

Department of Economics

A Bootstrap Invariance Principle for Highly Nonstationary Long Memory Processes

George Kapetanios

Working Paper No. 507

February 2004

ISSN 1473-0278



Queen Mary
University of London

A Bootstrap Invariance Principle for Highly Nonstationary Long Memory Processes

G. Kapetanios*
Queen Mary, University of London.

February 10, 2004

Abstract

This paper presents an invariance principle for highly nonstationary long memory processes, defined as processes with long memory parameter lying in $(1, 1.5)$. This principle provides the tools for showing asymptotic validity of the bootstrap in the context of such processes.

JEL Codes: C15, C22

Keywords: Long Memory, Bootstrap

1 Introduction

In recent years the popularity of the bootstrap has increased dramatically in econometrics and statistics. Decreases in computational costs and theoretical results showing that the bootstrap is valid under mild conditions and may improve upon asymptotic approximations of small sample probability distributions of various statistics have contributed to this popularity.

A parallel development in the econometric literature is the exploration of highly persistent time series processes. Chief among these are random walk processes. Another class of highly persistent processes that encompasses random walks are long memory processes. Specific instances of these processes

*Department of Economics, Queen Mary, University of London, Mile End Rd., London E1 4NS. Email: G.Kapetanios@qmul.ac.uk

were introduced in the statistical literature by Granger and Joyeux (1980) but were known in fields such as climatology and hydrology since the 1950's.

Despite rapid developments in both the study of the bootstrap and long memory processes, there is little theoretical work on applying the bootstrap to long memory processes. Andrews and Lieberman (2002) presents results which show that the bootstrap can provide improvements upon the asymptotic approximation to probability distributions of covariance parameter estimates for stationary long memory processes. Hidalgo (2003) suggests a new bootstrap technique based on resampling in the frequency domain which is valid for stationary long memory processes (see also Lazarova (2003)). This work concentrates on stationary long memory processes. There is no work to the best of the author's knowledge that either suggests new bootstrap procedures or deals with the validity of existing bootstrap procedures for nonstationary long memory processes.

This paper aims to provide a new bootstrap procedure valid for a large class of highly nonstationary long memory processes. The term 'highly nonstationary' is made exact by assuming that the long memory parameter d lies in $(1, 1.5)$. Our work is a simple extension of the work by Park (2002) on an invariance principle for the sieve bootstrap for Brownian motions. We extend that work to long memory processes. To achieve that we use a strong approximation result for fractional Brownian motions provided by Konstantopoulos and Sakhanenko (2003) extending earlier work by Sakhanenko (1980) used by Park (2002). Armed with this strong approximation result we prove the validity of the following bootstrap procedure for long memory processes. We estimate semiparametrically the long memory parameter of the process. Then, we apply the difference operator $(1 - L)^d$ to the data to obtain a stationary process. We then use the sieve bootstrap to resample from this stationary process. Finally, we cumulate the resampled data using $(1 - L)^{-d}$. The main technical issue with this rather intuitive resampling procedure is the presence of an estimate of d rather than the true d . We provide conditions under which estimating d does not matter for the validity of the bootstrap.

The relevance of our results is clear. It enables valid inference through the bootstrap for any statistic for highly nonstationary long memory processes. The structure of the paper is as follows: Section 2 presents initial

results. Section 3 presents the invariance principle. Section 4 concludes. The Appendix contains the necessary proofs.

2 Basic Results

We consider the following general fractionally integrated model

$$(1 - L)^d y_t = u_t^d, \quad 1 > d < 3/2, \quad t = 1, \dots, T \quad (1)$$

u_t^d is an $I(0)$ process. We define $I(0)$ processes, following Davidson and De-Jong (2000), to be processes whose partial sums converge weakly to Brownian motion. We have the following model for u_t

$$u_t^d = \pi(L)\epsilon_t, \quad \pi(z) = \sum_{i=1}^{\infty} \pi_i \epsilon_{t-i} \quad (2)$$

The model for y_t can be written as an infinite moving average in terms of u_t^d

$$y_t = \sum_{i=0}^{\infty} a_i^d u_{t-i}^d \quad (3)$$

where $a_i^d = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)}(-1)^i$. It can equivalently be written as an infinite autoregression given by

$$y_t = \sum_{i=1}^{\infty} b_i^d y_{t-i} + u_t^d \quad (4)$$

where $b_i^d = -\frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}$. These results follow from the binomial expansion for $(1 - L)^d$ which is formally valid for any real d . As a matter of definition the operator $(1 - L)^{-d}$, $d > 0$ applied to an $I(0)$ process will be used to denote cumulation of that process up to order d . More specifically, for $0 < d < 0.5$ the process obtained by applying $(1 - L)^{-d}$ to an $I(0)$ process is well defined and stationary. For $1 < d < 1.5$, $(1 - L)^{-d} = (1 - L)^{-1}(1 - L)^{-d+1}$ where $(1 - L)^{-1}$ is defined as a simple summation operator. This is a relatively general model. The main restriction for this model is the form of the long memory which is based on the binomial expansion of $(1 - L)^d$. Denote an estimate of d by \hat{d} . Strictly speaking \hat{d} should be denoted as \hat{d}_T but for the sake of notational brevity the subscript is dropped. We assume the following:

Assumption 1 $\hat{d} - d = O_p(T^{-\phi})$, $\phi > 0$.

Assumption 2 (a) ϵ_t are *i.i.d.* random variables such that $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, $E|\epsilon_t|^r < \infty$ for $r = \max(4, \frac{2+\phi}{2\phi})$

(b) $\pi(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{k=0}^{\infty} |k|^s |\pi_k| < \infty$ for some $s \geq 1$

We comment on the above assumptions: Assumption 1 is general and does not specify a particular estimator of the parameter d . It assumes consistency and some rate of convergence for the estimator. Note that most of the literature has concentrated on the properties of estimators of d when d lies in $(-0.5, 0.5)$. However, if we assume that d lies in $(1, 1.5)$ then by simply differencing the data we obtain an $I(d^*)$ process for which estimation of $d^* = d - 1$ has all the desirable properties which readily translate to d . However, Assumption 1 is not as general as it may appear at first. For example, it does not allow for estimators whose rate of convergence depends on d . This could introduce problems in the bootstrap similar in nature to those discussed in Inoue and Kilian (2003) which we wish to avoid by undertaking a simpler analysis¹. So, the estimator for $d \in (1.25, 1.50)$ suggested by Robinson (1994) would not satisfy Assumption 1. Of course, we can always restrict our analysis to an interval for d in which assumption 1 holds for a particular estimator. In the case of the semiparametric estimator by Robinson (1994) this would be $(1, 1.25)$. Note that a possible estimator for d may be based on a regression such as (19), in the following section, which would be jointly estimated with d in a similar way that $ARFIMA(p, d, 0)$ models with finite lag order are estimated. A conditional sum of squares estimator such as that proposed by Chung and Baillie (1993) can be straightforwardly applied. As estimation of $ARFIMA(p, d, 0)$ yield, in general, $T^{-1/2}$ consistent estimates of d , it is likely that such an estimator would satisfy assumption 1. Nevertheless a rigorous proof of this result is beyond the scope of this paper. Assumption 2 is a standard assumption for $I(0)$ time series. The only noteworthy feature is the requirement that $E|\epsilon_t|^r < \infty$ for $r = \max(4, \frac{2+\phi}{2\phi})$. This introduces a tradeoff between the rate of convergence of \hat{d} and the existence of higher moments of ϵ .

¹Note, however, that Inoue and Kilian (2003) clearly indicate that a rate of convergence that depends on d is not necessarily a cause for bootstrap invalidity.

We define $x_t^d = (1 - L)^{-d+1}\epsilon_t$. Hence x_t^d is $I(d - 1)$. We also define $z_t^d = (1 - L)^{-d+1}u_t^d$. Define the normalised sum

$$W_{T,d}(u) = \frac{1}{T^{d-1/2}} \sum_{i=1}^{[Tu]} x_t^d \equiv \frac{1}{T^{d-1/2}} (1 - L)^{-d} \epsilon_t, \quad 0 \leq u \leq 1 \quad (5)$$

Define also $W_d(r)$ as fractional Brownian motion. A representation for this process is given by

$$\frac{1}{\Gamma(d)U_d^{1/2}} \left(\int_0^\xi (\xi - s)^{d-1} dB(s) + \int_{-\infty}^0 [(\xi - s)^{d-1} - (-s)^{d-1}] dB(s) \right) \quad (6)$$

where

$$U_d = \frac{1}{\Gamma(d)^2} \left(\frac{1}{2d-1} + \int_0^\infty [(1 + \tau)^{d-1} \tau^{d-1}]^2 d\tau \right) \quad (7)$$

$\sigma_T^2 \sim \sigma_u^2 U_d T^{2d-1}$ and $B(u)$ is a standard Brownian motion for $d \geq 1/2$. Finally define

$$V_{T,d}(u) = \frac{1}{T^{d-1/2}} \sum_{i=1}^{[Tu]} z_t^d \equiv \frac{1}{T^{d-1/2}} (1 - L)^{-d} u_t \quad (8)$$

Then,

$$W_{T,d}(u) \xrightarrow{d} W_d(u) \quad (9)$$

For a proof of this see, e.g., Davidson and DeJong (2000). Lemma 6 shows that from this we also get via the Beveridge Nelson decomposition

$$V_{T,d}(u) \xrightarrow{d} V_d(u) \equiv \sigma\pi(1)W_d(u) \quad (10)$$

To construct our bootstrap invariance principle we need a stronger result than convergence in distribution. Firstly, we note that from the Skorohod representation theorem we have that there exists a probability space (Ω, \mathcal{F}, P) supporting a process $W_{T,d}(u)'$ and $W_d(u)$ such that $W_{T,d}(u)'$ and $W_{T,d}(u)$ have the same distribution and

$$W_{T,d}(u)' \xrightarrow{a.s.} W_d(u) \quad (11)$$

However, we have a stronger result that is needed for what follows. In particular, by Corollary 4 of Konstantopoulos and Sakhanenko (2003), we have that

$$\sup_u |W_{T,d}(u) - W_d(u)| = o_{a.s.}(T^{1/r-1/2}) + O_{a.s.}(T^{\max\{\beta,\gamma\}-d+1-1/2}(\log T)^{1/2}), \quad (12)$$

where the constants $\beta \leq 1/2 + d - 1$ and $\gamma \geq 0$ depend on the behaviour of the constants a_i of the MA representation of $(1 - L)y_t$. In particular for the case of long memory of the form we consider, i.e. based on the binomial expansion of $(1 - L)^d$ we have

$$\sup_u |W_{T,d}(u) - W_d(u)| = o_{a.s.}(T^{1/r-1/2}) + O_{a.s.}(T^{-d+1}(\log T)^{1/2}), \quad (13)$$

Note that this is a worse approximation than the best possible rates implied by (12). From this it follows

$$P\left(\sup_u |W_{T,d}(u) - W_d(u)|\right) = o(1) \quad (14)$$

3 The Bootstrap Procedure

The bootstrap algorithm we suggest consists of the following operations:

- Estimate d using any estimator that satisfies Assumption 1. Denote the estimate by \hat{d} .
- Difference y_t \hat{d} times. Denote the resulting series by $u_t^{\hat{d}}$.
- Fit a sieve $AR(p_T)$ model on this process. Denote the residual of this operation by $\hat{\epsilon}_{t,p_T}^{\hat{d}}$ and a bootstrap resample from $\left\{\hat{\epsilon}_{t,p_T}^{\hat{d}}\right\}_{t=1}^T$ by $\left\{\epsilon_t^{\hat{d}*}\right\}_{t=1}^T$.
- Finally, cumulate the bootstrap sample denoted $u_t^{\hat{d}*} = \hat{\alpha}^{\hat{d}}(L)\epsilon_t^{\hat{d}*}$ \hat{d} times. $\hat{\alpha}^{\hat{d}}(L)$ denotes the estimated parameters from the sieve regression where the last operation of the above algorithm is equivalent (apart from the normalisation) to constructing the partial sums

$$V_{T,\hat{d}}^*(r) = \frac{1}{T^{d-1/2}}(1-L)^{-\hat{d}}u_t^{\hat{d}*} \quad (15)$$

Define also

$$W_{T,\hat{d}}^*(r) = \frac{1}{T^{d-1/2}}(1-L)^{-\hat{d}}\epsilon_t^{\hat{d}*} \quad (16)$$

Before we discuss in detail the structure of the sieve bootstrap we make a comment on bootstrap formalisation. Once a realisation of the process $\{\epsilon_t\}_{t=1}^T$ has occurred, it constitutes a population for the bootstrap. Formally, this realisation defines a probability space denoted $(\Omega^*, \mathcal{F}^*, P^*)$ in which the random sequences $\{\epsilon_t^{\hat{d}*}\}_{t=1}^T$ are defined. In this probability space the estimate of d as a function of T is simply a sequence of constants and is not stochastic. Expectations defined in this probability space will be denoted as $E^*(\cdot)$ to distinguish them from expectations defined in the probability space on which $\{\epsilon_t\}_{t=1}^T$ is defined. The same will hold for convergence in probability and distribution as well as order in probability which will be denoted by $\xrightarrow{p^*}$, $\xrightarrow{d^*}$, $O_{p^*}(\cdot)$ and $o_{p^*}(\cdot)$ respectively.

We now discuss estimation of the sieve. We consider the following long autoregression to approximate (2)

$$u_t^d = \alpha_1 u_{t-1}^d + \alpha_2 u_{t-2}^d + \dots + \alpha_{p_T} u_{t-p_T}^d + \epsilon_{t,p_T}^d \quad (17)$$

Estimation of this would give residuals denoted $\hat{\epsilon}_{t,p_T}^d$. Also define the following

$$u_t^d = \alpha_{p_T,1} u_{t-1}^d + \alpha_{p_T,2} u_{t-2}^d + \dots + \alpha_{p_T,p_T} u_{t-p_T}^d + e_{t,p_T}^d \quad (18)$$

where the coefficients $\alpha_{p_T,j}$, $j = 1, \dots, p_T$ are defined so that e_{t,p_T}^d is uncorrelated with u_{t-j}^d , $j = 1, \dots, p_T$. This will be used in the Appendix. Given $u_t^{\hat{d}}$ we fit the model

$$u_t^{\hat{d}} = \alpha_1 u_{t-1}^{\hat{d}} + \alpha_2 u_{t-2}^{\hat{d}} + \dots + \alpha_{p_T} u_{t-p_T}^{\hat{d}} + \epsilon_{t,p_T}^{\hat{d}} \quad (19)$$

Estimation of this gives parameter estimates denoted $\hat{\alpha}_{p_T,j}$, $j = 1, \dots, p_T$. Residuals are denoted $\hat{\epsilon}_{t,p_T}^{\hat{d}}$. We estimate (19) by Yule-Walker, and resample from the residuals. Ordinary least squares could be used instead with minimal changes in the proofs of our results. We make the following assumption concerning the choice of the lag order

Assumption 3 $p_T \rightarrow \infty$ and $p_T = o(T^{\phi/2})$

This assumption is not restrictive and we again see a tradeoff between the lag order and the rate of convergence of \hat{d} . We do not place a lower bound on the rate of growth of p_T . A data dependent procedure for selecting p_T could be used, such as e.g. Akaike's information criterion or the Bayesian information criterion. Ng and Perron (1995) show that under a similar setup these criteria lead to $p_T = O_p(\ln T)$. The following theorems form the core of our results and establish validity of the bootstrap. They are proven in the Appendix.

Theorem 4 $W_{T,\hat{d}}^* \xrightarrow{d^*} W_d$ *in probability*

Theorem 5 $V_{T,\hat{d}}^* \xrightarrow{d^*} V_d$ *in probability*

To clarify the notation above the convergence $W_{T,\hat{d}}^* \xrightarrow{d^*} W_d$ occurs in the probability space $(\Omega^*, \mathcal{F}^*, P^*)$. However, the sieve estimation as well as estimation of d occurs in the probability space (Ω, \mathcal{F}, P) where $\{\epsilon_t\}_{t=1}^T$ is defined. Results for these stochastic quantities are of the 'in probability' nature, hence the 'in probability' characterisation in the theorems.

4 Conclusion

This paper has provided a bootstrap invariance principle for a sieve bootstrap applied to highly nonstationary long memory processes. Our basic approach is intuitive. We estimate the long memory parameter, d , of the sample and then translate the process to the $I(0)$ space where a standard sieve bootstrap is applied prior to cumulating the data to obtain a bootstrap resample of the long memory process. The main technical difficulty of our work lies in the fact that d is estimated rather than known. In the latter case our work would follow straightforwardly from the work of Park (2002) who provide the rudiments of the analysis of the $I(0)$ stationary sieve bootstrap and the work of Konstantopoulos and Sakhanenko (2003) who provide a strong approximation result for fractional Brownian motion. For the case of estimated d

we see that there are two tradeoffs between the rate of convergence of \hat{d} and (i) the existence of higher moments of the innovations of the $I(0)$ stationary linear process driving the long memory process and (ii) the lag order of the sieve bootstrap.

We conjecture that a strong approximation result should hold for fractional Brownian motion with Hurst coefficient $H < 1/2$ where $H = d - 1/2$. Such a result would enable the straightforward extension of our results to all nonstationary long memory processes ($0.5 < d < 1.5$). This is clearly a topic for future research.

Appendix

Proof of Theorem 4

We can obtain the result

$$W_{T,d}^* \xrightarrow{d} W_d \quad (20)$$

where $W_{T,d}^*(r) = \frac{1}{T^{d-1/2}}(1-L)^{-d}\epsilon_t^{\hat{d}*}$ from

$$P\left(\sup_r W_{T,d}^*(r) - W_d\right) = o(1) \quad (21)$$

This follows from Corrolary 4 of Konstantopoulos and Sakhanenko (2003) if $E^* \left| \epsilon_t^{\hat{d}*} \right|^r < \infty$ for some $r > 2$. But Lemma 7 provides this result. Then we need

$$\sup \left| W_{T,\hat{d}}^* - W_{T,d}^* \right| = o_p(1) \quad (22)$$

But, by Lemma 10 this holds and the theorem is proven.

Proof of Theorem 5

By Lemma 11

$$\frac{1}{T^{d-1/2}}(1-L)^{-d}u_t^{\hat{d}*} \xrightarrow{d} V_d \quad (23)$$

But by Lemma 10

$$\sup_t \left| V_{T,d}^* - \frac{1}{T^{d-1/2}} (1-L)^{-d} u_t^{\hat{d}^*} \right| = o_p(1) \quad (24)$$

Hence the result is proven.

Lemmas

Lemma 6 $V_{T,d} \xrightarrow{d} V_d \equiv \sigma\pi(1)W_d$

Proof. By using the Beveridge Nelson decomposition (cf. Theorem 3.3 of Park (2002)) we write

$$u_t = \pi(1)\epsilon_t + (\tilde{u}_{t-1} - \tilde{u}_t) \quad (25)$$

where $\tilde{u}_t = \sum_{k=0}^{\infty} \tilde{\pi}_k \epsilon_{t-k}$, and $\tilde{\pi}_k = \sum_{i=k+1}^{\infty} \pi_i$. Under assumption 1 we have that $\sum_{k=0}^{\infty} |\tilde{\pi}_k| < \infty$ as shown by Phillips and Solo (1992). This further implies that the process has absolutely summable autocovariances by, say, (3.3.19) of Hamilton (1994). Hence we need to show that

$$\frac{1}{T^{d-1/2}} (1-L)^{-d} (\tilde{u}_{t-1} - \tilde{u}_t) = \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_0 - \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_{[Tr]} = o_p(1) \quad (26)$$

But this follows if we show that

$$\max_t \frac{1}{T^{d-1/2}} \tilde{z}_t^d = \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_{[Tr]} \quad (27)$$

is $o_p(1)$. This readily follows by (20) of Phillips and Solo (1992) if $E[(\tilde{z}_t^d)^2] < \infty$. But $\sum_{i=0}^{\infty} (a_i^{d-1})^2 < \infty$ by stationarity of a $I(d-1)$ process with $d < 3/2$. Hence, by Lemma 1 of Wright (1995) $E[(\tilde{z}_t^d)^2] < \infty$ and the result is proven. ■

Lemma 7 $E^* \left| \epsilon_t^{\hat{d}^*} \right|^r < \infty$ for some $r > 2$

Proof. We wish to show that

$$\frac{1}{T} \sum_{t=1}^T \left| \hat{\epsilon}_{t,p_T}^{\hat{d}} - \frac{1}{T} \sum \hat{\epsilon}_{t,p_T}^{\hat{d}} \right|^r \xrightarrow{p} 0 \quad (28)$$

Write

$$\frac{1}{T} \sum_{t=1}^T \left| \hat{\epsilon}_{t,p_T}^{\hat{d}} - \frac{1}{T} \sum \hat{\epsilon}_{t,p_T}^{\hat{d}} \right|^r < c(A_T + B_T + 2C_T + D_T + E_T) \quad (29)$$

where

$$A_T = \frac{1}{T} \sum_{t=1}^T |\epsilon_t|^r \quad (30)$$

$$B_T = \frac{1}{T} \sum_{t=1}^T |\epsilon_{p_T,t}^d - \epsilon_t|^r \quad (31)$$

$$C_T = \frac{1}{T} \sum_{t=1}^T |\hat{\epsilon}_{p_T,t}^d - \epsilon_{p_T,t}^d|^r \quad (32)$$

$$D_T = \left| \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_{p_T,t}^d \right| \quad (33)$$

$$E_T = \frac{1}{T} \sum_{t=1}^T \left| \hat{\epsilon}_{p_T,t}^{\hat{d}} - \epsilon_{p_T,t}^d \right|^r \quad (34)$$

where We need to show that B_T, C_T, E_T are $o_p(1)$ and A_T, D_T are $O_p(1)$. All these results apart from that for E_T are shown to be true in Lemma 3.2 of Park (2002). E_T is $o_p(1)$ by Lemma 8 ■

Lemma 8 $\frac{1}{T} \sum_{t=1}^T \left| \hat{\epsilon}_{p,t}^{\hat{d}} - \epsilon_{p,t}^d \right|^r = o_p(1)$

Proof.

$$\hat{\epsilon}_{p,t}^{\hat{d}} - \epsilon_{p,t}^d = u_t^{\hat{d}} - \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} u_{t-k}^{\hat{d}} - u_t^d + \sum_{k=1}^{p_T} \alpha_k u_{t-k}^d = \quad (35)$$

$$u_t^{\hat{d}} - \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} u_{t-k}^{\hat{d}} + \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} u_{t-k}^d - \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} u_{t-k}^d - u_t^d + \quad (36)$$

$$\sum_{k=1}^{p_T} \alpha_k u_{t-k}^d - \sum_{k=1}^{p_T} \alpha_{p_T,k} u_{t-k}^d + \sum_{k=1}^{p_T} \alpha_{p_T,k} u_{t-k}^d \quad (37)$$

So

$$\frac{1}{T} \sum_{t=1}^T \left| \hat{\epsilon}_{p,t}^{\hat{d}} - \epsilon_{p,t}^d \right|^r \leq c \left(\frac{1}{T} \sum_{t=1}^T \left| u_t^{\hat{d}} - u_t^d \right|^r + \frac{1}{T} \sum_{t=1}^T \left| \sum_{k=1}^{p_T} (\hat{\alpha}_{p_T,k} - \alpha_{p_T,k}) u_{t-k}^d \right|^r \right) + \quad (38)$$

$$c \left(\frac{1}{T} \sum_{t=1}^T \left| \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} (u_{t-k}^{\hat{d}} - u_{t-k}^d) \right|^r + \frac{1}{T} \sum_{t=1}^T \left| \sum_{k=1}^{p_T} (\alpha_{p_T,k} - \alpha_k) u_{t-k}^d \right|^r \right) \quad (39)$$

We have that the first term of the RHS of 38 is $o_p(1)$ by (4.17) of Wright (1995). The second term is dominated by

$$\max_{1 \leq k \leq p} |\hat{\alpha}_{p_T,k} - \alpha_{p_T,k}|^r \frac{p_T}{T} \left(\sum_{t=1}^{T-1} |u_t^d|^r + \sum_{t=-1}^{1-p_T} |u_t^d|^r \right) = O_p \left(\frac{p_T}{T^\phi} \times \frac{p_T}{T} \times T \right) = o_p(1) \quad (40)$$

The above equality follows by Lemma 9. The fourth term of (39) is $o_p(1)$ by

$$E \left| \sum_{k=1}^{p_T} (\alpha_{p_T,k} - \alpha_k) u_{t-k}^d \right|^r \leq E |u_t^d|^r \left(\sum_{k=1}^{p_T} |\alpha_{p_T,k} - \alpha_k| \right)^r = o(p_T^{-rs}) \quad (41)$$

The third term can be written as

$$\frac{1}{T} \sum_{t=1}^T \left| (\hat{d} - d) \sum_{k=1}^{p_T} \hat{\alpha}_{p_T,k} \sum_{j=1}^{t-k} \bar{a}_j u_{t-k-j}^d \right|^r \quad (42)$$

by (4.18) of Wright (1995) where we have dropped a term in $(\hat{d} - d)^2$ which is dominated by the above term in probability. Also $\bar{a}_j \sim j^{-1}$ as $j \rightarrow \infty$. Now for $r = \frac{2+\phi}{2\phi}$.

$$\frac{1}{T} \sum_{t=1}^T \left| (\hat{d} - d) \sum_{k=1}^{p_T} \hat{\alpha}_{p_T, k} \sum_{j=1}^{t-k} \bar{a}_j u_{t-k-j}^d \right|^{\frac{2+\phi}{2\phi}} \leq \quad (43)$$

$$(\hat{d} - d)^{\frac{2+\phi}{2\phi}} \max_{k,j} \left[(\hat{\alpha}_{p_T, k} \bar{a}_j)^{\frac{2+\phi}{2\phi}} \right] \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^{p_T} \sum_{j=1}^{t-k} |u_{t-k-j}^d|^{\frac{2+\phi}{2\phi}} = \quad (44)$$

$$O_p(T^{-1-\phi/2}) O_p(\max(\hat{\alpha}_{p_T, k} \bar{a}_j)) O_p(p_T T) \quad (45)$$

But $\max_{k,j} \left[(\hat{\alpha}_{p_T, k} \bar{a}_j)^{\frac{2+\phi}{2\phi}} \right] = O_p(1)$ by Lemma 9 and $\max_{k,j} \left[(\alpha_{p_T, k} \bar{a}_j)^{\frac{2+\phi}{2\phi}} \right] = O_p(1)$. Also $O_p(p_T T) = o_p(T^{1+\phi/2})$. So, overall the above term is $o_p(1)$. Hence the result is proven. ■

Lemma 9 $\hat{\alpha}_{p_T, k} - \alpha_{p_T, k} = O_p(T^{-\phi})$ and $\max_{1 \leq k \leq p} |\hat{\alpha}_{p_T, k} - \alpha_{p_T, k}| = o_p(T^{-\phi/2})$.

Proof. We consider estimation of the long autoregression by Yule-Walker. If the regressors were u_t^d the result follows immediately by Hannan and Kavalieris (1983). Combining this with Theorem 1 of Wright (1995) which shows that

$$\frac{1}{T} \sum_{t=1}^{T-k} u_t^d u_{t+k}^d - \frac{1}{T} \sum_{t=1}^{T-k} u_t^d u_{t+k}^d = O_p(T^{-\phi}) \quad (46)$$

gives the first part of the Lemma. The second part readily follows if we note that

$$\max_{1 \leq k \leq p_T} |\hat{\alpha}_{p_T, k} - \alpha_{p_T, k}| \leq \sum_{k=1}^{p_T} |\hat{\alpha}_{p_T, k} - \alpha_{p_T, k}| \quad (47)$$

■

Lemma 10 $\max_t \frac{1}{T^{\hat{d}-1/2}} |(1-L)^{-\hat{d}} u_t - (1-L)^{-d} u_t| = o_{p^*}(1)$ where u_t is an $I(0)$ process defined in the bootstrap probability space $(\Omega^*, \mathcal{F}^*, P^*)$

Proof.

$$(1-L)^{-\hat{d}} u_t - (1-L)^{-d} u_t = (1-L)^{-d} \left((1-L)^{-(\hat{d}-d)} - 1 \right) u_t = (1-L)^{-d} (d-\hat{d}) v_t \quad (48)$$

where $v_t = \sum_{j=1}^{\infty} \bar{a}_j u_{t-j}$ by (4.18) of Wright (1995) where $u_t = 0, t \leq 0$. Note that $\bar{a}_j \sim j^{-1}$ as $j \rightarrow \infty$ and hence $\sum_{j=1}^{\infty} \bar{a}_j^2 < \infty$. Define

$$v_t^T = \sum_{j=1}^{\infty} \frac{\bar{a}_j}{\ln T} u_{t-j} = \sum_{j=1}^{\infty} \bar{a}_j^T u_{t-j} \quad (49)$$

Since $\sum_{j=1}^{\infty} |\bar{a}_j^T| < \infty$ this is a process with absolutely summable autocovariances $\forall T$. By the Marcinkiewicz-Zygmund inequality

$$E^* |v_t^T|^r \leq c \left(\sum_{j=1}^{\infty} (\bar{a}_j^T)^2 \right)^{r/2} E^* |u_t|^r < \infty \quad (50)$$

for some constant c . Define

$$\bar{v}_t^T = \sum_{j=1}^{\infty} \frac{a_j^d}{T^{d-1/2}} v_{t-j}^T = \sum_{j=1}^{\infty} a_j^{T,d} v_{t-j}^T \quad (51)$$

where again $v_t^T = 0, t \leq 0$. We have that $\sum_{j=1}^{\infty} (a_j^{T,d})^2 < \infty \forall T$ by the order of magnitude of the variance of an $I(d)$ process $1 < d < 1.5$. By a further application of the Marcinkiewicz-Zygmund inequality we have that

$$E^* |\bar{v}_t^T|^r \leq c \left(\sum_{j=1}^{\infty} (a_j^{T,d})^2 \right)^{r/2} E^* |v_t^T|^r < \infty \quad (52)$$

Now we have

$$P^* \left(\max_t \left| \ln T (\hat{d} - d) \bar{v}_t^T \right| > \delta \right) \leq \sum_t P^* \left(\left| \ln T (\hat{d} - d) \bar{v}_t^T \right| > \delta \right) = \quad (53)$$

$$TP^* \left(\left| \bar{v}_t^T \right| > \delta / (\ln T (\hat{d} - d)) \right) < \frac{1}{\delta^r} T (\ln T)^r (\hat{d} - d)^r E^* |\bar{v}_t^T|^r \quad (54)$$

For $r > \frac{1}{\phi}$ the above term is $o_{p^*}(1)$ and hence the result is proven. Note that we are allowed to move $(\hat{d} - d)$ on the RHS of the inequality in the probability above because for the probability space supporting the bootstrap probability measure P^* the term $(\hat{d} - d)$ is not stochastic but a sequence of constants implicitly depending on T . ■

Lemma 11 $\frac{1}{T^{d-1/2}}(1-L)^{-d}u_t^{\hat{d}*} \xrightarrow{d^*} V_d \equiv \sigma\pi(1)W_d$

Proof. Retracing the steps of Lemma 6 we have using the Beveridge Nelson decomposition

$$\tilde{u}_t^{\hat{d}^*} = \frac{1}{\hat{\alpha}(1)} \sum_{k=1}^{p_T} \left(\sum_{i=k}^{p_T} \hat{\alpha}_{p_T, i} \right) u_{t-k+1}^{\hat{d}^*} \quad (55)$$

Then, we need to show

$$\frac{1}{T^{d-1/2}} (1-L)^{-d} (\tilde{u}_{t-1}^{\hat{d}^*} - \tilde{u}_t^{\hat{d}^*}) = \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_0^{\hat{d}^*} - \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_{[Tr]}^{\hat{d}^*} = o_p(1) \quad (56)$$

This follows if we show

$$\max_t \frac{1}{T^{d-1/2}} \tilde{z}_t^{\hat{d}^*} \equiv \frac{1}{T^{d-1/2}} (1-L)^{-d+1} \tilde{u}_{[Tr]}^{\hat{d}^*} = o_p(1) \quad (57)$$

Now,

$$P^* \left(\max_t \left| \frac{1}{T^{d-1/2}} \tilde{z}_t^{\hat{d}^*} \right| > \delta \right) \leq \sum_t P^* \left(\left| \frac{1}{T^{d-1/2}} \tilde{z}_t^{\hat{d}^*} \right| > \delta \right) = \quad (58)$$

$$TP^* \left(\left| \frac{1}{T^{d-1/2}} \tilde{z}_t^{\hat{d}^*} \right| > \delta \right) < \frac{1}{\delta^2} T (T^{1/2-d})^2 E^* \left| \tilde{z}_t^{\hat{d}^*} \right|^2 \quad (59)$$

which will be $o_p(1)$ if $E^* \left| \tilde{u}_t^{\hat{d}^*} \right|^2 < \infty$ since $d-1 < 0.5$. But by (55) we have, by Minkowski's inequality

$$E^* \left| \tilde{u}_t^{\hat{d}^*} \right|^r \leq \left(\frac{1}{|\hat{\alpha}(1)|} \sum_{k=1}^{p_T} \left(\sum_{i=k}^{p_T} |\hat{\alpha}_{p_T, i}| \right) \right)^r E^* \left| u_t^{\hat{d}^*} \right|^r, \quad r > 2 \quad (60)$$

So we need to prove that $E^* \left| u_t^{\hat{d}^*} \right|^2 < \infty$. But by the Marcinkiewicz-Zygmund inequality we have that

$$E^* \left| u_t^{\hat{d}^*} \right|^r \leq c \left(\sum_{j=1}^{\infty} (\hat{\alpha}_{p_T, j})^2 \right)^{r/2} E^* \left| \epsilon_t^{\hat{d}^*} \right|^r \quad (61)$$

where we set $\hat{\alpha}_{p_T,i} = 0$ for $p_T < i$. By consistency of $\hat{\alpha}_{p_T,k}$ from Lemma 9 $\sum_{j=1}^{\infty} (\hat{\alpha}_{p_T,i})^2 \leq \infty$ and the result is proven. ■

References

- ANDREWS, D. W. K., AND O. LIEBEMAN (2002): “Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes,” Cowles Foundation Discussion Paper No. 1378.
- CHUNG, C. F., AND R. T. BAILLIE (1993): “Small Sample Bias in Conditional Sum of Squares Estimators of Fractionally Integrated ARMA Models,” *Empirical Economics*, 18, 791–806.
- DAVIDSON, J., AND R. DEJONG (2000): “The Functional Central Limit Theorem and Convergence to Stochastic Integrals II: Fractionally Integrated Processes,” *Econometric Theory*, 16, 621–642.
- GRANGER, C. W. J., AND R. JOYEUX (1980): “An Introduction to Long Memory Time Series Models and Fractional Differencing,” *Journal of Time Series Analysis*, 1, 15–39.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press.
- HANNAN, E. J., AND L. KAVALIERIS (1983): “The Convergence of Autocorrelations and Autoregressions,” *Australian Journal of Statistics*, 25, 287–297.
- HIDALGO, J. (2003): “An alternative bootstrap to moving blocks for time series regression models,” *Journal of Econometrics*, 117, 369–399.
- INOUE, A., AND L. KILIAN (2003): “The continuity of the Limit Distribution in the Parameter of Interest Is Not Essential for the Validity of the Bootstrap,” *Econometric Theory*, 19, 944–961.
- KONSTANTOPOULOS, T., AND A. SAKHANENKO (2003): “Convergence and Convergence Rate to Fractional Brownian Motion for Weighted Random Sums,” Mimeo, University of Texas at Austin.

- LAZAROVA, S. (2003): “Testing for structural change in regression with long memory processes,” Mimeo, London School of Economics.
- NG, S., AND P. PERRON (1995): “Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag,” *Journal of the American Statistical Association*, 90, 268–281.
- PARK, J. Y. (2002): “An Invariance Principle for Sieve Bootstrap in Time Series,” *Econometric Theory*, 18, 469–490.
- PHILLIPS, P. C. B., AND V. SOLO (1992): “Asymptotics for Linear Processes,” *Annals of Statistics*, 20, 971–1001.
- ROBINSON, P. M. (1994): “Semiparametric Analysis of Long Memory Time Series,” *Annals of Statistics*, 22, 515–539.
- SAKHANENKO, A. (1980): “On Unimprovable Estimates of the Rate of Convergence in Invariance Principle,” in *Nonparametric Statistical Inference*, no. 32 in *Colloquia Mathematica Societatis Janos Bolyai*, pp. 779–783.
- WRIGHT, J. H. (1995): “Stochastic Orders of Magnitude Associated with Two Stage Estimators of Fractional ARIMA Systems,” *Journal of Time Series Analysis*, 16, 119–126.

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2004 George Kapetanios
All rights reserved.**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096 or Fax: +44 (0)20 8983 3580
Email: j.conner@qmul.ac.uk
Website: www.econ.qmul.ac.uk/papers/wp.htm**