{kj,2} = \sum_{i=1}^{4} c_{kj,2}^{(i)}$, where

$$c_{kj,2}^{(1)} = \varphi_{2,n-1}(\lambda_k)\varphi_{2,n-1}(-\lambda_j)D_n(\lambda_k - \lambda_j) \tag{3.63}$$

$$c_{kj,2}^{(2)} = \sum_{t=1}^{n} \widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} \tag{3.64}$$

$$c_{kj,2}^{(3)} = -\varphi_{2,n-1}(\lambda_k)\sum_{t=1}^{n} \widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} \tag{3.65}$$

$$c_{kj,2}^{(4)} = -\varphi_{2,n-1}(-\lambda_j)\sum_{t=1}^{n} \widetilde{\varphi}_{2t}(\lambda_k)e^{it(\lambda_k-\lambda_j)} \,, \tag{3.66}$$

with $D_t(\lambda) = \sum_{j=1}^{t} e^{ij\lambda}$. Because

$$
\begin{aligned}
D_n(\lambda_k - \lambda_j) &= n\,,\, j = k & (3.67)\\
&= 0\,,\, j \neq k\,,\ \mathrm{mod}\, n\,, & (3.68)
\end{aligned}
$$

using also (3.44) we have

$$
\begin{aligned}
c^{(1)}_{kj,2} &= n\left|\varphi_{2,n-1}(\lambda_j)\right|^2 = O\left(\frac{n}{\lambda_j^{2\gamma_2}}\right)\,,\, j = k\,, & (3.69)\\
&= 0\,,\, j \neq k\,.
\end{aligned}
$$

To consider $c^{(2)}_{kj,2}$, note that from Lemma 3.1, for $0 \leq q \leq n-1$,

$$
\begin{aligned}
\left|c^{(2)}_{kj,2}\right| &\leq \sum_{t=1}^{n-q} |\widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)| + \sum_{t=n-q+1}^{n} |\widetilde{\varphi}_{2t}(\lambda_k)\widetilde{\varphi}_{2t}(-\lambda_j)|\\
&\leq \frac{C}{\lambda_j\lambda_k}\sum_{t=1}^{n-q}(n-t+2)^{2(\gamma_2-1)} + \frac{Cq}{(\lambda_j\lambda_k)^{\gamma_2}}\\
&\leq C\left(\frac{q^{2\gamma_2-1}}{\lambda_j\lambda_k} + \frac{q}{(\lambda_j\lambda_k)^{\gamma_2}}\right)\\
&\leq \frac{C}{(\lambda_j\lambda_k)^{\gamma_2+1/2}}\,, & (3.70)
\end{aligned}
$$

on picking $q = [(\lambda_j\lambda_k)^{-1/2}]$. Next, to consider $c^{(3)}_{kj,2}$, we write

$$
\sum_{t=1}^{n} \widetilde{\varphi}_{2t}(-\lambda_j)e^{it(\lambda_k-\lambda_j)} = e^{i(n+1)(\lambda_k-\lambda_j)}\sum_{t=1}^{n-1} \varphi_{2t}e^{-it\lambda_j}D_t(\lambda_j - \lambda_k)\,, \tag{3.71}
$$

which by (3.35) is bounded in modulus by

$$
\sum_{t=1}^{n-1} |\varphi_{2t}|\,|D_t(\lambda_j - \lambda_k)| \leq C\frac{n^{\gamma_2+1}}{1+n\,|\lambda_j - \lambda_k|}\,, \tag{3.72}
$$

for $1 \le j, k \le n/2$. Thus by (3.44)

$$\left|c_{kj,2}^{(3)}\right| \le \frac{Cn^{\gamma_2+1}}{\lambda_j^{\gamma_2}(1+n|\lambda_j-\lambda_k|)}\,. \tag{3.73}$$

Likewise

$$\left|c_{kj,2}^{(4)}\right| \le \frac{Cn^{\gamma_2+1}}{\lambda_j^{\gamma_2}(1+n|\lambda_j-\lambda_k|)}\,. \tag{3.74}$$

With reference to (3.58) we have from (3.61) and (3.69)

$$\begin{aligned}
\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(1)}\right| &\le \frac{C}{n^2}\sum_{j=m+1}^{M}\lambda_j^{-2(\gamma_1+\gamma_2)} \\
&\le Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{M} j^{-2(\gamma_1+\gamma_2)} \\
&\le C\frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{2(\gamma_1+\gamma_2-1/2)}} = o\left(n^{2(\gamma_1+\gamma_2-1)}\right)\,.
\end{aligned} \tag{3.75}$$

From (3.61) and (3.70)

$$\begin{aligned}
\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(2)}\right| &\le \frac{C}{n^3}\left(\sum_{j=m+1}^{M}\lambda_j^{-\gamma_1-\gamma_2-\frac{1}{2}}\right)^2 \\
&\le Cn^{2(\gamma_1+\gamma_2-1)}\left(\sum_{j=m+1}^{M} j^{-\gamma_1-\gamma_2-\frac{1}{2}}\right)^2 \\
&\le C\frac{n^{2(\gamma_1+\gamma_2-1)}}{m^{2(\gamma_1+\gamma_2-1/2)}} = o\left(n^{2(\gamma_1+\gamma_2-1)}\right)\,.
\end{aligned} \tag{3.76}$$

From (3.61) and (3.73)

$$\left|\frac{1}{n^4}\sum_{j=m+1}^{M}\sum_{k=m+1}^{M} c_{jk,1}c_{kj,2}^{(3)}\right| \le Cn^{\gamma_2-2}\sum_{j=m+1}^{M}\lambda_j^{-\gamma_1-\gamma_2}\sum_{k=m+1}^{M}\frac{1}{(1+n|\lambda_j-\lambda_k|)\lambda_k^{\gamma_1}}$$

$$\le \; Cn^{2(\gamma_1+\gamma_2-1)}\sum_{j=m+1}^{\infty} j^{-\gamma_1-\gamma_2}\left\{j^{-\gamma_1}+\sum_{k=j+1}^{2j}(k-j)^{-1}k^{-\gamma_1}+\sum_{k=2j+1}^{M}(k-j)^{-1}k^{-\gamma_1}\right\}$$

$$
\begin{aligned}
&\leq \quad Cn^{2(\gamma_1+\gamma_2-1)} \sum_{j=m+1}^{\infty} j^{-\gamma_1-\gamma_2} \left\{ j^{-\gamma_1} + j^{-\gamma_1} \sum_{k=1}^{j} k^{-1} + \sum_{k=j}^{\infty} k^{-\gamma_1-1} \right\} \\
&\leq \quad Cn^{2(\gamma_1+\gamma_2-1)} \sum_{j=m+1}^{\infty} j^{-2\gamma_1-\gamma_2} \log j \\
&\leq \quad Cn^{2(\gamma_1+\gamma_2-1)} \sum_{j=m+1}^{\infty} j^{\varepsilon-2\gamma_1-\gamma_2} \qquad (3.77)
\end{aligned}
$$

for $\varepsilon > 0$. Now because $\gamma_1+\gamma_2 > \frac{1}{2}$ and $\gamma_1 > \frac{1}{2}$ we can choose ε such that $2\gamma_1+\gamma_2-\varepsilon > 1$ in which case (3.77) is $o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$. In view of (3.74) the same bound is obtained on replacing $c_{kj,2}^{(3)}$ by $c_{kj,2}^{(4)}$. Thus we have shown that $a_1 = o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$.

Next,

$$
\begin{aligned}
a_2 \quad &= \quad \frac{1}{n^4}\widetilde{\sum} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1.n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j-i(r-q)\lambda_k} \\
&\quad \times e^{i(r-t)\lambda+i(q-s)\mu} f_{11}(\lambda)f_{22}(\mu)d\lambda d\mu \\
&= \quad \frac{1}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{j=m+1}^{M}\sum_{k=m+1}^{M} \Phi_1(\lambda_j,-\lambda)\Phi_2(-\lambda_j,-\mu) \\
&\qquad\qquad \times \Phi_1(-\lambda_k,\lambda)\Phi_2(\lambda_k,\mu)f_{11}(\lambda)f_{22}(\mu)d\lambda d\mu \qquad (3.78)
\end{aligned}
$$

and this is bounded in modulus by (3.55) $= o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$, in the same way as was shown for (3.54).

Finally

$$
\begin{aligned}
a_3 = &\frac{1}{n^4}\widetilde{\sum} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \varphi_{1,n-t}(\lambda_j)\varphi_{2,n-s}(-\lambda_j)\varphi_{1.n-r}(-\lambda_k)\varphi_{2,n-q}(\lambda_k)e^{i(t-s)\lambda_j-i(r-q)\lambda_k} \\
&\qquad \times e^{i(s-t)\mu_1+i(r-s)\mu_2+i(q-r)\mu_3} f(\mu_1,\mu_2,\mu_3)\prod_{i=1}^{3} d\mu_i \\
= &\frac{1}{n^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{j=m+1}^{M}\sum_{k=m+1}^{M} \Phi_1(\lambda_j,-\mu_1)\Phi_2(-\lambda_j,\mu_1-\mu_2)
\end{aligned}
$$

$$\times \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3) f(\mu_1, \mu_2, \mu_3) \prod_{i=1}^{3} d\mu_i \tag{3.79}$$

and in view of Assumption 3E this is bounded in modulus by

$$\frac{C}{n^4} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{j=m+1}^{M} \Phi_1(\lambda_j, -\mu_1)\Phi_2(-\lambda_j, \mu_1 - \mu_2) \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3) \prod_{i=1}^{3} d\mu_i$$

$$\leq \frac{C}{n^4} \left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{j=m+1}^{M} \Phi_1(\lambda_j, -\mu_1)\Phi_2(-\lambda_j, \mu_1 - \mu_2) \right|^2 \prod_{i=1}^{3} d\mu_i \right)^{1/2}$$

$$\times \left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left| \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3) \right|^2 \prod_{i=1}^{3} d\mu_i \right)^{\frac{1}{2}} . \tag{3.80}$$

The second integral is bounded by

$$2\pi \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sum_{k=m+1}^{M} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_2(\lambda_k, \mu_3)$$

$$\times \sum_{j=m+1}^{M} \Phi_1(\lambda_k, \mu_3 - \mu_2)\Phi_2(-\lambda_k, -\mu_3) \prod_{i=1}^{3} d\mu_i . \tag{3.81}$$

Because (cf (3.56))

$$\int_{-\pi}^{\pi} \Phi_1(-\lambda_k, \mu_2 - \mu_3)\Phi_1(\lambda_j, \mu_3 - \mu_2) d\mu_2 = 2\pi c_{jk,1} , \tag{3.82}$$

it follows that (3.81) $= O\left(\sum\sum_{i,j=m+1}^{M} c_{jk,1}c_{kj,2}\right)$. Treating the other integral in (3.80) in the same way we see that (3.80) is bounded by (3.58) $= o\left(n^{2(\gamma_1+\gamma_2-1)}\right)$, to complete the proof of (3.20) when $\gamma_2 > 0$. For $\gamma_2 = 0$, the same proof applies on substituting 1 for $\Phi_{2t}(\lambda)$ and $D_n(\lambda+\mu)$ for $\Phi_2(\lambda, \mu)$, to deduce that $a_j = o(1)$ for $j = 1, 2, 3$.

To prove (3.21) we start by bounding the left side by an analogous expression to (3.51), with $\sum_{j=m+1}^{M}\sum_{k=m+1}^{M}$ replaced by $\sum_{j=1}^{m}\sum_{k=1}^{m}$ in $\widetilde{\sum}$. Thus the revised a_1 is bounded by

$$Cn^{2(\gamma_1+\gamma_2-1)}\left\{\sum_{j=1}^{M} j^{-2(\gamma_1+\gamma_2)}+\left(\sum_{j=1}^{M} j^{-\gamma_1-\gamma_2-\frac{1}{2}}\right)^2+\sum_{j=1}^{M} j^{-\gamma_1-2\gamma_2}\log j\right\}=O\left(n^{2(\gamma_1+\gamma_2-1)}\right), \tag{3.83}$$

while the revised a_2 and a_3 are similarly easily seen to have the same bound. □

Proof of Theorem 3.3 In view of (3.11)/(3.12), and because

$$\mid I_{zz}^{ab}(\pi)\mid\leq (I_{zz}^{aa}(\pi))^{1/2}(I_{zz}^{bb}(\pi))^{1/2} \tag{3.84}$$

and for any M, $m < M \leq (n-1)/2$

$$\widehat{F}_{zz}^{ab}(m+1,M)\leq\left[\widehat{F}_{zz}^{aa}(m+1,M)\right]^{\frac{1}{2}}\left[\widehat{F}_{zz}^{bb}(m+1,M)\right]^{\frac{1}{2}}, \tag{3.85}$$

it is not necessary to look at the behaviour of the cross-periodogram. We have

$$\widehat{F}_{zz}^{aa}(m+1,M) \;\leq\; 2\widehat{F}_{\omega\omega}^{aa}(m+1,M)+2\widehat{F}_{\chi\chi}^{aa}(m+1,M) \tag{3.86}$$

$$I_{zz}^{aa}(\pi) \;\leq\; 2I_{\omega\omega}^{aa}(\pi)+2I_{\chi\chi}^{aa}(\pi)\,. \tag{3.87}$$

Define $\psi'_{a,t-k}$ the a-th row of Ψ_{t-k}, and for any $\lambda\in(0,\pi]$ consider

$$\begin{aligned}
I_{\omega\omega}^{aa}(\lambda) &= \frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1}^{n} e^{i(s-t)\lambda}\sum_{j=1}^{s}\sum_{k=1}^{t}\psi'_{a,s-j}\varepsilon_j\varepsilon'_k\psi_{a,t-k}\\
&= \frac{1}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n} e^{i(s-t)\lambda}\sum_{k=1}^{\min(s,t)}\psi'_{a,s-k}\varepsilon_k\varepsilon'_k\psi_{a,t-k}+ \qquad (3.88)\\
&\qquad \frac{1}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n} e^{i(s-t)\lambda}\sum_{j=1}^{s}\sum_{k=1,k\neq j}^{t}\psi'_{a,s-j}\varepsilon_j\varepsilon'_k\psi_{a,t-k}\,. \qquad (3.89)
\end{aligned}$$

The expected value of the above, non-negative scalar random variable is equal to the expected value of (3.88), which is bounded by

$$EI_{ww}^{aa}(\lambda) \le \frac{C}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n} e^{i(s-t)\lambda}\sum_{k=1}^{\min(s,t)}\sum_{b=1}^{p}\sum_{c=1}^{p}\psi_{ab,s-k}\psi_{ac,t-k} \,. \tag{3.90}$$

Defining $\psi_{ab.k}=0,\ k<0$, we get

$$\begin{aligned} EI_{ww}^{aa}(\lambda) &= O(\frac{1}{n}\sum_{k=1}^{n}|\sum_{b=1}^{p}\sum_{t=1}^{n}e^{it\lambda}\psi_{ab,t-k}|^2) \\ &= O(\frac{1}{n}\sum_{k=1}^{n}|\sum_{b=1}^{p}\sum_{t=1}^{n}e^{i(t-k)\lambda}\psi_{ab,t-k}|^2) \\ &= O(\frac{p}{n}\sum_{k=1}^{n}\sum_{b=1}^{p}|\sum_{t=1}^{n-k}e^{it\lambda}\psi_{ab,t}|^2)\,. \end{aligned} \tag{3.91}$$

On the other hand, for the deterministic χ_t

$$I_{\chi\chi}^{aa}(\lambda) = \frac{1}{2\pi n}|\sum_{t=1}^{n}e^{it\lambda}\chi_{at}|^2\,. \tag{3.92}$$

Clearly to estimate (3.91) and (3.92) we need to estimate $\sum_{t=u}^{v}e^{it\lambda}\kappa_t$ for a deterministic sequence κ_t and $0\le u\le v\le n$ such that

$$|\kappa_t|\le C(1+t)^{\rho-1}\,,\ |\kappa_t-\kappa_{t-1}|\le C\frac{|\kappa_t|}{t}\,,\ \rho>\frac{1}{2}\,. \tag{3.93}$$

For $\rho\ne 1$, this issue was considered in Lemma 3.1, where we showed that for $0\le\lambda\le\pi$,

$$\begin{aligned} |\sum_{t=u}^{v}e^{it\lambda}\kappa_t| &\le \frac{C}{\lambda^\rho}\,,\ \frac{1}{2}<\rho<1\,, &(3.94)\\ |\sum_{t=u}^{v}e^{it\lambda}\kappa_t| &\le C\frac{v^{\rho-1}}{|\lambda|}\,,\ \rho>1\,. &(3.95) \end{aligned}$$

For $\rho=1$

$$|\sum_{t=u}^{v}e^{its}\kappa_t|\le\frac{C}{|\lambda|}(\sum_{t=u}^{v-1}t^{-1}+1)\le C\frac{\log v}{|\lambda|}\,, \tag{3.96}$$

though this is not sharp when κ_t is identically equal to 1, i.e. for the simple random walk case for ω_t, or the equivalent case for χ_t, when $C/|\lambda|$ was obtained. Now in view of (3.24)/(3.25), it follows that

$$\begin{aligned}
EI^{aa}_{\omega\omega}(\lambda) &\leq \frac{C}{|\lambda|^{2\gamma_a}} , \ \frac{1}{2} < \gamma_a < 1 , && (3.97)\\
&\leq C\frac{(\log n)^2}{\lambda^2} , \ \gamma_a = 1 , && (3.98)\\
&\leq \frac{Cn^{2\gamma_a - 2}}{\lambda^2} , \ \gamma_a > 1 , && (3.99)\\
I^{aa}_{\chi\chi}(\lambda) &\leq \frac{C}{n|\lambda|^{2\delta_a+2}} , \ -\frac{1}{2} < \delta_a < 0 , && (3.100)\\
&\leq \frac{C}{n}\frac{(\log n)^2}{\lambda^2} , \ \delta_a = 0 , && (3.101)\\
&\leq \frac{Cn^{2\delta_a - 1}}{\lambda^2} , \ \delta_a > 0 , && (3.102)
\end{aligned}$$

for $0 < \lambda \leq \pi$. Thus

$$\begin{aligned}
E\widehat{F}^{aa}_{\omega\omega}(m+1, M) &\leq C(\frac{n}{m})^{2\gamma_a - 1} , \ \frac{1}{2} < \gamma_a < 1 , && (3.103)\\
&\leq \frac{Cn(\log n)^2}{m} , \ \gamma_a = 1 , && (3.104)\\
&\leq \frac{Cn^{2\gamma_a - 1}}{m} , \ \gamma_a > 1 , && (3.105)
\end{aligned}$$

and

$$\begin{aligned}
\widehat{F}^{aa}_{\chi\chi}(m+1, M) &\leq \frac{Cn^{2\delta_a}}{m^{2\delta_a+1}} , \ -\frac{1}{2} < \delta_a < 0 , && (3.106)\\
&\leq \frac{C(\log n)^2}{m} , \ \delta_a = 0 , && (3.107)\\
&\leq \frac{Cn^{2\delta_a}}{m} , \ \delta_a > 0 . && (3.108)
\end{aligned}$$

Thus under Assumption 3C

$$\widehat{F}^{aa}_{zz}(m+1, M) = o_p(n^{2\gamma_a - 1} + n^{2\delta_a}) \qquad (3.109)$$

for $\gamma_a \in (\frac{1}{2}, \infty)/1$, $\delta_a \in (-\frac{1}{2}, \infty)/0$,with obvious modifications for $\gamma_a = 1$, $\delta_a = 0$. $\square$

Chapter 4

SEMIPARAMETRIC FREQUENCY DOMAIN ANALYSIS OF FRACTIONAL COINTEGRATION[1]

4.1 INTRODUCTION

Several motivations for the generalizations of cointegration analysis to fractional circumstances were discussed in Chapter 1. We recall from Definition 1.1 that $z_t \sim FCI(d_1, ..., d_p; d_e)$ if its a-th element $z_{at} \sim I(d_a)$, $d_a > 0$, $a = 1, ..., p$, and there exists a $p \times 1$ vector $\alpha \neq \underline{0}$ such that $e_t = \alpha' z_t \sim I(d_e)$ where $0 \leq d_e < \min_{1 \leq a \leq p} d_a$; the more standard circumstances where $d_1 = ... = d_p = 1$, $d_e = 0$ were termed the $CI(1)$ case. Aspects of the $CI(1)$ methodology for cointegration analysis were investigated in Section 1.2; it is possible to imagine how these could be extended to more general $FCI(d_1, ..., d_p; d_e)$ situations where the d_a and d_e are, while not necessarily one

[1] **This chapter is based on the paper "Semiparametric Frequency Domain Analysis of Fractional Cointegration"; the paper is the outcome of joint work with Prof.P.M.Robinson.**

and zero, known values, especially in view of the limit theory for nonstationary fractional processes which we presented in Chapter 2; this line of study has been recently pursued by Dolado and Marmol (1996). However, when non-integral d_a and d_e are envisaged, assuming their values known seems somehow more arbitrary than stressing the $CI(1)$ case in an autoregressive setting. It seems of greater interest to study the problem in the context of unknown orders of integration d_a and d_e in the observed and cointegrated processes, possibly less or greater than unity. For example, in circumstances where $CI(1)$ cointegration has been rejected it may be possible to find evidence of $FCI(d_1, ..., d_p; d_e)$ cointegration for some $(d_1, ..., d_p; d_e) \neq (1, ..., 1; 0)$. Our allowance for some "memory" remaining in the cointegrating residual e_t (i.e., $d_e > 0$), is appealing, especially recalling how d_e can be linked to the speed of convergence to long run equilibrium (compare for instance Diebold and Rudebusch (1989)).

Cointegration is commonly thought of as a stationary relation between nonstationary variables (so that $d_a \geq \frac{1}{2}$, for all a, $d_e < \frac{1}{2}$). Other circumstances covered by Definition 1.1 are also worth entertaining. One case is $d_e \geq \frac{1}{2}$, when both z_t and e_t are nonstationary. Another is $0 \leq d_a < \frac{1}{2}$, for all a, when both z_t and e_t are stationary.

The latter situation was considered by Robinson (1994a), as an application of results that we reported in Theorem 1.16, which provides limit theory for averages of periodogram ordinates on a degenerating frequency band in stationary long memory series. Ordinary least squares estimates (OLS) (and other "full-band" estimates such as generalized least squares) are inconsistent due to the usual simultaneous equation bias. Robinson (1994a) showed, in case of bivariate z_t, that a narrow-band frequency domain least squares (FDLS) estimate of (a normalized) α can be consistent. It is possible that some macroeconomic time series that have been modelled as nonstationary with a unit root could arise from stationary $I(d)$ processes with d near $\frac{1}{2}$, say, and interest in the phenomenon of cointegration of stationary variates has recently emerged in a finance context. Moreover, it is likely to be extremely difficult in practice to distinguish a z_t with unit root from one, say, composed additively of a stationary autoregression with a root near the unit circle, and a stationary long memory process.

FDLS is defined in the following section, after which in Section 3 we extend Robinson's (1994a) results to a more general stationary vector setting, with rates of convergence. The results of Chapter 3 on the approximation of sample moments of nonstationary sequences by narrow-band periodogram averages are exploited in Section 4 to demonstrate the usefulness of FDLS for nonstationary z_t: we partition z_t into $(y_t, x_t')'$ and we show that correlation between x_t and e_t does not prevent consistency of OLS, but it produces a larger second order bias relative to FDLS in the $CI(1)$ case, and a slower rate of convergence in many circumstances in which z_t exhibits less-than-$I(1)$ nonstationarity. When e_t is itself nonstationary, the two estimates share a common limit distribution. The parametrization we shall adopt for z_t in the nonstationary case relies directly on the linear expansion of the fractional differencing operator, as in Assumption 2A; see Section 4. In this sense, our parametrization is equivalent to Dolado and Marmol's (1996) definition of a nonstationary fractionally integrated process only for $d_a > \frac{3}{2}$, while for $\frac{1}{2} < d_a < \frac{3}{2}$ these authors actually consider the partial sum of stationary, long memory innovations, (as in (1.129)), which leads in our terminology to "cointegration with fractionally integrated errors" (as in Jeganathan (1996), cf.Chapter 5). Because of its practical relevance, we support our theoretical result for the $CI(1)$ case in finite samples by Monte Carlo simulations in Section 5. Section 6 describes a semiparametric methodology for investigating the question of cointegration in possibly fractional conditions, and applies it to series that were studied in the early papers of Engle and Granger (1987) and Campbell and Shiller (1987). Section 7 mentions possibilities for further work. Proofs are collected in the Appendix.

OLS by no means represents the state of the art in $CI(1)$ analysis. In Chapter 1, we discussed more elaborate estimates which have been proposed and shown to have advantages over OLS, such as Phillips and Hansen's (1990) fully modified least squares, Phillips' (1991a) maximum likelihood estimate for the error-correction mechanism (ECM), and Phillips' (1991b) spectral regression for the ECM. Fully modified least squares and spectral regression estimates make use of OLS at an early stage, so one implication of our results for the $CI(1)$ case is that FDLS be substituted here. As noted in Chapter 1, these methods are all specifically designed for the $CI(1)$ case, and in more general settings the validity and optimality of the associated inference procedures will be lost, and they may have no obvious advantage over OLS. Moreover, like OLS, they will not even be consistent in the stationary case. For the computationally simple FDLS procedure, this chapter demonstrates a consistency-robustness not achieved by OLS and other procedures, a matching of limit distributional properties in some cases, and superiority in others, including the standard $CI(1)$ case.

4.2 FREQUENCY DOMAIN LEAST SQUARES

Suppose we observe vectors $z_t = (y_t, x_t')'$, $t = 1, ..., n$, where y_t is real-valued and x_t is a $(p-1) \times 1$ vector with real-valued elements. Consider, for various m and assuming the inverse exists, the statistic

$$\widehat{\beta}_m = \widehat{F}_{xx}(1, m)^{-1} \widehat{F}_{xy}(1, m) , \tag{4.1}$$

where $\widehat{F}_{xx}(1, m), \widehat{F}_{xy}(1, m)$ represent averaged periodogram matrices, as introduced in Chapter 1, (1.20). We can interpret $\widehat{\beta}_m$ as estimating the unknown β in the "regression model"

$$y_t = \beta' x_t + e_t,\ t = 1, 2, \dots . \tag{4.2}$$

Recall that

$$\widehat{F}_{xx}(1, n-1) = \frac{1}{n} \sum_{t=1}^{n-1} (x_t - \overline{x})(x_t - \overline{x})',\ \widehat{F}_{xy}(1, n-1) = \frac{1}{n} \sum_{t=1}^{n-1} (x_t - \overline{x})(y_t - \overline{y}). \tag{4.3}$$

Thus $\widehat{\beta}_{n-1}$ is the OLS estimate of β with allowance for a non-zero mean in the unobservable e_t. Our main interest is in cases $1 < m < n-1$, where, because the discrete Fourier transform $w_a(\lambda)$ has complex conjugate $\overline{w}_a(2\pi - \lambda)$, we restrict further to $1 < m < n/2$. Then we call $\widehat{\beta}_m$ an FDLS estimate. Properties of e_t will be discussed subsequently, but these permit it to be correlated with x_t as well as y_t, $\widehat{\beta}_m$ being consistent for β due to $\widehat{F}_{ee}(1, m)$ being dominated by $\widehat{F}_{xx}(1, m)$ in a sense to be indicated. This can happen when z_t is stationary with long memory and e_t is stationary with less memory, if

$$\frac{1}{m} + \frac{m}{n} \to 0, \text{ as } n \to \infty , \tag{4.4}$$

which rules out OLS. Under (4.4) $\widehat{\beta}_m$ can be termed a "narrow-band" FDLS estimator. As a consequence of results from Chapter 3, it can also happen when x_t is nonstationary while e_t is stationary or nonstationary with less memory, if only

$$m < n, \qquad m \to \infty, \text{ as } n \to \infty , \tag{4.5}$$

(cf. Assumption 3C), which includes OLS. In both situations $z_t \sim FCI(d_1, ..., d_p; d_e)$ and the focus on low frequencies is thus natural. Notice that when $\lim(m/n) = \theta \in (0, \frac{1}{2})$ (so that $\widehat{\beta}_m$ is not narrow-band), $\widehat{\beta}_m$ is a special case of the estimate introduced by Hannan (1963) and developed by Engle (1974) and others. However, while such m satisfy (4.5), our primary interest is in the narrow-band case (4.4) where $\widehat{\beta}_m$ is based on a degenerating band of frequencies and its superiority over OLS can be established under wider circumstances. It is the stationary case which we first discuss.

4.3 STATIONARY COINTEGRATION

The covariance stationary processes with which we shall be concerned will always be assumed to have absolutely continuous spectral distribution function. We impose the following condition on z_t introduced earlier. Generalizing the notation we adopt for scalar sequences, for two matrices M and N, of equal dimension and possibly complex-valued elements, we say that $M \approx N$ if, for each (a, b), the ratio of the (a, b)-th elements of M and N tends to unity.

Assumption 4A The vector process z_t is covariance stationary with

$$f_{zz}(\lambda) \approx \Lambda G \Lambda, \text{ as } \lambda \to 0^+ , \tag{4.6}$$

where G is a real matrix whose lower $(p-1) \times (p-1)$ submatrix has full rank and

$$\Lambda = diag\left\{\lambda^{-d_1}, ..., \lambda^{-d_p}\right\} , \tag{4.7}$$

for $0 < d_a < \frac{1}{2}$, $1 \leq a \leq p$, and there exists a $p \times 1$ vector $\alpha \neq 0$, and a $c \in (0, \infty)$ and $d_e \in [0, \widetilde{d})$, such that

$$\alpha' f_{zz}(\lambda)\alpha \approx c\lambda^{-2d_e}, \text{ as } \lambda \to 0^+ . \tag{4.8}$$

Assumption 4A is similar to that introduced by Robinson (1995a) (cf.Theorem 1.18), where it is shown to hold for vector stationary and invertible fractional ARIMA processes (we could allow here, as there, for negative orders of integration greater than $-\frac{1}{2}$). However, there G was positive definite, whereas if Assumption 4A is imposed it has reduced rank, because otherwise

$$\alpha' f_{zz}(\lambda)\alpha \approx (\alpha'\Lambda)\, G\, (\Lambda\alpha) \geq \widetilde{c}\lambda^{-2\widetilde{d}} \text{ as } \lambda \to 0^+ , \tag{4.9}$$

for $0 < \widetilde{c} < \infty$. Nevertheless G must be non-negative definite because $f_{zz}(\lambda)$ is, for all λ. The rank condition on G is a type of no-multicollinearity one on x_t. Notice that $z_{at} \sim I(d_a)$, $a = 1, ..., p$ and $e_t \sim I(d_e)$ if we adopt the stationary definition (replacing $\eta_t 1(t > 0)$ by η_t in (1.2)) of an $I(d)$ process. Notice then that Assumption 4A follows if $z_t \sim FCI(d_1, ..., d_p; d_e)$ with $d_e < \widetilde{d}$. We adopt the normalization given by $\alpha = (1, -\beta')'$, so the cointegrating relation is given by (4.2) if $e_t \sim I(d_e)$. We stress that Assumption 4A does not restrict the spectrum of e_t away from frequency zero, because it is only local properties that matter since here we consider $\widehat{\beta}_m$ under (4.4). Asymptotic properties of $\widehat{\beta}_m$ require an additional regularity condition, such as

Assumption 4B

$$z_t = \mu_z + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j} \, , \; \sum_{j=0}^{\infty} \|\Psi_j\|^2 < \infty \, , \tag{4.10}$$

where $\mu_z = Ez_0$, and the $p \times 1$ vectors ε_t satisfy

$$E\left(\varepsilon_t \mid \Im_{t-1}\right) = \underline{0} \, , \; E\left(\varepsilon_t \varepsilon_t' \mid \Im_{t-1}\right) = \sum \, , \text{ a.s.,} \tag{4.11}$$

for a constant, full rank matrix $\sum$, $\Im_t$ being the σ-field of events generated by ε_s, $s \leq t$, and the $\varepsilon_t \varepsilon_t'$ are uniformly integrable.

This assumption is a generalization of that of Theorem 1.16 (Robinson (1994a)), the square summability of the Ψ_j only confirming, in view of the other assumptions, the finite variance of z_t implied by Assumption 4A. We could replace the martingale difference assumption on ε_t and $\varepsilon_t \varepsilon_t' - \sum$ by fourth moment conditions, as in Robinson (1994c). Notice that it would be equivalent to replace z_t by $(e_t, x_t')'$ with e_t given by (4.2). When z_t satisfies both Assumptions 4A and 4B, the Ψ_j are restricted by the requirement that $\| \alpha' \Psi(e^{i\lambda}) \| \approx c^* \lambda^{-d_e}$ as $\lambda \to 0^+$, for $0 < c^* < \infty$. A very simple model covered by Assumptions 4A and 4B is (4.2) and $x_t = \varphi x_{t-1} + u_t$, with $p = 2$, $0 < \varphi < 1$, $u_t \sim I(d_1)$ (implying $x_t \sim I(d_1)$), and $e_t \sim I(d_e)$, $0 \leq d_e < d_1 < \frac{1}{2}$. When u_t and e_t are not orthogonal OLS is of course inconsistent for β, as indeed is any other standard cointegration estimator, notwithstanding the fact that for φ close enough to unity x_t is indistinguishable for any practical purpose from a unit root process.

Theorem 4.1 Under Assumption 4A with $\alpha = (1, -\beta')'$, Assumption 4B and (4.4), as $n \to \infty$

$$\widehat{\beta}_{am} - \beta_a = O_p\left(\left(\frac{n}{m}\right)^{d_e - d_a}\right) \ , a = 1, ..., p-1 \ , \tag{4.12}$$

where $\widehat{\beta}_{am}$ and β_a are the a-th elements of, respectively, $\widehat{\beta}_m$ and β.

It follows that if there is cointegration, so $\widetilde{d} > d_e$, $\widehat{\beta}_m$ is consistent for β. In case the d_a are identical there is a common stochastic order $O_p((n/m)^{-b})$, varying inversely with the strength $b = d_a - d_e$ of the cointegrating relation. We conjecture that Theorem 3.1 is sharp and that under suitable additional conditions the $(n/m)^{d_a - d_e}\left(\widehat{\beta}_{am} - \beta_a\right)$ will jointly converge in probability to a non-null constant vector. We conjecture also that after bias-correction and with a different normalization the limit distribution will be normal in some proper subset of stationary $(d_1, ..., d_{p-1}, d_e)$-space, and non-normal elsewhere (cf the derivation of Lobato and Robinson (1996) of the limit distribution of the scalar averaged periodogram). A proper study of this issue would take up considerable space, however, whereas our principle purpose here is to establish consistency, with rates, as an introduction to a study of $\widehat{\beta}_m$ in nonstationary environments.

4.4 NONSTATIONARY FRACTIONAL COINTEGRATION

For z_t nonstationary ($d_a > \frac{1}{2}$, $a = 1, 2, ..., p$), we find it convenient to stress a linear representation for $w_t = (e_t, x_t')'$ in place of that for $z_t = (y_t, x_t')'$ in Assumption 4B.

Assumption 4C The vector sequence w_t is given by

$$(w_t - \mu_w) = \Delta(L)\eta_t 1(t > 0) \ , \tag{4.13}$$

for

$$\Delta(L) = diag\left\{(1-L)^{-d_e}, (1-L)^{-d_1}, (1-L)^{-d_{p-1}}\right\} \ , \tag{4.14}$$

μ_w a fixed vector with p-th element zero, and

$$d_a > \frac{1}{2} \, , \, a = 1, ..., p-1 \, , \, d_\epsilon \geq 0 \, , \tag{4.15}$$

and (cf. Assumption 2A-2D)

$$\eta_t = A(L)\varepsilon_t \, , \quad A(L) = \sum_{j=0}^{\infty} A_j L^j \, , \tag{4.16}$$

$$rank\{A(1)\} = p \, , \tag{4.17}$$

$$\sum_{j=0}^{\infty} \left(\sum_{k=j}^{\infty} \|A_k\|^2 \right)^{\frac{1}{2}} < \infty \, , \tag{4.18}$$

where the ε_t are independent and identically distributed $p \times 1$ vectors such that

$$E\varepsilon_t = \underline{0} \, , \, E\varepsilon_t \varepsilon_t' = \sum \, , \, rank(\sum) = p \, , \tag{4.19}$$

$$E \|\varepsilon_t\|^\theta < \infty \, , \, \theta > \max\left(4, \frac{2}{2d_{\min} - 1}\right) . \tag{4.20}$$

Assumption 4C strengthens the requirements on ε_t of Assumption 4B. Under (4.18)

$$\sum_{j=0}^{\infty} \|A_j\| \leq \sum_{j=0}^{\infty} \left(\sum_{k=j}^{\infty} \|A_k\|^2 \right)^{\frac{1}{2}} < \infty \tag{4.21}$$

so (4.17) and (4.18) imply that all elements of η_t are in $I(0)$, whereas with reference to (1.2), for $a = 2, ..., p$ the a-th element of w_t (and thus of x_t) is in $I(d_a)$, while its first element, e_t, is in $I(d_e)$, so in particular w_t could be a vector fractional ARIMA process. We have allowed for an unknown intercept, μ_w, in w_t. Note that $\Delta(L)$ introduced in (4.14) is not the same as the fractional differencing operator introduced in Assumption 2A1, because we have allowed for $d_e < \frac{1}{2}$. Also, (4.20) is stronger than Assumption 2B. On the other hand, (4.16)-(4.18) are analogous to Assumptions 2A2-2A4 and are repeated here only for convenience.

From (4.13) and (4.16), we can write

$$w_t = \mu_w + \sum_{j=0}^{\infty} B_{jt}\varepsilon_{t-j} \ , \tag{4.22}$$

where

$$B_{jt} = \sum_{i=0}^{\min(j,t-1)} \Delta_i A_{j-i} \tag{4.23}$$

with Δ_j given by the formal (binomial) expansion $\Delta(L) = \sum_{j=0}^{\infty} \Delta_j L^j$. Defining the non-singular matrix

$$P = \begin{bmatrix} I_{p-1} & \underline{0}' \\ -\beta' & 1 \end{bmatrix} , \tag{4.24}$$

we find that (4.13) is equivalent to

$$z_t = \mu_z + \sum_{j=0}^{\infty} \Psi_{jt}\varepsilon_{t-j} \ , \tag{4.25}$$

where $\mu_z = P^{-1}\mu_w$, $\Psi_{jt} = P^{-1}B_{jt}$. The representation (4.25) can be compared with the time-invariant one (4.10) for the stationary case (in which $\Psi_j = P^{-1}B_{j\infty}$).

Take $d = (d_1, ..., d_{p-1})'$; for notational convenience, we modify slightly the functions $D(.)$ and $G(.)$ as defined in Chapter 2, and we introduce

$$D = diag\left\{\Gamma(d_1)n^{\frac{1}{2}-d_1}, ..., \Gamma(d_{p-1})n^{\frac{1}{2}-d_{p-1}}\right\} \ , \ G(r,d) = diag\left\{r^{d_1-1}, ..., r^{d_{p-1}-1}\right\} . \tag{4.26}$$

Let Ω_{22} be the right-hand lower $(p-1)\times(p-1)$ submatrix of $A(1)\sum A(1)'$, where Ω_{22} has full rank under Assumption 4C. Let

$$W(r;d,\Omega_{22}) = \int_0^r G(r-s;d)dB(s;\Omega_{22}) \ , \ W(d,\Omega_{22}) = \int_0^1 W(r;d,\Omega_{22})dr \ , \tag{4.27}$$

$$V(d,\Omega_{22}) = \int_0^1 \{W(r;d,\Omega_{22})W'(r;d,\Omega_{22}) - W(d,\Omega_{22})W(d,\Omega_{22})'\}\,dr\;. \tag{4.28}$$

$W(r;d,\Omega_{xx})$ is "type II" multivariate fractional Brownian motion, as introduced in Chapter 2, albeit with a different normalization. Let $\overline{x} = n^{-1}\sum_{t=1}^{n} x_t$.

Theorem 4.1 Under Assumption 4C, as $n \to \infty$

$$Dx_{[nr]} \Rightarrow W(r;d,\Omega_{22})\;,\; 0 < r \le 1\;, \tag{4.29}$$

$$D\overline{x} \Rightarrow W(d,\Omega_{22})\;, \tag{4.30}$$

$$D\widehat{F}_{xx}(1,n-1)D \Rightarrow V(d,\Omega_{22})\;. \tag{4.31}$$

The proof of (4.29) was given in Chapter 2 under somewhat milder conditions and the result is reported here only for convenience; (4.30) and (4.31) follow from the continuous mapping theorem. For $d_1 = ... = d_{p-1} = 1$, fractional Brownian motion reduces to classical Brownian motion and so (4.29) includes a multivariate invariance principle for $I(1)$ processes, as can be found for instance in Phillips and Durlauf (1986). (4.31) provides an invariance principle for the sample covariance matrix of x_t (see (4.3)), and due to the following lemma Theorems 3.1 and 3.2 can be applied to deduce one for $\widehat{F}_{xx}(1,m)$.

Lemma 4.1 Let Assumption 4C hold. Then with the choices

$$(a_{1t},a_{2t}) = (x_{at},e_t)\;,\;\gamma_1 = d_a\;,\;\gamma_2 = d_e\;,\;a = 1,...,p-1 \tag{4.32}$$

or

$$(a_{1t},a_{2t}) = (x_{at},x_{bt})\;,\;\gamma_1 = d_a\;,\;\gamma_2 = d_b\;,\;a,b = 1,...,p-1\;, \tag{4.33}$$

it follows that Assumptions 3A, 3B and 3D are satisfied.

Lemma 4.2 Under Assumption 4C and (4.5), as $n \to \infty$

$$D\widehat{F}_{22}(1,m)D \Rightarrow V(d,\Omega_{22}) \ .$$

We can now proceed to investigate asymptotic behaviour of OLS and FDLS in various of the cases that arise when x_t is nonstationary, and e_t has short memory or stationary or nonstationary long memory.

Case I: $d_a + d_e < 1$, $a = 1, ..., p-1$.

Here not only does x_t possess less-than-unit-root nonstationarity, but the collective memory in x_t and e_t is more limited than in the $CI(1)$ case. It corresponds to $\gamma_1 + \gamma_2 < 1$ of Section 4, and we require first a more precise result than Theorem 3.1 in case $m = n-1$. Let b'_{aj} be the a-th row of $B_j = B_{j\infty} = \sum_{i=0}^{j} \Delta_i A_{j-i}$ given by (4.23). Define

$$\xi_a = \sum_{j=0}^{\infty} b'_{aj} \sum b_{pj} \ , \ a = 1, ..., p-1 \ . \tag{4.34}$$

Lemma 4.3 Under Assumption 4C with $d_a + d_e < 1$, $d_a > \frac{1}{2}$, $a = 1, ..., p-1$,

$$\lim_{n \to \infty} E\widehat{F}^{(a)}_{xx}(1, n-1) = \xi_a \ , \ \ a = 1, ..., p-1 \ , \tag{4.35}$$

where the right hand side is finite.

Lemma 4.3 is of some independent interest in that it indicates how sample covariances between a nonstationary and a stationary sequence can be stochastically bounded and have the same structure as when both sequences are stationary, so long as the memory parameters sum to less than 1, as automatically applies in the fully stationary case.

Define ξ to be the $(p-1)\times 1$ vector with a-th element ξ_a if $d_a = d_{\min}$ and zero if $d_a > d_{\min}$, where $d_{\min} = \min\limits_{1\le a\le p-1} d_a$.

Theorem 4.2 Let Assumption 4C hold with $d_a + d_e < 1$, $d_e < d_a$, $a = 1, ..., p-1$, and

$$rank\left\{V(d, \Omega_{22})\right\} = p - 1 \text{ , a.s.} \tag{4.36}$$

Then as $n \to \infty$

$$n^{d_{\min}-\frac{1}{2}} D^{-1}\left(\widehat{\beta}_{n-1} - \beta\right) \Rightarrow V(d, \Omega_{22})^{-1}\xi \ , \tag{4.37}$$

and under (4.4)

$$n^{d_{\min}-\frac{1}{2}} D^{-1}\left(\widehat{\beta}_{m} - \beta\right) = O_p\left(\left(\frac{n}{m}\right)^{d_{\min}+d_e-1}\right) = o_p(1) \ . \tag{4.38}$$

Theorem 4.2 indicates that so long as ξ is non-null the $n^{d_{\min}+d_a-1}\left(\widehat{\beta}_{a,n-1} - \beta_a\right)$ have a nondegenerate limit distribution, whereas when the interval $(0, \lambda_m)$ degenerates $\widehat{\beta}_{am} - \beta_a = o_p\left(n^{1-d_{\min}-d_a}\right)$, $a = 1, ..., p-1$, so that FDLS converges faster than OLS. In view of the "global" nature of $\widehat{\beta}_{m-1}$ and the "local" nature of $\widehat{\beta}_m$ this outcome is at first sight surprising, but it is due to the bias of $\widehat{\beta}_m$ becoming negligible relative to that of $\widehat{\beta}_{n-1}$. Notice that the rate of convergence of $\widehat{\beta}_{n-1}$ is independent of d_e.

Case II: The $CI(1)$ case $(d_a = 1,\ a = 1, ..., p-1,\ d_e = 0)$.

Now we consider the case considered in the bulk of the cointegration literature, where z_t has a unit root and the cointegrating error is $I(0)$. Write ι for the $(p-1)\times 1$ vector of units, so that in the present case $d = \iota$ and $W(r; d, \Omega_{22}) = B(r; \Omega_{22})$. Let $B_*(r; \Omega)$ be p-dimensional Brownian motion with covariance matrix $\Omega = A(1)\sum A(1)'$, and thus write

$$B_*(r;\Omega) = \begin{pmatrix} B(r;\Omega_{22}) \\ B(r;\omega_{11}) \end{pmatrix} , \tag{4.39}$$

where ω_{11} is the $(1,1)$-th element of Ω, and in general $B(r;\Omega_{22})$ and $B(r;\omega_{11})$ are correlated and thus in effect depend not only on Ω_{22} and ω_{11} but on the other elements of Ω also. As in Chapter 1, (1.42), we write

$$\varkappa = \int_0^1 B(r;\Omega_{22})dB(r;\omega_{11}) . \tag{4.40}$$

Denote by γ_j the $(p-1)\times 1$ vector with $(a-1)$-th element $E\eta_{at}e_{t+j}$, recalling that $d_e = 0$ implies $e_t = \eta_{1t}$. Now define

$$\Gamma_j = \sum_{\ell=|j|}^{\infty} \gamma_{\ell sign(\ell)} , \ j = 0, \pm 1, \ldots , \tag{4.41}$$

so that $\Gamma_j = \sum_{\ell=j}^{\infty} \gamma_\ell$ for $j \geq 0$ and $\Gamma_j = \sum_{\ell=-\infty}^{j} \gamma_\ell$ for $j < 0$, and the sum (4.41) converges absolutely for all j under Assumption 4C. Let $h(\lambda)$ be the vector function with Fourier coefficients given by

$$\Gamma_{|j|} - \Gamma_{-|j|-1} = \int_{-\pi}^{\pi} h(\lambda)e^{ij\lambda}d\lambda \ , \ j = 0, \pm 1, \ldots . \tag{4.42}$$

Assumption 4D $h(\lambda)$ is continuous at $\lambda = 0$, and integrable.

Assumption 4D is implied by $\sum_{j=0}^{\infty} \| \Gamma_{|j|} - \Gamma_{-|j|-1} \| < \infty$, which is in turn implied by

$$\sum_{j=0}^{\infty} (j+1) \| \gamma_j - \gamma_{-j-1} \| < \infty , \tag{4.43}$$

in which case we may write

$$h(0) = \frac{1}{2\pi} \sum_{j=0}^{\infty} (2j+1)(\gamma_j - \gamma_{-j-1}) \,. \tag{4.44}$$

Of course (4.43) is itself true if $\sum_{j=-\infty}^{\infty}(|j|+1) \left\| \gamma_j \right\| < \infty$ for which a sufficient condition in terms of (4.16) is

$$\sum_{j=0}^{\infty} (j+1) \left\| A_j \right\| < \infty \,, \tag{4.45}$$

which is stronger than (4.18), while holding when η_t is a stationary ARMA process.

Lemma 4.4 Under Assumption 4D and (4.4)

$$\lim_{n \to \infty} E\left(\frac{n}{m} \widehat{F}_{xe}(1, m) \right) = \frac{1}{2} h(0) \,. \tag{4.46}$$

Theorem 4.3 Let Assumption 4C hold with $d = \iota$, $d_e = 0$. Then as $n \to \infty$

$$n\left(\widehat{\beta}_{n-1} - \beta\right) \Rightarrow V(\iota, \Omega_{22})^{-1} \left\{ \varkappa + \Gamma_0 \right\} \tag{4.47}$$

and if also Assumption 4D and (4.4) hold

$$n\left(\widehat{\beta}_m - \beta\right) \Rightarrow V(\iota, \Omega_{22})^{-1} \varkappa \,. \tag{4.48}$$

Thus in the $CI(1)$ case $\widehat{\beta}_{n-1}$ and $\widehat{\beta}_m$ have the same rate of convergence but under (4.4) $\widehat{\beta}_m$ does not suffer from the "second-order bias" term Γ_0 incurred by $\widehat{\beta}_{n-1}$. More precisely, as the proof of Theorem 4.3 indicates, there is a second-order bias of order $O(m/n^2)$ in $\widehat{\beta}_m$ which is thus too small to contribute to (4.48), by comparison with the $O(n^{-1})$ second-order bias in (4.47). As reported in Chapter 1, Theorem 1.10, Phillips

(1991b) considered a form of narrow-band spectral regression in the $CI(1)$ case, albeit stressing a system type of estimate which has superior limit distributional properties to $\widehat{\beta}_m$, assuming the $CI(1)$ hypothesis is correct. However his proof is based on weighted autocovariance spectrum estimates, rather than our averaged periodogram ones. As is well known, in many stationary environments these two types of estimate are very close asymptotically, but in the $CI(1)$ case the weighted autocovariance version of $\widehat{\beta}_m$ turns out to exhibit second-order bias due to correlation between u_t and e_t (specifically, to their cross-spectrum at zero frequency). Note that $V(d, \Omega_{22}) = \Xi$ in the terminology of Chapter 1, (1.43), when $d = \iota$.

Case III: The Case $d_a + d_e > 1$, $a = 1, ..., p-1$, $d_e < \frac{1}{2}$.

We now look at the case where the collective memory in each (x_{at}, e_t) combination exceeds that of the previous two cases, yet e_t is still stationary. Thus x_t could have less than unit root stationarity but in that case the memory in e_t must compensate suitably. On the other hand x_t could exhibit nonstationarity of arbitrarily high degree.

Theorem 4.4 Let Assumption 4C hold with $d_a + d_e > 1, 0 \leq d_e < \frac{1}{2} < d_a$, $a = 1, ..., p-1$, and let (4.36) hold. Then for $a = 1, ..., p-1$, as $n \to \infty$

$$\widehat{\beta}_{a,n-1} - \beta_a = O_p\left(n^{d_e - d_a}\right) , \tag{4.49}$$

and if also (4.5) holds

$$\widehat{\beta}_{am} - \widehat{\beta}_{a,n-1} = o_p\left(n^{d_e - d_a}\right) , \tag{4.50}$$

$$\widehat{\beta}_{am} - \beta_a = O_p\left(n^{d_e - d_a}\right) . \tag{4.51}$$

The results (4.49) and (4.51) only bound the rates of convergence of OLS and FDLS, and we have been unable to characterize even the exact rate of convergence of OLS in the present case, due to the fact that on the one hand e_t is stationary so that the continuous mapping theorem does not suffice, whereas on the other hand e_t cannot be approximated by a semi-martingale, unlike in the short-memory case $d_e = 0$ (where in fact an exact rate and limit distribution can be derived, as it was in the $CI(1)$ case). We conjecture, however, that at least under some additional conditions the rate in (4.49) is exact, whereupon (4.50) implies immediately that $\widehat{\beta}_m$ shares the same rate and limit distribution as $\widehat{\beta}_{n-1}$.

Case IV: The case $d_e > \frac{1}{2}$.

Now we suppose that cointegration does not account for all the nonstationarity in z_t, so that e_t is nonstationary, as is motivated by some of the empirical experience to be described in Section 6. Write $d_* = (d', d_e)'$ and

$$\begin{aligned}
D_* &= diag\left\{\Gamma(d_1)n^{\frac{1}{2}-d_1}, ..., \Gamma(d_{p-1})n^{\frac{1}{2}-d_{p-1}}, \Gamma(d_e)n^{\frac{1}{2}-d_e}\right\}, && (4.52)\\
G_*(r; d_*) &= diag\left\{r^{d_1-1}, ..., r^{d_{p-1}-1}, r^{d_e-1}\right\}, && (4.53)\\
W_*(r; d_*, \Omega) &= \int_0^r G_*(r-s; d_*)dB_*(s; \Omega) = \begin{bmatrix} W(r; d, \Omega_{22}) \\ w(r; d_e, \omega_{11}) \end{bmatrix}, && (4.54)\\
W_*(d_*, \Omega) &= \int_0^1 W_*(r; d_*, \Omega)dr, && (4.55)\\
U(d_*, \Omega) &= \int_0^1 \{W(r; d, \Omega_{22}) - W(d, \Omega_{22})\}\, w(r; d_e, \omega_{11})dr. && (4.56)
\end{aligned}$$

Let $\overline{w} = n^{-1}\sum_{t=1}^{n} w_t$. The following theorem is analogous to Theorem 4.1 and needs no additional explanation.

Theorem 4.5 Under Assumption 4C and $\frac{1}{2} < d_\epsilon < d_a$, $a = 1, ..., p-1$, as $n \to \infty$

$$D_* w_{[nr]} \Rightarrow W_*(r; d_*, \Omega) , \quad (4.57)$$

$$D_* \overline{w} \Rightarrow W(d_*, \Omega) , \quad (4.58)$$

$$n^{\frac{1}{2}-d_e} D \widehat{F}_{xe}(1, n-1) \Rightarrow U(d_*, \Omega) . \quad (4.59)$$

Theorem 4.6 Under Assumption 4C, $\frac{1}{2} < d_e < d_a$, $a = 1, ..., p-1$, and (4.36), as $n \to \infty$

$$n^{\frac{1}{2}-d_e} D^{-1} \left(\widehat{\beta}_{n-1} - \beta\right) \Rightarrow V(d, \Omega)^{-1} U(d_*, \Omega) , \quad (4.60)$$

and if also (4.5) holds

$$n^{\frac{1}{2}-d_e} D^{-1} \left(\widehat{\beta}_m - \beta\right) \Rightarrow V(d, \Omega)^{-1} U(d_*, \Omega) . \quad (4.61)$$

Now so long as m is regarded as increasing with n, the limit distribution is unaffected however many frequencies we omit from $\widehat{\beta}_m$. Notice that in case the d_a are all equal the rate of convergence reflects the cointegrating strength b defined in Section 1, such that $\widehat{\beta}_m$ is n^b-consistent.

4.5 MONTE CARLO EVIDENCE

Because OLS is often used as a preliminary step in $CI(1)$ analysis, the previous section suggests that even if fractional possibilities are to be ignored, FDLS might be substituted at this stage. To compare the performance of FDLS with OLS in moderate sample sizes a small Monte Carlo study in the $CI(1)$ case was conducted. The models we employed are as follows. For $a = 1, 2$, let η_{at} be a sequence of $N(0, a)$ random variables, independent across t.

Model A: AR(1) cointegrating error, $p = 2$, in (4.2) with

$$(1-L)x_t = \eta_{1t} , \tag{4.62}$$

$$(1-\varphi L)e_t = \eta_{2t} , \tag{4.63}$$

$$E\eta_{1t}\eta_{2t} = 1 , \ \varphi = 0.8, 0.6, 0.4, 0.2 . \tag{4.64}$$

Model B: AR(2) cointegrating error, $p = 2$, in (4.2) with

$$(1-L)x_t = \eta_{1t} , \tag{4.65}$$

$$(1-\varphi_1 L - \varphi_2 L^2)e_t = \eta_{2t} , \tag{4.66}$$

$$E\eta_{1t}\eta_{2t} = 1 , \ \varphi_2 = -0.9; \ \varphi_1 = .947, .34, -.34, -.947 . \tag{4.67}$$

We fix $\varphi_2 = -0.9$ in Model B to obtain a spectral peak for e_t in the interior of $(0, \pi)$, in particular at $\lambda^* = \arccos(-\varphi_1(1+\varphi_2)/4\varphi_2)$, that is at $\lambda^* = \pi/3, 4\pi/9, 5\pi/9$ and $2\pi/3$, respectively, for the four φ_1. On the other hand in Model A e_t always has a spectral peak at zero frequency.

Series of lengths $n = 64$, 128 and 256 were generated. $\widehat{\beta}_{n-1}$ and $\widehat{\beta}_m$, for $m = 3, 4, 5$ were computed, as were an estimate superior to OLS in the $CI(1)$ case (cf.Theorem 1.8), the fully-modified least squares estimate (FM-OLS, denoted $\widetilde{\beta}_{FM}$) of Phillips and Hansen (1990) which uses OLS residuals at a first step, and also a modified version of this (denoted $\widetilde{\beta}^*_{FM}$), using FDLS residuals. Bartlett nonparametric spectral estimation was used in $\widetilde{\beta}_{FM}$ and $\widetilde{\beta}^*_{FM}$, with lag numbers $\widetilde{m} = 4, 6, 8$, for $n = 64, 128, 256$ respectively.

Monte Carlo bias and mean squared error (MSE), based on 5000 replications, are reported in Tables 4.1 and 4.2 for Models A and B respectively. For each m, FDLS is superior to OLS in terms of both bias and MSE in every single case, often significantly. In fact $\widehat{\beta}_m$ is best for the smallest m, 3. Our modified version $\widetilde{\beta}^*_{FM}$ of FM-OLS improves on the standard one $\widetilde{\beta}_{FM}$ in 23 out of 24 cases in terms of bias, and 16 out of 24 in terms of MSE, with 8 ties. The intuition underlying FDLS is that on the smallest frequencies cointegration implies a high signal-to-noise ratio, so it is not surprising that FDLS performs better for AR(2) e_t than AR(1) e_t, especially as λ^* increases. It is possible to devise an e_t with such power around $\lambda = 0$ that, in finite samples, FDLS performs worse than OLS, for example when $\varphi_1 \simeq 2$, $\varphi_2 \simeq -1$, $\varphi_2+\varphi_1 < 1$, so e_t is "near-$I(2)$" and in small samples e_t dominates x_t. However, given the intuition underlying the concept of cointegration, we believe this could be described as a "pathological" case.

4.6 EMPIRICAL EXAMPLES

Our empirical work employs the data of Engle and Granger (1987) and Campbell and Shiller (1987). We consider seven bivariate series, denoting by y the variable chosen to be "dependent" and by x the "independent" one in (4.2), and by d_y, d_x integration orders. We describe the methodology used in three steps.

1) Memory of raw data

A necessary condition for cointegration is $d_x = d_y$, which can be tested using estimates of d_x and d_y. Three types of estimate were computed, and one test statistic. The estimates are all "semiparametric", based only on a degenerating band of frequencies around zero frequency and assuming only a local-to-zero model for the spectral density (cf Assumption 4A) rather than a parametric model over all frequencies. The semiparametric estimates are inefficient when the parametric model is correct, but are consistent more generally and seem natural in the context of the present paper. Their asymptotic properties were established by Robinson (1995a,b) and reviewed in Section 1.3, Theorems 1.18-1.20; there we imposed the assumptions of stationarity and invertibility (having integration order between -$\frac{1}{2}$ and $\frac{1}{2}$) and so because our raw series seem likely to be nonstationary, and quite possibly with integration orders between $\frac{1}{2}$ and $\frac{3}{2}$, we first-differenced them prior to d estimation, and then added unity. Although the stationarity assumption is natural in view of the motivation of these estimates, we mentioned in Chapter 1 that recently Hurvich and Ray (1995), Velasco (1997a,b) have shown that they can still be consistent and have the same limit distribution as in Robinson (1995a,b) under nonstationarity (although with a different definition of $I(d)$ nonstationarity from ours, namely (1.130)), at least if a data taper is used. We thus estimated d_x and d_y directly from the raw data also, but as the results were similar they are not reported.

Denote by Δz_t either Δx_t or Δy_t where Δ is the difference operator. We describe the estimation and testing procedures as follows

(i) Log-periodogram regression. For $z = x, y$ we report in the tables $\widetilde{d}_z = 1 + \widetilde{\delta}_z$, where $\widetilde{\delta}_z$ is the slope estimate obtained by regressing $\log\left(I_{\Delta z \Delta z}(\lambda_j)\right)$ on $-2\log(\lambda_j)$ and an intercept, for $j = 1, ..., \ell$, where ℓ is a bandwidth number, tending to infinity slower than n. This is the version proposed by Robinson (1995a) rather than the original one of Geweke and Porter-Hudak (1983); trimming out of low frequencies, which is suggested by Robinson (1995a), is not entertained, because recent evidence of Hurvich, Deo and Brodsky (1998) suggests that this is not necessary for nice asymptotic properties.

(ii) Test of $d_x = d_y$. We report the Wald statistic, denoted W in the tables, of Robinson (1995a,b), based on the difference $\widetilde{d}_x - \widetilde{d}_y = \widetilde{\delta}_z - \widetilde{\delta}_y$. The significance of W is judged by comparison with the upper tail of the χ_1^2 distribution, the 5% and 1% points being respectively 3.78 and 5.5

(iii) GLS log-periodogram regression. Given that $d_x = d_y$ we estimate the common value by $\widetilde{d}_{GLS} = 1 + \widetilde{\delta}_{GLS}$ where $\widetilde{\delta}_{GLS}$ is the generalized least squares (GLS) log-periodogram estimate of Robinson (1995a) based on the bivariate series $(\Delta x_t, \Delta y_t)$, using residuals from the regression in (i), $\widetilde{\delta}_{GLS}$ is asymptotically more efficient than $\widetilde{\delta}_x$ and $\widetilde{\delta}_y$ when $d_x = d_y$.

(iv) Gaussian estimation. For $z = x, y$ we report $\widehat{d}_z = 1 + \widehat{\delta}_z$ where $\widehat{\delta}_z$ minimizes

$$\log\left(\sum_{j=1}^{\ell} \lambda_j^{2d} I_{\Delta z \Delta z}(\lambda_j)\right) - \frac{2d}{\ell}\sum_{j=1}^{\ell}\log(\lambda_j)\,, \tag{4.68}$$

which is a concentrated narrow-band Gaussian pseudo-likelihood, see Theorem 1.20 and Künsch (1987), Robinson (1995b). As shown by Robinson (1996b), $\widehat{\delta}_z$ is asymptotically more efficient than $\widetilde{\delta}_z$.

For the estimates in (iii) and (iv) we report also approximate 95% confidence intervals (denoted CI in the tables) based on the (normal) asymptotic distribution theory developed by Robinson (1995a,b). From Section 1.3 we know that Robinson (1995a) assumed Gaussianity in establishing consistency and asymptotic normality of the estimates in (i) and (iii), but recent work of Velasco (1997b) suggests that this can be relaxed. Although progress is currently being made on the choice of bandwidth ℓ in log-periodogram and Gaussian estimation, we have chosen a grid of three arbitrary values for each data set analyzed in order to judge sensitivity to ℓ. Note that the estimates are $\ell^{\frac{1}{2}}$-consistent.

2) Cointegration analysis

We report $\widehat{\beta}_m$ and also a "high-frequency" estimate

$$\widehat{\beta}_{-m} = \frac{\widehat{F}_{xy}(m+1, [(n-1)/2])}{\widehat{F}_{xx}(m+1, [(n-1)/2])} \tag{4.69}$$

based on the remaining frequencies, substantial deviations between $\widehat{\beta}_m$ and $\widehat{\beta}_{-m}$ suggesting that a full-band estimate such as OLS could be distorted by misspecification at high frequencies which is irrelevant to the essentially low-frequency concept of cointegration.

The tables include results for three values of m for each data set. These are much smaller than the bandwidths ℓ used in inference on d_x and d_y due to the anticipation of nonstationarity in the raw data: for stationary x_t, y_t optimal rules of bandwidth choice would lead to m that are more comparable with the ℓ we have used. After computing residuals $\widehat{e}_t = y_t - \widehat{\beta}_m x_t$, we obtained the low- and high-frequency R^2 quantities

$$R_m^2 = 1 - \frac{\widehat{F}_{\widehat{e}\widehat{e}}(1,m)}{\widehat{F}_{yy}(1,m)}\, , \; R_{-m}^2 = 1 - \frac{\widehat{F}_{\widehat{e}\widehat{e}}(m+1,[(n-1)/2])}{\widehat{F}_{yy}(m+1,[(n-1)/2])}\, . \tag{4.70}$$

We can judge the fit of a narrow-band regression by R_m^2 and by comparing this with R_{-m}^2 see to what extent this semiparametric fit compares with a parametric one.

For each m we report also the fractions

$$r_{xx,m} = \frac{\widehat{F}_{xx}(1,m)}{\widehat{F}_{xx}(1,[(n-2)/2])}\, , \; r_{xy,m} = \frac{\widehat{F}_{xy}(1,m)}{\widehat{F}_{xy}(1,[(n-2)/2])}\, , \tag{4.71}$$

their closeness to unity indicating directly the empirical, finite sample relevance of Theorems 3.1 and 3.2 (though note that $r_{xy,m}$ need not lie in $[0,1]$.)

3) Memory of cointegrating error

We estimated d_e first by $\widetilde{d}_e$ and $\widehat{d}_e$, which are respectively the log-periodogram and Gaussian estimates of (i) and (iv) above, based on first differences of the $\widehat{e}_t$ and then adding unity. We also report $\widetilde{d}_e^*$ and $\widehat{d}_e^*$ which use the raw $\widehat{e}_t$ and do not add unity, because in general we have little prior reason for believing e_t is either stationary or nonstationary. In addition we report 95% confidence intervals based on the asymptotic theory of Robinson (1995a,b), though strictly this has not been justified in case of the residuals $\widehat{e}_t$.

Tables 4.3-4.9 report empirical results based on several data sets.

a) Consumption (y) and income (x) (quarterly data), 1947Q1-1981Q2

Engle and Granger (1987) found evidence of $CI(1)$ cointegration in these data. Table 4.3 tends to suggest an integration order very close to one for both variables, the estimates ranging from .89 to 1.08 for income and from 1.04 to 1.13 for consumption. The Wald statistic is at most 1.06, so we can safely not reject the $d_x = d_y$ null. Exploiting this information, one obtains GLS estimates ranging from .953 to 1.02; but with confidence intervals all so narrow as to exclude unity. The $\widehat{\beta}_m$ are about .232, which is close to OLS (.229), but the high frequency estimates $\widehat{\beta}_{-m}$ are closer to .20. The unexplained variability is four times smaller around frequency zero $(1 - R_m^2)$ than at short run frequencies $(1 - R_{-m}^2)$. Variability concentrates rapidly around frequency zero, 85.1% of the variance of income being accounted for by the three smallest periodogram ordinates, less than 5% of the total. This proportion rises to 92.6% for 6 frequencies, and is even greater for the cross-periodogram, confirming the high coherency of the two series at low frequencies. The residual diagnostics are less clear-cut, but in only one case out of 12 does the confidence interval for d_e include zero, providing strong evidence against weak dependence. The estimates of d_e vary quite noticeably with ℓ and the procedure adopted, ranging from .2 to .87.

b) Stock prices (y) and dividends (x) (annual data), 1871-1986.

The idea that these might be cointegrated follows mainly from a present value model, which asserts that an asset price is linear in the present discounted value of future dividends, $y_t = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^i E_t(x_{t+i}) + c$, where δ is the discount factor; see Campbell and Shiller (1987). In Table 4.4, the estimates of d_x, d_y appear close to unity, although now the hypothesis that dividends are mean-reverting ($d_x < 1$) appears to be supported. The Wald statistics for testing $d_x = d_y$ are always manifestly insignificant. A marked difference between $\widehat{\beta}_m$ and $\widehat{\beta}_{-m}$ is found, the former oscillating around 33 and the latter below 24. The spectral R^2 still indicate a much better fit at low frequencies, but empirical evidence of cointegration is extremely weak. Notice in particular that if y and x are not cointegrated, $d_e = \max(d_x, d_y)$, as is amply confirmed by the Gaussian estimates, where one gets identical estimates of d_x and $d_e = (1.04, .91, .90)$ for $\ell = 22, 30, 40$. The results of Campbell and Shiller on this data set were, in their own words, inconclusive;

our findings confirm those of Phillips and Ouliaris (1988), who were unable to reject the null of no cointegration at the 10% level.

c) Log prices (y) and wages (x) (monthly data), 1960M1-1979M12

The results in Table 4.5 tend to develop those of Engle and Granger (1987) by supporting an absence of a cointegrating relationship of any order. Where our conclusions differ is in the integration orders of x and y, in particular of log prices, which appear not to be unity, ranging from 1.54 to 1.60, while confidence intervals never include unity. This is not very surprising in that the inflation rate might plausibly be characterized as a stationary long memory process. W is always above 5.8, so we reject also at the 1% level the hypothesis that $d_x = d_y$, and so because this necessary condition for cointegration is not satisfied the analysis is taken no further.

d) Quantity theory of money (quarterly data): log M1, M2, M3 or L (y) and log GNP (x), where L denotes total liquid assets, 1959Q1 - 1981Q2.

Engle and Granger (1987) found the classical equation $MV = PY$ of the quantity theory of money to hold for $M = M2$, but not $M1, M3, L$. This is somewhat unsatisfactory since the latter monetary aggregates are linked with M2 in the long run, so that there might exist cointegration (albeit of different orders) between more than one of these aggregates and GNP. For log L in Table 4.6 we reject at the 1% level the hypothesis that GNP shares the same integration order. For M2, in Table 4.7, the necessary condition for cointegration is met, GLS confidence intervals tending to suggest integration orders around 1.3, which seems unsurprising since both aggregates are nominal. The $\widehat{\beta}_m$ are not noticeably influenced by m and are indeed the same as OLS (.99). Estimates of the d_e are strongly inconsistent with stationarity, ranging from 1.02 to 1.23, the confidence intervals excluding values below .88. Overall, it seems very difficult to draw reliable conclusions about the existence of fractional cointegration between these variables given such a small sample. The relationship of nominal GNP with M1 and M3, in Tables 4.8 and 4.9, appears much closer to that with M2 than Engle and Granger concluded, exploiting the greater flexibility of our framework. In particular, the common integration order for the bivariate raw data is again estimated via GLS to be 1.31, 1.39, 1.29, (for $m = 16$, 22, 30) for nominal GNP and M1 and 1.33, 1.44, 1.42 for nominal GNP and M3; estimates of d_e range from .76 to 1.20 for the former case and from .88 to 1.08 for the latter.

4.7 FINAL COMMENTS

This chapter demonstrates that OLS estimates of a cointegrating vector are asymptotically matched or bettered in a variety of stationary and nonstationary cases by a narrow-band frequency domain estimate, FDLS. The overall superiority of FDLS relies on correlation between the cointegrating errors and regressors; in the absence of such correlation FDLS is inferior to OLS for stationary data, and comparable for nonstationary data. The finite-sample advantages of FDLS in correlated situations are observed in a small Monte Carlo study. FDLS is incorporated in a semiparametric methodology for investigating the possibility of fractional cointegration, which is applied to bivariate macroeconomic series.

This chapter leaves open numerous avenues for further research. It is possible that the whole of x_t does not satisfy the conditions of either Section 3 or one of the cases I, II or III/IV of Section 4, but rather that subsets of x_t are classified differently. It is straightforward to extend our results to cover such situations, and we have not done so for the sake of simplicity, and because the case $p = 2$ is itself of practical importance. A more challenging development would cover such omitted cases as when x_t has integration order $\frac{1}{2}$, on the boundary between stationarity and nonstationarity, though this can be thought of as occupying a measure-zero subset of the parameter space. From a practical viewpoint a significant deficiency of our treatment of nonstationarity is the lack of allowance for deterministic trends, such as (possibly nonintegral) powers of t, but if these are suitably dominated by the stochastic trends Theorem 3.3 suggests that the results of Sections 4 continue to hold.

A more challenging area for study is the extent to which we can improve on FDLS in our semiparametric context, with unknown integration orders, to mirror the improvement of OLS by various estimates in the $CI(1)$ case. There is also a need, as in the $CI(1)$ case, to allow for the possibility of more than one cointegrating relation, where we might wish to permit these to have different integration orders. Certainly it seems clear that results such as Lemma 4.2 can be established for more general quadratic forms, and so the extensive asymptotic theory for quadratic forms of stationary long memory series can be significantly extended in a nonstationary direction. The choice of bandwidth m in $\widehat{\beta}_m$ seems less crucial under nonstationarity than under stationarity, but nevertheless some criterion must be given to practitioners. For the stationary case, which seems of interest in financial applications, bandwidth theory of Robinson (1994c) can be developed, but there is a need also to develop asymptotic distribution theory for FDLS, useful application of which is likely to require bias-correction due to correlation between x and e. For the relatively short macroeconomic series analyzed in this chapter the semiparametric approach employed, while based on very mild assumptions, will not produce as reliable estimates of integration orders as correctly specified parametric time series models, and it is possible to analyze narrow-band β estimates in such a parametric framework also.

APPENDIX

Proof of Theorem 3.1 From (4.1), (4.2) we have

$$\widehat{\beta}_m - \beta = \widehat{F}_{xx}(1,m)^{-1}\widehat{F}_{xe}(1,m) \ . \tag{4.72}$$

By the Cauchy inequality, as in Robinson (1994a)

$$\mid \widehat{F}_{xe}^{(a)}(1,m) \mid \leq \left\{ \widehat{F}_{xx}^{(aa)}(1,m)\widehat{F}_{ee}(1,m) \right\}^{\frac{1}{2}} \ . \tag{4.73}$$

For any non-null $p \times 1$ vectors γ and δ, by Assumption 4A

$$\gamma' \left\{ \widehat{F}_{zz}(1,m) - F_{zz}(\lambda_m) \right\} \delta = o_p\left(\{\gamma' F_{zz}(\lambda_m)\gamma\delta' F_{zz}(\lambda_m)\delta\}^{\frac{1}{2}} \right), \text{ as } n \to \infty \tag{4.74}$$

by a straightforward multivariate extension of Theorem 1.16 (Robinson (1994a)), with

$$F_{zz}(\lambda) = \int_0^{\lambda} \mathrm{Re}\left\{ f_{zz}(\mu) \right\} d\mu \approx G(\lambda), \text{ as } \lambda \to 0^+ \ , \tag{4.75}$$

where $G(\lambda)$ has (a,b)-th element

$$G_{ab}(\lambda) = \frac{G_{ab}\lambda^{1-d_a-d_b}}{1-d_a-d_b} \ , \tag{4.76}$$

G_{ab} being the (a,b)-th element of G. Applying (4.74), (4.75) and (4.4),

$$\Lambda_m^{-1} \left\{ \widehat{F}_{zz}(1,m) - F_{zz}(\lambda_m) \right\} \Lambda_m^{-1} = o_p(\lambda_m) \ , \tag{4.77}$$

where $\Lambda_m = diag\{\lambda_m^{-d_1}, ..., \lambda_m^{-d_p}\}$, so

$$\widehat{F}_{xx}^{(aa)}(1,m) = G_{ab}(\lambda_m) + o_p\left(\lambda_m^{1-d_a-d_b}\right) \ , \tag{4.78}$$

$$\begin{aligned}\widehat{F}_{ee}(1,m) &= \alpha'\widehat{F}_{zz}(1,m)\alpha = \alpha' F_{zz}(\lambda_m)\alpha + o_p(\alpha' F_{zz}(\lambda_m)\alpha) \\ &= O_p(\lambda_m^{1-2d_e}) . \end{aligned} \tag{4.79}$$

because

$$\alpha' F_{zz}(\lambda)\alpha = \int_0^\lambda \alpha' f_{zz}(\mu)\alpha d\mu \approx c\int_0^\lambda \mu^{-2d_e} d\mu = c\frac{\lambda^{1-2d_e}}{1-2d_e}, \text{ as } \lambda \to 0^+ . \tag{4.80}$$

Denote by $\widetilde{\Lambda}_m$, $\widetilde{G}(\lambda)$ the right-hand lower $(p-1)\times(p-1)$ submatrices of Λ_m, $G(\lambda)$. For any $(p-1)\times 1$ non-null vector $\nu = (\nu_1, ..., \nu_{p-1})'$

$$\nu'\widetilde{\Lambda}_m^{-1} F_{xx}(\lambda_m)\widetilde{\Lambda}_m^{-1}\nu \approx \nu'\widetilde{\Lambda}_m^{-1}\widetilde{G}(\lambda_m)\widetilde{\Lambda}_m^{-1}\nu \geq \omega \int_0^{\lambda_m} \sum_{j=1}^{p-1} \left\{(\lambda_m/\lambda)^{d_j}\nu_j\right\}^2 d\lambda \tag{4.81}$$

$$= \omega\lambda_m \sum_{j=1}^{p-1} \frac{\nu_j^2}{1-2d_j} , \tag{4.82}$$

where ω is the smallest eigenvalue of the right-hand lower $(p-1)\times(p-1)$ submatrix of G, which is positive definite by Assumption 4A. It follows that $\widetilde{\Lambda}_m \widehat{F}_{xx}^{-1}(\lambda_m)\widetilde{\Lambda}_m = O_p(n/m)$, whence the proof is completed by elementary manipulation. □

Proof of Lemma 4.1 From (4.13)

$$x_{at} = (1-L)^{-d_a}\{\eta_{at}1(t>0)\} , \; a = 1, ..., p-1 , \tag{4.83}$$

$$e_t = (1-L)^{-d_e}\left\{\eta_{pt}1(t>0)\right\} , \tag{4.84}$$

where η_{at} is the a-th element of η_t. We take

$$\varphi_k(\gamma) = \frac{\Gamma(k+\gamma)}{\Gamma(\gamma)\Gamma(k+1)} , \tag{4.85}$$

since this is the coefficient of L^k in the Taylor expansion of $(1-L)^{-\gamma}$, and choose $\varphi_{1k} =$

$\varphi_k(d_a)$ in both cases (4.32) and (4.33), with $\varphi_{2k} = \varphi_k(d_e)$ in (4.32) and $\varphi_{2k} = \varphi_k(d_b)$ in (4.33). Now

$$\varphi_k(\gamma) = O\left((1+k)^{\gamma-1}\right) \tag{4.86}$$

from Abramowitz and Stegun (1970). Also

$$\varphi_k(\gamma) - \varphi_{k+1}(\gamma) = -\varphi_{k+1}(\gamma - 1) = O\left((1+k)^{\gamma-2}\right) \;, \tag{4.87}$$

to check Assumption 3B. Next we take $\eta_{1t} = \eta_{at}$ in each case and $\eta_{2t} = \eta_{pt}$ and $\eta_{2t} = \eta_{bt}$ in (4.32) and (4.33) respectively. Now the spectral density matrix of η_t is $(2\pi)^{-1}A(e^{i\lambda})\sum A(e^{i\lambda})^*$ whose modulus is bounded by $C\left(\sum_{j=0}^{\infty}\|A_j\|\right)^2 < \infty$ from (4.21). Thus Assumption 3A is satisfied. Finally the fourth cumulant of $\eta_{a0}, \eta_{ai}, \eta_{b,i+j}, \eta_{b,i+j+k}$ is, for $i, j, k \geq 0$

$$cum\left(\sum_{d=-\infty}^{\infty}\alpha'_{a,-d}\varepsilon_d, \sum_{e=-\infty}^{\infty}\alpha'_{a,i-e}\varepsilon_e, \sum_{f=-\infty}^{\infty}\alpha'_{b,i+j-f}\varepsilon_f, \sum_{g=-\infty}^{\infty}\alpha'_{b,i+j+k-g}\varepsilon_g\right) \;, \tag{4.88}$$

where α'_{aj} is the a-th row of A_j. This is bounded in absolute value by

$$C\sum_{d=-\infty}^{\infty}\|\alpha_{a,-d}\|\,\|\alpha_{a,i-d}\|\,\|\alpha_{a,i+j-d}\|\,\|\alpha_{a,i+j+k-d}\| \;. \tag{4.89}$$

Because the sum of this, over all a, b, c, is finite due to (4.21), it follows that the Fourier coefficients of the fourth cumulant spectrum of $\eta_{at}, \eta_{at}, \eta_{bt}, \eta_{bt}$ are absolute summable, so that their spectrum is indeed bounded and Assumption 3D is satisfied. □

Proof of Lemma 4.2 We write

$$\begin{aligned} D\widehat{F}_{xx}(1,m)D &= D\widehat{F}_{xx}(1,n-1)D \\ &\quad -D\left\{\widehat{F}_{xx}(m+1,n-1) - E\widehat{F}_{xx}(m+1,n-1)\right\}D \\ &\quad -DE\widehat{F}_{xx}(m+1,n-1)D \;. \end{aligned} \tag{4.90}$$

In view of Lemma 4.1 the last two components are $o_p(1)$ and $o(1)$ from Theorems 3.2 and 3.1 respectively. The proof is completed by appealing to Theorem 4.1. $\square$

Proof of Lemma 4.3 We begin by estimating the b_{ajt}. First

$$\|A_j\| \leq \left(\sum_{\ell=j}^{\infty} \|A_\ell\|^2 \right)^{\frac{1}{2}} \tag{4.91}$$

$$\leq \frac{2}{j} \sum_{k=[\frac{j}{2}]}^{j} \left(\sum_{\ell=k}^{\infty} \|A_\ell\|^2 \right)^{\frac{1}{2}} \tag{4.92}$$

$$= o(j^{-1}) \text{ as } j \to \infty , \tag{4.93}$$

where (4.92) is due to monotonic decay of the right hand side of (4.91), and (4.93) follows from (4.18). The a-th diagonal element of Δ_k is $\varphi_k(d_a)$ for $a = 1, ..., p-1$ and $\varphi_k(d_e)$ for $a = p$, where $\varphi_k(\gamma)$ is given by (4.85). Now from (4.23), for $a = 1, ..., p-1$

$$\begin{aligned} \|b_{ajt}\| &\leq \sum_{\ell=0}^{r} \varphi_\ell(d_a) \|A_{j-\ell}\| + \sum_{\ell=r+1}^{\min(j,t-1)} \varphi_\ell(d_a) \|A_{j-\ell}\| \\ &\leq C \max_{j-r\leq\ell\leq j} \|A_\ell\| \sum_{\ell=0}^{r} \ell^{d_a-1} + Cr^{d_a-1} \sum_{\ell=0}^{\infty} \|A_\ell\| \\ &\leq Cr^{d_a}(j-r)^{-1} + Cr^{d_a-1} \end{aligned} \tag{4.94}$$

for $1 \leq r < \min(j, t-1)$. It follows that for $j \leq 2t$

$$\|b_{ajt}\| \leq Cj^{d_a-1} , i = 1, ..., p-1 , \tag{4.95}$$

on taking $r \approx j/2$. For $j > 2t$ we have more immediately

$$\|b_{ajt}\| \leq \sum_{\ell=0}^{t-1} |\varphi_\ell(d_a)| \parallel A_{j-\ell} \parallel \leq C(j-t)^{-1}t^{d_a} , a = 1, ..., p-1 . \tag{4.96}$$

Similarly

$$
\begin{aligned}
\|b_{pjt}\| &\leq O\left(j^{d_e-1}\right) \;, j \leq 2t \;. &(4.97)\\
&= O\left((j-t)^{-1}t^{d_e}\right) \;, j > 2t \;. &(4.98)
\end{aligned}
$$

Next notice that $b_{aj} = b_{ajt}$ for $0 \leq j < t$, so from (4.22)

$$
\begin{aligned}
E(x_{at}e_t) &= \sum_{j=0}^{t-1} b'_{aj} \sum b_{pj} + \sum_{j=t}^{\infty} b'_{ajt} \sum b_{pjt} \\
&= \xi_a - \sum_{j=t}^{\infty} b'_{aj} \sum b_{pj} + \sum_{j=t}^{2t-1} b'_{ajt} \sum b_{pjt} + \sum_{j=2t}^{\infty} b'_{ajt} \sum b_{pjt} \\
&= \xi_a + O\left(\sum_{j=t}^{\infty} j^{d_a+d_e-2} + t.t^{d_a+d_e-2} + t^{d_a+d_e} \sum_{j=t}^{\infty} j^{-2}\right) \\
&= \xi_a + O\left(t^{d_a+d_e-1}\right) &(4.99)
\end{aligned}
$$

because $d_a + d_e < 1$, where this and (4.95), (4.97) imply that $|\xi_a| < \infty$. On the other hand

$$
\begin{aligned}
|E(\overline{x}_a\overline{e})| &\leq \frac{C}{n^2} \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \|b_{ajs}\| \, \|b_{pkt}\| \\
&\leq \frac{C}{n^2} \left\{ \sum_{s=1}^{n} \sum_{t=1}^{n} \sum_{\max(0,s-t)}^{\min(2s,s+t)} j^{d_i-1}(j+t-s)^{d_e-1} + \sum_{s=1}^{n} \sum_{t=1}^{n} t^{2d_e} \sum_{s+t}^{2s} j^{d_i-1}(j-s)^{-1} \right. \\
&\quad \left. + \sum_{s=1}^{n} \sum_{t=1}^{n} s^{d_a} \sum_{2s}^{s+t} (j+t-s)^{d_e-1}(j-s)^{-1} + \sum_{s=1}^{n} \sum_{t=1}^{n} s^{d_a} t^{d_e} \sum_{\min(2s,s+t)}^{\infty} (j-s)^{-2} \right\} \\
&= O\left(n^{d_a-d_e-1}\right) \;. &(4.100)
\end{aligned}
$$

The proof is routinely completed in view of (4.3). □

Proof of Theorem 4.2 It is convenient to introduce the abbreviating notation

$$\widetilde{A} = \widehat{F}_{xx}(1, n-1),\ \widetilde{b} = \widehat{F}_{x\epsilon}(1, n-1),\ \widehat{A} = \widehat{F}_{xx}(1, m),\ \widehat{b} = \widehat{F}_{x\epsilon}(1, m)\ . \tag{4.101}$$

Thus

$$\widehat{\beta}_{n-1} - \beta = D(D\widetilde{A}D)^{-1}D\widetilde{b}\ , \tag{4.102}$$

$$\widehat{\beta}_{m} - \beta = D(D\widehat{A}D)^{-1}D\widehat{b}\ . \tag{4.103}$$

Now

$$D(\widetilde{A} - \widehat{A})D = D\left\{(\widetilde{A} - \widehat{A}) - E(\widetilde{A} - \widehat{A})\right\}D + DE(\widetilde{A} - \widehat{A})D \to_p 0 \tag{4.104}$$

from Theorems 3.1, 3.2, Assumption 4C and Lemma 4.1, so that $D\widetilde{A}D, D\widehat{A}D \Rightarrow V(d, \Omega)$ by Theorem 4.1. Now denote by $\widetilde{b}_a, \widehat{b}_a$ the a-th elements of $\widetilde{b}, \widehat{b}$. From Theorem 3.2, Assumption 4C and Lemma 4.1

$$\widetilde{b}_a = E\widetilde{b}_a + (\widetilde{b}_a - E\widetilde{b}_a) = E\widetilde{b}_a + O_p(n^{d_a+d_e-1}),\ a = 1, ..., p-1\ , \tag{4.105}$$

whereas from Lemma 4.3

$$\lim_{n\to\infty} n^{d_{\min}-\frac{1}{2}} DE\widetilde{b} = \xi\ . \tag{4.106}$$

Then (4.37) follows from (4.36). Finally

$$\widehat{b}_a = E\widehat{b}_a + \left\{\widehat{b}_a - E(\widehat{b}_a)\right\} = O\left(\left(\frac{n}{m}\right)^{d_a+d_e-1}\right) + O_p\left(n^{d_a+d_e-1}\right)\ , \tag{4.107}$$

from Theorems 3.1 and 3.2, Assumption 4C and Lemma 4.1 so the a-th element of (4.103) is

$$\begin{aligned} O_p\left(n^{\frac{1}{2}-d_a} \max_{1\le a<p} n^{\frac{1}{2}-d_a}\left(\frac{n}{m}\right)^{d_a+d_e-1}\right) &= O_p\left(n^{1-d_a-d_{\min}}\left(\frac{n}{m}\right)^{d_{\min}+d_e-1}\right) \\ &= o_p\left(n^{1-d_a-d_{\min}}\right)\ , \end{aligned} \tag{4.108}$$

since $d_{min} + d_e < 1$, to complete the proof of (4.38). $\square$

Proof of Lemma 4.4 For $1 \leq j \leq m$, writing $\widetilde{\Gamma}_j = \sum_{\ell \geq j}^{\infty} \gamma_\ell$,

$$\begin{aligned} EI_{xe}(\lambda_j) &= \frac{1}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n}(\gamma_{s-1} + \ldots + \gamma_{s-t})e^{i(t-s)\lambda_j} \\ &= \frac{1}{2\pi n}\sum_{s=1}^{n}\sum_{t=1}^{n}\left(\widetilde{\Gamma}_{s-t} - \widetilde{\Gamma}_{s-1}\right)e^{i(t-s)\lambda_j} \\ &= \frac{1}{2\pi}\sum_{\ell=1-n}^{n-1}\left(1 - \frac{|\ell|}{n}\right)\widetilde{\Gamma}_\ell e^{-i\ell\lambda_j} \end{aligned} \tag{4.109}$$

from (3.67). Now for $\ell \geq 0$ $\widetilde{\Gamma}_\ell = \Gamma_\ell$, whereas for $\ell < 0$, $\widetilde{\Gamma}_\ell = \Gamma_0 + \Gamma_{-1} - \Gamma_{\ell-1}$, so (4.109) has real part

$$\frac{1}{2\pi}\sum_{\ell=0}^{n-1}\left(1 - \frac{\ell}{n}\right)\Gamma_\ell \cos \ell\lambda_j + \frac{1}{2\pi}\sum_{\ell=1-n}^{-1}\left(1 + \frac{\ell}{n}\right)(\Gamma_0 + \Gamma_{-1} - \Gamma_{\ell-1})\cos \ell\lambda_j\,. \tag{4.110}$$

The first term can be written

$$\frac{1}{4\pi}\sum_{\ell=1-n}^{n-1}\left(1 - \frac{|\ell|}{n}\right)\Gamma_{|\ell|}\cos \ell\lambda_j + \frac{\Gamma_0}{4\pi}\,. \tag{4.111}$$

To deal with the second term of (4.110) note that for $1 \leq j \leq n-1$

$$\sum_{\ell=0}^{n-1}\ell e^{i\ell\lambda_j} = \frac{e^{i\lambda_j} - 1}{(1 - e^{i\lambda_j})^2} - \frac{(n-1)}{1 - e^{i\lambda_j}} = \frac{-n}{1 - e^{i\lambda_j}}\,, \tag{4.112}$$

which has real part

$$-\frac{n}{2}\left(\frac{1}{1 - e^{i\lambda_j}} + \frac{1}{1 - e^{-i\lambda_j}}\right) = -\frac{n}{2}\left(\frac{2 - 2\cos\lambda_j}{2 - 2\cos\lambda_j}\right) = -\frac{n}{2}\,. \tag{4.113}$$

Thus, the second term in (4.110) is

$$\frac{(\Gamma_0+\Gamma_{-1})}{2\pi}\sum_{\ell=0}^{n-1}\left(1-\frac{\ell}{n}\right)\cos\ell\lambda_j-\frac{\Gamma_0+\Gamma_{-1}}{2\pi}$$
$$-\frac{1}{4\pi}\sum_{\ell=1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\Gamma_{-|\ell|-1}\cos\ell\lambda_j+\frac{\Gamma_{-1}}{4\pi}$$

$$=-\frac{1}{4\pi}\sum_{\ell=1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\Gamma_{-|\ell|-1}\cos\ell\lambda_j+\frac{\Gamma_0}{4\pi}\,. \tag{4.114}$$

It follows that (4.109) has real part

$$\frac{1}{4\pi}\sum_{\ell=1-n}^{n-1}\left(1-\frac{|\ell|}{n}\right)\left(\Gamma_{-|\ell|}-\Gamma_{-|\ell|-1}\right)\cos\ell\lambda_j\,, \tag{4.115}$$

which is the Cesaro sum, to $n-1$ terms, of the Fourier series of $h(\lambda_j)/2$. Equivalently we can write

$$E\left\{\frac{n}{m}\widehat{F}_{xe}(1,m)\right\}=\frac{1}{4\pi nm}\sum_{j=1}^{m}\int_{-\pi}^{\pi}|D_n(\lambda-\lambda_j)|^2\,h(\lambda)d\lambda\,. \tag{4.116}$$

Fix $\varepsilon>0$. There exists $\delta>0$ such that $\|h(\lambda)-h(0)\|<\varepsilon$ for $0<|\lambda|\le\delta$. Let n be large enough that $2\lambda_m<\delta$. Then the difference between the right hand side of (4.116) and $h(0)/2$ is bounded in absolute value by

$$\frac{1}{4\pi nm}\sum_{j=1}^{m}\int_{-\pi}^{\pi}|D_n(\lambda-\lambda_j)|^2\,\|h(\lambda)-h(0)\|\,d\lambda$$
$$\le\ \frac{1}{4\pi nm}\left\{\varepsilon\max_{1\le j\le m}\int_{-\delta}^{\delta}|D_n(\lambda-\lambda_j)|^2\,d\lambda+\sup_{\frac{\delta}{2}<|\lambda|<\pi}|D_n(\lambda)|^2\left(\int_{-\pi}^{\pi}\|h(\lambda)\|\,d\lambda+2\pi\,\|h(0)\|\right)\right\}$$
$$=\ O\left(\varepsilon+\frac{1}{n}\right)\,, \tag{4.117}$$

because of Assumption 4D, (??) and

$$\int_{-\pi}^{\pi} |D_n(\lambda)|^2 \, d\lambda = 2\pi n \, . \tag{4.118}$$

Because ε is arbitrary, the proof is complete. $\square$

Proof of Theorem 4.3 (4.47) is familiar under somewhat different conditions from ours (see e.g. Theorem 1.2 and the other references mentioned in Section 1.2), but we briefly describe its proof in order to indicate how the outcome differs from (4.48). We have

$$n\left(\widehat{\beta}_{n-1} - \beta\right) = \left(n^{-1}\widetilde{A}\right)^{-1} \{(\widetilde{a} - E\widetilde{a}) + E\widetilde{a}\} \, . \tag{4.119}$$

Now

$$n^{-1}\widetilde{A} \Rightarrow V(\iota, \Omega) \, , \; \widetilde{a} - E\widetilde{a} \Rightarrow U(\Omega) \, , \text{ as } n \to \infty \tag{4.120}$$

from Assumption 4C, Theorem 4.1 and the continuous mapping theorem. Because $E\widetilde{a} \to \Gamma_0$ by elementary calculations and $V(\iota, \Omega)$ is a.s. of full rank by Phillips and Hansen (1990), (4.47) is proved. Next

$$\begin{aligned} n(\widehat{\beta}_m - \beta) &= \left[n^{-1}\widetilde{A} + n^{-1}\left\{(\widehat{A} - \widetilde{A}) - E(\widehat{A} - \widetilde{A})\right\} + n^{-1}E(\widehat{A} - \widetilde{A})\right]^{-1} \\ &\quad \times [\widetilde{a} - E\widetilde{a} + \{(\widehat{a} - \widetilde{a}) - E(\widehat{a} - \widetilde{a})\} + E\widetilde{a}] \, . \end{aligned} \tag{4.121}$$

Since $n^{-1}\left\{\widehat{A} - \widetilde{A} - E(\widehat{A} - \widetilde{A})\right\} \to_p 0$, $n^{-1}E(\widehat{A} - \widetilde{A}) \to 0$ and $\widehat{a} - \widetilde{a} - E(\widehat{a} - \widetilde{a}) \to_p 0$ from Theorems 3.1, 3.2, and Lemma 4.1, the proof of (4.48) is completed by invoking (4.120), Lemma 4.4 and (4.4). $\square$

Proof of Theorem 4.4 From (4.102), Theorem 4.1, (4.36), (3.17) of Theorem 3.1 and

(3.21) of Theorem 3.2 we deduce (4.49). Next, using (4.103),

$$\widehat{\beta}_{n-1} - \widehat{\beta}_m = D(D\widehat{A}D)^{-1}\left\{D(\widehat{A} - \widetilde{A})D\right\}(D\widetilde{A}D)^{-1}D\widetilde{a} \tag{4.122}$$

$$-(D\widehat{A}D)^{-1}D\left\{(\widehat{a} - \widetilde{a}) - E(\widehat{a} - \widetilde{a})\right\} . \tag{4.123}$$

The a-th element of the right side of (4.122) is $o_p\left(n^{d_a+d_e}\right)$ by arguments used in the previous proof and (4.104), while the a-th element of (4.123) is also $o_p\left(n^{d_a+d_e}\right)$ on applying also (3.16) of Theorem 3.1 and (3.20) of Theorem 3.2, to prove (4.50). Then (4.51) is a consequence of (4.49) and (4.50). □

Proof of Theorem 4.6 The proof of (4.60) follows routinely from (4.102), (4.31), (4.36) and (4.59). Then (4.61) is a consequence of (4.50) and (4.60), because in view of (4.123) it is clear that (4.50) holds for all $d_e < d_{\min}$. □

TABLE 4.1: MONTE CARLO BIAS AND MSE FOR MODEL A

	BIAS, n=64						MSE, n=64					
φ_1	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.194	.210	.229	.295	.171	.154	.128	.123	.130	.154	.094	.091
.6	.068	.083	.096	.177	.062	.041	.033	.033	.034	.059	.027	.024
.4	.031	.037	.046	.125	.036	.014	.015	.014	.014	.031	.011	.009
.2	.017	.020	.026	.097	.024	.003	.009	.007	.008	.019	.007	.006

	BIAS, n=128						MSE, n=128					
φ_1	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.074	.087	.110	.175	.075	.062	.034	.034	.037	.057	.026	.024
.6	.020	.023	.035	.096	.018	.006	.008	.008	.008	.018	.005	.006
.4	.008	.010	.015	.066	.010	-1e-4	.004	.003	.003	.009	.002	.002
.2	.003	.005	.007	.051	.008	-4e-4	.001	.001	.001	.003	.001	.001

	BIAS, n=256						MSE, n=256					
φ_1	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
.8	.038	.048	.053	.097	.026	.019	.008	.009	.009	.018	.006	.005
.6	.008	.013	.015	.050	.003	-.002	.002	.002	.002	.005	.001	.001
.4	.004	.005	.006	.034	.003	-.001	7e-4	7e-4	7e-4	.002	5e-4	5e-4
.2	.001	.003	.004	.026	.002	-.001	4e-4	3e-4	4e-4	.001	3e-4	3e-4

TABLE 4.2: MONTE CARLO BIAS AND MSE FOR MODEL B

	BIAS, n=64						MSE, n=64					
λ^*	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-.008	-.010	-.010	.089	.025	.007	.007	.007	.006	.025	.010	.007
$\frac{4\pi}{9}$	-.005	-.006	-.007	.057	.032	.022	.003	.002	.002	.011	.006	.004
$\frac{5\pi}{9}$	-.002	-.005	-.006	.040	.011	.003	.001	.001	.001	.007	.002	.002
$\frac{2\pi}{3}$	-.003	-.003	-.004	.030	.031	.026	.001	.001	7e-4	.005	.004	.003

	BIAS, n=128						MSE, n=128					
λ^*	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_6$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-.001	-.003	-.004	.0.44	.007	6e-4	.002	.001	.001	.005	.002	.001
$\frac{4\pi}{9}$	-.001	-.001	-.002	.026	.005	5e-4	5e-4	5e-4	4e-4	.002	5e-4	4e-4
$\frac{5\pi}{9}$	-.001	-.001	-.001	.020	.014	.011	3e-4	3e-4	2e-4	.001	7e-4	5e-4
$\frac{2\pi}{3}$	-.001	-.001	-.001	.015	.011	.009	2e-4	2e-4	2e-4	9e-4	5e-4	4e-4

	BIAS, n=256						MSE, n=256					
λ^*	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$	$\widehat{\beta}_6$	$\widehat{\beta}_8$	$\widehat{\beta}_{10}$	$\widehat{\beta}_{n-1}$	$\widetilde{\beta}_{FM}$	$\widetilde{\beta}^*_{FM}$
$\frac{\pi}{3}$	-5e-4	-.001	-.002	.022	-.002	-.004	3e-4	3e-4	3e-4	.001	3e-4	3e-4
$\frac{4\pi}{9}$	-.001	-7e-4	-7e-4	.013	.005	.004	1e-4	1e-4	1e-4	4e-4	2e-4	1e-4
$\frac{5\pi}{9}$	-3e-4	-6e-4	-.001	.009	.005	.003	5e-5	5e-5	5e-5	2e-4	1e-4	7e-5
$\frac{2\pi}{3}$	-3e-4	-5e-4	-.001	.007	.007	.006	3e-5	3e-5	3e-5	2e-4	1e-4	9e-5

TABLE 4.3: CONSUMPTION (y) AND INCOME (x)

(n=138, $\widehat{\beta}_{n-1}$=.229, $1-R^2$=.009)

1) **Memory of Raw Data**

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widetilde{d}_{GLS}$	CI	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
22	.89	1.13	1.06	.95	.94, .97	.99	.78, 1.20	1.13	.92, 1.34
30	.95	1.04	.02	.98	.97, .99	1.03	.84, 1.21	1.10	.93, 1.29
40	1.02	1.04	.02	1.02	1.02, 1.03	1.08	.92, 1.24	1.12	.96, 1.28

2) **Cointegration Analysis**

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.231	.219	.85	.86	.003	.013
4	.232	.210	.88	.89	.004	.013
6	.232	.201	.93	.93	.004	.013

3) **Memory of Cointegrating Error**

ℓ	$\widetilde{d}^*_e$	CI	$\widetilde{d}_e$	CI	$\widehat{d}^*_e$	CI	$\widehat{d}_e$	CI
22	.20	-.05, 46	.56	.29, .84	.44	.22, .65	.62	.41, .83
30	.57	.27, .87	.84	.60, 1.07	.68	.49, .86	.78	.60, .96
40	.61	.38, .84	.86	.66, 1.06	.76	.60, .92	.87	.71, 1.02

TABLE 4.4: STOCK PRICES (y) AND DIVIDENDS (x)

(n=116, $\widehat{\beta}_{n-1}$=30.99, $1-R^2$=.15)

1) Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widetilde{d}_{GLS}$	CI	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
22	.91	.96	.07	.94	.92, .96	.36	.15, .57	1.04	.83, 1.25
30	.86	.83	.04	.84	.83, .85	.48	.30, .66	.91	.73, 1.09
40	.91	.84	.36	.87	.86, .87	.70	.54, .86	.90	.74, 1.06

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	33.16	23.24	.78	.84	.076	.215
4	33.55	21.49	.79	.85	.093	.210
6	32.47	22.81	.85	.89	.114	.190

3) Memory of Cointegrating Error

ℓ	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI	$\widehat{d}^*_e$	$\widehat{d}_e$	CI
22	.73	.74	.47, 1.01	.95	1.04	.83, 1.26
30	.60	.60	.36, .83	.85	.91	.73, 1.09
40	.64	.66	.46, .86	.84	.90	.74, 1.06

TABLE 4.5: LOG PRICES (y) AND LOG WAGES (x)

(n=360, $\widehat{\beta}_{n-1}$=.706, $1-R^2$=.033)

Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
30	1.16	1.60	5.84	1.07	.89, 1.25	1.24	1.06, 1.42
40	1.03	1.54	11.1	1.07	.92, 1.23	1.25	1.09, 1.41
60	.99	1.54	19.9	1.07	.94, 1.20	1.27	1.14, 1.40

TABLE 4.6: LOG L (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=1.039, $1-R^2$=.00085)

Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
16	1.29	1.61	5.51	1.23	.98, 1.48	1.46	1.21, 1.71
22	1.36	1.68	6.30	1.25	1.03, 1.46	1.56	1.35, 1.77
30	1.29	1.68	9.99	1.22	1.04, 1.40	1.60	1.42, 1.68

TABLE 4.7: LOG M2 (y) AND LOG NOMINAL GNP (x)

($n=90$, $\widehat{\beta}_{n-1}=.99$, $1-R^2=.0026$)

1) Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widetilde{d}_{GLS}$	CI	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
16	1.29	1.53	.69	1.29	1.28, 1.30	1.23	.98, 1.48	1.35	1.10, 1.60
22	1.36	1.56	.83	1.38	1.37, 1.39	1.25	1.03, 1.46	1.47	1.25, 1.69
30	1.29	1.64	3.67	1.33	1.32, 1.34	1.22	1.04, 1.40	1.59	1.41, 1.78

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.99	.98	.83	.84	.002	.003
4	.99	.99	.87	.87	.002	.003
6	.99	.99	.91	.91	.003	.003

3) Memory of Cointegrating Error

ℓ	$\widetilde{d}^*_x$	$\widetilde{d}_e$	CI	$\widehat{d}^*_e$	$\widehat{d}_e$	CI
16	1.15	1.19	.88, 1.52	1.20	1.23	.98, 1.48
22	1.10	1.16	.89, 1.43	1.04	1.10	.89, 1.31
30	1.10	1.15	.92, 1.38	1.02	1.09	.91, 1.27

TABLE 4.8: LOG M1 (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=.643, $1-R^2$=.00309)

1) Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widetilde{d}_{GLS}$	CI	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
16	1.29	1.53	7.70	1.31	1.30, 1.32	1.23	.98, 1.48	1.39	1.14, 1.64
22	1.36	1.42	.346	1.39	1.38, 1.40	1.25	.97, 1.46	1.33	1.12, 1.55
30	1.29	1.29	.001	1.29	1.28, 1.30	1.22	1.04, 1.40	1.27	1.09, 1.45

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1-R^2_m$	$1-R^2_{-m}$
3	.645	.632	.83	.84	.003	.003
4	.645	.630	.87	.88	.003	.003
6	.645	.629	.91	.91	.003	.003

3) Memory of Cointegrating Error

ℓ	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI	$\widehat{d}^*_e$	$\widehat{d}_e$	CI
16	.99	1.20	.88, 1.52	.97	1.06	.81, 1.31
22	.76	1.07	.80, 1.34	.77	.92	.71, 1.13
30	.78	.88	.64, 1.11	.76	.85	.67, 1.03

TABLE 4.9: LOG M3 (y) AND LOG NOMINAL GNP (x)

(n=90, $\widehat{\beta}_{n-1}$=1.0997, $1 - R^2$=.0023)

1) Memory of Raw Data

ℓ	$\widetilde{d}_x$	$\widetilde{d}_y$	W	$\widetilde{d}_{GLS}$	CI	$\widehat{d}_x$	CI	$\widehat{d}_y$	CI
16	1.29	1.45	.79	1.33	1.33, 1.34	1.23	.98, 1.48	1.34	1.08, 1.60
22	1.36	1.62	2.01	1.44	1.43, 1.44	1.25	1.03, 1.46	1.50	1.28, 1.71
30	1.29	1.71	7.04	1.42	1.41, 1.43	1.22	1.04, 1.40	1.65	1.47, 1.83

2) Cointegration Analysis

m	$\widehat{\beta}_m$	$\widehat{\beta}_{-m}$	$r_{xx,m}$	$r_{xy,m}$	$1 - R^2_m$	$1 - R^2_{-m}$
3	1.10	1.10	.83	.83	.002	.002
4	1.10	1.10	.87	.87	.002	.002
6	1.10	1.10	.91	.91	.002	.002

3) Memory of Cointegrating Error

ℓ	$\widetilde{d}^*_e$	$\widetilde{d}_e$	CI	$\widehat{d}^*_e$	$\widehat{d}_e$	CI
16	.88	.89	.57, 1.21	1.02	1.02	.76, 1.28
22	.97	1.00	.73, 1.27	1.05	1.08	.87, 1.29
30	.96	1.01	.78, 1.21	.98	1.04	.86, 1.22

Chapter 5

COINTEGRATED TIME SERIES WITH LONG MEMORY INNOVATIONS

5.1 Introduction

Consider the two-dimensional observations (y_t, x_t), $t = 1, 2, \ldots.$, where

$$\begin{cases} y_t = \beta x_t + e_t \\ x_t = \phi x_{t-1} + u_t \end{cases} , \ u_t \sim I(d_u) \ , \ e_t \sim I(d_e) \ , \ |\phi| \leq 1 \ , \tag{5.1}$$

with $0 \leq d_e, d_u < \frac{1}{2}$. Here, as in Section 4.3, we adopt for u_t, e_t the stationary definition of fractional integration, i.e. we identify u_t (e_t) with z_t and we substitute η_t for $\eta_t 1(t > 0)$ on the right hand side of (1.2). We allow for nonstationarity in x_t, y_t by including the possibility that ϕ equals unity, thus covering partial sums of long memory innovations as in (1.129); it would be tempting in this case to indicate $x_t \sim I(d_x)$, and hence to term the bivariate process (y_t, x_t) cointegrated of fractional order (d_x, d_e), with $d_x = d_u + 1$; indeed this is the terminology adopted by Dolado and Marmol (1996) in similar circumstances, as mentioned in Section 4.1. Note however that this terminology does not agree with Section 4.4, where the definition of fractional cointegration in the nonstationary case $d > \frac{1}{2}$ is reserved for processes which generalize (1.2) via Assumption 2A or Assumption 4C (cf. also Chapter 2, Section 3); to prevent ambiguities, it may be better when $\phi = 1$ to term x_t a unit root process with long memory innovations, and to define (5.1) a cointegrated vector with long memory (or fractionally integrated) innovations, as in Jeganathan (1996), where this model is investigated under the assumption that the distribution of (e_t, u_t) is known. For $\phi = 1$, (5.1) can hence be considered an alternative avenue for the generalization of the $CI(1)$ case to fractional circumstances.

For fractional cointegration analysis, in Chapter 4 we advocated a form of narrow-band estimates that in the bivariate case could be considered the solution of the following minimization problem:

$$\widehat{\beta}_m = Re\left\{\arg\min_{\beta} \sum_{j=1}^{m} |w_y(\lambda_j) - \beta w_x(\lambda_j)|^2\right\} \ , \ \lambda_j = \frac{2\pi j}{n} \ , \ j = 1, 2, ...m \ , \tag{5.2}$$

for $w_y(.), w_x(.)$ discrete Fourier tranforms and under the bandwidth condition

$$m < n \ , \ \frac{1}{m} + \frac{m}{n} \to 0 \text{ as } n \to \infty \ . \tag{5.3}$$

The solution of (5.2) can be written in closed form as

$$\widehat{\beta}_m = Re\left\{\frac{\sum_{j=1}^{m} w_x(\lambda_j) w_y(\lambda_j)^*}{\sum_{j=1}^{m} w_x(\lambda_j) w_x(\lambda_j)^*}\right\} = \frac{\widehat{F}_{xy}(1,m)}{\widehat{F}_{xx}(1,m)} \ , \tag{5.4}$$

where $\widehat{F}(1,m)$ denotes as in (1.20) the real part of the discretely averaged periodogram. We recall from Chapter 1 that for stationary, short range dependent vector sequences $z_t = (y_t, x_t)$, the statistic

$$\widehat{f}_{zz}(0) = \lambda_m^{-1} \widehat{F}_{zz}(1,m) \tag{5.5}$$

provides under (5.3) and regularity conditions a consistent estimate of the spectral density matrix at the origin; hence we can rewrite formally

$$\widehat{\beta}_m = \frac{\widehat{F}_{xy}(1,m)/\lambda_m}{\widehat{F}_{xx}(1,m)/\lambda_m} = \frac{\widehat{f}_{xy}(0)}{\widehat{f}_{xx}(0)} . \tag{5.6}$$

The behaviour of (5.4) under (5.1) is difficult to establish in general, although some possibilities considered in Chapter 4 are relevant here, for instance the stationary and $CI(1)$ cases. Rather than pursuing a complete characterization of FDLS under cointegration with long memory innovations, in this chapter we find more interesting to investigate instead the behaviour of a different frequency-domain semiparametric procedure, entailing minimization of a functional of the continuously averaged periodogram; namely we consider

$$\widetilde{\beta}_M = Re\left\{\arg\min_{\beta} \int_{-\pi}^{\pi} K_M(\lambda)|w_y(\lambda) - \beta w_x(\lambda)|^2 d\lambda\right\} , \tag{5.7}$$

which in closed form becomes

$$\widetilde{\beta}_M = \frac{\int_{-\pi}^{\pi} K_M(\lambda) I_{xy}(\lambda) d\lambda}{\int_{-\pi}^{\pi} K_M(\lambda) I_{xx}(\lambda) d\lambda} , \tag{5.8}$$

where for $M = 1, 2, \ldots$ $K_M(\lambda)$ represents a frequency-domain kernel such that $K_M(-\lambda) = K_M(\lambda)$, M representing a bandwidth parameter such that

$$M < n \, , \, \frac{1}{M} + \frac{M}{n} \to 0 \text{ as } n \to \infty \, . \tag{5.9}$$

For short range dependent processes an estimate of the spectral density matrix at frequency ω alternative to (5.5) is given by (cf. (1.22))

$$\widetilde{f}(\omega) = \int_{-\pi}^{\pi} K_M(\lambda) I(\omega - \lambda) d\lambda \, , \tag{5.10}$$

whence we can rewrite $\widetilde{\beta}_M$ (formally) as

$$\widetilde{\beta}_M = \frac{\widetilde{f}_{xy}(0)}{\widetilde{f}_{xx}(0)}, \; \widetilde{f}_{ab}(0) = \int_{-\pi}^{\pi} K_M(\lambda) I_{ab}(\lambda) d\lambda, \; a, b = x, y \, , \tag{5.11}$$

so that, analogously to $\widehat{\beta}_m$, $\widetilde{\beta}_M$ can be interpreted as resulting from a form of spectral regression carried over the components of x_t and y_t corresponding to the smallest frequencies. Under regularity conditions (Brillinger (1981)), $\widetilde{f}_{ab}(0)$ is equivalent to

$$\widetilde{f}_{ab}(0) = \sum_{\tau=-M}^{M} k_M(\tau) c_{ab}(\tau) \ , \tag{5.12}$$

where $k(.)$ is the lag window defined by $k_M(\tau) = (2\pi)^{-1} \int_{-\infty}^{\infty} K_M(\alpha) e^{i\tau\alpha} d\alpha$, and

$$c_{ab}(\tau) = \begin{cases} \sum_{t=1}^{n-\tau} a_t b_{t+\tau} \ , & \tau \geq 0 \\ \sum_{t=|\tau|+1}^{n} a_t b_{t-\tau} \ , & \tau < 0 \end{cases} \ . \tag{5.13}$$

Therefore we can rewrite for (5.8)

$$\widetilde{\beta}_M = \frac{\sum_{\tau=-M}^{M} k_M(\tau) c_{xy}(\tau)}{\sum_{\tau=-M}^{M} k_M(\tau) c_{xx}(\tau)} \ , \ 1 \leq M \leq n-1 \ , \tag{5.14}$$

and adopt for $\widetilde{\beta}_M$ the natural definition of Weighted Covariance Estimate (WCE). Note that (5.6) can also be approximated by (5.14), if we take $k_M(\tau) = \sin(\pi v)/\pi v$, $v = \tau/M$ and $M = n-1$. However, for convenience we shall impose the bandwidth condition (5.9), which rules out (5.6). For short range dependent (e_t, u_t) and $\phi = 1$, i.e. in the $CI(1)$ case, $\widetilde{\beta}_M$ is the special case for $\widehat{f}_{vv}(0) \equiv 1$ of the estimates (1.83) considered by Phillips (1991b) and reviewed in Chapter 1, cf. Theorem 1.10; the following sections analyzes the behaviour of $\widetilde{\beta}_M$ in the fractional circumstances considered in this chapter.

We shall show that under suitable conditions (including $d_u > d_e$ in the stationary case, considered in Section 2) WCE is consistent, even if x_t and e_t are not orthogonal. In other words, given long memory behaviour in the innovations, the presence or absence of a unit root in the data generating process (DGP) of x_t is no longer crucial to derive consistent estimates for β - a conclusion that mirrors the results from Chapter 4. Hence we can again overcome the sharp distinction between stationarity and nonstationarity which characterizes the $CI(1)$ case. For some aspects we are also able to derive here sharper results than in Chapter 4. and in particular we characterize the limit distribution of OLS and WCE estimates when the DGP of x_t does include a unit root and e_t is stationary long memory. Companion to this derivation is a functional central limit theorem for a class of quadratic forms in nonstationary fractionally integrated variables, a result which may have some independent interest and can be extended to more general quadratic forms, as shown at the end of Section 3. In Section 4 some numerical evidence to motivate the procedure and investigate its finite sample performance is presented; most proofs are collected in the Appendix.

Throughout the chapter, we restrict our attention to the bivariate case for simplicity; multivariate generalizations require in the nonstationary case suitable extensions of functional central limit theorems from Gorodetskii (1977) and Chan and Terrin (1995); such extensions are still under investigation.

5.2 The Stationary Case

When ϕ is in absolute value smaller than unity, we find it notationally convenient to specify a model for the covariance stationary sequence $(e_t, x_t) = ((e_t, 1-\phi L)^{-1}u_t)$ rather than for (e_t, u_t), and to write d_x for d_u.

Assumption 5A Assume that (5.1) hold, where $|\phi| < 1$ and

$$(e_t, x_t)' = \Psi(L)\varepsilon_t \ , \ \Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k \ , \tag{5.15}$$

with $\Psi_0 = I_p$, and for $k = 1, 2, ...$ Ψ_k has (a, b)-th element

$$\psi_{1bk} \approx c_{1b} k^{d_e - 1} \ , \ \psi_{2bk} \approx c_{2b} k^{d_x - 1} \ , \text{ as } k \to \infty \ , \ 0 < c_{ab} < \infty \ , \tag{5.16}$$

for $a, b = 1, 2$, and where $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ represents a zero-mean, *i.i.d.* sequence that satisfies $E||\varepsilon_t||^4 < \infty$.

The meaning and interpretation of Assumption 5A has been largely discussed in the previous chapters, cf. Assumption 4A and 4C. As a consequence of (5.16), as $\tau \to \pm\infty$,

$$Ex_t x_{t+\tau} \approx g_x |\tau|^{2d_x - 1},\ 0 < g_x < \infty,\ Ex_t e_{t+\tau} \approx g_{xe} |\tau|^{d_x + d_e - 1}, \tag{5.17}$$

cf (1.87) and Bladt (1994); $g_{xe} \equiv 0$ if $E\varepsilon_{1t}\varepsilon_{2t} = 0$.

In the sequel, we find it convenient to set $k_M(\tau) \triangleq k(\frac{\tau}{M})$, and to introduce

Assumption 5B The kernel $k(.)$ is a real-valued, Lebesgue-measurable function that for $v \in R$ satisfies

$$\int_{-1}^{1} k(v)dv = 1,\ 0 \le k(v) \le C\ ,\ k(v) = 0,\ |v| > 1\ . \tag{5.18}$$

Assumption 5B is common for spectral estimates, and it is satisfied by (normalized versions of) truncated kernels such as the Bartlett, modified Bartlett, Parzen, and many others; see Brillinger (1981) for a review.

Lemma 5.1 Under Assumptions 5A and 5B, as $M \to \infty$ for $M = o(n^2)$ we have

$$\left\{ \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\gamma_{xx}(\tau) \right\}^{-1} \sum_{\tau=-M}^{M} k(\frac{\tau}{M}) c_{xx}(\tau) \quad = \quad 1 + o_p(1)\ , \tag{5.19}$$

$$\left\{ \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\gamma_{xe}(\tau) \right\}^{-1} \sum_{\tau=-M}^{M} k(\frac{\tau}{M}) c_{xe}(\tau) \quad = \quad 1 + o_p(1)\ . \tag{5.20}$$

For $d_x > 0$, the spectral density of x_t has a singularity at frequency zero and cannot be estimated there; as reported in Chapter 1, Theorem 1.16, it was shown by Robinson (1994a) under (5.3) and regularity conditions weaker than 5A that

$$p\lim \frac{\widehat{F}(1,m)/\lambda_m}{F(1,m)/\lambda_m} = 1 \text{ , as } n \to \infty \text{ .} \tag{5.21}$$

Lemma 5.1 is similar to (5.21), relating however to the case when the continuously averaged periodogram is considered.

Under Assumption 5A and in view of (5.17), (5.18), as $M \to \infty$ we have, by the dominated convergence theorem

$$M^{-2d_x} \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\gamma_{xx}(\tau) = \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\frac{\gamma_{xx}(\tau)}{M^{2d_x-1}}\frac{1}{M} \approx B_{xx}, \tag{5.22}$$

$$M^{-d_x-d_e} \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\gamma_{xe}(\tau) = \sum_{\tau=-M}^{M} k(\frac{\tau}{M})\frac{\gamma_{xe}(\tau)}{M^{d_x+d_e-1}}\frac{1}{M} \approx B_{xe}, \tag{5.23}$$

for $\gamma_{ab}(\tau) = Ea_0b_\tau$, $a, b = x, e$, and

$$B_{xx} = g_x \int_{-1}^{1} k(v)v^{1-2d_x}dv \text{ , } B_{xe} = g_{xe} \int_{-1}^{1} k(v)v^{1-d_x-d_e}dv \text{ .} \tag{5.24}$$

Hence B_{xe} can be equal to zero if g_{xe} is, in which case the left-hand side of (5.23) is $o_p(1)$. As an application of Lemma 5.1, we consider the statistic

$$\ln |\sum_{\tau=-M}^{M} k(\frac{\tau}{M})c_{xx}(\tau)| - 2d_x \ln M = \ln B_{xx} + o_p(1) \text{ as } n \to \infty, \tag{5.25}$$

whence a consistent estimate of the parameter d_x can be obtained under Assumptions 5A, 5B and (5.9) by

$$\widehat{d_x} = \frac{\ln |\sum_{\tau=-M}^{M} k(\frac{\tau}{M})c_{xx}(\tau)|}{2\ln M} \text{ .} \tag{5.26}$$

This estimate is likely to be severely biased in finite samples, though, and rather than investigating in more detail its properties we concentrate on (5.1), for which we introduce the following result.

Theorem 5.1 Under (5.1), Assumptions 5A-5B and $M^2 = o(n)$, as $M \to \infty$

$$M^{d_x - d_e}(\widetilde{\beta}_M - \beta) = \frac{B_{xe}}{B_{xx}} + o_p(1) \,. \tag{5.27}$$

Theorem 5.1 suggests that the presence of correlation between x_t and e_t does not prevent consistency of $\widetilde{\beta}_M$, the rate of convergence being determined by the strength of the cointegrating relationship $d_x - d_e$ (cf. Theorem 4.1). We delay to future research the investigation of issues such as: the determination of optimal bandwidth parameter M (Robinson (1994c); the choice of an optimal kernel $k(.)$; the estimation of the bias term B_{xe}/B_{xx}, the implementation of bias reduction techniques, and the derivation of the asymptotic distribution for the adjusted estimate; the comparison between performance of WCE and FDLS in the case of cointegration between stationary variables. We focus instead on the unit root case, which is dealt with in the next section.

5.3 The Unit Root Case

The unit root case is characterized by the identification $\phi = 1$ in (5.1), so to obtain (after the initialization $x_0 = 0$)

$$x_t = \sum_{s=1}^{t} u_s \;,\; t = 1, 2, \ldots \,. \tag{5.28}$$

We consider first the $CI(1)$ case. For convenience, we write $\omega_{ab} = 2\pi f_{ab}(0)$, $a, b = u, e$, with $|\omega_{ab}| < \infty$, because of course (cf. Chapter 1)

$$0 < f(0) < \infty \,. \tag{5.29}$$

The following result is proved by Phillips (1991b), cf. Theorem 1.10 in Chapter 1.

Lemma 5.2 (*Phillips 1991b*) Let (5.1) hold for $\phi = 1$, and assume that as $n \to \infty$

$$\frac{1}{n^2}\sum_{t=1}^{n}(\sum_{s=1}^{t} u_s)^2 \Rightarrow \int_0^1 B^2(r;\omega_{22})dr \ , \tag{5.30}$$

$$\frac{1}{n}\sum_{t=1}^{n}(\sum_{s=1}^{t} u_s)e_t \Rightarrow \int_0^1 B(r;\omega_{22})dB(r;\omega_{11}) + \sum_{\tau=0}^{\infty}\gamma_{ue}(\tau) \ , \tag{5.31}$$

Assume also that Assumption 5B holds. Then under (5.9), as $n \to \infty$

$$n(\tilde{\beta}_M - \beta) \Rightarrow \left(\int_0^1 B^2(r;\omega_{22})dr\right)^{-1}\left(\int_0^1 B(r;\omega_{22})dB(r;\omega_{11}) + \sum_{\tau=0}^{\infty}\gamma_{ue}(\tau)\right). \tag{5.32}$$

Regularity conditions under which (5.30)/(5.31) hold are discussed in Chapter 1, Theorems 1.2-1.6, and hence need not be reviewed here.

When (5.29) fails, e_t and u_t are not short range dependent and the asymptotics for $\tilde{\beta}_M$ depends on functional central limit theorems for normalized partial sums of long memory innovations. As discussed in Chapter 2, Section 2, such results have now been given under a variety of different conditions, major references including Taqqu (1975, 1979), Dobrushin and Major (1979), Davydov (1970), Gorodetskii (1977), and more recently Chan and Terrin (1995), Csorgo and Mielniczuk (1995). For our purposes, we introduce the following

Assumption 5C (5.28) holds, where for $-\frac{1}{2} < d_u < \frac{1}{2}$

$$u_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j} \ , \ \sum_{j=0}^{\infty}|\psi_j|^2 < \infty \ , \ \psi_j \approx cj^{d_u-2} \text{ as } j \to \infty \ , \ 0 < c < \infty \ , \tag{5.33}$$

$$\varepsilon_t \equiv i.i.d.(0,\sigma_\varepsilon^2) \ , \ \sigma_\varepsilon^2 < \infty \ , \ E|\varepsilon_t|^\delta < \infty \ , \ \delta > \frac{1}{2d_u+1} \ . \tag{5.34}$$

We have allowed here for the possibility that u_t is antipersistent, i.e. $d_u < 0$ (the condition on δ is clearly redundant if $d_u \geq 0$). Some of the results of this section need somewhat stronger assumptions than 5C, and therefore we introduce also

Assumption 5D Assume (5.1) holds with $\phi = 1$, and for $\vartheta_u(L)$ and $\vartheta_e(L)$ lag polynomials such that $0 < |\vartheta_u(1)|, |\vartheta_e(1)| < \infty$, we have

$$a_t = \int_{-\pi}^{\pi} \exp(it\lambda) f_{aa}(\lambda)^{1/2} dM_a(\lambda) \,, \tag{5.35}$$

$$f_{aa}(\lambda) = \frac{1}{2\pi} |\lambda|^{-2d_a} |\vartheta_a(e^{i\lambda})|^2 \,,\ 0 < \lambda \leq \pi,\ a = u, e \,, \tag{5.36}$$

where $M_u(.)$, $M_e(.)$ are complex-valued, Gaussian random measure which satisfy

$$dM_a(\lambda) = \overline{dM_a(-\lambda)} \tag{5.37}$$

$$EdM_a(\lambda) = 0 \tag{5.38}$$

$$EdM_a(\lambda)\overline{dM_b(\lambda)} = \begin{cases} 0, & \lambda \neq \mu \\ d\lambda, & \lambda = \mu \end{cases} \,,\ a, b = u, e \,. \tag{5.39}$$

Because by Wold representation theorem any Gaussian covariance stationary sequence can be viewed as a linear process with *i.i.d.* innovations, Assumption 5D entails stricter conditions on u_t than Assumption 5C. In the sequel, for notational convenience we shall occasionally use the identification $d_x = d_u + 1$; we stress again, however, that x_t does not satisfy (1.2).

Lemma 5.3 Let $u_t = \sum_{j=0}^{\infty} \alpha_j \xi_{t-j}$, for $t = 1, 2, \ldots$, where

$$\sum_{j=0}^{\infty} |\alpha_j|^2 < \infty \,,\ \sum_{j=0}^{\infty} \alpha_j \alpha_{j+\tau} \approx c\tau^{2d_u - 1},\ 0 < c < \infty,\ \text{as } \tau \to \infty, \tag{5.40}$$

$$E\xi_t = 0 \, , \; E\xi_t^2 < C \, , \; E\xi_t\xi_s = 0 \, , \; t \neq s \, . \tag{5.41}$$

Under (5.9), (5.28) and for $k(.)$ such that (5.18) holds, we have

$$c_{xx}(0) - \frac{1}{M}\sum_{\tau=-M}^{M} k(\frac{\tau}{M})c_{xx}(\tau) = o_p(n^{2d_x-1}) \, . \tag{5.42}$$

Because ξ_t need not be stationary nor satisfy any moment condition of order greater than two, (5.40)/(5.41) are weaker than Assumption 5C.

We noted already in Chapter 4, Section 4, Case III that quadratic forms involving at the same time stationary and nonstationary variates are typically more difficult to deal with than others, where both variables are other stationary or nonstationary. This dichotomy is met also under the circumstances of the present chapter, and for the following result we need to narrow the focus and impose Assumption 5D.

Lemma 5.4 Under Assumption 5D, $d_x + d_e > 1$, (5.9) and (5.18)

$$c_{xe}(0) - \frac{1}{M}\sum_{\tau=-M}^{M} k(\frac{\tau}{M})c_{xe}(\tau) = o_p(n^{d_x+d_e-1}) \, . \tag{5.43}$$

In Chapter 2 we introduced "type I" fractional Brownian motion, which for convenience we parameterize here by d_a, $a = u, e$, for $-\frac{1}{2} < d_a < \frac{1}{2}$, so that the moving average representation (2.3) becomes, up to multiplicative constants

$$B(r; d_a) = \int_R \{(r-s)_+\}^{d_a} - \{(-s)_+\}^{d_a} \, dB(s) \, , \tag{5.44}$$

while the harmonizable representation (2.15), on which we focus, is

$$B(r; d_a) = K(d_a) \int_R \frac{\exp(i\lambda r) - 1}{i\lambda} f_{aa}(\lambda)^{1/2} dM_a(\lambda) \, , \tag{5.45}$$

$$K(d_a) = \left(\pi^{-1}(d_a + \frac{1}{2})\Gamma(2d_a + 1)\sin\pi(d_a + \frac{1}{2})\right)^{1/2} . \tag{5.46}$$

Here we consider the compound processes, for $d_a + d_b > 0$, $a, b = u, e$

$$P(d_a) = \int_0^1 B^2(r; d_a)dr , \tag{5.47}$$

$$Q(d_a, d_b) = \int_0^1 B(r; d_a)dB(r; d_b) + C(d_a, d_b) , \tag{5.48}$$

where

$$C(d_a, d_b) = K(d_a)K(d_b)\int_R [\int_0^s \frac{1 - \exp(-it\mu)}{i\mu}dt] f_{ab}(\mu)d\mu \tag{5.49}$$

The stochastic integral on the right-hand side of (5.48) is defined only in a formal sense to be equal to

$$\int_{R^2}'' [\int_0^s \exp(it\lambda)\frac{\exp(it\mu) - 1}{i\mu}] f_{uu}(\mu)^{1/2} f_{ee}(\lambda)^{1/2} dM_u(\mu)dM_e(\lambda) , \tag{5.50}$$

where $\int_{R^2}''$ signifies that the integral excludes the diagonals $\mu = \pm\lambda$. (5.50) is a multiple Wiener-Ito stochastic integral in the sense of Major (1981), but it cannot be defined as an Ito integral with respect to $B(r; d_e)$ because fractional Brownian motion is not a semimartingale.

Lemma 5.5 *(Gorodetskii (1977)), Chan and Terrin (1995))* As $n \to \infty$, under Assumptions 5C and 5D

$$\left(n^{2d_x}\vartheta_u(1)^2\right)^{-1}\sum_{t=1}^{n} x_t^2 \Rightarrow P(d_u) , \tag{5.51}$$

Also, under Assumption 5D, $d_x + d_e > 1$

$$\left(n^{d_x+d_e}\vartheta_u(1)\vartheta_e(1)\right)^{-1}\sum_{t=1}^{n} x_t e_t \Rightarrow Q(d_u, d_e) . \tag{5.52}$$

Proof (5.51) follows under Assumption 5C from Gorodetskii (1977) and the continuous mapping theorem; (5.52) is given in Chan and Terrin (1995). □

We learned from Lemma 5.2 that in the $CI(1)$ case $\widetilde{\beta}_M$ and $\widehat{\beta}_{n-1}$ share the same asymptotic distribution, as opposed to FDLS estimates; when the innovation are long memory a stronger result holds, namely the difference between the two estimates is asymptotically $o_p(n^{d_e-d_x})$. More precisely,

Theorem 5.2 Under Assumption 5D, (5.1), (5.9) and (5.18)

$$n^{d_x-d_e}(\widehat{\beta}_{n-1}-\beta)=\frac{\vartheta_e(1)}{\vartheta_u(1)}P(d_u)^{-1}Q(d_u,d_e)\ , \tag{5.53}$$

and

$$\begin{aligned} |\widetilde{\beta}_M-\widehat{\beta}_{n-1}| &= o_p(n^{d_e-d_x})\ , &(5.54)\\ n^{d_x-d_e}(\widetilde{\beta}_M-\beta) &= \frac{\vartheta_e(1)}{\vartheta_u(1)}P(d_u)^{-1}Q(d_u,d_e)\ . &(5.55)\end{aligned}$$

Proof (5.53) follows from Lemma 5.5 and the continuous mapping theorem. For (5.54) we can rewrite (cf. the proof of Theorem 4.4)

$$\begin{aligned}\widetilde{\beta}_M-\widehat{\beta}_{n-1} &= \left\{\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})c_{xx}(\tau)\right\}^{-1}\left\{\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})c_{xe}(\tau)-c_{xe}(0)\right\}\\ &\quad+\left\{\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})c_{xx}(\tau)\right\}^{-1}\left\{\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})c_{xx}(\tau)-c_{xx}(0)\right\}(c_{xx}(0))^{-1}c_{xe}(0)\\ &= O_p(n^{1-2d_x})o_p(n^{d_x+d_e-1})+O_p(n^{1-2d_x})o_p(n^{2d_x-1})O_p(n^{d_e-d_x})\\ &= o_p(n^{d_e-d_x}) &(5.56)\end{aligned}$$

where the stochastic orders of magnitude follow from Lemmas 5.3 and 5.4, so that the proof of (5.54) is completed. (5.55) follows immediately. □

The constant $C(d_u, d_e)$ at the numerator in (5.53) is due to the non-zero correlation between u_t and e_t. The left-hand side of (5.53) generalizes in an intuitive way the rate of convergence and the asymptotic distribution of the $CI(1)$ case , which is provided by Lemma 5.2.

Theorem 5.2 might be extended to allow for $d_u < 0$, provided $d_x + d_e > 1$, i.e. $d_u + d_e > 0$. However we refrain from the analysis of this case here, both for the sake of brevity and to mantain simmetry with the stationary case where $d_u < 0$ was ruled out. The possibility of an "antipersistent" behaviour in the innovation sequence u_t seems in any case less relevant from the point of view of practical applications.

From Lemmas 5.3-5.5 we learn that under (5.9)

$$\frac{n^{1-2d_x}}{M} \sum_{\tau=-M}^{M} k(\tau/M) c_{xx}(\tau) \Rightarrow \vartheta_u(1)^2 P(d_u) + o_p(1) \tag{5.57}$$

and

$$\frac{n^{1-d_x-d_e}}{M} \sum_{\tau=-M}^{M} k(\tau/M) c_{xe}(\tau) \Rightarrow \vartheta_u(1)\vartheta_e(1) Q(d_u, d_e) + o_p(1) \,. \tag{5.58}$$

In view of (5.12), (5.57)/(5.58) provide the asymptotic distribution of the weighted covariance estimate of the (cross-) spectral density at zero frequency for the variables x_t and e_t when the former is nonstationary. This result can have some independent interest, for istance for estimates of the differencing parameter $d_u + 1$ under the same circumstances as in Hurvich and Ray (1995), Velasco (1996), (1997b) and (1.129) of Chapter 1. Moreover, the same argument as in Lemma 5.3-5.4 can be exploited in the analysis of the behaviour of more general quadratic forms in nonstationary variables (cp. Fox and Taqqu (1985),(1987), Giraitis and Surgailis (1990) and the other references mentioned in Chapter 1, Section 3). Consider the quadratic form

$$\sum_{t=1}^{n} \sum_{s=1}^{n} b_{M,n}(t-s) x_t x_s \,, \tag{5.59}$$

where $b_{M,n}(\tau) = k^*(\tau/M)$ (say), for $k^*(.)$ such that Assumption 5B holds and x_t satisfying Assumption 5D; hence

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n} b_{M,n}(t-s)x_t x_s = \sum_{\tau=-n+1}^{n-1} b_{M,n}(\tau)\frac{1}{n}\sum_{t=1}^{n-|\tau|} x_t x_{t+|\tau|}$$
$$= \sum_{\tau=-M}^{M} k^*(\frac{\tau}{M})c_{xx}(\tau) \ . \tag{5.60}$$

Then under (5.9) we have

$$\frac{n^{-2d_x-1}}{M}\sum_{t=1}^{n}\sum_{s=1}^{n} b_{M,n}(t-s)x_t x_s \Rightarrow \vartheta_u(1)^2\int_0^1 B^2(r;d_u)dr \ , \text{ as } n\to\infty. \tag{5.61}$$

The main difference between (5.61) and the results for the stationary long memory case we reviewed in Section 1.3 lies in the functional dependence of the weight function $b(.)$ from n and M; this dependence, however, appears an intrinsic consequence of the notion of nonstationarity.

5.4 Numerical Evidence

We argued earlier here and in the previous chapter about the practical difficulty to distinguish, on the basis of a finite sample of observations, between a unit root process and a stationary ARMA with long memory innovations and roots close to the unit circle. In order to provide numerical evidence to support this claim, 5000 replications of stationary series of length $n = 64$ and $n = 128$ were generated according to the following model:

$$x_t = (1-L)^{-d}(1-\phi L)^{-1}u_t \ , \ u_t \equiv n.i.d.(0,1) \ , \tag{5.62}$$

with $\phi = .5, .6, .7, .8, .9$ and $d = .1, .2, .3, .4$. For each series, then, the classical Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF(1)) tests for unit roots were carried out, choosing 5% as the size of the test (implying a rejection value of -1.95 for both procedures). The results are reported in Tables 5.1 and 5.2. While the power properties of these tests are satisfactory for small values of both d and ϕ, when these two parameters approach the nonstationarity region the probability to reject the null $H_0 : \phi = 1$ decays very rapidly. For instance, for $n = 128$ and $\phi = .8$ the probability to reject the null using a DF test is 99.2% when $d = .1$, 83.98% for $d = .2$, 44.04% for $d = .3$ and only 14.12% for $d = .4$; using an ADF(1) test the probability to reject the null is around 5% even when ϕ is as small as .6. Of course, the probability of a type II error would increase further with smaller sample sizes or for lower significance level, i.e., 2.5% or 1%.

As noted earlier, one of the main advantages of the WCE procedure is its robustness to the presence/absence of unit roots in the DGP of the raw data. In order to investigate the finite sample behaviour of this estimate, then, we have considered the following DGP:

$$y_t = x_t\beta + e_t \ , \ e_t \equiv n.i.d.(0, 18) \ , \ Eu_t e_t = 3 \ , \tag{5.63}$$

with x_t generated according to (5.62) but now $\phi = .6, .7, .8, .9, 1$, i.e. also the nonstationary case is included. Again series of length $n = 64$ and $n = 128$ were considered, with 5000 replications; for each sample, $\widetilde{\beta}_M$ was then calculated for $M = 4, 6$, resp., ($M \simeq \sqrt{n}/2$) and $k(v)$ the truncated kernel $k(v) = 1(|v| < 1)$. The mean squared error (MSE) and the bias for WCE are reported in Table 5.3; as expected, the performance of the estimate increases rapidly as ϕ and d grow, i.e., exactly when it becomes more and more difficult to distinguish stationarity and nonstationarity. For instance the MSE for $n = 64$ and $d = .3$ decreases steadily from .144 to .011 when ϕ increases from .6 to 1. Note that there is a discontinuity in the asymptotic rate of convergence of the estimate when we pass from $\phi = .9$ to $\phi = 1$; however, although this is somewhat mirrored in the result of our simulations, for a fixed sample the improvement in the MSE for increasing values of ϕ is relatively smooth.

APPENDIX

Proof of Lemma 5.1 Recall that

$$E\sum_{p=-M}^{M} k(\frac{p}{M})c_{xx}(p) = \sum_{p=-M}^{M} k(\frac{p}{M})(1-\frac{p}{n})\gamma_{xx}(p)\ , \tag{5.64}$$

$$E\sum_{p=-M}^{M} k(\frac{p}{M})c_{xe}(p) = \sum_{p=-M}^{M} k(\frac{p}{M})(1-\frac{p}{n})\gamma_{xe}(p)\ , \tag{5.65}$$

where, under (5.9), (5.22) and (5.23)

$$\lim_{M\to\infty}\frac{\sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xx}(p)}{\sum_{p=-M}^{M}k(\frac{p}{M})\gamma_{xx}(p)} = \lim_{M\to\infty}\frac{\sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xe}(p)}{\sum_{p=-M}^{M}k(\frac{p}{M})\gamma_{xe}(p)} = 1\ , \tag{5.66}$$

by the dominated convergence theorem. Hence it is enough to prove that under Assumption 5A we have

$$\sum_{p=-M}^{M} k(\frac{p}{M})\left\{c_{xx}(p)-(1-\frac{p}{n})\gamma_{xx}(p)\right\} = o_p(M^{2d_x})\ , \tag{5.67}$$

$$\sum_{p=-M}^{M} k(\frac{p}{M})\left\{c_{xe}(p)-(1-\frac{p}{n})\gamma_{xe}(p)\right\} = o_p(M^{d_x+d_e})\ . \tag{5.68}$$

For (5.67)/(5.68), it is sufficient to show that

$$\sum_{p=-M+1}^{M-1}\sum_{q=-M+1}^{M-1}|Cov(c_{xa}(p),c_{xa}(q))| = o(M^{2d_x+2d_a})\ ,\ a=x,e\ . \tag{5.69}$$

From Hannan (1970), p.209, we have that $Cov(c_{xa}(p),c_{xa}(q))$ is equal to

$$\frac{1}{n}\sum_{r=-n+1}^{n-1}(1-\frac{|r|}{n})(\gamma_{xx}(r)\gamma_{aa}(r+q-p)+\gamma_{xa}(r+q)\gamma_{ax}(r-p) \tag{5.70}$$

$$+\frac{1}{n^2}\sum_{r=-n+1}^{n-1}\sum_{s=1-r}^{n-r} cum_{xaxa}(s,s+p,s+r,s+r+q) \tag{5.71}$$

Now for (5.70), when $d_x + d_a < \frac{1}{2}$

$$
\begin{aligned}
\left|\sum_{r=-n+1}^{n-1} \gamma_{xx}(r)\gamma_{aa}(r+p-q) + \gamma_{xa}(r+p)\gamma_{ax}(r-q)\right| &\le C \sum_{r=-\infty}^{\infty} \left\{\gamma_{xx}^2(r) + \gamma_{xx}^2(r)\right\} \\
&< \infty \qquad (5.72)
\end{aligned}
$$

because $\gamma(r)\gamma(s) \le \frac{1}{2}\{\gamma^2(r) + \gamma^2(s)\}$, $r, s = 0, \pm1, \pm2, \dots$, and hence

$$
\begin{aligned}
& \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \left| \sum_{r=-n+1}^{n-1} (1 - \frac{|r|}{n}) \{\gamma_{xx}(r)\gamma_{aa}(r+p-q) + \gamma_{xa}(r+p)\gamma_{ax}(r-q)\} \right| \\
= \ & O(\frac{M^2}{n}) = o(1) \ . \qquad (5.73)
\end{aligned}
$$

For $d_x + d_a \ge \frac{1}{2}$, and in view of (5.17),

$$
\begin{aligned}
& \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \left| \sum_{r=-n+1}^{n-1} (1 - \frac{|r|}{n}) \gamma_{xx}(r)\gamma_{aa}(r+p-q) \right| \\
\le \ & \frac{M}{n} \sum_{\tau=M-1}^{M-1} \sum_{r=-n+1}^{n-1} |\gamma_{xx}(r)\gamma_{aa}(r+\tau)| \le C\frac{M}{n} \sum_{\tau=M-1}^{M-1} \sum_{r=-n+1}^{n-1} |r|^{2d_x-1}|r+\tau|^{2d_a-1} \\
= \ & O(\frac{M}{n} M^{2d_x+2d_a}) = o(M^{2d_x+2d_a}) \ . \qquad (5.74)
\end{aligned}
$$

Also, by Cauchy-Schwarz and elementary inequalities

$$
\begin{aligned}
& \sum_{p=-M}^{M} \sum_{q=-M}^{M} \frac{1}{n} \left| \sum_{r=-n+1}^{n-1} (1 - \frac{|r|}{n}) \gamma_{xa}(r+p)\gamma_{ax}(r-q) \right| \qquad (5.75) \\
\le \ & C\frac{M^2}{n} \sum_{r=-n+1}^{n-1} \gamma_{xa}^2(r) = O(M^2 n^{2d_x+2d_a-2}) = o(M^{2d_x+2d_x}) \ . \qquad (5.76)
\end{aligned}
$$

For (5.71), by Hannan (1970), p.211 and Assumption 5A we have that

$$
cum_{xaxa}(p,q,r,s) \le C \sum_{d=0}^{\infty} g(p+d)g(d+q-p)g(d+r-p)g(d+s-p) \qquad (5.77)
$$

with $g(u) = (|u|+1)^{d_x-1}$. Hence the left-hand side of (5.71) is bounded by

$$\begin{aligned}
& \frac{C}{n^2}\sum_{r=0}^{n-1}\sum_{s=1-r}^{n-r}\sum_{d=0}^{\infty} g(d)g(d+p)g(d+r)g(d+r+q) \\
\leq\ & \frac{C}{n}\sum_{r=-n+1}^{n-1}\sum_{d=-\infty}^{\infty} g(d)g(d+p)g(d+r)g(d+r+q) \\
\leq\ & \frac{C}{n}\sum_{r=-n+1}^{n-1}\sum_{d=-\infty}^{\infty} 4g^2(d) = O(1)\ , \qquad (5.78)
\end{aligned}$$

where the last inequality follows from $ABCD \leq A^4 + B^4 + C^4 + D^4$, which holds for real-valued A, B, C, D. □

Proof of Theorem 5.1 We can rewrite

$$\begin{aligned}
\tilde{\beta}_M - \beta &= \left(\sum_{p=-M}^{M} k(\frac{p}{M})c_{xx}(p)\right)^{-1} \sum_{p=-M}^{M} k(\frac{p}{M})c_{xe}(p) \\
&= A^{-1}b \qquad (5.79)
\end{aligned}$$

for

$$\begin{aligned}
A &= \sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xx}(p) + \sum_{p=-M}^{M} k(\frac{p}{M})\left(c_{xx}(p) - (1-\frac{p}{n})\gamma_{xx}(p)\right) \qquad (5.80) \\
b &= \sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xe}(p) + \sum_{p=-M}^{M} k(\frac{p}{M})\left(c_{xe}(p) - (1-\frac{p}{n})\gamma_{xe}(p)\right) \qquad (5.81)
\end{aligned}$$

Now by Lemma 5.1 and Assumption 5A

$$\begin{aligned}
\sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xx}(p)A^{-1} &= 1 + o_p(1)\ , \qquad (5.82) \\
\left(\sum_{p=-M}^{M}(1-\frac{p}{n})k(\frac{p}{M})\gamma_{xe}(p)\right)^{-1} b &= 1 + o_p(1)\ , \qquad (5.83)
\end{aligned}$$

and hence the result follows by Slutsky's theorem and (5.66) . □

Proof of Lemma 5.3 Because by Assumption 5B $M^{-1}\sum_{\tau=-M}^{M} k(\frac{\tau}{M}) \approx 1$ as $M \to \infty$, it is sufficient to prove that

$$\frac{1}{M}\sum_{\tau=-M}^{M} k(\frac{\tau}{M})\left\{c_{xx}(0) - c_{xx}(\frac{\tau}{M})\right\} = o_p(n^{2d_x-1}) \, . \tag{5.84}$$

The left-hand side of (5.84) is equal to $2/M$ times

$$\frac{1}{n}\sum_{\tau=1}^{M} k(\frac{\tau}{M})\left\{\sum_{t=1}^{n} x_t^2 - \sum_{t=\tau+1}^{n} x_t x_{t-\tau}\right\} \tag{5.85}$$

$$= \frac{1}{n}\sum_{\tau=1}^{M} k(\frac{\tau}{M})\sum_{t=1}^{\tau} x_t^2 \tag{5.86}$$

$$+\frac{1}{n}\sum_{\tau=1}^{M} k(\frac{\tau}{M})\sum_{t=\tau+1}^{n} x_t(x_t - x_{t-\tau}) \, . \tag{5.87}$$

For (5.86), we have easily

$$\sum_{t=1}^{M}\left\{\sum_{\tau=t}^{M} k(\frac{\tau}{M})\right\} x_t^2 \le CM\sum_{t=1}^{M} x_t^2 \, , \tag{5.88}$$

where

$$\sum_{t=1}^{M} x_t^2 = \sum_{t=1}^{M}(\sum_{s=1}^{t}\sum_{j=0}^{\infty}\alpha_j\xi_{t-j})^2 = \sum_{t=1}^{M}\sum_{s=1}^{t}\sum_{k=1}^{t}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\alpha_j\alpha_i\xi_{s-j}\xi_{k-i} \, . \tag{5.89}$$

To bound the expected value of the above (non-negative) random variable, we use (5.40)/(5.41) to obtain

$$\begin{aligned}\sum_{t=1}^{M}\sum_{s=1}^{t}\sum_{k=1}^{t}\sum_{j=0}^{\infty}\alpha_j\alpha_{k-s+j} &\le C\sum_{t=1}^{M}\sum_{s=1}^{t}\sum_{k=1}^{t}(|k-s|+1)^{2d_u-1} \\ &= C\sum_{t=1}^{M} t\sum_{v=0}^{t-1}(1-\frac{|v|}{t})v^{2d_u-1} \le CM^{2d_x},\end{aligned} \tag{5.90}$$

whence it follows that $\sum_{t=1}^{M} x_t^2 = o_p(n^{2d_x})$. By Cauchy-Schwarz inequality, (5.87) is bounded by

$$\frac{1}{n}\left\{\sum_{t=1}^{n}x_t^2\right\}^{\frac{1}{2}}\sum_{\tau=1}^{M}k(\frac{\tau}{M})\left\{\sum_{t=\tau+1}^{n}(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}} . \tag{5.91}$$

The last element has stochastic order of magnitude

$$\begin{aligned} & O_p(\frac{1}{n}\left\{\sum_{t=1}^{n}x_t^2\right\}^{\frac{1}{2}})O_p(E\sum_{\tau=1}^{M}k(\frac{\tau}{M})\left\{\sum_{t=\tau+1}^{n}(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}}) \\ = \; & O_p(n^{d_x-1}E\sum_{\tau=1}^{M}\left\{\sum_{t=\tau+1}^{n}(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}}) , \end{aligned} \tag{5.92}$$

in view of Assumption 5B and because $\{n^{-1}\sum_{t=1}^{n}x_t^2\}^{\frac{1}{2}} = O_p(n^{d_x-1})$ follows from previous calculations. From Jensen's inequality

$$E\sum_{\tau=1}^{M}\left\{\sum_{t=\tau+1}^{n}(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}} \leq C\sum_{\tau=1}^{M}\left\{\sum_{t=\tau+1}^{n}E(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}} , \tag{5.93}$$

where

$$\begin{aligned} \sum_{t=\tau+1}^{n}E(x_t-x_{t-\tau})^2 &= \sum_{t=\tau+1}^{n}E(\sum_{s=t-\tau+1}^{t}u_s)^2 \\ &= \sum_{t=\tau+1}^{n}E\sum_{s=t-\tau+1}^{t}\sum_{k=t-\tau+1}^{t}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\alpha_j\alpha_i\xi_{s-j}\xi_{k-i} \\ &< C\sum_{t=\tau+1}^{n}\sum_{s=0}^{\tau-1}\sum_{k=0}^{\tau-1}\sum_{j=0}^{\infty}\alpha_j\alpha_{k-s+j} \\ &\leq C\tau\sum_{t=\tau+1}^{n}\sum_{v=0}^{\tau-1}v^{2d_u-1} \leq Cn\tau^{2d_x-1} . \end{aligned} \tag{5.94}$$

It follows that

$$\sum_{\tau=1}^{M}\left\{\sum_{t=\tau+1}^{n}E(x_t-x_{t-\tau})^2\right\}^{\frac{1}{2}} \leq C\sum_{\tau=1}^{M}(n\tau^{2d_x-1})^{\frac{1}{2}} = o(n^{1/2}M^{d_x+1/2}) ; \tag{5.95}$$

hence in view of (5.92), (5.87) is $o_p(n^{2d_x})$, which completes the proof of Lemma 5.3. □

Proof of Lemma 5.5 In the sequel, we repeatedly use the inequality

$$|Ea_t b_s| \le C(1+|t-s|)^{d_a+d_b-1},\ a,b=u,e\ , \tag{5.96}$$

which holds under Assumption 5D. We have

$$\begin{aligned}
c_{xe}(0)-\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})c_{xe}(\tau) &= \frac{1}{M}\sum_{\tau=1}^{M}k(\frac{\tau}{M})\left\{c_{xe}(0)-c_{xe}(\tau)\right\} \\
&\quad +\frac{1}{M}\sum_{\tau=-M}^{-1}k(\frac{\tau}{M})\left\{c_{xe}(0)-c_{xe}(\tau)\right\} \\
&\quad +c_{xe}(0)\left\{1-\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})\right\} \\
&= I+II+III+IV+V
\end{aligned} \tag{5.97}$$

with

$$I = \frac{1}{Mn}\sum_{\tau=1}^{M}k(\frac{\tau}{M})\left\{\sum_{t=\tau+1}^{n}x_t(e_t-e_{t-\tau})\right\}\ ,\ II=\frac{1}{Mn}\sum_{\tau=1}^{M}k(\frac{\tau}{M})\sum_{t=1}^{\tau}x_te_t \tag{5.98}$$

$$III = \frac{1}{Mn}\sum_{\tau=-M}^{-1}k(\frac{\tau}{M})\sum_{t=\tau+1}^{n}(x_t-x_{t-\tau})e_t\ ,\ IV=\frac{1}{Mn}\sum_{\tau=-M}^{-1}k(\frac{\tau}{M})\sum_{t=1}^{\tau}x_te_t\ , \tag{5.99}$$

$$V = c_{xe}(0)\left\{1-\frac{1}{M}\sum_{\tau=-M}^{M}k(\frac{\tau}{M})\right\}\ . \tag{5.100}$$

Define

$$\widetilde{\sum}=\sum_{t=1}^{M}\sum_{s=1}^{M}\sum_{v=s}^{M}\sum_{\tau=t}^{M}\ . \tag{5.101}$$

In view of Isserlis formula (Brillinger (1981), p.21) which for zero-mean Gaussian variables gives

$$Ex_1x_2x_3x_4 = Ex_1x_2Ex_3x_4+Ex_1x_3Ex_3x_2+Ex_1x_4Ex_3x_2\ , \tag{5.102}$$

the expected value of the square of (II) is bounded by

$$E\left\{\sum_{t=1}^{M}\left\{\sum_{\tau=t}^{M}k(\frac{\tau}{M})\right\}x_te_t\right\}^2 \leq CE\widetilde{\sum}x_te_tx_se_s = C\left\{\Gamma_1+\Gamma_2+\Gamma_3\right\} , \tag{5.103}$$

for

$$\Gamma_1 = \widetilde{\sum}Ex_te_tEx_se_s,\ \Gamma_2 = \widetilde{\sum}Ex_tx_sEe_te_s,\ \Gamma_3 = \widetilde{\sum}Ex_se_tEx_te_s . \tag{5.104}$$

Now in view of (5.96), Γ_1 is bounded by

$$\widetilde{\sum}E\sum_{j=1}^{t}u_je_tE\sum_{i=1}^{s}u_ie_s \leq C\widetilde{\sum}t^{d_u+d_e}s^{d_u+d_e} = O(M^{2d_u+2d_e}) = o(M^2n^{2d_u+2d_e+2}) , \tag{5.105}$$

and Γ_2 is bounded by

$$\begin{aligned} & CM^2\sum_{t=1}^{M}\sum_{s=1}^{M}(|t-s|+1)^{2d_e-1}E\sum_{i=1}^{t}\sum_{j=1}^{s}u_iu_j \\ \leq\ & CM^2\sum_{t=1}^{M}\sum_{s=1}^{M}(|t-s|+1)^{2d_e-1}M\sum_{k=1}^{M}|k|^{2d_u-1} \\ =\ & O(M^{2d_u+1}\sum_{t=1}^{M}\sum_{s=1}^{M}(|t-s|+1)^{2d_e-1}) = o(M^2n^{2d_x+2d_e}) , \end{aligned} \tag{5.106}$$

As far as Γ_3 is concerned we have that

$$\begin{aligned} \Gamma_3 \ \leq\ & CM^2\sum_{t=1}^{M}\sum_{s=1}^{M}\sum_{j=1}^{M}\sum_{i=1}^{M}(|s-j|+1)^{d_u+d_e-1}(|t-i|+1)^{d_u+d_e-1} \\ =\ & O(M^{2d_u+2d_e+4}) = o(M^2n^{2d_x+2d_e}) , \end{aligned} \tag{5.107}$$

because

$$|Ex_se_t| = |E\sum_{i=1}^{s}u_ie_t| \leq C\sum_{i=1}^{s}(|t-i|+1)^{d_u+d_e-1} . \tag{5.108}$$

Hence $\Gamma_i = o(M^2n^{2d_x+2d_e})$, $i=1,2,3$, and (II) is $o_p(n^{d_x+d_e-1})$; same argument can be applied to IV. For (I) is concerned, we can write

$$\begin{aligned}
&\sum_{\tau=1}^{M} k(\frac{\tau}{M}) \left\{ \sum_{t=\tau+1}^{n} x_t e_t - \sum_{t=\tau+1}^{n} (x_t - x_{t-\tau}) e_{t-\tau} - \sum_{t=\tau+1}^{n} x_{t-\tau} e_{t-\tau} \right\} \\
= \quad & \Delta_1 - \Delta_2 - \Delta_3 \ , \qquad (5.109)
\end{aligned}$$

where

$$\begin{aligned}
\Delta_1 &= \sum_{\tau=1}^{M} k(\frac{\tau}{M}) \left\{ \sum_{t=N-\tau+1}^{n} x_t e_t \right\}, \ \Delta_2 = \sum_{\tau=1}^{M} k(\frac{\tau}{M}) \left\{ \sum_{t=\tau+1}^{n} (x_t - x_{t-\tau}) e_{t-\tau} \right\}, \\
\Delta_3 &= \sum_{\tau=1}^{M} k(\frac{\tau}{M}) \left\{ \sum_{t=1}^{\tau} x_t e_t \right\} . \qquad (5.110)
\end{aligned}$$

Now Δ_1 and Δ_2 are, apart from a change of index, proportional to (II) which we analyzed before. On the other hand, Δ_3 can be dealt exactly as (III), again with a change of index; therefore we analyze (III). In view of (5.28), we have

$$\sum_{t=\tau+1}^{n} (x_t - x_{t-\tau}) e_t = \sum_{t=\tau+1}^{n} \sum_{i=t-\tau+1}^{t} u_i e_t \ . \qquad (5.111)$$

Also

$$\begin{aligned}
(\sum_{t=\tau+1}^{n} \sum_{i=t-\tau+1}^{t} u_i e_t)^2 &= \sum_{s=\tau+1}^{n} \sum_{t=\tau+1}^{n} \sum_{i=t-\tau+1}^{t} \sum_{j=k-\tau+1}^{s} e_t u_i e_s u_j \\
&= \widehat{\sum} e_t u_s e_s u_j \ , \qquad (5.112)
\end{aligned}$$

with

$$\sum_{s=\tau+1}^{n} \sum_{t=\tau+1}^{n} \sum_{i=t-\tau+1}^{t} \sum_{j=s-\tau+1}^{s} = \widehat{\sum} . \qquad (5.113)$$

Hence we obtain $E\widehat{\sum} e_t u_i e_s u_j = \Theta_1 + \Theta_2 + \Theta_3$, for

$$\Theta_1 = \widehat{\sum} E e_t e_s E u_i u_j, \ \Theta_2 = \widehat{\sum} E e_t u_i E e_s u_j, \ \Theta_3 = \widehat{\sum} E e_t u_j E u_i e_s \ . \qquad (5.114)$$

Thus

$$\begin{aligned}\Theta_1 &= E\sum_{s=\tau+1}^{n}\sum_{t=\tau+1}^{n} e_t e_s E\sum_{i=t-\tau+1}^{t}\sum_{j=s-\tau+1}^{s} u_i u_j \\ &= O(n(n-\tau)^{2d_e})O(\tau^{2d_u+1}) = o(n^{2d_x+2d_e})\ , \end{aligned} \tag{5.115}$$

$$\begin{aligned}\Theta_2 &= \sum_{t=\tau+1}^{n}\sum_{i=t-\tau+1}^{t} Ee_t u_i \sum_{s=\tau+1}^{n}\sum_{j=s-\tau+1}^{s} Ee_s u_j \\ &\leq C\sum_{t=\tau+1}^{n}\sum_{i=t-\tau+1}^{t}(|t-i|+1)^{d_u+d_e-1}\sum_{s=\tau+1}^{n}\sum_{j=s-\tau+1}^{s}(|s-j|+1)^{d_u+d_e-1} \\ &= O(\left\{(n-\tau)\tau^{d_u+d_e}\right\}^2) = o(n^{2d_x+2d_e})\ , \end{aligned} \tag{5.116}$$

$$\begin{aligned}\Theta_3 &= \sum_{t=\tau+1}^{n}\sum_{j=s-\tau+1}^{s} Ee_t u_j \sum_{s=\tau+1}^{n}\sum_{i=t-\tau+1}^{t} Eu_i e_s \\ &= O(\left\{(n-\tau)\tau^{d_u+d_e}\right\}^2) = o(n^{2d_x+2d_e})\ . \end{aligned} \tag{5.117}$$

Because the expected value of the square of (III) is bounded by $Cn^{-2}\left\{\Theta_1+\Theta_2+\Theta_3\right\}$, we have easily $(III) = o_p(n^{d_x+d_e-1})$.

Finally, from of Assumption B we have $(V) = o_p(c_{xe}(0)) = o_p(n^{d_x+d_e})$, the last equality following from (5.52). □

TABLE 5.1: POWER OF THE DF TEST

d,ϕ	n=64 .5	.6	.7	.8	.9	n=128 .5	.6	.7	.8	.9
.1	100	99.48	95.36	66.70	20.24	100	100	100	99.20	56.54
.2	99.22	95.06	74.08	34.56	6.46	100	100	99.50	83.98	19.92
.3	92.12	73.16	41.24	13.90	2.06	99.92	98.92	88.02	44.04	4.45
.4	67.12	41.72	18.12	4.18	.56	95.76	83.14	52.00	14.12	.64

TABLE 5.2: POWER OF THE ADF(1) TEST

d,ϕ	n=64 .5	.6	.7	.8	.9	n=128 .5	.6	.7	.8	.9
.1	69.76	44.08	18.18	4.18	.58	99.34	92.42	58.72	12.42	.52
.2	41.40	20.56	7.12	1.26	.16	87.84	59.50	21.30	2.44	.08
.3	19.12	8.02	1.96	.36	.10	50.36	22.86	5.02	.42	.00
.4	7.36	24.4	.58	.12	.08	18.92	5.3	.9	.04	.00

TABLE 5.3A: MSE AND BIAS FOR THE WCE ESTIMATES, n=64

d,ϕ	BIAS .6	.7	.8	.9	1	MSE .6	.7	.8	.9	1
.1	-.043	-.035	-.036	-.023	-.009	.221	.179	.125	.072	.031
.2	-.041	-.038	-.035	-.027	-.001	.180	.142	.088	.055	.020
.3	-.026	-.030	-.023	-.022	-.001	.144	.098	.066	.037	.011
.4	-.039	-.025	-.019	-.015	3E-4	.104	.070	.047	.024	.006

TABLE 5.3B: MSE AND BIAS FOR THE WCE ESTIMATES, n=128

d,ϕ	BIAS .6	.7	.8	.9	1	MSE .6	.7	.8	.9	1
.1	-.049	-.046	-.048	-.030	-.016	.114	.082	.051	.027	.006
.2	-.051	-.045	-.038	-.031	-.011	.082	.057	.036	.017	.003
.3	-.055	-.040	-.036	-.025	-.007	.062	.041	.024	.011	.002
.4	-.051	-.040	-.031	-.020	-.003	.043	.027	.016	.007	.001

REFERENCES

Abramovitz, M. and I.Stegun (1970) *Handbook of Mathematical Functions*, Dover

Adenstedt, R. (1974) "On Large Sample Estimation for the Mean of a Stationary Random Sequence", *Annals of Mathematical Statistics,* 2, 1095-1107

Akonom, J. and Gourieroux, C. (1987) "A Functional Central Limit Theorem for Fractional Processes" mimeo, CEPREMAP

Anderson, T.W. (1993) "Goodness of Fit Tests for Spectral Distributions", *Annals of Statistics,* 21, 830-847

Andrews, D.W.K. (1984) "Non-Strong Mixing Autoregressive Processes", *Journal of Applied Probability,* 21, 930-934

Andrews, D.W.K. (1991) "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica,* 59, 817-858

Andrews, D.W.K. and J.C.Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica,* 60, 953-966

Baillie, R.T. (1996) "Long-Memory Processes and Fractional Integration in Econometrics", *Journal of Econometrics,* 73, 5-59

Banerjee, A., Dolado, J., Galbraith, J.W. and Hendry, D.F. (1993) *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford, Oxford University Press

Beran, J.A. and N.Terrin (1994) "Estimation of the Long Memory Parameter, Based on a Multivariate Central Limit Theorem", *Journal of Time Series Analysis,* 15, 269-278

Billingsley, P. (1968) *Convergence of Probability Measures,* J.Wiley, New York

Bingham, N.H., Goldie, C.M. and J.L.Teugels (1989) *Regular Variation,* 2nd ed., Cambridge University Press, Cambridge

Box, G.E.P. and G.M.Jenkins (1976) *Time Series Analysis: Forecasting and Control,* San Francisco: Holden Day

Brillinger, D.R. (1979) "Confidence Intervals for the Crosscovariance Function", *Selecta Statistica Canadiana,* 5, 3-16

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory,* Holt, Rinehart & Winston, New York

Brockwell, P.J, and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition, New York, Springer and Verlag

Burkholder, D.L. (1973) "Martingale Transforms", *Annals of Mathematical Statistics,* 1, 19-42

Campbell, J.Y. and R.J. Shiller (1987) "Cointegration and Tests of Present Value Models", *Journal of Political Economy,* vol.95, no.5

Cappuccio, N. and D.Lubian (1997) "Spurious Regression Between $I(1)$ Processes with Long Memory Errors", *Journal of Time Series Analysis,* 18, 341-354

Chan, N.H. and C.Z.Wei (1988) "Limiting Distribution of Least Squares Estimates of Unstable Autoregressive Processes" *Annals of Statistics,* 16, 367-401

Chan, N.H. and N.Terrin (1995) "Inference for Unstable Long-Memory Processes with Applications to Fractional Unit Roots Autoregressions", *Annals of Statistics* 23, 1662-1683

Csőrgo, S. and Mielniczuk, J. (1995) "Distant Long-Range Dependent Sums and Regression Estimation", *Stochastic Processes and their Applications,* 59, 143-155

Dahlhaus, R. (1988) "Empirical Spectral Processes and their Applications to Time Series Analysis", *Stochastic Processes and their Applications,* 30, 69-83

Dahlhaus, R. (1989) "Efficient Parameter Estimation for Self-similar Processes" *Annals of Statistics,* 16, 808-841

Dahlhaus, R. (1995) "Efficient Location and Regression Estimation for Long Range Dependent Regression Models", *Annals of Statistics,* 1029-1047

Davidson ,J, and R.deJong (1997) "The Functional Central Limit Theorem and

Weak Convergence to Stochastic Integrals: Results for Weakly Dependent and Fractionally Integrated Processes", preprint

Davydov, Y.A. (1970) "The Invariance Principle for Stationary Processes", *Theory of Probability and Its Applications,* 15, 487-498

Dickey, D.A. and W.A.Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association,* 74, 427-431

Diebold, F.X. and G.Rudebusch (1989) "Long Memory and Persistence in Aggregate Output", *Journal of Monetary Economics,* 24, 189-209

Dobrushin, R.L. (1979) "Gaussian and their Subordinated Self-Similar Random Generalized Fields", *Annals of Probability,* 7, 1-28

Dobrushin, R.L. and P.Major (1979) "Non-Central Limit Theorems for Non-Linear Functionals of Gaussian Fields", *Z. Wahrscheinlichkeitstheorie verw. Gebiete,* 50, 27-52

Dolado, J. and F.Marmol (1996) "Efficient Estimation of Cointegrating Relationships Among Higher Order and Fractionally Integrated Processes", Documento de Trabajo n.9617, Banco de Espana

Dunsmuir, W. (1979) "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise", *Annals of Statistics,* 7, 490-506

Eicker, F. (1967) "Limit Theorems for Regression with Unequal and Dependent Errors", in *Proc.Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol.1, Berkeley, University of California Press

Einmahl, U. (1989) "Extensions of Results of Komlos, Major and Tusnady to the Multivariate Case", *Journal of Multivariate Analysis,* 28, 20-48

Engle, R. (1974) "Band Spectrum Regression", *International Economic Review,* 15, 1-11

Engle, R. and C.W.J.Granger, (1987) "Cointegration and Error Correction:

Representation, Estimation and Testing" *Econometrica* 55, 251-276

Fox, R. and M.S.Taqqu (1985) "Non-Central Limit Theorems for Quadratic Forms in Random Variables Having Long-Range Dependence", *Annals of Probability,* 13, 428-436

Fox, R. and M.S.Taqqu (1986) "Large Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series", *Annals of Statistics,* 17, 517-532

Fox, R. and M.S.Taqqu (1987) "Central Limit Theorems for Quadratic Forms in Random Variables Having Long-Range Dependence", *Probability Theory and Related Fields* 74, 213-240

Geweke, J. and S. Porter-Hudak (1983) "The Estimation and Application of Long Memory Time Series Models", *Journal of Time Series Analysis,* 4, 221-238

Ginovian, M.S. (1994) "On Toeplitz Type Quadratic Functionals of Stationary Gaussian Proceses", *Probability Theory and Related Fields,* 100, 395-406

Giraitis, L. and D.Surgailis (1990) "A Central Limit theorem for Quadratic forms in strongly Dependent Linear Variables and Its Application to Asymptotic Normality of Whittle's Estimate", *Probability Theory and Related Fields* 86, 87-104

Giraitis, L., Robinson, P.M. and A.Samarov (1997) "Rate Optimal Semiparametric Estimation of the Memory Parameter of a Gaussian Time Series with Long Range Dependence", *Journal of Time Series Analysis,* 18, 49-60

Giraitis, L., Robinson, P.M. and D.Surgailis (1997) "Variance-Type Estimators of Long Memory", preprint

Gorodetskii, V.V. (1977) "On Convergence to Semi-Stable Gaussian Processes", *Theory of Probability and its Applications,* 22, 498-508

Gourieroux, C., Maurel, F. and A.Monfort (1989) "Least Squares and Fractionally Integrated Regressors", Working Paper #8913, INSEE

Granger, C.W.J. (1980) "Long Memory Relationships and the Aggregation of Dynamic Models", *Journal of Econometrics,* 14, 227-238

Granger, C.W.J. and R.Joyeux (1980) "An Introduction to Long-Memory Time Series Models" *Journal of Time Series Analysis,* 1, 15-29

Grenander, U. (1954) "On Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance", *Annals of Mathematical Statistics,* 25, 252-272

Grenander, U. and M.Rosenblatt (1984) "Statistical Analysis of Stationary Time Series", 2nd edition, Chelsea, New York

Hall, P., Koul, H.L. and B.A.Turlach (1996), "On Estimating the Self-Similarity Index for a Long-Range Dependence Parameter", preprint

Hannan, E.J. (1963) "Regression for Time Series with Errors of Measurement", *Biometrika,* 50, 293-302

Hannan, E.J. (1970) *Multiple Time Series,* New York, J.Wiley

Hannan, E.J. (1971) "Nonlinear Time Series Regression", *Journal of Applied Probability,* 8, 767-780

Hannan, E.J. and P.M.Robinson (1973) "Lagged Regression with Unknown Lags", *Journal of the Royal Statistical Society,* Series B, 35, 252-267

Hansen, B.E. (1992a) "Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes", *Econometrica* 60, 967-972

Hansen, B.E. (1992b) "Convergence to Stochastic Integrals for Dependent Heterogenous Processes", *Econometrica,* 60, 967-972

Herrndorf, N. (1984) "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables", *Annals of Probability* 12, 141-153

Hidalgo, J. (1997) "Estimation of the Pole of the Long Range Process", preprint

Hidalgo, J. and Y.Yajima (1997) "Semiparametric Estimation of the Long Range Parameter", preprint

Hosking, J.R.M (1981) "Fractional Differencing", *Biometrika,* 68, 165-176

Hurvich, C.M and B.K.Ray (1995) "Estimation of the Memory Parameter for NonStationary or NonInvertible Fractionally Integrated Processes", *Journal of Time Series Analysis,* 16, 17-40

Hurvich, C.M., R.Deo and J.Brodsky (1998) "The Mean Squared Error of Geweke and Porter-Hudak's Estimates of the Memory Parameter of a Long memory Time Series", *Journal of Time Series Analysis,* 19, 19-46

Ibragimov, I.A. and Y.A.Rozanov (1977) *Gaussian Random Processes*, Springer and Verlag, New York

Janacek, G.J (1982) "Determining the Degree of Differencing for Time series via the Log Spectrum", *Journal of Time Series Analysis,* 3, 177-183

Jeganathan, P. (1980) "An Extension of a Result of LeCam Concerning Asymptotic Normality", *Sankhya Series A,* 42, 146-160

Jeganathan, P. (1988) "On the Strong Approximation of the Distributions of Estimators in Linear Stochastic Models, I and II: Stationary and Explosive Models", *Annals of Statistics,* 16, 1283-1314

Jeganathan, P. (1996) "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors", University of Michigan, Dep. of Statistics, *Technical Report #277*

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economics Dynamics and Control,* 12, 231-254

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregression Models", *Econometrica* 59, 1551-1580

Kluppelberg, C. and T. Mikosch (1998) "The Integrated Periodogram for Stable Processes", *Annals of Statistics,* forthcoming

Kokoszka, P. and T.Mikosch (1997) "The Integrated Periodogram for Long-Memory Processes with Finite or Infinite Variance", *Stochastic Processes and their Applications,* 66, 55-78

Kokoszka, P. and Taqqu, M.S. (1995) "Fractional ARIMA with Stable Innovations", *Stochastic Processes and their Applications,* 60, 19-47

Kokoszka, P. and Taqqu, M.S. (1996) "Parameter Estimation for Infinite Variance Fractional ARIMA", *Annals of Statistics,* 24, 1880-1913

Kolmogorov, A.N. (1940) "Wienersche Spiralen und Inige Andere Interessante Kurven in Hilbertschen Raum", *Comptes Rendues (Doklady) URSS*, 26, 115-118

Kolmogorov, A.N. (1941) "Local Structure of Turbulence in an Incompressible Liquid for Very Large Reynolds Numbers", *Comptes Rendue (Doklady) URSS (N.S.)*, 30, 299-303

Komlos, J., Major, P. and Tusnady, G. (1976) "An Approximation of Partial Sums of Independent R.V.'s and the Sample Distribution Function II", *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34, 33-58

Künsch, H.R. (1987) "Statistical Aspects of Self-Similar Processes", *Proc.1st World Congress of the Bernoulli Society*, 67-74, VNU Science Press

Kurtz, T.G. and P.Protter (1991) "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations", *Annals of Probability*, 19, 1035-1070

LeCam, L. (1986) *Asymptotic Methods in Statistical Decision Theory*, New York, Springer and Verlag

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

Lobato, I. (1995) *Multivariate Analysis of Long Memory Series in the Frequency Domain*, Ph.D. Thesis, University of London

Lobato, I. (1997) "Consistency of the Averaged Cross-Periodogram in Long Memory Series", *Journal of Time Series Analysis*, 18, 137-156

Lobato, I. and P.M.Robinson (1996) "Averaged Periodogram Estimation of Long Memory", *Journal of Econometrics*, 73, 303-324

Lubian, D. (1996) "Long Memory Errors in Time Series Regressions with a Unit Root", preprint, University of Padova

Magnus, J. and H.Neudecker (1991) *Matrix Differential Calculus with Applications to Statistics and Econometrics*, 2nd edition, J.Wiley, Chichester

Major, P. (1976) "The Approximation of Partial Sums of Independent R.V.'s"

Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete, 35, 213-220

Major, P. (1981) "Limit Theorems for Non-Linear Functionals of Gaussian Sequences", *Zeitschrift fur Wahrscheinlichkeitstheorie verw. Gebiete,* 57, 129-158

Mandelbrot, B.B. and Van Ness J.W. (1968) "Fractional Brownian Motions, Fractional Noises and Applications" *SIAM Review,* 10, 422-437

Mandelbrot, B.B. and J.R.Wallis (1969), "Noah, Joseph and Operational Hydrology", *Water Resources Research,* 5, 228-267

Marinucci, D. and P.M. Robinson (1997a) "Alternative Definitions of Fractional Brownian Motion", forthcoming on *Journal of Statistical Planning and Inference*

Marinucci, D. and P.M. Robinson (1997b) "A Functional Central Limit Theorem for Multivariate Fractional Time Series", preprint

McLeish, D.L. (1977) "On the Invariance Principle for Nonstationary Mixingales", *Annals of Probability,* 5, 616-621

Mikosch, T. and R.Norvaisa (1995) "Uniform Convergence of the Empirical Spectral Distribution Function", preprint

Mohr, D. (1981) *Modelling Data as Fractional Gaussian Noise,* Ph.D. Thesis, University of Princeton

Nelson, C.R. and C.I.Plosser (1982) "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics,* 10, 139-162

Newey, W.K. and K.West (1987) "A Simple Positive Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica,* 55, 703-708

Park, J.Y. and P.C.B. Phillips (1988) "Statistical Inference in Regressions with Integrated Processes, Part1", *Econometric Theory* 5, 95-132

Pham, T.D. and Tran, L.T. (1985) "Some Mixing Properties of Time Series Models", *Stochastic Processes and their Applications,* 19, 297-303

Phillips, P.C.B. (1988a) "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$", *Journal of Multivariate Analysis* 24, 252-264

Phillips, P.C.B. (1988b) "Weak Convergence of Sample Covariance Matrix to Stochastic Integrals via Martingale Approximation", *Econometric Theory* 4, 528-533

Phillips, P.C.B. (1991a) "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306

Phillips, P.C.B (1991b) "Spectral Regression for Cointegrated Time Series", *Advances in Econometrics*, 5th World Congress, Cambridge, Cambridge University Press

Phillips, P.C.B. and N.Durlauf (1986) "Multiple Time Series Regression with Integrated Processes", *Review of Economics and Statistics* 53, 473-495

Phillips, P.C.B and B.E.Hansen (1990) "Statistical Inference in Instrumental Variable Regression with I(1) Variables", *Review of Economic Studies* 57, 99-125

Phillips, P.C.B. and S.Ouliaris (1988) "Testing for Cointegration Using Principal Components Methods", *Journal of Economics Dynamics and Control,* 12, 205-230

Phillips, P.C.B. and Solo, V. (1992) "Asymptotics for Linear Processes", *Annals of Statistics,* 20, 971-1001

Pollard, D. (1984) *Convergence of Stochastic Processes,* New York, Springer and Verlag

Press, H. and J.W. Tukey (1956) "Power Spectral Methods of Analysis and their Application to Problems in Airline Dynamics", *Bell Systems Monographs* 2606

Prokhorov, Y.V. (1956) "Convergence of Random Processes and Limit Theorems in Probability Theory", *Theory of Probability and its Applications,* 1, 157-213

Robinson, P.M. (1972) "Nonlinear Vector Time Series Regression", *Journal of Applied Probability,* 9,

Robinson, P.M. (1978) "Statistical Inference for a Random Coefficient Autoregressive Model", *Scandinavian Journal of Statistics,* 5, 163-168

Robinson, P.M. (1991) "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models" *Econometrica,* 59, 1329-1354

Robinson, P.M. (1994a) "Semiparametric Analysis of Long-Memory Series", *Annals of Statistics,* 22, 515-539

Robinson, P.M. (1994b) "Time Series with Strong Dependence", *Advances in Econometrics*, 6th World Congress, Cambridge, Cambridge University Press

Robinson, P.M. (1994c) "Rates of Convergence and Optimal Spectral Bandwidth for Long Range Dependence", *Probability Theory and Related Fields* 99, 443-473

Robinson, P.M. (1994d) "Efficient Tests of Nonstationary Hypothesis", *Journal of the American Statistical Association*, 89, 1420-1437

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

Robinson, P.M (1995b) "Gaussian Semiparametric Estimation of Long-Range Dependence" *Annals of Statistics*, 23, 1630-1661

Robinson, P.M. (1997) "Inference-Without-Smoothing in the Presence of Nonparametric Autocorrelation", *Econometrica,* forthcoming

Robinson, P.M. and F.J.Hidalgo (1997) "Time Series Regression with Long Range Dependence", *Annals of Statistics*, 25, 77-104

Robinson, P.M. and D.Marinucci (1997) "Semiparametric Frequency-Domain Analysis of Fractional Cointegration", preprint

Robinson, P.M. and D.Marinucci (1998) "Narrow-Band Approximations of Sample Moments in the Presence of Deterministic and Syochastic Trends", in preparation

Samorodnitsky, G. and M.Taqqu (1994) *Stable Non-Gaussian Random Processes,* Chapman and Hall, New York

Silveira, G. (1991) *Contributions to Strong Approximations in Time Series with Applications in Nonparametric Statistics and Functional Central Limit Theorems* Ph.D. Thesis, University of London

Sims, C. (1988) "Bayesian Skepticism on Unit Root Econometrics", *Journal of Economics Dynamics and Control*, 12, 463-474

Sinai, Y.G. (1976) "Self-Similar Probability Distributions", *Theory of Probability and its Applications,* 21, 64-80

Skorokhod, A.V. (1956) "Limit Theorems for Stochastic Processes", *Theory of Probability and its Applications* 1, 261-290

Sowell, F.B. (1990) "The Fractional Unit Root Distribution", *Econometrica,* 58, 495-505

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

Taqqu, M. S. (1975) "Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process", *Zeitschrift fur Wahrscheinlichkeistheorie. verw. Gebiete,* 31, 287-302

Taqqu, M.S. (1979) "Convergence of Integrated Processes of Arbitrary Hermite Rank", *Zeitscrift fur Wahrscheinlichkeitstheorie verw. Gebiete,* 50, 53-83

Taqqu, M.S., Teverovsky, V. and W.Willinger, (1995), "Estimators for Long-Range Dependence: an Empirical Study", *Fractals,* 3, 785-798

Terrin, N. and M.S.Taqqu (1991) "Convergence in Distribution of Sums of Bivariate Appell Polynomials with Long-Range Dependence", *Probability Theory and Related Fields,* 90, 57-81

Teverovsky, V. and M.S.Taqqu (1996) "Testing for Long-Range Dependence in the Presence of Shifting Mean or a Slowly Declining Trend, Using a Variance-Type Estimator", *Journal of Time Series Analysis,* forthcoming

Truong-Van, B. (1995) "Invariance Principles for Semi-Stationary Sequence of Linear Processes and Applications to ARMA Process", *Stochastic Processes and their Applications,* 58, 155-172

Yajima, Y. (1988) "On Estimation of a Regression Model with Long-Memory Stationary Errors", *Annals of Statistics,* 16, 791-807

Yajima, Y. (1989) "A Central Limit Theorem of Fourier Transforms of Strongly Dependent Stationary Processes", *Journal of Time Series Analysis,* 10, 375-383

Yajima, Y. (1991) "Asymptotic Properties of the LSE in a Regression Model with Long-Memory Stationary Errors", *Annals of Statistics,* 19, 158-177

Yong, C.H. (1974) *"Asymptotic Behaviour of Trigonometric Series"*, Chinese University of Hong Kong

Velasco, C. (1996) "NonStationary Log-Periodogram Regression", preprint, London School of Economics & Political Science

Velasco, C. (1997a) "Non-Gaussian Log-Periodogram Regression", preprint, London School of Economics

Velasco, C. (1997b) "Gaussian Semi-Parametric Estimation of Non-Stationary Time Series", preprint, University of Oxford

Verwaat, W. (1985) "Sample Paths of Self-Similar Process with Stationary Increments", *Annals of Probability,* 13, 1-27

Weyl, H. (1917) "Bemerkungen zum Begriff der Differential-Quotenten gebrochener Ordnung", *Vierteljschr. Naturfosch. Ges. Zurich*, 62, 296-302

Whittle, P. (1951) *Hypothesis Testing in Time Series Analysis,* Uppsala, Almqvist and Wicksell

Wooldridge, J.M. and H.White (1988) "Some Invariance Principles and Central Limit Theorems for Dependent Heterogenous Processes", *Econometric Theory* 4, 210-230

Zygmund, A. (1977) *Trigonometric Series,* Cambridge: Cambridge University Press

MISPRINTS

- page 17, 9↑, $I_{uv}\lambda)$ should be $I_{uv}(\lambda)$

- page 18, line 10↓, $x_t'B$ should be $B'x_t$

- page 18, line 3↑, $F_{xx}(1, n-1)$ should be $\widehat{F}_{xx}(1, n-1)$, $\widehat{F}_{xy}(1, n-1)$ should be $\widehat{F}_{xy}(1, n-1)$

- page 18, line 2↑, $F_{xx}(1, n-1)$ should be $\widehat{F}_{xx}(1, n-1)$

- page 19, 6↑, $\widetilde{B}$ should be $\widehat{B}$

– page 23, line 3↓, $X'X$ should be $\sum_{t=1}^{n} x_t x_t'$

- page 27, line 10↓, Granger (1987) should be Engle and Granger (1987)

– page 53, line 3↓, $H_j(x)$ should be $H_j^i(x)$

- page 60, (2.59), $\widehat{v}$ should be $\widehat{\nu}$

- page 65, 2↑, Assumption 2B should be Assumption 2A

- page 70, 7↑, contant should be constant

- page 71, $[ny_2]$ should be $[n\rho_2]$ and $[nr_1]$ should be $[n\rho_1]$

- page 73, line 10↓, $r_2 - r_1 > \frac{1}{n}$ should be $r_2 - r_1 < \frac{1}{n}$

– page 87, line 5↓, add $\kappa_t = 1(t = 0)$ for $\rho = 0$

- page 87, line 7↓, add $\rho = 0$ after $0 \leq |\lambda| \leq \pi$

- page 106, line 4 ↓, real matrix should be real symmmetric matrix

- page 107, line 2↑, Theorem 4.1 should be Theorem 3.1

- page 108, line 1↑, dots are missing after $(1 - L)^{-d_1}$

- page 110, equation (4.24), $\underline{0}'$ should be $\underline{0}$

- page 113, line 8↑, $\widehat{\beta}_{m-1}$ should be $\widehat{\beta}_{n-1}$

- page 121, line 8↑, $\widetilde{\delta}_z$ should be $\widetilde{\delta}_x$

- page 128, line 4↓, add a.s.

- page 150, line 8↑, $(e_t, 1 - \phi L)^{-1}...)$ should be $(e_t, (1 - \phi L)^{-1}...)$

– page 151, line 4↑, $M = o(n^2)$ should be $M^2 = o(n)$

- page 152, equation (5.24), v^{1-2d_x} should be $|v|^{2d_x-1}$ and $v^{1-d_x-d_e}$ should be $!v|^{d_x+d_e-1}$

– page 159, line 7↓, simmetry should be symmetry

– add to references Bladt, M. (1994) "Multivariate Self-Similar Processes: Second-Order Theory", *Journal of Applied Probability,* 31, 139-147

-add to references Hannan, E.J. (1979) "The Central Limit Theorem in Time Series Regression", *Stochastic Processes and their Applications,* 9, 281-289