Econometric Theory, **18**, 2002, 51–78. Printed in the United States of America.

SAMPLE MEANS, SAMPLE AUTOCOVARIANCES, AND LINEAR REGRESSION OF STATIONARY MULTIVARIATE LONG MEMORY PROCESSES

CHING-FAN CHUNG
Academia Sinica

We develop an asymptotic theory for the first two sample moments of a stationary multivariate long memory process under fairly general conditions. In this theory the convergence rates and the limits (the fractional Brownian motion, the Rosenblatt process, etc.) all depend intrinsically on the degree of long memory in the process. The theory of the sample moments is then applied to the multiple linear regression model. An interesting finding is that, even though all the regressors and the disturbance are stationary and ergodic, the joint long memory in one single regressor and in the disturbance can invalidate the usual asymptotic theory for the ordinary least squares (OLS) estimation. Specifically, the convergence rates of the OLS estimators become slower, the limits are not normal, and the standard t- and F-tests all collapse.

1. INTRODUCTION

In this paper we develop an asymptotic theory for the ordinary least squares (OLS) estimator in the time series regression in the presence of long-range dependence in both stochastic regressors and disturbance. We define the long-range dependence, or long memory, in a stationary process x_t based on the behavior of the coefficients ψ_j of its infinite-order moving average representation $x_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, where a_t are innovations. More specifically, we assume the order of ψ_j is $O(j^{d-1})$, where d is a memory parameter whose value lies between 0 and 0.5. That is, ψ_j converges to zero slowly at a hyperbolic rate that depends on the memory parameter d.

The main conclusion from our analysis is as follows: if in a multiple linear regression model one of the regressors has long memory with a memory parameter d_1 and the disturbance also has long memory with a memory parameter d^* such that $d_1 + d^* > 0.5$, then the OLS estimator will converge at the $T^{d_1+d^*-1}$ rate (which is slower than the usual $T^{-0.5}$ rate with T being the sample size) to

Address correspondence to: Ching-Fan Chung, Institute of Economics, Academia Sinica, Nankang, Taipei, Taiwan; e-mail: cfchung@econ.sinica.edu.tw.

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

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

a nonnormal distribution. In fact, the condition on d_1 and d^* is not only sufficient but also necessary for this nonstandard result. That is, if none of the regressors has sufficiently strong long memory to add to the long memory of the disturbance, then the standard $\sqrt{T}$-consistency theory will apply and the OLS estimator will have a normal limiting distribution. One striking feature of our result is that even though all the regressors and the disturbance are stationary and ergodic, the OLS estimator may converge slowly to a nonnormal limiting distribution and the conventional *t*- and *F*-tests may all collapse simply as a result of joint long memory in one regressor and in the disturbance. We note that Robinson (1994) and Robinson and Hidalgo (1997), when proposing a class of estimators for the long memory regression model that is $\sqrt{T}$-consistent and asymptotically normal, point out the possibility that the OLS estimator no longer needs to be $\sqrt{T}$-consistent if sufficiently strong long-range dependence exists in both regressors and disturbance. We also note that Sowell (1990) demonstrates the fast convergence rate and the nonnormal asymptotic distribution of the OLS estimator in a first-order autoregression model of a nonstationary long memory process with a memory parameter that lies between 0.5 and 1.5. Other related studies, particularly on estimation efficiency in the multiple linear regression models with *nonstochastic* regressors, include research by Yajima (1988, 1991) and Dahlhaus (1995), who consider the least squares estimator and the best linear unbiased estimator; and by Koul (1992), Koul and Mukherjee (1993), and Giraitis, Koul, and Surgailis (1996), who consider a number of robust estimators.

To analyze the multiple linear regression model of long memory processes, we first derive some asymptotic results for multivariate long memory processes that generalize their univariate counterparts, particularly Hosking's (1996) results on sample autocovariances and Fox and Taqqu's (1987) theory on the product of two univariate long memory processes. In so doing we also relax the stringent Gaussian assumption underlying Hosking's and Fox and Taqqu's theory. The plan of this paper is as follows. After defining the multivariate long memory processes based on the matrix moving average coefficients in Section 2, we then prove the functional central limit theorems for the sample mean and sample autocovariance of multivariate long memory processes, respectively, in Section 3. These theorems for the sample moments are then employed to develop the asymptotic theory for the multiple linear regression model and to analyze the conventional *t*- and *F*-test statistics in Section 4. Section 5 concludes. The proofs appear in the Appendix.

2. THE MULTIVARIATE LONG MEMORY PROCESS

Consider a stationary K-dimensional process $\mathbf{x}_t = [x_{1t} \ldots x_{K^t}]'$ that has a mean $\boldsymbol{\mu} = [\mu_1 \ldots \mu_K]'$ and an infinite-order moving average representation as follows:

$$\mathbf{x}_t - \boldsymbol{\mu} = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{a}_{t-j}, \tag{1}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

where the innovations $\mathbf{a}_t = [a_{1t} \ldots a_{K^t}]'$ are K-dimensional martingale differences with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$ such that for some $\epsilon > 0$, $\sup_t E(|a_{it}|^{2+\epsilon} \| \mathcal{F}_{t-1}) < \infty$, a.s., for all $i = 1, \ldots, K$. Let $E(\mathbf{a}_t \mathbf{a}_t' \| \mathcal{F}_{t-1}) = \mathbf{\Sigma}_a$, a.s. The $K \times K$ matrix coefficients $\mathbf{\Psi}_j$ are often referred to as impulse responses. The main characterization of the process $\mathbf{x}_t$ considered in this paper is that the impulse responses $\mathbf{\Psi}_j$ converge at slow hyperbolic rates as $j \to \infty$. More precisely, there are K memory parameters $d_1, d_2, \ldots, d_K$, whose values lie in (0, 0.5) such that the impulse responses $\mathbf{\Psi}_j$ can be approximated by

$$\mathbf{\Psi}_j \sim \mathbf{D}\left[\frac{1}{\Gamma(d)} j^{d-1}\right]\mathbf{\Pi}, \quad \text{as } j \to \infty, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function and $\mathbf{\Pi}$ is a nonsingular $K \times K$ matrix of constants that are independent of j and may be functions of a smaller set of unknown parameters. The notation $\mathbf{D}[\, j^{d-1}/\Gamma(d)]$ represents a $K \times K$ diagonal matrix with $j^{d_1-1}/\Gamma(d_1), \ldots, j^{d_K-1}/\Gamma(d_K)$ on the diagonal. In fact, for any univariate function f of a single variable, the notation $\mathbf{D}[\, f(d)]$ represents a $K \times K$ diagonal matrix with $f(d_1), \ldots, f(d_K)$ on the diagonal. Also, the notation $\sim$ is defined as follows: given two sequences of matrices $\mathbf{U}_j$ and $\mathbf{V}_j$ of the same dimensions, $\mathbf{U}_j \sim \mathbf{V}_j$, as $j \to \infty$, if $u_{ik,j}/v_{ik,j} \to 1$, as $j \to \infty$, for all i and k, where $u_{ik,j}$ and $v_{ik,j}$ are the (i,k)th elements of $\mathbf{U}_j$ and of $\mathbf{V}_j$, respectively. Let $\boldsymbol{\psi}_{i,j}'$ and $\boldsymbol{\pi}_i'$ be the ith *rows* of $\mathbf{\Psi}_j$ and $\mathbf{\Pi}$, respectively; then (2) implies that $\boldsymbol{\psi}_{i,j}' \sim j^{d_i-1}\Gamma(d_i)^{-1}\boldsymbol{\pi}_i'$, as $j \to \infty$, for all $i = 1, \ldots, K$. Note that the conditions on the memory parameters $d_i \in (0, 0.5)$, for $i = 1, \ldots, K$, ensure that the impulse responses are square-summable and the infinite sum in (1) exists.

The typical stochastic process that satisfies conditions (1) and (2) is the K-dimensional stationary and invertible vector fractionally integrated autoregressive moving average process, or the VARFIMA process, $\tilde{\mathbf{x}}_t$ that is defined by

$$\mathbf{\Phi}(B)\mathbf{D}[(1-B)^d](\tilde{\mathbf{x}}_t - \boldsymbol{\mu}) = \mathbf{\Theta}(B)\mathbf{a}_t, \tag{3}$$

where B is the lag operator and the autoregressive moving average (ARMA) operators $\mathbf{\Phi}(B) = \mathbf{\Phi}_0 - \sum_{j=1}^{p} \mathbf{\Phi}_j B^j$ and $\mathbf{\Theta}(B) = \mathbf{\Theta}_0 + \sum_{j=1}^{q} \mathbf{\Theta}_j B^j$ are $K \times K$ matrix polynomials in B. The matrix coefficients of polynomials $\mathbf{\Phi}(B)$ and $\mathbf{\Theta}(B)$ are assumed to satisfy the standard stationarity and invertibility conditions for the vector ARMA model and are functions of some estimable parameters after normalization and identification restrictions are imposed. The operator $\mathbf{D}[(1-B)^d]$ is a $K \times K$ diagonal matrix with $(1-B)^{d_1}, \ldots, (1-B)^{d_K}$ on the diagonal. When the differencing parameters $d_1, d_2, \ldots, d_K$ are fractional numbers, then $(1-B)^{d_i}$ is defined by the expansion $\sum_{j=0}^{\infty} \Gamma(j - d_i)[\Gamma(j+1)\Gamma(-d_i)]^{-1}B^j$. Finally, the innovations $\mathbf{a}_t$ are assumed to be K-dimensional martingale differences that satisfy the same moment conditions as those listed under (1). Hosking (1981) shows that if $-0.5 < d_i < 0.5$, for all $i = 1, \ldots, K$, then the process $\tilde{\mathbf{x}}_t$ is both stationary and invertible. We shall call $\tilde{\mathbf{x}}_t$ the VARFIMA$(d_1, \ldots, d_K)$ process to stress the central role of the fractional differencing parameters $d_1, d_2, \ldots, d_K$. Also,

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

the univariate version with $K = 1$ is referred to as the ARFIMA(p, d_1, q) process, which has been extensively studied by Hosking (1981), Granger (1980, 1981), Granger and Joyeux (1980), and Sowell (1992a, 1992b). A stationary and invertible VARFIMA$(d_1, \ldots, d_K)$ process $\tilde{\mathbf{x}}_t$ has an infinite-order moving average representation as in (1). It is straightforward to derive $\boldsymbol{\Psi}_j$ for the VARFIMA$(d_1, \ldots, d_K)$ process, and Chung (1999) has demonstrated the approximation (2) with $\boldsymbol{\Pi} = \boldsymbol{\Phi}(1)^{-1}\boldsymbol{\Theta}(1)$.

Given the hyperbolic convergence rates (2) of the impulse responses, the cumulative impulse responses also progress at hyperbolic rates as indicated in the following lemma.

LEMMA 1. *Given the approximation (2), then*

$$\sum_{k=0}^{j} \boldsymbol{\Psi}_k \sim \mathbf{D}\left[\frac{1}{\Gamma(d+1)} j^d\right]\boldsymbol{\Pi}, \quad as\ j \to \infty. \tag{4}$$

Note that it is because the cumulative impulse response $\sum_{k=0}^{j} \boldsymbol{\Psi}_k$ diverges hyperbolically that we say the process $\mathbf{x}_t$ has long memory. These hyperbolically divergent cumulative impulse responses are very different from the geometrically convergent cumulative impulse responses that characterize the stationary and invertible vector ARMA model.

The following lemma indicates that the autocovariances $\boldsymbol{\Gamma}(j) \equiv \mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_{t+j})$ of the long memory process $\mathbf{x}_t$ must also converge at hyperbolic rates.

LEMMA 2. *Given the stationary process* $\mathbf{x}_t$ *defined by (1) and (2), we have*

$$\boldsymbol{\Gamma}(j) \sim \mathbf{D}(j^{d-0.5}) \cdot \mathbf{A} \cdot \mathbf{D}(j^{d-0.5}), \quad as\ j \to \infty,$$

where the (i,k)*th element of the* $K \times K$ *matrix* $\mathbf{A}$ *is* $\Gamma(1 - d_i - d_k)/[\Gamma(d_k) \times \Gamma(1 - d_k)] \cdot \boldsymbol{\pi}_i' \boldsymbol{\Sigma}_a \boldsymbol{\pi}_k$.

Hence, as $j \to \infty$, not only do the autocovariances $\gamma_{ii}(j)$ of each x_{it} die out slowly at a hyperbolic rate, i.e., $\gamma_{ii}(j) \sim j^{2d_i - 1} \cdot \Gamma(1 - 2d_i)/[\Gamma(d_i)\Gamma(1 - d_i)] \cdot \boldsymbol{\pi}_i' \boldsymbol{\Sigma}_a \boldsymbol{\pi}_i$, the covariances $\gamma_{ik}(j)$ between the current x_{it} and the future $x_{k,t+j}$, for $i \neq k$, also taper off at hyperbolic rates, i.e., $\gamma_{ik}(j) \sim j^{d_i + d_k - 1} \cdot \Gamma(1 - d_i - d_k)/[\Gamma(d_k)\Gamma(1 - d_k)] \cdot \boldsymbol{\pi}_i' \boldsymbol{\Sigma}_a \boldsymbol{\pi}_k$. Hosking (1996) presents a similar result for the univariate case.

3. THE SAMPLE MEAN AND THE SAMPLE AUTOCOVARIANCES

We first present a limit theorem for the partial sum of the process $\mathbf{x}_t$ defined by (1) and (2) and then examine the asymptotic behavior of the sample mean.

THEOREM 1. *Given the stationary process* $\mathbf{x}_t$ *defined by (1) and (2), if* $0 < d_i < 0.5$, *for all* $i = 1, \ldots, K$, *then*

$$\mathbf{D}(T^{-0.5-d}) \sum_{t=1}^{[Ts]} (\mathbf{x}_t - \boldsymbol{\mu}) \Rightarrow \mathbf{B}_d(s), \tag{5}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

where $\Rightarrow$ *denotes weak convergence,* $[z]$ *means the largest integer that is smaller than or equal to* z*, and* $\mathbf{B}_d(s)$ *is a* K*-dimensional fractional Brownian process with* K *different memory parameters* $d_1, \ldots, d_K$ *and is defined as* $\mathbf{B}_d(s) = [B_{d_1}(s) B_{d_2}(s) \ldots B_{d_K}(s)]'$*. Here each* $B_{d_i}(s)$ *is a one-dimensional fractional Brownian motion that is defined by*

$$B_{d_i}(s) \equiv \frac{1}{\Gamma(d_i + 1)} \times \left\{ \int_0^s (s-t)^{d_i} dB_0^{(i)}(t) + \int_{-\infty}^0 [(s-t)^{d_i} - (-t)^{d_i}] dB_0^{(i)}(t) \right\}, \tag{6}$$

where $B_0^{(i)}(t)$ *is the* i*th element of a* K*-dimensional Brownian motion with the variance-covariance matrix* $\mathbf{\Pi \Sigma}_a \mathbf{\Pi}'$*.*

Note that the individual components of the preceding multivariate convergence result are

$$\frac{1}{T^{0.5+d_i}} \sum_{t=1}^{[Ts]} (x_{it} - \mu_i) \Rightarrow B_{d_i}(s), \qquad i = 1, \ldots, K,$$

where the univariate fractional Brownian motion $B_{d_i}(s)$ was first proposed by Mandelbrot and Van Ness (1968), from whose work it is easy to derive the expectation of $B_{d_i}(t)$, which is zero, and the covariance between $B_{d_t}(t)$ and $B_{d_j}(t)$, which is $\eta_{ij} \cdot (\boldsymbol{\pi}_i' \mathbf{\Sigma}_a \boldsymbol{\pi}_j)$ where

$$\eta_{ij} \equiv \frac{1}{\Gamma(d_i + 1)\Gamma(d_j + 1)} \times \left\{ \frac{1}{1 + d_i + d_j} + \int_{-\infty}^0 [(1-t)^{d_i} - (-t)^{d_i}][(1-t)^{d_j} - (-t)^{d_j}]\, dt \right\}.$$

Consequently, the variance of the sample mean $\bar{\mathbf{x}} = T^{-1} \sum_{t=1}^T \mathbf{x}_t$ can be approximated by

$$\mathrm{Var}(\bar{\mathbf{x}}) \sim \mathbf{D}(T^{d-0.5}) \cdot \mathbf{\Lambda} \cdot \mathbf{D}(T^{d-0.5}), \quad \text{as } T \to \infty, \tag{7}$$

where $\mathbf{\Lambda}$ is a $K \times K$ matrix of which the (i, j)th element is $\eta_{ij} \cdot (\boldsymbol{\pi}_i' \mathbf{\Sigma}_a \boldsymbol{\pi}_j)$. The unusual convergence rate here implies that the sample mean of the long memory process $\mathbf{x}_t$ converges to the population mean more slowly than the usual $T^{-0.5}$ rate. Slower convergence rates in the univariate case were first noted by Adenstedt (1974).

There are extensive studies of the partial sum and the sample mean of the univariate long memory process. The earliest one seems to be work by Davydov (1970, Theorem 2 and Remark 4), who presents a result that is similar to the univariate version of Theorem 1, although his assumptions are slightly more stringent than ours. More recently, Avram and Taqqu (1987, Theorem 2 with

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

$n = 1$) give another univariate version based on different assumptions. Davidson and de Jong (1998) have extended the functional central limit theorem for long memory processes of the form such as (1) but with innovations $\mathbf{a}_t$ that are more general than the martingale differences as we have assumed here.

Let us now consider the "square" of the long memory process $\mathbf{x}_t$. Specifically, we will first derive the limiting distributions of the sample autocovariances with known mean (in Theorem 2) and then the ones with unknown means (in Corollary 1). Our asymptotic theory requires more stringent moment conditions and narrower ranges for the memory parameters d_i as indicated in the following theorem.

THEOREM 2. *Given the stationary process* $\mathbf{x}_t$ *defined by (1) and (2), we assume that for some* $\epsilon > 0$, $\sup_t E(|a_{it}|^{4+\epsilon} \| \mathcal{F}_{t-1}) < \infty$, *a.s. If* $0.25 < d_i < 0.5$, *for all* $i = 1, \ldots, K$, *then for any finite integer* j *we have*

$$\mathbf{D}(T^{-d}) \left\{ \sum_{t=1}^{[Ts]-j} [(\mathbf{x}_t - \boldsymbol{\mu})(\mathbf{x}_{t+j} - \boldsymbol{\mu})' - \boldsymbol{\Gamma}(j)] \right\} \mathbf{D}(T^{-d}) \Rightarrow \mathbf{R}(s), \tag{8}$$

where $\boldsymbol{\Gamma}(j)$ *are population autocovariances and* $\mathbf{R}(s)$ *is a* $K \times K$ *matrix of processes of which the* (ℓ, k)*th element is defined by*

$$R^{(\ell k)}(s) \equiv \frac{1}{\Gamma(d_\ell)\Gamma(d_k)} \times \int_{-\infty}^{s} \int_{-\infty}^{s} \left[\int_0^s (t - r_\ell)_+^{d_\ell - 1} (t - r_k)_+^{d_k - 1} \, dt \right] dB_0^{(\ell)}(r_\ell) dB_0^{(k)}(r_k). \tag{9}$$

Here $(z)_+^\delta \equiv 0$ *if* $z \leq 0$, *and* $(z)_+^\delta \equiv z^\delta$ *if* $z > 0$; *and* $B_0^{(k)}(r)$ *is the kth element of a K-dimensional Brownian motion with the variance-covariance matrix* $\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'$. *Furthermore, we have the joint weak convergence of (5) and (8).*

The most interesting finding is that the limit does not depend on j. The conditions $0.25 < d_i < 0.5$, for all $i = 1, \ldots, K$, imply that the process $\mathbf{x}_t$ has *strong* long memory. The univariate version of this result was first shown by Rosenblatt (1979) under the Gaussian assumption. Fox and Taqqu (1985, Theorem 1) present a general theory for the Gaussian long memory process that includes the result of Theorem 2 as a special case. All these previous works on the univariate case rely on the Gaussian assumption, whereas ours does not.

The univariate process $R^{(\ell k)}(t)$ for $\ell \neq k$ was first suggested by Fox and Taqqu (1987, Theorem 6.1) in an analysis of the product of two univariate series x_{1t} and x_{2t} that relies on the stronger assumption that x_{1t} and x_{2t} are stationary joint Gaussian processes, whereas the process $R^{(\ell\ell)}(s)$ is the so-called Rosenblatt (1961, 1979) process, for which Avram and Taqqu (1987, Theorem 2 with $n = 2$) have derived the sample variance under the assumptions that the series $x_{\ell t}$ is expressible as an infinite-order moving average of independently and identically distributed innovations with finite fourth mo-

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

ments and that the moving average coefficients converge hyperbolically. Avram and Taqqu do not make the Gaussian assumption, and so their theory is more or less equivalent to the univariate version of Theorem 2. It should be pointed out that in the definition of $R^{(\ell k)}(s)$ in (9) the integration on t cannot be interchanged with the integration on $B_0^{(\ell)}(r_\ell)$ because $(t - r_\ell)^{d_\ell - 1}$ is not square-integrable. Here $R^{(\ell k)}(s)$ is self-similar with the self-similarity parameter $d_\ell + d_k$; i.e., for any $\alpha > 0$, $R^{(\ell k)}(\alpha s) \overset{d}{=} \alpha^{d_\ell + d_k} R^{(\ell k)}(s)$. Unlike the fractional Brownian motion, $R^{(\ell k)}(s)$ is not Gaussian. Fox and Taqqu (1987, Lemma 6.4) and Hosking (1996) have derived cumulants for the univariate version of $R^{(\ell k)}(s)$ with $\ell \neq k$ and with $\ell = k$, respectively. In particular, their results imply that $E[R^{(\ell k)}(s)] = 0$ for all ℓ and k, from which we conclude that the mean of the limiting process $\mathbf{R}(s)$ in Theorem 2 is also zero.

Let us now consider the sample autocovariances that are defined as

$$\begin{aligned}\hat{\mathbf{\Gamma}}(j) &\equiv \frac{1}{T-j}\sum_{t=1}^{T-j}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t+j} - \bar{\mathbf{x}})' \\ &\approx \frac{1}{T}\sum_{t=1}^{T-j}(\mathbf{x}_t - \boldsymbol{\mu})(\mathbf{x}_{t+j} - \boldsymbol{\mu})' - (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})',\end{aligned}$$

for $j = 1,2,\ldots < \infty$. To derive the limiting distribution of $\hat{\mathbf{\Gamma}}(j)$, we first note that Theorems 1 and 2 imply that if the process $\mathbf{x}_t$ satisfies the conditions of Theorem 2, then

$$\begin{aligned}&\mathbf{D}(T^{-d})\left\{\sum_{t=1}^{[Ts]-j}[(\mathbf{x}_t - \boldsymbol{\mu})(\mathbf{x}_{t+j} - \boldsymbol{\mu})' - \mathbf{\Gamma}(j)]\right\}\mathbf{D}(T^{-d}) \\ &\quad - \mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}(\mathbf{x}_t - \boldsymbol{\mu})\cdot\sum_{t'=1}^{[Ts]}(\mathbf{x}_{t'} - \boldsymbol{\mu})'\mathbf{D}(T^{-0.5-d}) \\ &\Rightarrow \mathbf{R}(s) - \mathbf{B}_d(s)\mathbf{B}_d(s)',\end{aligned}$$

where the processes $\mathbf{B}_d(s)$ and $\mathbf{R}(s)$ are defined in Theorems 1 and 2, respectively. Consequently, we have the first part of the following corollary. The second part of the corollary considers the complementary case with $0 < d_i < 0.25$, for all $i = 1,\ldots,K$. The result is a direct consequence of Hannan's (1976, p. 397) central limit theorem for the sample autocovariances of a stationary process. The two parts of this corollary illustrate the sharp contrast between two cases of long memory processes.

COROLLARY 1. *Given the stationary process* $\mathbf{x}_t$ *defined by (1) and (2), we assume that for some* $\epsilon > 0$, $\sup_t E(|a_{it}|^{4+\epsilon}\|\mathcal{F}_{t-1}) < \infty$, *a.s.*

(1) If $0.25 < d_i < 0.5$, *for all* $i = 1,\ldots,K$, *then, for any finite integer* j, *we have*

$$\mathbf{D}(T^{0.5-d})[\hat{\mathbf{\Gamma}}(j) - \mathbf{\Gamma}(j)]\mathbf{D}(T^{0.5-d}) \Rightarrow \mathbf{R}(1) - \mathbf{B}_d(1)\mathbf{B}_d(1)'. \quad \textbf{(10)}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

(2) If $0 < d_i < 0.25$*, for all* $i = 1, \ldots, K$*, then*

$$\sqrt{T}\,\mathrm{vec}[\hat{\mathbf{\Gamma}}(j) - \mathbf{\Gamma}(j)] \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{W}_{xx}^{(j)}),$$

for a $K^2 \times K^2$ *positive-definite matrix* $\mathbf{W}_{xx}^{(j)}$ *whose definition can be found in Hannan (1976, p. 397).*

The univariate version of these results was first suggested by Hosking (1996, Theorem 4) under the Gaussian assumption. It is interesting to note that, as the values of d_i decline from 0.5 toward 0.25, the convergence rate of the sample variance grows more and more quickly toward the standard $T^{-0.5}$ rate. When the values of d_i pass 0.25, then the convergence rate of the sample variance stops changing and settles at the $T^{-0.5}$ rate, whereas the limiting distribution changes from a nonnormal distribution to a normal one. The case where $d_i = 0.25$ is not available under the present assumptions. We note that Hosking (1996, Theorem 4) has derived a result for the univariate case under the Gaussian innovations assumption that $\sqrt{T/\ln(T)}[\hat{\gamma}_{ii}(j) - \gamma_{ii}(j)]$ will have a Gaussian limiting distribution given $d_i = 0.25$.

When $0.25 < d_i < 0.5$, for all $i = 1, \ldots, K$, we have

$$E[\hat{\mathbf{\Gamma}}(j)] - \mathbf{\Gamma}(j) \sim -\mathbf{D}(T^{d-0.5}) \cdot \mathbf{\Lambda} \cdot \mathbf{D}(T^{d-0.5}) \tag{11}$$

as $T \to \infty$ because $E[\hat{\mathbf{\Gamma}}(j)]$ is bounded for all T. That is, although the sample autocovariance is asymptotically unbiased, the negative-definite discrepancy between the expectation of the sample autocovariance and the population autocovariance shrinks at a rate slower than the usual $T^{-0.5}$ rate. The empirical implication is that the sample autocovariance of the highly persistent long memory process $\mathbf{x}_t$ with $d_i > 0.25$ tends to underestimate the population autocovariance when the sample size is small, a conclusion first explicitly suggested by Newbold and Agiakloglou (1993). Also note that because the bias in the sample autocovariance $\hat{\mathbf{\Gamma}}(j)$ is independent of the lag j, Hosking (1996, Theorem 3) also points out the possibility that the difference between two sample autocovariances $\hat{\mathbf{\Gamma}}(j) - \hat{\mathbf{\Gamma}}(j')$, for $j \neq j'$, may contain no bias at all. Hosking (1996, Theorem 5) goes one step further to prove that when the memory parameter is greater than 0.25 so that the sample autocovariances have nonstandard limiting distributions, the difference between two sample autocovariances can nevertheless have the usual normal asymptotic distribution. This interesting result in turn has been employed by Chung and Schmidt (1996) to design a strategy for estimating the ARFIMA model. Here, without going into detail, we simply say that Hosking's result on the difference in sample autocovariances can be naturally generalized to the multivariate case.

The downward bias in the sample autocovariance has profound impacts on the estimation of the long memory model. It is widely reported in the ARFIMA literature (e.g., Chung and Baillie, 1993; Cheung and Diebold, 1994; Chung and Schmidt, 1996) that when the mean of the ARFIMA process is unknown and the fractional differencing parameter is close to 0.5, then both the exact

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

and the approximate maximum likelihood estimation (MLE) of the ARFIMA model under the Gaussian assumption yield substantial biases in small-sample applications. Because the MLE under the Gaussian assumption relies completely on the first two moments of the process, it seems obvious that the bias in the MLE can be traced back to the bias in the sample autocovariances as suggested here. Incidentally, Chung and Baillie (1993) and Cheung and Diebold (1994) also report that the bias in the MLE can be greatly reduced when the mean of the process is known. This finding too has its root in the sample second moments. To see this let us replace the sample mean $\bar{\mathbf{x}}$ by the population mean $\boldsymbol{\mu}$ in the sample autocovariance formula. Theorem 2 implies that the expectation of such sample autocovariance will be approximately equal to the population autocovariance and no slowly converging asymptotic bias will occur. In other words, the downward bias in the sample autocovariance, and also the bias in the MLE, is to some extent due to the operation of estimating the unknown mean, which results in the extra term $-\mathbf{B}_d(1)\mathbf{B}_d(1)'$ in the limiting process (see expression (10)).

4. THE LINEAR REGRESSION MODEL

The theory we develop in the previous sections can be used to establish the asymptotic distributions of the OLS estimators in a multiple linear regression with long memory processes. Consider the multiple linear regression model of a univariate process y_t on a K-dimensional process $\mathbf{x}_t$ defined by (1) and (2): $y_t = \alpha + \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t$, where the unobservable disturbance ε_t is assumed to be a univariate process defined by

$$\varepsilon_t = \sum_{j=0}^{\infty} \psi_j^* b_{t-j}, \quad \text{where } \psi_j^* \sim \frac{1}{\Gamma(d^*)} j^{d^*-1}\pi^*, \tag{12}$$

where d^* lies between 0 and 0.5, π^* is a nonzero constant, and the innovations b_t are martingale differences with respect to the same sequence of σ-fields $\{\mathcal{F}_t\}$ that are associated with $\mathbf{a}_t$. Moreover, let $E(b_t^2 \| \mathcal{F}_{t-1}) = \sigma_b^2$, a.s. We also assume that the disturbance ε_t is uncorrelated with the regressors $\mathbf{x}_t$ so that their covariance $\mathbf{m}_{x\varepsilon} = \mathbf{0}$. Given these specifications, we can then use the results from previous sections to derive the limiting distributions for the OLS estimators of $\boldsymbol{\beta}$, α, and σ_ϵ^2 (the variance of the disturbance), which can be written, respectively, as $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \hat{\mathbf{M}}_{xx}^{-1}\hat{\mathbf{m}}_{x\varepsilon}$, $\hat{\alpha} = \alpha + \bar{\varepsilon} + \bar{\mathbf{x}}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, and $s^2 = \hat{m}_{\varepsilon\varepsilon} - \hat{\mathbf{m}}_{x\varepsilon}'\hat{\mathbf{M}}_{xx}^{-1}\hat{\mathbf{m}}_{x\varepsilon}$. Here the $K \times K$ matrix $\hat{\mathbf{M}}_{xx}$ and the scalar $\hat{m}_{\varepsilon\varepsilon}$ are the sample variances of $\mathbf{x}_t$ and ε_t, respectively, and the K-vector $\hat{\mathbf{m}}_{x\varepsilon}$ is the sample covariance between $\mathbf{x}_t$ and ε_t.

THEOREM 3. *Given that for some* $\epsilon > 0$, $\sup_t E(|a_{it}b_t|^{2+\epsilon} \| \mathcal{F}_{t-1}) < \infty$, *a.s., for all* $i = 1, \ldots, K$, *let us assume that the first regressor* x_{1t} *has a memory parameter* d_1 *that is strictly greater than* d_i*'s associated with all other* $K - 1$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

regressors and that x_{1t} is correlated with all other regressors. If $d_1 + d^* > 0.5$, where d^* is the memory parameter associated with the disturbance ε_t, then $T^{1-d_1-d^*}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \mathbf{c}\cdot z$, where $\mathbf{c}$ is the first column of the inverse of the variance of $\mathbf{x}_t$, denoted by $\mathbf{M}_{xx}$, and is a K-vector of constants, whereas $z \equiv R^{(1*)}(1) - B_{d_1}(1)B_{d^*}(1)$ is a univariate random variable. Here $R^{(1*)}(s)$ is defined by

$$R^{(1*)}(s) \equiv \frac{1}{\Gamma(d_1)\Gamma(d^*)} \times \int_{-\infty}^{s}\int_{-\infty}^{s}\left[\int_0^s (t-r)_+^{d_1-1}(t-r^*)_+^{d^*-1}\,dt\right] dB_0^{(1)}(r)\,dB_0^*(r^*).$$

Also, $B_{d_1}(s)$ is a fractional Brownian motion defined by

$$B_{d_1}(s) \equiv \frac{1}{\Gamma(d_1+1)}\left\{\int_0^s (s-t)^{d_1} dB_0^{(1)}(t) + \int_{-\infty}^0 [(s-t)^{d_1} - (-t)^{d_1}]dB_0^{(1)}(t)\right\},$$

and $B_{d^*}(s)$ is a fractional Brownian motion defined by

$$B_{d^*}(s) \equiv \frac{1}{\Gamma(d^*+1)}\left\{\int_0^s (s-t)^{d^*} dB_0^*(t) + \int_{-\infty}^0 [(s-t)^{d^*} - (-t)^{d^*}]dB_0^*(t)\right\}.$$

In the last three definitions $B_0^{(1)}(t)$ is the first element of K-dimensional Brownian motion with the variance-covariance matrix $\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'$, and $B_0^*(t)$ is a one-dimensional Brownian motion with the variance $\pi^{*2}\sigma_b^2$. The terms $B_0^{(1)}(t)$ and $B_0^*(t)$ are uncorrelated.

The convergence rate $T^{d_1+d^*-1}$ of the OLS slope estimator $\hat{\boldsymbol{\beta}}$ is slower than the usual $T^{-0.5}$ rate (because $d_1 + d^* > 0.5$), and the limit is not a normal random vector but is proportional to a peculiar random variable z. We note that z has a zero mean (because $B_{d_1}(1)$ and $B_{d^*}(1)$ are uncorrelated) and a finite variance. Consequently, the mean of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is $\boldsymbol{\beta}$, whereas $\mathrm{Var}(\hat{\boldsymbol{\beta}})$ can be approximated by $T^{2(d_1+d^*-1)}\mathrm{Var}(z)\cdot\mathbf{cc}'$.

It is quite surprising to see that such a nonstandard result can come from a linear regression model in which all the variables are stationary. What is required here is simply that one of the regressors is highly persistent with a large memory parameter d_1, whereas the disturbance exhibits some long memory with a memory parameter d^* just large enough to make $d_1 + d^* > 0.5$. We note that other regressors may have various degrees of long memory or no long memory at all but the OLS slope estimators $\hat{\beta}_k, k = 1,\ldots,K$, all have nonstandard limits. The intuition for all $\hat{\beta}_k, k = 1,\ldots,K$, to have the same convergence rate $T^{d_1+d^*-1}$ is as follows: as a result of the correlations between x_{1t} and other regressors, each $\hat{\beta}_k - \beta_k$ is a linear combination of K random variables that converge to zero at different $T^{d_1+d^*-1}$ rates. Consequently, it is the random variable with the slowest convergence rate $T^{d_1+d^*-1}(>T^{d_i+d^*-1},$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

for $i \neq 1$) that dominates the asymptotic behavior of $\hat{\beta}_k - \beta_k$, for all $k = 1,\ldots,K$. It is readily seen that if regressors are uncorrelated with each other, then $\hat{\beta}_k - \beta_k$ will converge at different rates.

When none of $d_i + d^*$ is greater than 0.5, then we return to the standard asymptotic theory as indicated in the following corollary, which is a straightforward consequence of part 2 of Corollary 1.

COROLLARY 2. *Given that for some* $\epsilon > 0$, $\sup_t E(|a_{it}b_t|^{2+\epsilon}\|\mathcal{F}_{t-1}) < \infty$, *a.s., for all* $i = 1,\ldots,K$, *if* $d_i + d^* < 0.5$, *for all* $i = 1,\ldots,K$, *then*

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{M}_{xx}^{-1}\mathbf{W}_{x\varepsilon}\mathbf{M}_{xx}^{-1}),$$

where $\mathbf{M}_{xx}$ *is the variance of* $\mathbf{x}_t$ *and* $\mathbf{W}_{x\varepsilon}$ *is the asymptotic variance of* $\hat{\mathbf{m}}_{x\varepsilon}$.

We now present the limits of the other two OLS estimators.

COROLLARY 3. *If the assumptions of Theorem 3 hold, in particular, if there is at least one of* d_i, $i = 1,\ldots,K$, *say* d_1, *such that* $d_1 + d^* > 0.5$, *then*

$$T^{0.5-d^*}(\hat{\alpha} - \alpha) \Rightarrow B_{d^*}(1) \stackrel{d}{=} \mathcal{N}(0, \lambda^*\pi^{*2}\sigma_b^2),$$

$$\text{with } \lambda^* = \frac{\Gamma(1-2d^*)}{(1+2d^*)\Gamma(1+d^*)\Gamma(1-d^*)},$$

and $s^2 \xrightarrow{p} \sigma_\varepsilon^2$, *where the fractional Brownian motion* $B_{d^*}(s)$ *is defined in Theorem* 3.

It is also easy to show that, under the assumptions of Theorem 3, the coefficient of determination $R^2 \xrightarrow{p} 1 - \sigma_\varepsilon^2/\mathrm{Var}(y_t)$ and the Durbin–Watson statistic $\mathrm{DW} \xrightarrow{p} 2 - 2\rho_\varepsilon(1)$, where $\rho_\varepsilon(1)$ is the autocorrelation of ε_t at lag one, just as in the standard linear regression model.

For any given finite constants $\beta_\circ$ and $\alpha_\circ$, the t-statistics for testing $\beta_i = \beta_\circ$ and $\alpha = \alpha_\circ$ are $t_i \equiv T^{0.5}(\hat{\beta}_i - \beta_\circ)/\sqrt{s^2\hat{\mathbf{M}}_{xx}^{ii}}$ and $t_\alpha \equiv T^{0.5}(\hat{\alpha} - \alpha_\circ)/\sqrt{s^2(1 - \bar{\mathbf{x}}'\hat{\mathbf{M}}_{xx}^{-1}\bar{\mathbf{x}})}$, respectively, where $\hat{\mathbf{M}}_{xx}^{ii}$ is the ith diagonal element of the matrix $\hat{\mathbf{M}}_{xx}^{-1}$. Also, to test the general linear hypothesis on the entire slope coefficient $\mathbf{H}\boldsymbol{\beta} = \mathbf{g}$, where $\mathbf{H}$ is a $J \times K$ matrix of known constants of full row rank and $\mathbf{g}$ is a J-vector of known constants, the standard F-test statistic is $F = T\cdot(\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{g})'(\mathbf{H}\hat{\mathbf{M}}_{xx}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{g})/(Js^2)$.

COROLLARY 4. *Suppose the assumptions of Theorem 3 hold, in particular, if there is at least one of* d_i, $i = 1,\ldots,K$, *say,* d_1, *such that* $d_1 + d^* > 0.5$, *then*

$$T^{0.5-d_1-d^*}t_i \Rightarrow \frac{\mathbf{M}_{xx}^{1i}}{\sqrt{\sigma_\varepsilon^2\cdot\mathbf{M}_{xx}^{ii}}}\,z, \qquad T^{-d^*}t_\alpha \Rightarrow \frac{1}{\sqrt{\sigma_\varepsilon^2(1 - \boldsymbol{\mu}'\mathbf{M}_{xx}^{-1}\boldsymbol{\mu})}}\,B_{d^*}(1),$$

$$T^{1-2d_1-2d^*}F \Rightarrow \frac{\mathbf{c}'\mathbf{H}'(\mathbf{H}\mathbf{M}_{xx}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{c}/J}{\sigma_\varepsilon^2}\,z^2,$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

under their respective null hypotheses, where $\mathbf{M}_{xx}^{1i}$ *is the* $(1,i)$*th element of* $\mathbf{M}_{xx}^{-1}$, $\mathbf{M}_{xx}^{ii}$ *is the ith diagonal element of* $\mathbf{M}_{xx}^{-1}$, $\boldsymbol{\mu}$ *is the mean of the regressor* $\mathbf{x}_t$, *and* σ_ε^2 *is the variance of the disturbance* ε_t, *whereas the random variable* z, *the fractional Brownian motion* $B_{d^*}(s)$, *and the vector* $\mathbf{c}$ *are all defined in Theorem 3.*

Because $0.5 - d_1 - d^* < 0$, $-d^* < 0$, and $1 - 2d_1 - 2d^* < 0$, the t- and F-test statistics t_i, t_α, and F are all divergent. The consequence is that when these test statistics are employed, the rejection probability grows as the sample size so that the size distortion of these tests deteriorates as the sample size increases no matter whether the null hypothesis of the respective test is true or not. The reason for this divergence result is as follows. Although the OLS estimators are consistent, their convergence rates are slower than the $T^{-0.5}$ rate if at least one $d_i + d^* > 0.5$. In contrast, the corresponding standard errors given by the usual formulas converge at the usual $T^{-0.5}$ rate. As a result, the t-statistics (and also the F-statistic) all become divergent, and the tests collapse. Note that for the case where $d_i + d^* < 0.5$, for all $i = 1, \ldots, K$, the t- and F-statistics will converge, but not to the standard normal or χ^2 distribution as the standard theory of the OLS estimation suggests. This is simply because the disturbance term is autocorrelated.

It is interesting to compare our results with those of Phillips and Park (1988), Phillips and Durlauf (1986), and Park and Phillips (1988, 1989), who consider a linear regression model in which regressors are integrated processes whereas the disturbance is a weakly dependent process. It is found that in their model, although the OLS estimators converge to nonnormal random variables at the fast T^{-1} rate, the t- and F-test statistics still converge. Moreover, when regressors and the disturbance are assumed to be independent, the limits of the t- and F-statistics can even be the usual standard normal and χ^2 random variables, respectively, as shown by Phillips and Park (1988). The reason why their results are so different from ours lies in the convergence rates of the sample variance $\hat{\mathbf{M}}_{xx}$ and the sample covariance $\hat{\mathbf{m}}_{x\varepsilon}$.

We first note that, because in our model regressors are assumed to be stationary and, more important, ergodic, $\hat{\mathbf{M}}_{xx}$ converges in probability to the population variance $\mathbf{M}_{xx}$. But in Phillips and his coauthors' framework regressors are nonstationary and nonergodic integrated (or near-integrated) processes so that the population variance $\mathbf{M}_{xx}$ does not even exist. Moreover, a closer inspection of the formulas of the t- and F-statistics reveals that the convergence rates of these statistics are determined by the term $T^{0.5}\hat{\mathbf{M}}_{xx}^{-1/2}\hat{\mathbf{m}}_{x\varepsilon} \equiv \mathbf{v}$. In our model, because $\hat{\mathbf{M}}_{xx}^{-1/2}$ converges in probability to $\mathbf{M}_{xx}^{-1/2}$, the convergence rate of $\mathbf{v}$ and therefore the convergence rates of the t- and F-statistics are completely determined by that of $T^{0.5}\hat{\mathbf{m}}_{x\varepsilon}$ (or more precisely, the slowest convergence rate among the K rates of its K elements). But Phillips and his coauthors are able to show that in their model both $(T^{0.5}\hat{\mathbf{M}}_{xx}^{-1/2})^{-1}$ and $\hat{\mathbf{m}}_{x\varepsilon}$ converge at the same rate so that the t- and F-statistics must converge. The conclusion is that, even though

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

regressors are integrated and completely persistent, the t- and F-tests can work as usual so long as the disturbance does not have long memory. The success of the t- and F-tests in Phillips and his coauthors' model lies on the sharp difference in the persistence levels of the regressors and of the disturbance.

Finally, we note that when the true value of the regression coefficient $\boldsymbol{\beta}$ is zero, the hypothesis of insignificant slope coefficient estimates will be overly rejected by the conventional t- and F-tests. This type of spurious effect is similar to the one considered by Granger and Newbold (1974) and Phillips (1986) in regressing an integrated process on another uncorrelated integrated process. Our results here indicate that the potential spurious effects in t- and F-tests may be more pervasive because they can occur even in a regression model of stationary processes. A more detailed analysis of the spurious effect is presented in Tsay and Chung (2000).

5. CONCLUSION

We have developed a unified theory for the first two sample moments of a stationary multivariate long memory process without the Gaussian assumption. One interesting conclusion from this theory is that the convergence rates and the limits (the fractional Brownian motion, the Rosenblatt process, etc.) all depend intrinsically on the degree of long memory in the process. The theory for the sample moments is then applied to the multiple linear regression model. It is found that, even though all the regressors and the disturbance are stationary, the joint long memory in one regressor and in the disturbance can wreck the $\sqrt{T}$-convergence theory of the OLS estimation.

It should be pointed out that the functional central limit theorems we establish are not only necessary for the study of the multiple linear regression model but also interesting for their own sake. From these results we are able to illustrate that the sample autocovariances of a highly persistent long memory process tend to underestimate their population counterparts, a problem that does not seem to happen to the short memory ARMA processes but is a serious issue in the long memory literature because it has profound repercussions on the estimation accuracy when long memory models are fitted to the data.

One issue we have not pursued in this paper is how to apply a generalized least squares (GLS) estimation to utilize the autocovariances of the disturbance to improve estimation efficiency. We note that in our approach only the long run behavior of the disturbance is specified. To implement the GLS estimation, we must make additional assumptions on the disturbance. In this regard Robinson and Hidalgo (1997) have proposed a class of GLS estimators that rely on a more precise specification of the spectral density of the disturbance. The main conclusion from their study is that, in contrast to the possibly slower convergence rate of the OLS estimator, their GLS estimators are $\sqrt{T}$-consistent under general regularity conditions. It appears that the GLS estimator, if implemented correctly, will follow the standard asymptotic theory. To understand this, let us

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

consider a simple example. Suppose in our model the disturbance ε_t is fully specified as an ARFIMA$(0, d^*, 0)$ process, whereas x_{1t} is the only regressor that has long memory and is also assumed to be an ARFIMA$(0, d_1, 0)$ process such that $d^* + d_1 > 0.5$. Theorem 3 says that the OLS slope estimator in such a case will converge slowly at the $T^{d_1+d^*-1}$ rate. Now let us apply the fractional differencing operator $(1 - B)^{-d^*}$ to the entire regression model. Then the resulting disturbance will be a martingale difference, whereas the regressor that is transformed from x_{1t} becomes an ARFIMA$(0, d_1 - d^*, 0)$ process. Because $d_1 - d^*$ is necessarily smaller than 0.5, the OLS estimator of the transformed model, which is nothing but the GLS estimator of the original model, is $\sqrt{T}$-consistent. Obviously, a correct GLS strategy should help in solving the slow convergence problem with the OLS estimator.

REFERENCES

Adenstedt, R.K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095–1107.

Avram, F. & M.S. Taqqu (1987) Noncentral limit theorems and Appell polynomials. *Annals of Probability* 15, 767–775.

Cheung, Y.-W. & F.X. Diebold (1994) On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics* 62, 301–316.

Chung, C.-F. (1999) Calculating and Analyzing Impulse Responses for the Vector ARFIMA Model. Preprint.

Chung, C.-F. & R.T. Baillie (1993) Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models. *Empirical Economics* 18, 791–806.

Chung, C.-F. & P. Schmidt (1996) The Minimum Distance Estimator for Fractionally Integrated ARMA Models. Preprint.

Dahlhaus, R. (1995) Efficient location and regression estimation for long range dependent regression models. *Annals of Statistics* 23, 1029–1047.

Davidson, J. & R. de Jong (1998) The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals II: Fractionally Integrated Processes. Preprint.

Davydov, Y.A. (1970) The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15, 487–489.

Fox, R. & M.S. Taqqu (1985) Non-central limit theorems for quadratic forms in random variables having long-range dependence. *Annals of Probability* 13, 428–446.

Fox, R. & M.S. Taqqu (1987) Multiple stochastic integrals with dependent integrators. *Journal of Multivariate Analysis* 21, 105–127.

Giraitis, L., H.L. Koul, & D. Surgailis (1996) Asymptotic normality of regression estimators with long memory errors. *Statistics and Probability Letters* 29, 317–335.

Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14, 227–238.

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

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

Granger, C.W.J. & P. Newbold (1974) Spurious regression in econometrics. *Journal of Econometrics* 2, 111–120.

Hannan, E.J. (1976) The asymptotic distribution of serial covariances. *Annals of Statistics* 4, 396–399.

Helland, I.S. (1982) Central limit theorems for martingales with discrete or continuous time. *Scandinavia Journal of Statistics* 9, 79–94.

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

Hosking, J.R.M. (1981) Fractional differencing. *Biometrika* 68, 165–176.
Hosking, J.R.M. (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics* 73, 261–284.
Koul, H.L. (1992) M-estimators in linear models with long range dependent errors. *Statistics and Probability Letters* 14, 153–164.
Koul, H.L. & K. Mukherjee (1993) Asymptotics of R-, MD-, and LAD-estimators in linear regression models with long range dependent errors. *Probability Theory and Related Fields* 95, 535–553.
Mandelbrot, B.B. & J.W. Van Ness (1968) Fractional Brownian motions, fractional noises, and applications. *SIAM Review* 10, 422–437.
Newbold, P. & C. Agiakloglou (1993) Bias in the sample autocorrelations of fractional noise. *Biometrika* 80, 698–702.
Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory* 4, 468–497.
Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* 5, 95–131.
Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.
Phillips, P.C.B. & S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.
Phillips, P.C.B. & J.Y. Park (1988) Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association* 83, 111–115.
Robinson, P.M. (1994) Time series with strong dependence. In C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress.* Cambridge: Cambridge University Press.
Robinson, P.M. & F.J. Hidalgo (1997) Time series regression with long range dependence. *Annals of Statistics* 25, 77–104.
Rosenblatt, M. (1961) Independence and dependence. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 431–443. Berkeley: University of California Press.
Rosenblatt, M. (1979) Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 49, 125–132.
Sowell, F.B. (1990) The fractional unit root distribution. *Econometrica* 58, 495–505.
Sowell, F.B. (1992a) Modelling long-run behavior with the fractional ARIMA model. *Journal of Monetary Economics* 29, 277–302.
Sowell, F.B. (1992b) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53, 165–188.
Tsay, W.-J. & C.-F. Chung (2000) The spurious regression of fractionally integrated processes. *Journal of Econometrics* 96, 155–182.
Yajima, Y. (1988) On estimation of a regression model with long-memory stationary errors. *Annals of Statistics* 16, 791–807.
Yajima, Y. (1991) Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Annals of Statistics* 19, 158–177.

APPENDIX

The proof of Lemma 1 is straightforward and is omitted here. To prove Lemma 2, we need the following lemma.

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

LEMMA A.1. *Given the assumptions of Lemma 2, we have*

$$\mathbf{D}(j^{0.5-d})\left(\sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\Sigma}_a\boldsymbol{\Psi}'_{k+j}\right)\mathbf{D}(j^{0.5-d})$$

$$= \mathbf{D}(j^{0.5-d})\left\{\sum_{k=1}^{\infty}\mathbf{D}\left[\frac{1}{\Gamma(d)}k^{d-1}\right]\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d)}(k+j)^{d-1}\right]\right\}\mathbf{D}(j^{0.5-d})$$

$$+ o(1), \quad as\, j \to \infty.$$

Proof of Lemma A.1. Given that $\psi'_{i,j}$ and $\boldsymbol{\pi}'_i$ are the ith rows of $\boldsymbol{\Psi}_j$ and $\boldsymbol{\Pi}$, respectively, it is sufficient to show that

$$j^{1-d_i-d_\ell}\sum_{k=0}^{\infty}\boldsymbol{\psi}'_{i,k}\boldsymbol{\Sigma}_a\boldsymbol{\psi}_{\ell,k+j} = j^{1-d_i-d_\ell}\frac{1}{\Gamma(d_i)\Gamma(d_\ell)}\boldsymbol{\pi}'_i\boldsymbol{\Sigma}_a\boldsymbol{\pi}_\ell\sum_{k=1}^{\infty}k^{d_i-1}(k+j)^{d_\ell-1}$$

$$+ o(1), \quad \text{as } j \to \infty,$$

for all $i, \ell = 1, \ldots, K$. Let $\psi^\circ_{i,k}$ be a scalar such that $\boldsymbol{\psi}_{i,k} \equiv \psi^\circ_{i,k}\cdot\Gamma(d_i)^{-1}\boldsymbol{\pi}_i$ and $\psi^\circ_{i,k} \sim k^{d_i-1}$. It is sufficient to show that

$$j^{1-d_i-d_\ell}\sum_{k=0}^{\infty}\psi^\circ_{i,k}\psi^\circ_{\ell,k+j} = j^{1-d_i-d_\ell}\sum_{k=1}^{\infty}k^{d_i-1}(k+j)^{d_\ell-1} + o(1), \quad \text{as } j \to \infty, \tag{A.1}$$

for all $i, \ell = 1, \ldots, K$. It is possible to find a sufficiently large but finite constant A such that $|\psi^\circ_{i,k}| < Ak^{d_i-1}$ and $|\psi^\circ_{\ell,k+j}| < A\cdot(k+j)^{d_\ell-1}$. Note that k^{d_i-1} is a decreasing sequence of positive numbers. Furthermore, for any $\varepsilon > 0$ there is an integer N such that $(1-\varepsilon)k^{d_i-1} < \psi^\circ_{i,k} < (1+\varepsilon)k^{d_i-1}$ and $(1-\varepsilon)(k+j)^{d_\ell-1} < \psi^\circ_{\ell,k+j} < (1+\varepsilon)\times(k+j)^{d_\ell-1}$, for all $k \geq N$. Consequently, for any $\varepsilon > 0$ we have

$$j^{1-d_i-d_\ell}\left|\sum_{k=0}^{\infty}\psi^\circ_{i,k}\psi^\circ_{\ell,k+j} - \sum_{k=1}^{\infty}k^{d_1-1}(k+j)^{d_\ell-1}\right|$$

$$\leq j^{1-d_i-d_\ell}\left\{\sum_{k=0}^{N-1}|\psi^\circ_{i,k}\psi^\circ_{\ell,k+j}| + \sum_{k=1}^{N-1}k^{d_i-1}(k+j)^{d_\ell-1}\right.$$

$$\left. + \sum_{k=N}^{\infty}|\psi^\circ_{i,k}\psi^\circ_{\ell,k+j} - k^{d_i-1}(k+j)^{d_\ell-1}|\right\}$$

$$\leq j^{1-d_i-d_\ell}\cdot N(A^2+1)\cdot j^{d_\ell-1} + (2\varepsilon+\varepsilon^2)\cdot j^{1-d_i-d_\ell}\sum_{k=N}^{\infty}k^{d_i-1}(k+j)^{d_\ell-1}.$$

On the right-hand side of the preceding inequality, the first term converges to 0 as $j \to \infty$, whereas in the second term $j^{1-d_i-d_\ell}\sum_{k=N}^{\infty}k^{d_i-1}(k+j)^{d_\ell-1}$ is bounded. These results imply (A.1). ■

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

Proof of Lemma 2. From (1), (2), and Lemma A.1, we have for the autocovariances $\mathbf{\Gamma}(j) \equiv \mathrm{Cov}(\mathbf{x}_t, \mathbf{x}_{t+j})$ that

$$\begin{aligned}
&\mathbf{D}(j^{0.5-d})\cdot\mathrm{Cov}(\mathbf{x}_t,\mathbf{x}_{t+j})\cdot\mathbf{D}(j^{0.5-d}) \\
&\quad= \mathbf{D}(j^{0.5-d})\cdot\mathrm{Cov}\left(\sum_{k=0}^{\infty}\mathbf{\Psi}_k\mathbf{a}_{t-k},\sum_{k=-j}^{\infty}\mathbf{\Psi}_{k+j}\mathbf{a}_{t-k}\right)\cdot\mathbf{D}(j^{0.5-d}) \\
&\quad= \mathbf{D}(j^{0.5-d})\left(\sum_{k=0}^{\infty}\mathbf{\Psi}_k\mathbf{\Sigma}_a\mathbf{\Psi}_{k+j}'\right)\mathbf{D}(j^{0.5-d}) \\
&\quad= \mathbf{D}(j^{0.5-d})\left\{\sum_{k=1}^{\infty}\mathbf{D}\left[\frac{1}{\Gamma(d)}k^{d-1}\right]\mathbf{\Pi}\mathbf{\Sigma}_a\mathbf{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d)}(k+j)^{d-1}\right]\right\}\mathbf{D}(j^{0.5-d})+o(1) \\
&\quad= \frac{1}{j}\sum_{k=1}^{\infty}\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{k}{j}\right)^{d-1}\right]\mathbf{\Pi}\mathbf{\Sigma}_a\mathbf{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{k}{j}+1\right)^{d-1}\right]+o(1) \\
&\quad\to \int_0^{\infty}\mathbf{D}\left[\frac{1}{\Gamma(d)}z^{d-1}\right]\mathbf{\Pi}\mathbf{\Sigma}_a\mathbf{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d)}(z+1)^{d-1}\right]dz \quad \text{as } j\to\infty \\
&\quad= \left[(\boldsymbol{\pi}_i'\mathbf{\Sigma}_a\boldsymbol{\pi}_k)\cdot\frac{1}{\Gamma(d_i)\Gamma(d_k)}\int_0^{\infty}z^{d_i-1}(z+1)^{d_k-1}\,dz,\quad i,k=1,\ldots,K\right]. \qquad \blacksquare
\end{aligned}$$

In the following proofs convergence is all taken as the sample size $T\to\infty$. To prove Theorem 1, we need the following lemma.

LEMMA A.2. *Given the assumptions of Theorem 1, we have that as* $T\to\infty$

$$\mathbf{D}(T^{-0.5-d})\sum_{t=1}^{T}\left(\sum_{j=0}^{t}\mathbf{\Psi}_j\right)\mathbf{a}_t = \mathbf{D}(T^{-0.5-d})\sum_{t=1}^{T}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}t^d\right]\mathbf{\Pi}\mathbf{a}_t+o_p(1) \tag{A.2}$$

and

$$\begin{aligned}
&\mathbf{D}(T^{-0.5-d})\sum_{t=-\infty}^{0}\left(\sum_{j=0}^{T-t}\mathbf{\Psi}_j-\sum_{j=0}^{-t}\mathbf{\Psi}_j\right)\mathbf{a}_t \\
&\quad= \mathbf{D}(T^{-0.5-d})\sum_{t=-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}[(T-t)^d-(-t)^d]\right\}\mathbf{\Pi}\mathbf{a}_t+o_p(1).
\end{aligned} \tag{A.3}$$

Proof of Lemma A.2. Given that $\boldsymbol{\psi}_{i,j}'$ and $\boldsymbol{\pi}_i'$ are the ith rows of $\mathbf{\Psi}_j$ and $\mathbf{\Pi}$, respectively, for the proof of (A.2) it is sufficient to show

$$T^{-0.5-d_i}\sum_{t=1}^{T}\sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t = T^{-0.5-d_i}\sum_{t=1}^{T}\frac{1}{\Gamma(d_i+1)}t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t+o_p(1),\quad \text{for } i=1,\ldots,K. \tag{A.4}$$

From Lemma 1 we have $\sum_{j=0}^{t}\mathbf{\Psi}_j\sim\mathbf{D}[\Gamma(d+1)^{-1}t^d]\mathbf{\Pi}$, which implies that we can find a sufficiently large but finite constant A such that

$$\left|\sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t\right| < A\cdot\frac{1}{\Gamma(d_i+1)}t^{d_i}\cdot|\boldsymbol{\pi}_i'\mathbf{a}_t|.$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

Furthermore, for any $\varepsilon > 0$ there is an integer N such that

$$(1-\varepsilon)\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t < \sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t < (1+\varepsilon)\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t, \quad \text{for all } t > N.$$

Consequently, for any $\varepsilon > 0$ and a sufficiently large T that is greater than N, we have

$$\begin{aligned}
&T^{-0.5-d_i}\left|\sum_{t=1}^{T}\sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t - \sum_{t=1}^{T}\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t\right| \\
&\quad\le T^{-0.5-d_i}\left\{\sum_{t=1}^{N}\left(\left|\sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t\right| + \frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot|\boldsymbol{\pi}_i'\mathbf{a}_t|\right)\right. \\
&\qquad\qquad \left. + \sum_{t=N+1}^{T}\left|\sum_{j=0}^{t}\boldsymbol{\psi}_{i,j}'\mathbf{a}_t - \frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t\right|\right\} \\
&\quad\le T^{-0.5-d_i}(A+1)\sum_{t=1}^{N}\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot|\boldsymbol{\pi}_i'\mathbf{a}_t| + \varepsilon\cdot T^{-0.5-d_i}\sum_{t=N+1}^{T}\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t.
\end{aligned}$$

The first term on the right-hand side converges to 0, a.s., when $T \to \infty$. Also note that

$$\frac{1}{T}\sum_{t=N+1}^{T}\frac{1}{\Gamma(d_i+1)^2}\left(\frac{t}{T}\right)^{2d_i} < \frac{1}{\Gamma(d_i+1)^2}\int_0^1 t^{2d_i}\,dt < \infty.$$

This result together with the assumption $\sup_t E(|a_{it}|^{2+\epsilon}\|\mathcal{F}_{t-1}) < \infty$, a.s. implies that the (conditional) variance of the sum

$$T^{-0.5-d_i}\sum_{t=N+1}^{T}\frac{1}{\Gamma(d_i+1)}\,t^{d_i}\cdot\boldsymbol{\pi}_i'\mathbf{a}_t$$

is bounded because $\mathbf{a}_t$ are (conditionally) uncorrelated and the preceding (conditional) variance equals the weighted sum of (conditional) variances of $\mathbf{a}_t$ whose maximum is bounded. This result in turn implies that the preceding sum is bounded in probability based on Chebyshev inequality. This completes the proof of (A.4). Equation (A.3) can be proved by a similar argument. ■

Proof of Theorem 1. From (1) we can write the partial sum of $\mathbf{x}_t$ as follows:

$$\begin{aligned}
&\mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}(\mathbf{x}_t - \boldsymbol{\mu}) \\
&\quad= \mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}\left(\sum_{j=0}^{t-1}\boldsymbol{\Psi}_j\mathbf{a}_{t-j} + \sum_{j=t}^{\infty}\boldsymbol{\Psi}_j\mathbf{a}_{t-j}\right) \\
&\quad= \mathbf{D}(T^{-0.5-d})\left(\sum_{t=1}^{[Ts]}\sum_{j=0}^{t-1}\boldsymbol{\Psi}_j\mathbf{a}_{t-j} + \sum_{t=1}^{[Ts]}\sum_{j=t}^{\infty}\boldsymbol{\Psi}_j\mathbf{a}_{t-j}\right) \\
&\quad= \mathbf{D}(T^{-0.5-d})\left[\sum_{t=1}^{[Ts]}\left(\sum_{j=0}^{[Ts]-t}\boldsymbol{\Psi}_j\right)\mathbf{a}_t + \sum_{t=-\infty}^{0}\left(\sum_{j=0}^{[Ts]-t}\boldsymbol{\Psi}_j - \sum_{j=0}^{-t}\boldsymbol{\Psi}_j\right)\mathbf{a}_t\right].
\end{aligned}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

Lemma A.2 thus implies

$$\begin{aligned}
&\mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}(\mathbf{x}_t-\boldsymbol{\mu})\\
&\quad=\mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}([Ts]-t)^d\right]\boldsymbol{\Pi}\mathbf{a}_t\\
&\qquad+\mathbf{D}(T^{-0.5-d})\sum_{t=-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}[([Ts]-t)^d-(-t)^d]\right\}\boldsymbol{\Pi}\mathbf{a}_t+o_p(1)\\
&\quad=\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d\right]\boldsymbol{\Pi}\mathbf{a}_t\\
&\qquad+\frac{1}{\sqrt{T}}\sum_{t=-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}\left[\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d-\left(-\frac{t}{T}\right)^d\right]\right\}\boldsymbol{\Pi}\mathbf{a}_t+o_p(1).
\end{aligned}$$

The two summands are both martingale differences with respect to an increasing sequence of σ-fields $\{\mathcal{F}_t\}$. We now consider the sum of their conditional variances. For the first term we have the following convergence:

$$\begin{aligned}
&\sum_{t=1}^{[Ts]}\mathrm{Var}\left\{\frac{1}{\sqrt{T}}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d\right]\boldsymbol{\Pi}\mathbf{a}_t\middle\|\mathcal{F}_{t-1}\right\}\\
&\quad\overset{\text{a.s.}}{=}\frac{1}{T}\sum_{t=1}^{[Ts]}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d\right]\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d+1)}\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d\right]\\
&\quad\to\int_0^s\mathbf{D}\left[\frac{1}{\Gamma(d+1)}(s-t)^d\right]\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'\mathbf{D}\left[\frac{1}{\Gamma(d+1)}(s-t)^d\right]dt<\infty.
\end{aligned}$$

Here the limit is a $K\times K$ matrix of integrals of which the (i,j)th element is

$$(\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_j)\cdot\frac{1}{\Gamma(d_i+1)\Gamma(d_j+1)}\int_0^s(s-t)^{d_i}(s-t)^{d_j}\,dt<\infty. \tag{A.5}$$

Similarly, for the second term we have the following convergence result:

$$\begin{aligned}
&\sum_{t=-\infty}^{0}\mathrm{Var}\left\{\frac{1}{\sqrt{T}}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}\left[\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d-\left(-\frac{t}{T}\right)^d\right]\right\}\boldsymbol{\Pi}\mathbf{a}_t\middle\|\mathcal{F}_{t-1}\right\}\\
&\quad\overset{\text{a.s.}}{=}\frac{1}{T}\sum_{t=-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}\left[\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d-\left(-\frac{t}{T}\right)^d\right]\right\}\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'\\
&\qquad\times\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}\left[\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d-\left(-\frac{t}{T}\right)^d\right]\right\}\\
&\quad\to\int_{-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}[(s-t)^d-(-t)^d]\right\}\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'\\
&\qquad\times\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}[(s-t)^d-(-t)^d]\right\}dt<\infty.
\end{aligned}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

This limit is also a $K \times K$ matrix of integrals of which the (i,j)th element is

$$(\boldsymbol{\pi}_i' \boldsymbol{\Sigma}_a \boldsymbol{\pi}_j) \cdot \frac{1}{\Gamma(d_i+1)\Gamma(d_j+1)} \int_{-\infty}^{0} [(s-t)^{d_i} - (-t)^{d_i}][(s-t)^{d_j} - (-t)^{d_j}]\, dt < \infty. \tag{A.6}$$

We now show that for some $\epsilon > 0$

$$\sum_{t=1}^{[Ts]} E\left[\left| \frac{1}{\sqrt{T}} \frac{1}{\Gamma(d_i+1)} \left(\frac{[Ts]}{T} - \frac{t}{T} \right)^{d_i} \boldsymbol{\pi}_i' \mathbf{a}_t \right|^{2+\epsilon} \middle\| \mathcal{F}_{t-1} \right] \to 0, \quad \text{as } T \to \infty. \tag{A.7}$$

The left-hand side of (A.7) can be written as

$$\begin{aligned} &\frac{1}{T^{\epsilon/2}} \frac{1}{\Gamma(d_i+1)^{2+\epsilon}} \frac{1}{T} \sum_{t=1}^{[Ts]} \left(\frac{[Ts]}{T} - \frac{t}{T} \right)^{d_i(2+\epsilon)} E(|\boldsymbol{\pi}_i' \mathbf{a}_t|^{2+\epsilon} \| \mathcal{F}_{t-1}) \\ &\quad \le \frac{1}{T^{\epsilon/2}} \frac{1}{\Gamma(d_i+1)^{2+\epsilon}} \sup_t E(|\boldsymbol{\pi}_i' \mathbf{a}_t|^{2+\epsilon} \| \mathcal{F}_{t-1}) \frac{1}{T} \sum_{t=1}^{[Ts]} \left(\frac{[Ts]}{T} - \frac{t}{T} \right)^{d_i(2+\epsilon)} \\ &\quad \sim \frac{1}{T^{\epsilon/2}} \frac{1}{\Gamma(d_i+1)^{2+\epsilon}} \sup_t E(|\boldsymbol{\pi}_i' \mathbf{a}_t|^{2+\epsilon} \| \mathcal{F}_{t-1}) \int_0^s (s-t)^{d_i(2+\epsilon)}\, dt \to 0, \quad \text{as } T \to \infty. \end{aligned}$$

Similarly, as $T \to \infty$,

$$\begin{aligned} &\sum_{t=-\infty}^{0} E\left\{ \left| \frac{1}{\sqrt{T}} \frac{1}{\Gamma(d_i+1)} \left[\left(\frac{[Ts]}{T} - \frac{t}{T} \right)^{d_i} - \left(-\frac{t}{T} \right)^{d_i} \right] \boldsymbol{\pi}_i' \mathbf{a}_t \right|^{2+\epsilon} \middle\| \mathcal{F}_{t-1} \right\} \\ &\quad \le \frac{1}{T^{\epsilon/2}} \frac{1}{\Gamma(d_i+1)^{2+\epsilon}} \sup_t E(|\boldsymbol{\pi}_i' \mathbf{a}_t|^{2+\epsilon} \| \mathcal{F}_{t-1}) \int_{-\infty}^{0} [(s-t)^{d_i} - (-t)^{d_i}]^{2+\epsilon}\, dt \to 0. \end{aligned} \tag{A.8}$$

With results (A.5)–(A.8) we can then apply a functional central limit theorem by Helland (1982, Theorem 3.2 and equation (2.9)), which implies

$$\frac{1}{T^{0.5+d_i}} \sum_{t=1}^{[Ts]} (x_{it} - \mu_i) \Rightarrow B_{d_i}(s),$$

for all $i = 1, \ldots, K$.

The joint convergence of the vector $\mathbf{x}_t$ follows from the facts that its elements are all linear functions of the same martingale differences $\mathbf{a}_t$ with the conditional variance-covariance matrix $\boldsymbol{\Sigma}_a$ and that their respective limits, i.e., the fractional Brownian motions $B_{d_i}(s)$, for $i = 1, \ldots, K$, are all defined from the same standard K-dimensional Brownian motion with the variance-covariance matrix $\boldsymbol{\Pi}\boldsymbol{\Sigma}_a\boldsymbol{\Pi}'$. ■

To prove Theorem 2, we need the following two lemmas.

LEMMA A.3. *Given the assumptions of Theorem 2, we consider K sequences of scalars $\psi_{i,t}^{\circ}$, for $i = 1, \ldots, K$ and $t = 1, 2, \ldots$, such that $\psi_{i,t}^{\circ} \sim t^{d_i - 1}$. Define $\psi_{i,(t)}^{\circ} = \psi_{i,t}^{\circ}$ if $t > 0$; and $\psi_{i,(t)}^{\circ} = 0$ if $t \le 0$. Then, for any integers k and ℓ, we have*

$$T^{1-d_i-d_j} \sum_{t=1}^{T} \psi_{i,(t-\ell)}^{\circ} \psi_{j,(t-k)}^{\circ} = T^{1-d_i-d_j} \sum_{t=1}^{T} (t-\ell)_+^{d_i-1} (t-k)_+^{d_j-1} + o(1), \quad \text{as } T \to \infty,$$

for all $i, j = 1, \ldots, K$, where $(z)_+^{\delta} \equiv 0$ if $z \le 0$ and $(z)_+^{\delta} \equiv z^{\delta}$ if $z > 0$.

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

The proof of Lemma A.3 is parallel to that of (A.1) and is omitted here.

LEMMA A.4. *Given the assumptions of Theorem 2, we have that as* $T \to \infty$

$$\begin{aligned}
&\mathbf{D}(T^{-d}) \sum_{t=1}^{T} \left[\left(\sum_{\ell=-\infty}^{t} \boldsymbol{\Psi}_{t-\ell} \mathbf{a}_\ell \right) \left(\sum_{k=-\infty}^{t} \boldsymbol{\Psi}_{t-k} \mathbf{a}_k \right)' - \sum_{\ell=-\infty}^{t} \boldsymbol{\Psi}_{t-\ell} \boldsymbol{\Sigma}_a \boldsymbol{\Psi}'_{t-\ell} \right] \mathbf{D}(T^{-d}) \\
&\quad = \frac{1}{\sqrt{T}} \sum_{k=-\infty}^{[Ts]} \frac{1}{\sqrt{T}} \sum_{\ell=-\infty}^{[Ts]} \left\{ \frac{1}{T} \sum_{t=1}^{[Ts]} \mathbf{D} \left[\frac{1}{\Gamma(d)} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d-1} \right] \boldsymbol{\Pi} (\mathbf{a}_\ell \mathbf{a}'_k - \delta_{\ell k} \boldsymbol{\Sigma}_a) \boldsymbol{\Pi}' \right. \\
&\qquad \left. \times \mathbf{D} \left[\frac{1}{\Gamma(d)} \left(\frac{t}{T} - \frac{k}{T} \right)_+^{d-1} \right] \right\} + o_p(1),
\end{aligned}$$

where $\delta_{\ell k} = 1$ *if* $\ell = k$ *and* $\delta_{\ell k} = 0$ *if* $\ell \neq k$. *Also,* $(z)_+^\delta \equiv 0$ *if* $z \leq 0$, *and* $(z)_+^\delta \equiv z^\delta$ *if* $z > 0$.

Proof of Lemma A.4. Given that $\boldsymbol{\psi}'_{i,t}$ and $\boldsymbol{\pi}'_i$ are the ith rows of $\boldsymbol{\Psi}_t$ and $\boldsymbol{\Pi}$, respectively, let $\psi^\circ_{i,t}$ be a scalar such that $\boldsymbol{\psi}_{i,t} \equiv \psi^\circ_{i,t} \cdot \Gamma(d_i)^{-1} \boldsymbol{\pi}_i$ and $\psi^\circ_{i,t} \sim t^{d_i - 1}$. We note that

$$\begin{aligned}
&T^{-d_i - d_j} \sum_{t=1}^{T} \left[\left(\sum_{\ell=-\infty}^{t} \boldsymbol{\psi}'_{i,t-\ell} \mathbf{a}_\ell \right) \left(\sum_{k=-\infty}^{t} \boldsymbol{\psi}'_{j,t-k} \mathbf{a}_k \right)' - \sum_{\ell=-\infty}^{t} \boldsymbol{\psi}'_{i,t-\ell} \boldsymbol{\Sigma}_a \boldsymbol{\psi}_{j,t-\ell} \right] \\
&\quad = \frac{1}{\Gamma(d_i)\Gamma(d_j)} T^{-d_i - d_j} \sum_{t=1}^{T} \left[\left(\sum_{\ell=-\infty}^{t} \psi^\circ_{i,t-\ell} \boldsymbol{\pi}'_i \mathbf{a}_\ell \right) \left(\sum_{k=-\infty}^{t} \psi^\circ_{j,t-k} \boldsymbol{\pi}'_j \mathbf{a}_k \right)' \right. \\
&\qquad \left. - \sum_{\ell=-\infty}^{t} \psi^\circ_{i,t-\ell} \psi^\circ_{j,t-\ell} \boldsymbol{\pi}'_i \boldsymbol{\Sigma}_a \boldsymbol{\pi}_j \right] \\
&\quad = \frac{1}{\Gamma(d_i)\Gamma(d_j)} T^{-d_i - d_j} \sum_{k=-\infty}^{T} \sum_{\ell=-\infty}^{T} \left[\sum_{t=1}^{T} \psi^\circ_{i,(t-\ell)} \psi^\circ_{j,(t-k)} \cdot \boldsymbol{\pi}'_i (\mathbf{a}_\ell \mathbf{a}'_k - \delta_{\ell k} \boldsymbol{\Sigma}_a) \boldsymbol{\pi}_j \right] \\
&\quad = \frac{1}{\Gamma(d_i)\Gamma(d_j)} \frac{1}{T} \sum_{k=-\infty}^{T} \sum_{\ell=-\infty}^{T} \left[T^{1-d_i-d_j} \sum_{t=1}^{T} \psi^\circ_{i,(t-\ell)} \psi^\circ_{j,(t-k)} \cdot \boldsymbol{\pi}'_i (\mathbf{a}_\ell \mathbf{a}'_k - \delta_{\ell k} \boldsymbol{\Sigma}_a) \boldsymbol{\pi}_j \right],
\end{aligned}$$

where $\psi^\circ_{i,(t)} = \psi^\circ_{i,t}$ if $t > 0$ and $\psi^\circ_{i,(t)} = 0$ if $t \leq 0$. So it is sufficient to prove that

$$\begin{aligned}
&\frac{1}{T} \sum_{k=-\infty}^{T} \sum_{\ell=-\infty}^{T} \left[T^{1-d_i-d_j} \sum_{t=1}^{T} \psi^\circ_{i,(t-\ell)} \psi^\circ_{j,(t-k)} \cdot \boldsymbol{\pi}'_i (\mathbf{a}_\ell \mathbf{a}'_k - \delta_{\ell k} \boldsymbol{\Sigma}_a) \boldsymbol{\pi}_j \right] \\
&\quad = \frac{1}{T} \sum_{k=-\infty}^{T} \sum_{\ell=-\infty}^{T} \left[T^{1-d_i-d_j} \sum_{t=1}^{T} (t-\ell)_+^{d_i - 1} (t-k)_+^{d_j - 1} \cdot \boldsymbol{\pi}'_i (\mathbf{a}_\ell \mathbf{a}'_k - \delta_{\ell k} \boldsymbol{\Sigma}_a) \boldsymbol{\pi}_j \right] \\
&\qquad + o_p(1), \qquad \textbf{(A.9)}
\end{aligned}$$

for all $i, j = 1, \ldots, K$. Lemma A.3 implies that we can find a sufficiently large but finite constant A such that

$$\left| T^{1-d_i-d_j} \sum_{t=1}^{T} \psi^\circ_{i,(t-\ell)} \psi^\circ_{j,(t-k)} \right| < A \cdot T^{1-d_i-d_j} \sum_{t=1}^{T} (t-\ell)_+^{d_i - 1} (t-k)_+^{d_j - 1}.$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

Furthermore, for any $\epsilon > 0$ there are integers N and N' such that

$$(1-\varepsilon)\cdot T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1} < T^{1-d_i-d_j}\sum_{t=1}^{T}\psi^\circ_{i,(t-\ell)}\psi^\circ_{j,(t-k)} < (1+\varepsilon)\cdot T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1},$$

for all $\ell \le -N$, $k \le -N$, $\ell \ge N$, $k \ge N$, and $T \ge N'$. Consequently, for any $\varepsilon > 0$ and a sufficiently large T such that $T \ge N'$, we have

$$\begin{aligned}
&\frac{1}{T}\left|\sum_{k=-\infty}^{T}\sum_{\ell=-\infty}^{T}T^{1-d_i-d_j}\sum_{t=1}^{T}\psi^\circ_{i,(t-\ell)}\psi^\circ_{j,(t-k)}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right.\\
&\quad\left.-\sum_{k=-\infty}^{T}\sum_{\ell=-\infty}^{T}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\\
&\le\frac{1}{T}\left\{\sum_{k=-N+1}^{N-1}\sum_{\ell=-N+1}^{N-1}T^{1-d_i-d_j}\left|\sum_{t=1}^{T}\psi^\circ_{i,(t-\ell)}\psi^\circ_{j,(t-k)}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\right.\\
&\quad+\sum_{k=-N+1}^{N-1}\sum_{\ell=-N+1}^{N-1}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\left|\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\\
&\quad+\sum_{k=N}^{T}\sum_{\ell=N}^{T}\left|T^{1-d_i-d_j}\sum_{t=1}^{T}\psi^\circ_{i,(t-\ell)}\psi^\circ_{j,(t-k)}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right.\\
&\quad\left.-T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\\
&\quad+\sum_{k=-\infty}^{-N}\sum_{\ell=-\infty}^{-N}\left|T^{1-d_i-d_j}\sum_{t=1}^{T}\psi^\circ_{i,(t-\ell)}\psi^\circ_{j,(t-k)}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right.\\
&\quad\left.\left.-T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\right\}+S\\
&\le\frac{1}{T}(A+1)\sum_{k=-N+1}^{N-1}\sum_{\ell=-N+1}^{N-1}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\\
&\quad\cdot\left|\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\right|\\
&\quad+\varepsilon\cdot\frac{1}{T}\sum_{k=N}^{T}\sum_{\ell=N}^{T}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j\\
&\quad+\varepsilon\cdot\frac{1}{T}\sum_{k=-\infty}^{-N}\sum_{\ell=-\infty}^{N}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j+S,
\end{aligned}$$

where S represents those terms corresponding to the summations $\sum_{k=-N+1}^{N-1}\sum_{\ell=N}^{T}$, $\sum_{k=-N+1}^{N-1}\sum_{\ell=-\infty}^{-N}$, $\sum_{k=N}^{T}\sum_{\ell=-N+1}^{N-1}$, $\sum_{k=N}^{T}\sum_{\ell=-\infty}^{-N}$, $\sum_{k=-\infty}^{-N}\sum_{\ell=-N+1}^{N-1}$, and $\sum_{k=-\infty}^{-N}\sum_{\ell=N}^{T}$. On the right-hand side of the preceding inequality, the first term converges to 0, a.s., when $T\to\infty$. Also note that

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

$$\frac{1}{T^2}\sum_{k=-\infty}^{T}\sum_{\ell=-\infty}^{T}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k}{T}\right)_+^{d_j-1}\right]^2$$
$$<\int_{-\infty}^{1}\int_{-\infty}^{1}\left[\int_0^1 (t-r_1)_+^{d_i-1}(t-r_2)_+^{d_j-1}\,dt\right]^2 dr_1\,dr_2<\infty.$$

This result together with the assumption $\sup_t E(|a_{it}|^{4+\epsilon}\|\mathcal{F}_{t-1})<\infty$, a.s. implies that the (conditional) variances of the two sums

$$\frac{1}{T}\sum_{k=N}^{T}\sum_{\ell=N}^{T}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j$$

$$\frac{1}{T}\sum_{k=-\infty}^{-N}\sum_{\ell=-\infty}^{-N}T^{1-d_i-d_j}\sum_{t=1}^{T}(t-\ell)_+^{d_i-1}(t-k)_+^{d_j-1}\cdot\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_j$$

are bounded, which in turn implies that the preceding two sums are bounded in probability based on Chebyshev inequality. By the same token, we can show that the terms in S are also negligible. This completes the proof of (A.9). ■

Proof of Theorem 2. The autocovariances $\boldsymbol{\Gamma}(j)$ of the process $\mathbf{x}_t$ can be expressed as follows:

$$\boldsymbol{\Gamma}(j)=\sum_{\ell=-\infty}^{t}\boldsymbol{\Psi}_{t-\ell}\boldsymbol{\Sigma}_a\boldsymbol{\Psi}_{t+j-\ell}'.$$

This result together with the infinite-order moving average representation of $\mathbf{x}_t$ and Lemma A.4 yields

$$\mathbf{D}(T^{-d})\left\{\sum_{t=1}^{[Ts]-j}[(\mathbf{x}_t-\boldsymbol{\mu})(\mathbf{x}_{t+j}-\boldsymbol{\mu})'-\boldsymbol{\Gamma}(j)]\right\}\mathbf{D}(T^{-d})$$
$$=\mathbf{D}(T^{-d})\sum_{t=1}^{[Ts]-j}\left[\left(\sum_{\ell=-\infty}^{t}\boldsymbol{\Psi}_{t-\ell}\mathbf{a}_\ell\right)\left(\sum_{k=-\infty}^{t+j}\boldsymbol{\Psi}_{t+j-k}\mathbf{a}_k\right)'-\sum_{\ell=-\infty}^{t}\boldsymbol{\Psi}_{t-\ell}\boldsymbol{\Sigma}_a\boldsymbol{\Psi}_{t+j-\ell}'\right]\mathbf{D}(T^{-d})$$
$$=\frac{1}{\sqrt{T}}\sum_{k=-\infty}^{[Ts]}\frac{1}{\sqrt{T}}\sum_{\ell=-\infty}^{[Ts]-j}\left\{\frac{1}{T}\sum_{t=1}^{[Ts]-j}\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d-1}\right]\boldsymbol{\Pi}(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\Pi}'\right.$$
$$\left.\cdot\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d-1}\right]\right\}+o_p(1),$$

where $\delta_{\ell k}=1$ if $\ell=k$ and $\delta_{\ell k}=0$ if $\ell\neq k$.

Let us consider the (i,h)th element of the preceding $K\times K$ matrix, which is

$$\frac{1}{\Gamma(d_i)\Gamma(d_h)}\frac{1}{\sqrt{T}}\sum_{k=-\infty}^{[Ts]}\frac{1}{\sqrt{T}}\sum_{\ell=-\infty}^{[Ts]-j}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]$$
$$\times\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_h,$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

for $i, h = 1, \ldots, K$. We note that the terms $\mathbf{a}_\ell \mathbf{a}_k' - \delta_{\ell k}\boldsymbol{\Sigma}_a$ inside the parentheses are martingale differences with respect to $\mathcal{F}_{\ell-1} \cap \mathcal{F}_{k-1}$, which can be considered a filtration with two indices ℓ and k and is nondecreasing with respect to each index. To see this, we note that $E(\mathbf{a}_\ell \mathbf{a}_k' - \delta_{\ell k}\boldsymbol{\Sigma}_a \| \mathcal{F}_{\ell-1} \cap \mathcal{F}_{k-1})$ is equal a.s. to $E[E(\mathbf{a}_\ell \| \mathcal{F}_{\ell-1})\mathbf{a}_k' \| \mathcal{F}_{k-1}] \overset{\text{a.s.}}{=} \mathbf{0}$ if $\ell > k$; is equal a.s. to $E[\mathbf{a}_\ell E(\mathbf{a}_k' \| \mathcal{F}_{k-1}) \| \mathcal{F}_{\ell-1}] \overset{\text{a.s.}}{=} \mathbf{0}$ if $\ell < k$; and is equal a.s. to $E(\mathbf{a}_\ell \mathbf{a}_\ell' \| \mathcal{F}_{\ell-1}) - \boldsymbol{\Sigma}_a \overset{\text{a.s.}}{=} \mathbf{0}$ if $\ell = k$. Let us now consider the sum of the conditional variances of the preceding expression:

$$
\begin{aligned}
&\sum_{k=-\infty}^{[Ts]} \sum_{\ell=-\infty}^{[Ts]-j} \operatorname{Var}\left\{ \frac{1}{T}\left[\frac{1}{T} \sum_{t=1}^{[Ts]-j} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d_i-1} \left(\frac{t}{T} - \frac{k-j}{T} \right)_+^{d_h-1} \right] \right. \\
&\qquad \left. \times \boldsymbol{\pi}_i'(\mathbf{a}_\ell \mathbf{a}_k' - \delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_h \,\middle\|\, \mathcal{F}_{\ell-1} \cap \mathcal{F}_{k-1} \right\} \\
&\overset{\text{a.s.}}{=} (\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_i)(\boldsymbol{\pi}_h'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_h) \frac{1}{T} \sum_{k=-\infty}^{[Ts]} \frac{1}{T} \\
&\quad \times \sum_{\ell=-\infty}^{[Ts]-j} \left[\frac{1}{T} \sum_{t=1}^{[Ts]-j} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d_i-1} \left(\frac{t}{T} - \frac{k-j}{T} \right)_+^{d_h-1} \right]^2 \\
&\quad + [\operatorname{Var}(\boldsymbol{\pi}_i'\mathbf{a}_\ell \mathbf{a}_\ell'\boldsymbol{\pi}_h \| \mathcal{F}_{\ell-1}) - (\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_i)(\boldsymbol{\pi}_h'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_h)] \\
&\quad \times \frac{1}{T^2} \sum_{\ell=-\infty}^{[Ts]} \left[\frac{1}{T} \sum_{t=1}^{[Ts]-j} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d_i-1} \left(\frac{t}{T} - \frac{\ell-j}{T} \right)_+^{d_h-1} \right]^2 \\
&\to (\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_i)(\boldsymbol{\pi}_h'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_h) \int_{-\infty}^{s} \int_{-\infty}^{s} \left[\int_0^s (t-r_1)_+^{d_i-1} (t-r_2)_+^{d_h-1} \, dt \right]^2 dr_1 \, dr_2 \\
&\quad + 0 < \infty, \qquad \textbf{(A.10)}
\end{aligned}
$$

for $0.5 < d_i + d_h < 1$. We note that no matter what value j is, as long as it is finite, we get the same convergence result. To explain why the term

$$
\frac{1}{T^2} \sum_{\ell=-\infty}^{[Ts]} \left[\frac{1}{T} \sum_{t=1}^{[Ts]-j} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d_i-1} \left(\frac{t}{T} - \frac{\ell-j}{T} \right)_+^{d_h-1} \right]^2
$$

converges to zero when $0.5 < d_i + d_h < 1$, here we only consider the case with $j = 0$ and note that the general case with $j > 0$ can be proved similarly. Let us first point out that when $j = 0$, then the preceding expression is equal to

$$
\frac{1}{T^{2(d_i+d_h)}} \sum_{\ell=0}^{T-1} \left(\sum_{t=1}^{T-\ell} t^{d_i+d_h-2} \right)^2 + \frac{1}{T} \frac{1}{T} \sum_{\ell=1}^{\infty} \left[\frac{1}{T} \sum_{t=1}^{T} \left(\frac{t}{T} + \frac{\ell}{T} \right)^{d_i+d_h-2} \right]^2,
$$

when $s = 1$. We now show that if $0.5 < d_i + d_h < 1$, then both terms in the present expression converge to zero. For the first term, we note

$$
\begin{aligned}
\frac{1}{T^{2(d_i+d_h)}} \sum_{\ell=0}^{T-1} \left(\sum_{t=1}^{T-\ell} t^{d_i+d_h-2} \right)^2 &\le \frac{1}{T^{2(d_i+d_h)}} \sum_{\ell=0}^{T-1} \left(\sum_{t=1}^{T} t^{d_i+d_h-2} \right)^2 \\
&= \frac{1}{T^{2(d_i+d_h)-1}} \left(\sum_{t=1}^{T} t^{d_i+d_h-2} \right)^2,
\end{aligned}
$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

which converges to zero because $1/T^{2(d_i+d_h)-1} \to 0$ whereas the term inside the parentheses is finite (and is equal to Riemann's zeta function) as long as $d_i + d_h - 2 < -1$. To see why the second term converges to zero, we simply note that

$$\frac{1}{T}\sum_{\ell=1}^{\infty}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{t}{T}+\frac{\ell}{T}\right)^{d_i+d_h-2}\right]^2 \to \int_0^{\infty}\left[\int_0^1 (t+r)^{d_i+d_h-2}\,dt\right]^2 dr$$

$$= \frac{1}{(d_i+d_h-1)^2}\int_0^{\infty}[(1+r)^{d_i+d_h-1} - r^{d_i+d_h-1}]^2\,dr < \infty,$$

because $-0.5 < d_i + d_h - 1 < 0$ (a similar integral appears earlier in the proof of Theorem 1).

Let us now show that for some $\epsilon > 0$ we have

$$\sum_{k=-\infty}^{[Ts]}\sum_{\ell=-\infty}^{[Ts]-j} E\left\{\left|\frac{1}{\Gamma(d_i)\Gamma(d_h)}\frac{1}{T}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]\right.\right.$$

$$\left.\left.\cdot\,\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k' - \delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_h\right|^{2+\epsilon}\right\| \mathcal{F}_{\ell-1}\cap\mathcal{F}_{k-1}\right\} \to 0, \qquad \textbf{(A.11)}$$

as $T \to \infty$. The left-hand side of (A.11) can be expressed as

$$\frac{1}{T^\epsilon}\frac{1}{[\Gamma(d_i)\Gamma(d_h)]^{2+\epsilon}}\frac{1}{T}\sum_{k=-\infty}^{[Ts]}\frac{1}{T}\sum_{\ell=-\infty}^{[Ts]-j}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]^{2+\epsilon}$$

$$\cdot E[|\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k' - \delta_{\ell k}\boldsymbol{\Sigma}_a)\boldsymbol{\pi}_h|^{2+\epsilon}\|\mathcal{F}_{\ell-1}\cap\mathcal{F}_{k-1}]$$

$$\le \frac{1}{T^\epsilon}\frac{1}{[\Gamma(d_i)\Gamma(d_h)]^{2+\epsilon}}\sup_\ell E(|\boldsymbol{\pi}_i'\mathbf{a}_\ell|^{2+\epsilon}\|\mathcal{F}_{\ell-1})\cdot\sup_k E(|\boldsymbol{\pi}_h'\mathbf{a}_k|^{2+\epsilon}\|\mathcal{F}_{k-1})$$

$$\cdot\frac{1}{T}\sum_{k=-\infty}^{[Ts]}\frac{1}{T}\sum_{\ell=-\infty}^{[Ts]-j}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]^{2+\epsilon}$$

$$+\frac{1}{T^\epsilon}\frac{1}{[\Gamma(d_i)\Gamma(d_h)]^{2+\epsilon}}\left[\sup_\ell E(|\boldsymbol{\pi}_i'\mathbf{a}_\ell\mathbf{a}_\ell'\boldsymbol{\pi}_h|^{2+\epsilon}\|\mathcal{F}_{\ell-1}) + |\boldsymbol{\pi}_i'\boldsymbol{\Sigma}_a\boldsymbol{\pi}_h|^{2+\epsilon}\right]$$

$$\cdot\frac{1}{T^2}\sum_{\ell=-\infty}^{[Ts]-j}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{\ell-j}{T}\right)_+^{d_h-1}\right]^{2+\epsilon}$$

$$\sim \frac{1}{T^\epsilon}\frac{1}{[\Gamma(d_i)\Gamma(d_h)]^{2+\epsilon}}\sup_\ell E(|\boldsymbol{\pi}_i'\mathbf{a}_\ell|^{2+\epsilon}\|\mathcal{F}_{\ell-1})\cdot\sup_k E(|\boldsymbol{\pi}_h'\mathbf{a}_k|^{2+\epsilon}\|\mathcal{F}_{k-1})$$

$$\cdot\int_{-\infty}^{s}\int_{-\infty}^{s}\left[\int_0^s (t-r_1)_+^{d_i-1}(t-r_2)_+^{d_h-1}\,dt\right]^{2+\epsilon} dr_1\,dr_2 + 0 \to 0, \quad \text{as } T\to\infty,$$

for $0.5 < d_i + d_h < 1$. Note that the boundedness of $\sup_\ell E(|\boldsymbol{\pi}_i'\mathbf{a}_\ell\mathbf{a}_\ell'\boldsymbol{\pi}_h|^{2+\epsilon}\|\mathcal{F}_{\ell-1})$ is guaranteed by the assumption that $\sup_\ell E(|a_{i\ell}|^{4+\epsilon}\|\mathcal{F}_{\ell-1}) < \infty$. The convergence results (A.10) and (A.11) satisfy the conditions of Helland's (1982, Theorem 3.2) functional central limit theorem, which implies the following weak convergence to the Rosenblatt process

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

$$\frac{1}{T^{d_i+d_h}}\sum_{t=1}^{[Ts]-j}[(x_{it}-\mu_i)(x_{h,t+j}-\mu_h)-\mathbf{\Gamma}_{ih}(j)]\Rightarrow R^{(ih)}(s),$$

for all $i,h=1,\ldots,K$.

The joint convergence of the matrix $(\mathbf{x}_t-\boldsymbol{\mu})(\mathbf{x}_{t+j}-\boldsymbol{\mu})'$ follows from the facts that its elements are all linear functions of the same martingale differences $\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\mathbf{\Sigma}_a$ and that their respective limits $R^{(ih)}(s)$, for $i,h=1,\ldots,K$, are all defined from the same K-dimensional Brownian motion with the variance-covariance matrix $\mathbf{\Pi\Sigma}_a\mathbf{\Pi}'$.

To prove the joint weak convergence of

$$\begin{aligned}&\mathbf{D}(T^{-0.5-d})\sum_{t=1}^{[Ts]}(\mathbf{x}_t-\boldsymbol{\mu})\\&\quad=\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]}\mathbf{D}\left[\frac{1}{\Gamma(d+1)}\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d\right]\mathbf{\Pi a}_t\\&\qquad+\frac{1}{\sqrt{T}}\sum_{t=-\infty}^{0}\mathbf{D}\left\{\frac{1}{\Gamma(d+1)}\left[\left(\frac{[Ts]}{T}-\frac{t}{T}\right)^d-\left(-\frac{t}{T}\right)^d\right]\right\}\mathbf{\Pi a}_t+o_p(1)\end{aligned}$$

and

$$\begin{aligned}&\mathbf{D}(T^{-d})\left\{\sum_{t=1}^{[Ts]-j}[(\mathbf{x}_t-\boldsymbol{\mu})(\mathbf{x}_{t+j}-\boldsymbol{\mu})'-\mathbf{\Gamma}(j)]\right\}\mathbf{D}(T^{-d})\\&\quad=\frac{1}{\sqrt{T}}\sum_{k=-\infty}^{[Ts]}\frac{1}{\sqrt{T}}\sum_{\ell=-\infty}^{[Ts]-j}\left\{\frac{1}{T}\sum_{t=1}^{[Ts]-j}\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d-1}\right]\mathbf{\Pi}(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\mathbf{\Sigma}_a)\mathbf{\Pi}'\right.\\&\qquad\left.\times\mathbf{D}\left[\frac{1}{\Gamma(d)}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d-1}\right]\right\}+o_p(1),\end{aligned}$$

Helland's (1982) Theorem 3.3 indicates that it is sufficient to show that the following sums of conditional covariances:

$$\begin{aligned}&\sum_{k=-\infty}^{[Ts]}\sum_{\ell=1}^{[Ts]-j}\mathrm{Cov}\left\{\frac{1}{\sqrt{T}}\left(\frac{[Ts]}{T}-\frac{\ell}{T}\right)^{d_m}\boldsymbol{\pi}_m'\mathbf{a}_\ell,\right.\\&\qquad\frac{1}{T}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]\\&\qquad\left.\times\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\mathbf{\Sigma}_a)\boldsymbol{\pi}_h\,\middle\|\,\mathcal{F}_{\ell-1}\cap\mathcal{F}_{k-1}\right\}\end{aligned}\tag{A.12}$$

$$\begin{aligned}&\sum_{k=-\infty}^{[Ts]}\sum_{\ell=-\infty}^{0}\mathrm{Cov}\left\{\frac{1}{\sqrt{T}}\left[\left(\frac{[Ts]}{T}-\frac{\ell}{T}\right)^{d_m}-\left(-\frac{\ell}{T}\right)^{d_m}\right]\boldsymbol{\pi}_m'\mathbf{a}_\ell,\right.\\&\qquad\frac{1}{T}\left[\frac{1}{T}\sum_{t=1}^{[Ts]-j}\left(\frac{t}{T}-\frac{\ell}{T}\right)_+^{d_i-1}\left(\frac{t}{T}-\frac{k-j}{T}\right)_+^{d_h-1}\right]\\&\qquad\left.\times\boldsymbol{\pi}_i'(\mathbf{a}_\ell\mathbf{a}_k'-\delta_{\ell k}\mathbf{\Sigma}_a)\boldsymbol{\pi}_h\,\middle\|\,\mathcal{F}_{\ell-1}\cap\mathcal{F}_{k-1}\right\}\end{aligned}\tag{A.13}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

are both convergent to zero, for $m, i, h = 1, \ldots, K$. It is straightforward to see that (A.12) is a.s. smaller than the product of $\sup_\ell E(\boldsymbol{\pi}_m' \mathbf{a}_\ell \cdot \boldsymbol{\pi}_i' \mathbf{a}_\ell \cdot \boldsymbol{\pi}_h' \mathbf{a}_\ell \| \mathcal{F}_{\ell-1})$, which is finite, and

$$\begin{aligned}
&\frac{1}{T^{1.5}} \sum_{\ell=1}^{[Ts]} \left(\frac{[Ts]}{T} - \frac{\ell}{T} \right)^{d_m} \left[\frac{1}{T} \sum_{t=1}^{[Ts]} \left(\frac{t}{T} - \frac{\ell}{T} \right)_+^{d_i+d_h-2} \right] \\
&\quad = \frac{1}{T^{d_m+d_i+d_h+0.5}} \sum_{\ell=1}^{T} (T-\ell)^{d_m} \sum_{t=1}^{T} (t-\ell)_+^{d_i+d_h-2} \\
&\quad = \frac{1}{T^{d_m+d_i+d_h+0.5}} \sum_{\ell=1}^{T} (T-\ell)^{d_m} \sum_{t=1}^{T-\ell} t^{d_i+d_h-2} \\
&\quad = \frac{1}{T^{d_m+d_i+d_h+0.5}} \sum_{\ell=0}^{T-1} \ell^{d_m} \sum_{t=1}^{\ell} t^{d_i+d_h-2} \\
&\quad = \frac{1}{T^{d_m+d_i+d_h+0.5}} O(T^{d_m+d_i+d_h}) = O(T^{-0.5}),
\end{aligned}$$

when $s = 1$. This proves that (A.12) converges to zero, whereas the convergence of (A.13) can be proved by a similar argument. ■

Proof of Corollary 1. Part 1 is straightforward. Part 2 is based on the central limit theorem of Hannan (1976, p. 397). We note that when $0 < d_i < 0.25$, for all $i = 1, \ldots, K$, then Lemma 2 implies the infinite sum of the squared autocovariances

$$\sum_{j=0}^{\infty} \boldsymbol{\Gamma}(j)\boldsymbol{\Gamma}(j)' \sim \sum_{j=0}^{\infty} \mathbf{D}(j^{d-0.5}) \cdot \mathbf{A} \cdot \mathbf{D}(j^{d-0.5}) \cdot \mathbf{D}(j^{d-0.5}) \cdot \mathbf{A} \cdot \mathbf{D}(j^{d-0.5}) < \infty.$$

Consequently, the spectral density of the process $\mathbf{x}_t$ is square-integrable, and Hannan's theory is applicable. The details of the variance-covariance matrix $\mathbf{W}_{xx}^{(j)}$ can be found in Hannan (1976, p. 397). ■

Proof of Theorem 3. From Corollary 1 we know that $\hat{\mathbf{M}}_{xx}$ converges in probability to $\mathbf{M}_{xx}$, with which Theorem 3 is then a direct consequence (note that the covariance $\mathbf{m}_{x\varepsilon}$ between the regressor $\mathbf{x}_t$ and the disturbance ε_t is $\mathbf{0}$ in the present linear regression framework):

$$\begin{aligned}
T^{1-d_1-d^*}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \hat{\mathbf{M}}_{xx}^{-1} \cdot T^{1-d_1-d^*} \mathbf{D}(T^{-1+d+d^*}) \mathbf{D}(T^{1-d-d^*}) \cdot \hat{\mathbf{m}}_{x\varepsilon} \\
&= \hat{\mathbf{M}}_{xx}^{-1} \mathbf{D}(T^{d-d_1}) \mathbf{D}(T^{1-d-d^*}) \cdot \hat{\mathbf{m}}_{x\varepsilon} \\
&= \hat{\mathbf{M}}_{xx}^{-1} \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & T^{d_2-d_1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & T^{d_K-d_1} \end{bmatrix} \mathbf{D}(T^{1-d-d^*}) \cdot \hat{\mathbf{m}}_{x\varepsilon} \\
&\Rightarrow \mathbf{M}_{xx}^{-1} \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix} [\mathbf{R}_{x\varepsilon}(1) - \mathbf{B}_d(1) B_{d^*}(1)] \\
&= \mathbf{c} \cdot [R^{(1^*)}(1) - B_{d_1}(1) B_{d^*}(1)].
\end{aligned}$$

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press

In the preceding analysis we assume that $d_i + d^* > 0.5$ for all i to simplify the expression. However, it is readily seen that the convergence and the resulting limit will not change even if some $d_i + d^* < 0.5$.

We should point out that the assumptions in Theorem 2 and Corollary 1 are that all d_i are greater than 0.25 so that the sum of each pair $d_i + d_j$ is greater than 0.5. But if we inspect the proofs of those theorems, we will find that it is actually the latter condition that is required to establish those theorems. In the preceding application of Theorem 2 and Corollary 1 we have directly assumed that the sum of each pair of d_i and d_j^* is greater than 0.5. ■

Proof of Corollary 2. Corollary 2 is a direct consequence of Corollary 1. ■

Proof of Corollary 3. From Theorem 1 we know that $\bar{\mathbf{x}}$ converges in probability to $\boldsymbol{\mu}$ and from Theorem 3 we know $T^{1-d_1-d^*}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges weakly. Consequently, $T^{0.5-d^*}\bar{\mathbf{x}}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = T^{d_1-0.5}\bar{\mathbf{x}}' \cdot T^{1-d_1-d^*}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to zero in probability because $T^{d_1-0.5}$ converges to zero. Now, from (5) we get

$$T^{0.5-d^*}(\hat{\alpha} - \alpha) = T^{0.5-d^*}\bar{\varepsilon} - T^{0.5-d^*}\bar{\mathbf{x}}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow B_{d^*}(1),$$

whereas $B_{d^*}(1) \stackrel{d}{=} \mathcal{N}(0, \pi^{*2}\lambda^*)$. The convergence $s^2 \xrightarrow{p} \sigma_\varepsilon^2$ is a direct result of Corollary 1 and the assumption that $\mathbf{m}_{x\varepsilon} = \mathbf{0}$. ■

Proof of Corollary 4. Corollary 4 is a direct consequence of Corollaries 1 and 3 and Theorem 3. ■

https://doi.org/10.1017/S0266466602181047 Published online by Cambridge University Press