UNIVERSITY OF EXETER

Economics Department Discussion Papers Series

ISSN 1473 – 3307

Type I and Type II Fractional Brownian Motions: a Reconsideration

James Davidson and Nigar Hashimzade

Paper number 08/16

URL: http://business-school.exeter.ac.uk/economics/papers/

Type I and Type II Fractional Brownian Motions: a Reconsideration

James Davidson* University of Exeter
Nigar Hashimzade University of Reading

Final Version, November 2008
Keywords: Fractional Brownian motion, long memory, ARFIMA, simulation.

Abstract

The so-called type I and type II fractional Brownian motions are limit distributions associated with the fractional integration model in which pre-sample shocks are either included in the lag structure, or suppressed. There can be substantial differences between the distributions of these two processes and of functionals derived from them, so that it becomes an important issue to decide which model to use as a basis for inference. Alternative methods for simulating the type I case are contrasted, and for models close to the nonstationarity boundary, truncating infinite sums is shown to result in a significant distortion of the distribution. A simple simulation method that overcomes this problem is described and implemented. The approach also has implications for the estimation of type I ARFIMA models, and a new conditional ML estimator is proposed, using the annual Nile minima series for illustration.

1 Introduction

The literature on long memory processes in econometrics (for recent examples, see *inter alia* Johansen and Nielsen (2008), Caporale and Gil-Alana (2008), Coakley, Dollery and Kellard (2008), Haldrup and Nielsen (2007)) has adopted two distinct models as a basis for the asymptotic analysis, the limit processes specified being known respectively as type I and type II fractional Brownian motion (fBM). These processes have been carefully examined and contrasted by Marinucci and Robinson (1999). When considered as real continuous processes on the unit interval, they can be defined respectively by

$$X(r) = \frac{1}{\Gamma(d+1)}\int_0^r (r-s)^d dB(s) + \frac{1}{\Gamma(d+1)}\int_{-\infty}^0 [(r-s)^d - (-s)^d] dB(s) \tag{1.1}$$

and

$$X^*(r) = \frac{1}{\Gamma(d+1)}\int_0^r (r-s)^d dB(s) \tag{1.2}$$

where $-\frac{1}{2} < d < \frac{1}{2}$ and B denotes regular Brownian motion. In other words, in the type II case the second term in (1.1) is omitted. It will be convenient to write the decomposition

$$X = X^* + X^{**} \tag{1.3}$$

*Department of Economics, University of Exeter, Streatham Court, Rennes Drive, Exeter EX4 4PU, UK. james.davidson@exeter.ac.uk, Tel. 01392 264517, Fax 07092 812113

where $X^{**}(r)$ is defined as the second of the two terms in (1.1). The processes X^* and X^{**} are Gaussian, and independent of each other, so we know that the variance of (1.1) will exceed that of (1.2). As shown by Marinucci and Robinson (1999), the increments of (1.1) are stationary, whereas those of (1.2) are not.

These processes are commonly motivated by postulating realizations of size n of discrete processes and considering the weak limits of normalized partial sums, as $n \to \infty$. Define

$$x_t = (1-L)^{-d} u_t \tag{1.4}$$

where we assume for the sake of exposition that $\{u_t\}_{-\infty}^{\infty}$ is an i.i.d. process with mean 0 and variance σ^2, and

$$(1-L)^{-d} = \sum_{j=0}^{\infty} b_j L^j \tag{1.5}$$

where, letting $\Gamma(\cdot)$ denote the gamma function,

$$b_j = \frac{\Gamma(d+j)}{\Gamma(d)\Gamma(1+j)}. \tag{1.6}$$

Defining the partial sum process

$$X_n(r) = \frac{1}{\sigma n^{1/2+d}} \sum_{t=1}^{[nr]} x_t \tag{1.7}$$

it is known that $X_n \xrightarrow{d} X$, where $\xrightarrow{d}$ denotes weak convergence in the space of measures on $D_{[0,1]}$, the space of cadlag functions of the unit interval equipped with the Skorokhod topology. (See for example Davidson and de Jong 2000). On the other hand, defining

$$u_t^* = 1(t>0)u_t \tag{1.8}$$

and x_t^* as the case corresponding to x_t in (1.4) when u_t^* replaces u_t, and then defining X_n^* like (1.7) with x_t^* replacing x_t, it is known that $X_n^* \xrightarrow{d} X^*$ (Marinucci and Robinson 2000).

The model in (1.8) is one that is often used in simulation exercises to generate fractionally integrated processes, as an alternative to the procedure of setting a fixed, finite truncation of the lag distribution in (1.4), common to every t. However, from the point of view of modelling real economic or financial time series, model (1.8) is obviously problematic. There is, in most cases, nothing about the date when we start to observe a series which suggests that we ought to set all shocks preceding it to 0. Such truncation procedures are common in time series modelling, but are usually justified by the assumption that the effect is asymptotically negligible. In this case, however, where the effect is manifestly not negligible in the limit, the choice of model becomes a critical issue.

The setting for this choice is the case where a Monte Carlo simulation is to be used to construct the null distribution of a test statistic postulated to be a functional of fBM. If model (1.8) is used to generate the artificial data, then the distribution so simulated will be the Type II case. However, if the observed data ought to be treated as drawn from (1.4), then the estimated critical values will be incorrect even in large samples. It then becomes of importance to know how large this error is.

Section 2 of the paper reviews and contrasts the main properties of these models. A leading difficulty in working with the type I model is to simulate it effectively, and as we show in Section 3, the fixed lag truncation strategy is not generally effective, except by expending a dramatically

large amount of computing resources. Since type I fBM has a harmonizable representation, another suggestion has been to use this to simulate the model, and then use a fast Fourier transform to recover the data in the time domain. However, we also show that this method cannot function effectively without large resources. Methods for generating type I processes do exist, for example circulant embedding and wavelet aproximations, but these are relatively difficult to implement in an econometric context. In Section 4 we suggest a new simulation method for type I processes, whose computational demands are trivial, and being implemented in the time domain adapts naturally to econometric modelling applications. The method is highly accurate when the data are Gaussian, and is always asymptotically valid.

Finally, we point out in Section 5 how the same approximation technique can be used to estimate ARFIMA time series models under the assumption that the true processes are of type I. This is in contrast to the usual time domain estimation by least squares, or conditional maximum likelihood, where the necessity of truncating lag distributions to match the observed data series implicitly (and perhaps inappropriately) imposes restrictions appropriate to the type II case. The method entails fitting some constructed regressors, whose omission will potentially bias the estimates in finite samples. The technique is illustrated with an application to the well-known series of annual Nile minima. Section 6 concludes the paper. Proofs are contained in Appendix A, and Appendix B exhibits some simulations of representative fractional Brownian functionals, under the two definitions.

The computations in this paper were carried out using the package Time Series Modelling 4 (Davidson 2008) which runs under the Ox 4/5 matrix programming system (Doornik 2006).

2 Properties of Fractional Brownian Motions

Our first task is to identify and contrast the distributions represented by (1.1) and (1.2). Since these are Gaussian with means of zero, this is simply a matter of determining variances and covariances of increments, and since

$$X(r_1)X(r_2) = \tfrac{1}{2}\left[X(r_1)^2 + X(r_2)^2 - (X(r_2) - X(r_1))^2\right]$$

a formula for the variance of an increment $X(r_2) - X(r_1)$ is sufficient to identify the complete covariance structure. It will further suffice, to motivate our discussion, to consider just the cases $r_1 = 0$ and $r_2 = r \in (0, 1]$. The formula

$$EX(r)^2 = V(d)r^{2d+1}$$

where

$$V(d) = \frac{1}{\Gamma(d+1)^2}\left(\frac{1}{2d+1} + \int_0^\infty ((1+\tau)^d - \tau^d)^2 d\tau\right) \tag{2.1}$$

is given by Mandelbrot and Van Ness (1968). However, for this formula to be operational a closed form for the integral in the second term is necessary. As we remark in the sequel, conventional numerical evaluations may suffer major inaccuracies. A proof of the closed-form representation

$$V(d) = \frac{\Gamma(1-2d)}{(2d+1)\Gamma(1+d)\Gamma(1-d)} \tag{2.2}$$

is given in Davidson and Hashimzade (2008). By contrast, the variance in the type II case is found by elementary arguments as

$$EX^*(r)^2 = V^*(d)r^{2d+1}$$

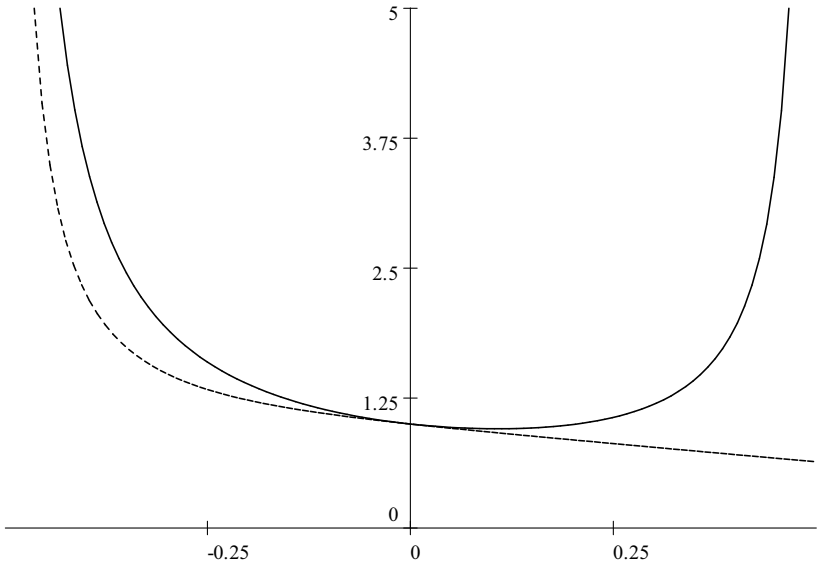

Figure 1: Plots of V (solid line) and V^* (dashed line) over $(-0.5, 0.5)$

where

$$V^*(d) = \frac{1}{(2d+1)\Gamma(d+1)^2}.$$

Plotting these formulae as functions of d (Figure 1) is the easiest way to see their relationship, and it is clear that, particularly for values of d close to 0.5, the differences can be substantial. While V diverges as $d \to 0.5$, V^* is declining monotonically over the same range, so that the variance of the second term in (1.1) comes to dominate that of the first term to an arbitrary degree.

It is easy to see how the distributions of functionals such as $\int_0^1 X dr$ and $\int_0^1 X^2 dr$ will differ correspondingly for these two models. The other important random variables arising in the asymptotic theory of estimators are stochastic integrals. Expressions of the form $\int_0^1 X_1 dX_2$ arise in the limit distributions of regression errors-of-estimate in models involving nonstationary series and possible long-memory error terms. The location parameter of this random variable is an important contributor to the degree of bias in the regression. The distribution theory for these random variables is studied in Davidson and Hashimzade (2008, 2009). For type I processes, the expected value is given in Davidson and Hashimzade (2008) Proposition 4.1 as

$$E\int_0^1 X_1 dX_2 = \sigma_{12}\frac{\Gamma(1-d_1-d_2)\sin\pi d_2}{\pi(d_1+d_2)(1+d_1+d_2)},$$

where $\sigma_{12} = E(X_1(1)X_2(1))$. On the other hand, by constructing the expectation as the limit of the normalized finite sum, we can quite easily show the following for type II processes X_1^* and X_2^*, where σ_{12} is defined analogously.

Proposition 2.1 $E\int_0^1 X_1^* dX_2^* = \dfrac{\sigma_{12} d_2}{(1+d_1+d_2)(d_1+d_2)\Gamma(1+d_1)\Gamma(1+d_2)}.$

In Figure 2 we show plots of these expressions, as d_2 varies over the interval $[0, \frac{1}{2})$, for $\sigma_{12} = 1$ and fixed $d_1 = 0.4$.

These large discrepancies clearly pose a very important issue - which of these models is the more appropriate for use in econometric inference? Marinucci and Robinson (1999) remark:

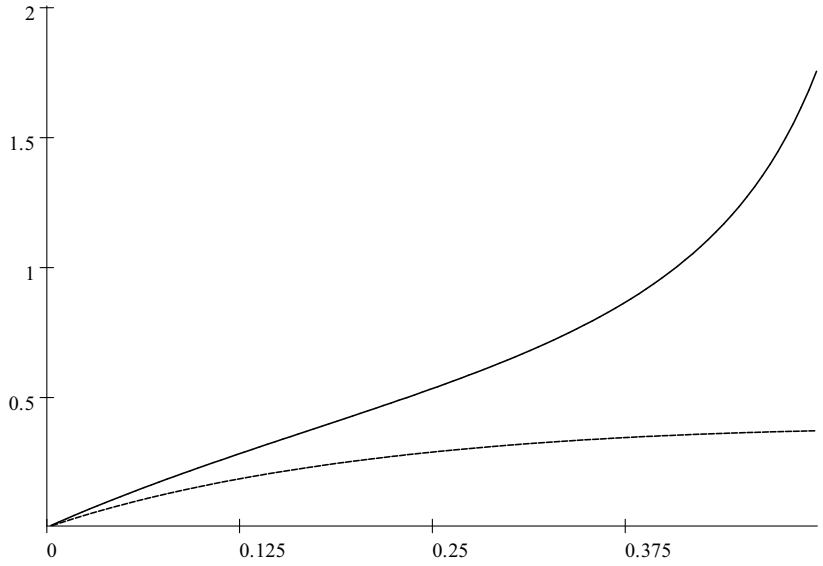

Figure 2: $E\int_0^1 X_1 dX_2$ (solid line) and $E\int_0^1 X_1^* dX_2^*$ (dashed line) as functions of d_2, with $\sigma_{12} = 1$, $d_1 = 0.4$

> "It is of some interest to note that [type II fBM] is taken for granted as the proper definition of fractional Brownian motion in the bulk of the econometric time series literature, whereas the probabilistic literature focuses on [type I fBM] This dichotomy mirrors differing definitions of nonstationary fractionally integrated processes..."

The feature of the type II model this last remark evidently refers to is that it incorporates the conventional integer integration models (I(1), I(2), etc.) neatly into a general framework. Letting d increase from 0 up to 1, and then 2 and beyond, yields a continuum of models, all nonstationary, but with continuously increasing 'memory'. An $I(1)$ model cannot be allowed to have an infinitely remote starting date, but must be conceived as a cumulation of increments initiated at date $t = 1$, with an initial condition x_0 that must be generated by a *different* mechanism. The view that this construction should apply seamlessly to the whole class of I(d) models leads naturally to the type II framework.

On the other hand, the type I framework requires us to keep cumulation (integer integration) and stationary long memory in conceptually separate compartments. In this view, a cumulation process must be assigned a finite start date, but its increments should then form a stationary sequence. In any autoregressive process, whether weakly or strongly dependent, stationarity combined with the assumption that the distribution is generated from i.i.d. shocks imposes a dependence on all past shocks, and in the case of long memory it implies non-negligible dependence on the 'infinite' past. This might appear unrealistic. However, we should not lose sight of the fact that this simple mathematical model is not necessarily the actual mechanism that 'nature' chooses to create sequences of data. The shocks are no more than the fictional concomitant of writing a stationary generation process in linear form. In practice, all we have to account for is the joint distribution of the finite sequence of sample increments. Stationarity of this distribution, in the linear fractional framework, implies the type I model.

By contrast, the type II model implies nonstationary increments of which the marginal distributions are dependent on the date *relative to the start of the observed sample*. This gives rise to a much less appealing data description. All realizations of the process would have to be found

at, or very close to, the unconditional mean of the process (i.e., 0) whenever we start to observe them. The obvious counterfactual is provided by discarding some initial observations from any process so generated which, obviously, produces a process requiring a description *different* from type 2. In practice there are plenty of nonlinear mechanisms that might generate stationary fractional processes well described by models of the ARFIMA class, of which the leading examples all involve some type of aggregation across units/agents. See for example Davidson and Sibbertsen (2004), Granger (1981), and also Byers, Davidson and Peel (1996, 2001) for an application.

The considerations we discuss here are a universal feature of time series modelling, but it has generally been possible to neglect them because the effects are of small order relative to sample size. This is true both in weakly dependent processes and in simple integrated processes. Long memory models are different, since choosing the wrong descriptive framework has asymptotic consequences, and exposes us to the hazard of incorrect inferences even in large samples. The practical value of asymptotic theory for fractionally integrated processes can only be to derive test statistics that must, in practice, be tabulated by simulation. This effort is of course compromised if the distributions we tabulate are different from those generated by 'nature'. If it is believed that the latter should realistically be treated as of type I, a suitable simulation algorithm becomes an essential prerequisite of useful research in this area. In the next section we review existing simulation methods, considering in particular the type of processes that they generate, and then go on to propose a new strategy which is simple to implement and appears very effective in practice.

3 Simulation Strategies For Type I Processes

Beran (1994) offers a number of suggestions for simulating long memory processes, in such a way as to reproduce the correct autocorrelation structure. However, he does not address the issues of stationarity and the role of presample influences. In this section we examine these methods, and others, with these issues in mind.

3.1 Using Presample Lags

A general procedure for generating a fractionally integrated series of length n is to apply, for $t = 1, \ldots, n$ and some fixed m, the formula

$$x_t = \sum_{j=0}^{m+t-1} b_j u_{t-j}, \tag{3.1}$$

where $\{u_{-m}, \ldots, u_n\}$ is a random sequence of suitable type, and $\{b_j, j = 0, \ldots, m\}$ is defined by (1.6). In the experiments reported in this paper, $\{u_t\}$ is always i.i.d. standard Gaussian. Choosing $m = 0$ and taking the formula in (1.7) to the limit will yield a type II process, as noted above. On the other hand, by choosing m large enough we should be able to approximate the type I process to any desired degree of accuracy. Note that the fixed lag length strategy of replacing $m + t - 1$ by m as the upper limit in (3.1) yields a stationary process, which might be viewed as desirable when attempting to approximate the true case $m = \infty$. However, it is clear that when m is large enough to achieve a good approximation, it is also large enough that the difference between the two cases is negligible. Therefore we do not consider this latter case explicitly.

Table 1 shows the standard deviations in 10,000 replications of the terminal points $X_n(1)$ of the process in (1.7) where x_t is generated by the model in (3.1) where $d = 0.4$, and $n = 1000$. For comparison, note the theoretical values: $\sqrt{V(0.4)} = 1.389$ and $\sqrt{V^*(0.4)} = 0.8401$. The

m	0	1000	3000	6000	9000
SD	0.843	0.996	1.036	1.108	1.137

Table 1: SDs of Terminal Values: Extended Lag Representation

coefficients converge so slowly, for values of d in this region, that the length of the presample needed for a close approximation to the type I process is infeasibly large.

3.2 Harmonic Representation

When u_t is i.i.d. Gaussian, the process x_t defined by (1.4) has the harmonic representation

$$x_t = \frac{\sigma}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{i\lambda t} \left(1 - e^{-i\lambda}\right)^{-d} W(d\lambda), \tag{3.2}$$

where i is the imaginary unit and W is a complex-valued Gaussian random measure with the properties

$$\begin{aligned} W(-d\lambda) &= \overline{W(d\lambda)} \\ E(W(d\lambda)) &= 0 \\ E(W(d\lambda)\overline{W(d\mu)}) &= \begin{cases} d\lambda, & \mu = \lambda \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

This process is stationary by construction. It is also shown in Davidson and Hashimzade (2007, Theorem 2.2) that the weak limit defined by (1.7) applied to the process (3.2) is type I fractional Brownian motion.

Therefore, we investigate a discrete form of (3.2) as a framework for simulation. Letting

$$g(\lambda) = \left(1 - e^{-i\lambda}\right)^{-d} \tag{3.4}$$

denote the transfer (frequency response) function of the process, define a sequence g_k by evaluating g at $\lambda_k = \pi k/m$, where $m \geq n$ is a suitably chosen power of 2. In principle, we can use the fast Fourier transform (FFT) to evaluate

$$x_t = \frac{\sigma}{\sqrt{2\pi m}} \sum_{k=1-m}^{m-1} e^{i\lambda_k t} g_k W_k, \ t = 0, \ldots, m-1 \tag{3.5}$$

after setting

$$W_k = \begin{cases} U_k + iV_k, & k \geq 0 \\ U_k - iV_k, & k < 0 \end{cases}$$

where $(U_k, V_k, k = 0, \ldots, m-1)$ are independent standard Gaussian pairs. Then take x_t for $t = m-n, \ldots, m-1$ to provide the generated sample of length n. Note that the model is easily generalized to include (e.g.) ARMA components, by simply augmenting g with multiplicative factors. While the sequence g_k from (3.4) can be evaluated in closed form as

$$g_k = \left(2\sin\frac{\lambda_k}{2}\right)^{-d} \left[\cos\left(\frac{(\pi - \lambda_k)d}{2}\right) - i\sin\left(\frac{(\pi - \lambda_k)d}{2}\right)\right] \tag{3.6}$$

for $|k| > 0$, there is an evident difficulty due to the singularity at zero. A natural way to achieve a discrete approximation is to replace (3.4) with its series expansion

$$g(\lambda) = \sum_{j=0}^{\infty} b_j e^{-i\lambda j},$$

where b_j is defined by (1.6). Evaluating (3.5) by replacing this infinite sum with the sum truncated at m terms will approach the limit (3.2) in just the right way, and the FFT can be used here too, for speedy evaluation. By taking m large enough we should, in principle, be able to compute type I fBM to any desired degree of accuracy.

However, Table 2 shows the standard deviation of $X_n(1)$ in 10,000 replications of this simulation method for the case $d = 0.4$, also setting $\sigma^2 = 1$ and $n = 1000$. As before, we find that

m	1000	5000	10,000	20,000
SD	1.106	1.128	1.166	1.200

Table 2: SDs of Terminal Values, Harmonic Representation

the increase in the SD as m is increased is extremely slow, and remains a long way from the type 1 SD of 1.389, even with infeasibly large m. This method evidently suffers from an essentially similar problem to the time domain moving average method.

3.3 Choleski Method and Circulant Embedding

Another approach is to base the simulation on reproducing the known autocorrelation structure of the increments. Let $\boldsymbol{\Omega}_n$ denote the covariance matrix of the vector $\boldsymbol{x}_n = (x_1, \dots x_n)'$. Given formulae for $\gamma(k) = E(x_t x_{t-k})$ for $k = 0, ..., n-1$, $\boldsymbol{\Omega}_n$ is easily constructed as the Toeplitz matrix with kth diagonals set to $\gamma(k)$. If $\boldsymbol{K}_n$ represents the Choleski decomposition, such that $\boldsymbol{\Omega}_n = \boldsymbol{K}_n \boldsymbol{K}_n'$, and $\boldsymbol{z}_n = (z_1, \dots z_n)'$ is a standard normal vector, then $\boldsymbol{K}_n \boldsymbol{z}_n$ is a stationary sequence having the same distribution as $\boldsymbol{x}_n$ (exactly) in the case that u_t is Gaussian. The process (1.7) must therefore converge to type I fBM.

For the case where x_t is generated by (1.4) where u_t is i.i.d$(0, \sigma^2)$, we have the well-known formula:

$$\gamma(k) = \sigma^2 \frac{\Gamma(1-2d)\,\Gamma(k+d)}{\Gamma(k+1-d)} \frac{\sin(\pi d)}{\pi}. \tag{3.7}$$

see e.g. Granger and Joyeux (1980), Sowell (1992). It is straightforward to extend this calculation to the ARFIMA(p, d, q) case using the formulae given by Sowell (1992). It could even be extended to the multivariate case by computing the block-Toeplitz matrix of the cross-autocorrelations, but for large samples this procedure would be computationally challenging.

An alternative way to base the simulation on the covariances is by the circulant embedding method, as described in Davies and Harte (1987) for example. Let $\boldsymbol{v}$ $(2n + 1 \times 1)$ denote the discrete Fourier transform (DFT) of the sequence

$$\gamma(0), \dots, \gamma(n-1), \gamma(n), \gamma(n-1), \dots, \gamma(1).$$

The generated data are then taken as the first n elements of the inverse DFT of the vector generated as $\text{diag}(\boldsymbol{v})\boldsymbol{z}$, where $\boldsymbol{z}$ is a complex-valued Gaussian vector, scaled by $n^{1/2}$ – see Davies and Harte (1987) or Beran (1994) for the complete details of the algorithm. Davies and Harte simulate the so-called fractional Gaussian noise, which has a different autocorrelation structure from (1.4) except in the tail, but the method is easily adapted as described. By use of the fast Fourier transform, this method is substantially more economical of time and memory than the Choleski method, and once again it yields type I fBM in the limit. We have checked the properties of both of these algorithms by simulation of the (1.4) model with $d = 0.4$ ($n = 1000$, and 10,000 replications). For comparison with the theoretical value (1.389, from above) the standard deviation obtained for the Choleski replicates in these experiments was 1.394, and for the circulant embedding method, 1.396.

It is striking that these successful methods of simulating type I work by reproducing the characteristics of the observed data directly, not by invoking the linear representation as is done explicitly in **3.1** and implicitly in **3.2**. However, while they can be generalized to any stationary process whose covariance sequence is known, such as the ARFIMA class, or fractional Gaussian noise, and can in principle be generalized to the multivariate case, they will have difficulty in accommodating non-Gaussianity, nonlinear short-run dependence and other important data features. They are generally too inflexible for implementation in econometric models, which tend to rely heavily on the 'independent shock' paradigm for their construction.

3.4 Wavelets

Simulation of fractional processes using wavelet methods has been quite extensively researched, see among other references Abry and Sellan (1996), Meyer, Sellan and Taqqu (1999), and Pipiras (2004, 2005). While many variants of the method are possible, the basic algorithm described by Meyer, Sellan and Taqqu (1999) is representative. Defining the Hurst coefficient $H = d + \frac{1}{2}$, they show that fractional Brownan motion $B_H(t)$, $t \in \mathbb{R}$ can be represented almost surely on compact intervals in the form

$$B_H(t) = \sum_{k=-\infty}^{\infty} \Phi_H(t-k) S_k^{(H)} + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} 2^{-jH} \Psi_H(2^j t - k) \varepsilon_{jk} - b_0,$$

where $\varepsilon_{jk} \sim \text{iid}\, \mathrm{N}(0,1)$, b_0 is the initial condition to ensure $B_H(0) = 0$, Ψ_H is the chosen wavelet function, and Φ_H is a described as a biorthogonal scaling function satisfying $\Phi_H(t) = O(|t|^{-2-2H})$. The key component is $S_k^{(H)}$, a discrete time fractionally integrated Gaussian process such as an ARFIMA$(0, H-\frac{1}{2}, 0)$, independent of $\{\varepsilon_{jk}\}$, and designed to capture the low frequency variations. The wavelets fill in the high frequency "details", at successively smaller scales as j increases. See the above-cited references for the details.

On a compact interval such as $t \in [0,1]$, replacing the infinite sums in j and k by finite sums is shown to provide a highly accurate approximation to fBM. The sequence $S_k^{(H)}$ can of course be generated by the Choleski or circulant embedding algorithms. Exploiting the self-similarity of the fractional process allows relatively short realizations to effectively mimic the 'large' deviations in fBM. Therefore, we can view the wavelet method as exploiting the benefits of the Choleski and related methods. In particular, it should reproduce the type I distribution at reduced computational cost. (We have not been able to check this assertion ourselves by direct calculation, but the belief that the method should inherit the properties of the approximating ARFIMA appears a reasonable one.) However, it suffers the same disadvantages as those methods in having no straightforward extension to the multivariate framework, and also being difficult to adapt to the context of econometric modelling in the time domain.

3.5 Simulation by Aggregation

Beran (1994) also suggests using the Granger (1981) aggregation scheme. Summing a large number of independently generated stable AR(1) processes, whose coefficients are randomly generated in the interval $[0,1)$ as $\sqrt{\alpha}$, where α is a drawing from the Beta(a,b) distribution, Granger showed that the resulting aggregate series x_t would possess the attributes of a fractional sequence with $d = 1-b$; for example, with $d < \frac{1}{2}$ the autocovariances $E(x_t x_{t-k})$ will decrease at the rate k^{2d-1}. The 'long memory' attribute can be identified with the incidence, in a suitable proportion of the aggregate, of AR roots close to 1. A related method is proposed by Enriquez (2004), in which discrete processes are drawn repeatedly from a distribution inducing the required persistence structure, and aggregated.

These procedures certainly generate processes with the correct autocorrelation structure in the limit, but this alone is not sufficient to ensure that the normalized partial sums converge to fBM. For a further discussion of related issues see Davidson and Sibbertsen (2004). These authors prove convergence to type I fBM under a different aggregation procedure, that of micro-processes undergoing random regime shifts following a power law. However, in this result the aggregated micro-series are stationary processes, with implicitly remote starting dates. Although this issue is not dealt with explicitly in the cited paper, it is a plausible conjecture that aggregating truncated processes, with presample shocks suppressed, would yield the type II case.

A formal proof of weak convergence to fBM still appears wanting for the Granger aggregation case. Enriquez (2004) provides a proof that his limit process is Gaussian and a.s. continuous, and that its increments possesses the requisite correlation structure. However, the issue of type I versus type II is not addressed, and neither of the formulae (1.1) and (1.2) are cited as limit processes. We can plausibly conjecture that, in either case, the limit is of type I or type II depending on the treatment of the presample shocks. In the Granger aggregation case, note that for a type I limit the AR series components need to be stationary, an attribute only attained asymptotically as they advance from their starting points. This approach to stationarity will be rapid in most cases, but the long memory attribute of the aggregate depends upon the incidence of components with roots close to 1. These may have low probability, but they are correspondingly influential in the aggregate, and require a large number of steps to attain their stationary distributions. In other words, the problem that arose in Sections 3.1 and 3.2 recurs here. One might also consider simply drawing x_0 from the relevant marginal distributions, as an alternative to long lead-ins, but here there is the difficulty that the variances are tending to diverge in those influential cases with α close to 1; a phenomenon not unrelated to the singularity at the origin in (3.6).

4 An Alternative Simulation Strategy

The last section considered a number of methods of generating discrete time series with fractional characteristics. It is noteworthy that while some yield approximations closer to the type I distribution, and others closer to the type II distribution, this distinction is not significantly discussed in the literature we have cited, despite its obvious importance in applications. For econometric modelling, the need to simulate complex and often multivariate processes with possibly nonlinear features strongly favours the the pre-sample lag method, for its evident flexibility and adaptability. The question we consider is whether these benefits can be reconciled with the need to simulate the type I model. The method we describe in this section is designed to meet these requirements, and is also computationally very economical.

4.1 The Univariate Case

Consider the MA(∞) representation (i.e. Wold representation) of the linear time series process x_t with weight sequence $\{b_j\}$, given for example by (1.6) in the case of (1.4). For $t = 1, \ldots, n$, write $x_t = x_t^* + x_t^{**}$ where

$$x_t^* = \sum_{j=1}^{t-1} b_j u_{t-j}, \qquad x_t^{**} = \sum_{j=t}^{\infty} b_j u_{t-j}. \tag{4.1}$$

In the representation (1.1), X^* and X^{**} are the weak limits of the partial sum processes X_n^* and X_n^{**} derived from x_t^* and x_t^{**} respectively. As such, each is Gaussian, and they are independent of each other. The problem noted is that to approximate X_n^{**} adequately by a finite sum may require taking the x_t^{**} to an infeasibly large number of terms.

Assume at this point that the u_t process is i.i.d. Gaussian. Then, x_t^* and x_t^{**} are independent of one another, and the vector $\boldsymbol{x}^{**} = (x_1^{**}, \ldots, x_n^{**})'$ is Gaussian with a known covariance matrix. A convenient fact is that the autocovariance formula has the alternative representation

$$E(x_0 x_{-k}) = \sigma^2 \sum_{j=0}^{\infty} b_j b_{j+k}. \tag{4.2}$$

Therefore, for any $t, s > 0$,

$$\begin{aligned} E(x_t^{**} x_s^{**}) &= E \sum_{j=0}^{\infty} b_{j+t} u_{-j} \sum_{k=0}^{\infty} b_{k+s} u_{-k} \\ &= \sigma^2 \sum_{j=0}^{\infty} b_{j+t} b_{j+s} \\ &= E(x_0 x_{-|t-s|}) - \sigma^2 \sum_{j=0}^{\min(t,s)-1} b_j b_{j+|t-s|}. \end{aligned} \tag{4.3}$$

Assuming that the sequence $\{b_j\}$ is easily constructed, the $n \times n$ covariance matrix

$$\boldsymbol{C}_n = E(\boldsymbol{x}^{**} \boldsymbol{x}^{**\prime})$$

can therefore be constructed with minimal computational effort.

This suggests an easy way to simulate the distribution of $\boldsymbol{x}^{**}$, by simply making an appropriate collection of Gaussian drawings. Let $\boldsymbol{x}^{**}$ by constructed, by any means whatever, to be independent of $\boldsymbol{x}^*$ and Gaussian with the correct covariance structure. If

$$X_n^{**}(r) = \frac{1}{n^{1/2+d}} \sum_{t=1}^{[nr]} x_t^{**}$$

denotes the corresponding partial sum process, the following result is easily established

Theorem 4.1 $X_n^{**} \xrightarrow{d} X^{**}$.

Thus, let the vector $\boldsymbol{x}^* = (x_1^*, \ldots, x_n^*)'$ be computed by the usual moving truncation method so that, by standard arguments, $X_n^* \xrightarrow{d} X^*$. It then follows by the continuous mapping theorem that $X_n = X_n^* + X_n^{**} \xrightarrow{d} X$, in other words, Type I fBM.

If u_t is either not Gaussian, or is weakly dependent but not i.i.d., this simulation strategy will be inexact in small samples. However, it will still be asymptotically valid under the usual conditions for the invariance principle, noting that the limiting Gaussianity is here induced directly in the simulation, not by a limit argument. Note, incidentally, that it would be perfectly possible to simulate the vector $\boldsymbol{x}^*$ in the same manner, instead of using (1.4) and (1.8) in conjunction with the random generation of $u_1, \ldots, u_n$. This approach would lead us, in a roundabout fashion, to the Choleski simulation method. The asymptotic distributions would be the same, but there are of course numerous advantages in terms of modelling flexibility with the dynamic simulation approach and little is lost, in this case, in terms of computing resources.

It turns out that $\boldsymbol{C}_n$ tends rapidly to singularity as n increases, which is not surprising in view of the fact that $\boldsymbol{x}^{**}$ basically combines the common set of random components $\{u_t,\, t < 1\}$ with changing weights. This means that in practice only a comparative handful of Gaussian drawings are needed to generate the complete sequence. If n is small enough that $\boldsymbol{C}_n$ can be diagonalized

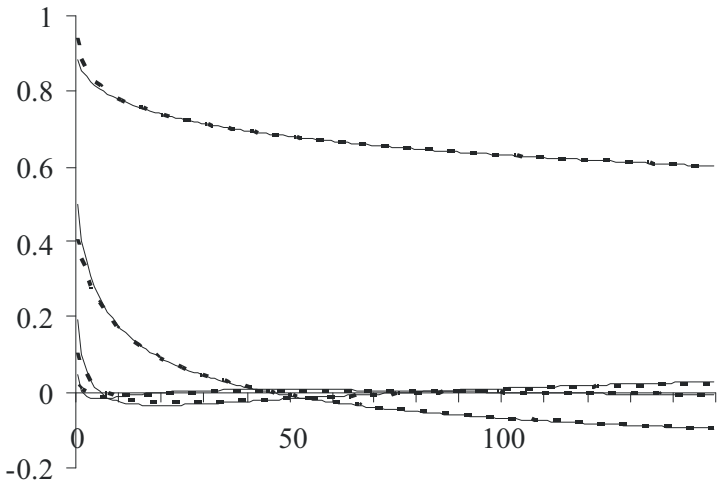

Figure 3: Columns of V_n, $n = 150$: Actual (solid line); interpolated from $p = 50$ (dashed line).

numerically (in practice, this appears to set $n \leq 150$ approximately, using the requisite Ox function) then it is a simple matter to obtain the decomposition

$$\boldsymbol{C}_n = \boldsymbol{V}_n \boldsymbol{V}_n' \tag{4.4}$$

where $\boldsymbol{V}_n$ is a $n \times s$ matrix, and s is chosen as the rank of the smallest positive eigenvalue. Then, it is only necessary to draw an independent standard Gaussian vector $\boldsymbol{z}$ $(s \times 1)$, and compute $\boldsymbol{x}^{**} = \boldsymbol{V}_n \boldsymbol{z}$. Note that in a Monte Carlo experiment, $\boldsymbol{V}_n$ only has to be computed once, and can then be stored for use in each replication. This means that generating a type I series has virtually the same computational cost as that of a type II series.

So much is straightforward, but we also need to deal with the case where n is too large to perform the required diagonalization. In practice, we treat $n = 150$ as a convenient cut-off point. To construct a suitable $\boldsymbol{V}_n$ matrix for cases with $n > 150$, we note the fact that the squared length of its tth row is $E(x_t^{**2})$, which we can obtain from (4.3) as before. We also have the fact that the columns of $\boldsymbol{V}_n$ are orthogonal and accordingly have a characteristic structure. We combine these pieces of information by constructing and diagonalizing $\boldsymbol{C}_p$, where p is chosen as the largest whole divisor of n not exceeding 150. $\boldsymbol{V}_n$ matrices are now constructed as follows: for $t = 1, [n/p], 2[n/p], \ldots, p[n/p]$, set the tth row of $\boldsymbol{V}_n$ by taking the $[pt/n]$th row of $\boldsymbol{V}_p$, renormalized to have squared norm equal to $E(x_t^{**2})$. Then, the missing rows are then filled in by linear interpolation, followed by renormalization such that $\boldsymbol{v}_{nt}' \boldsymbol{v}_{nt} = E(x_t^{**2})$. This procedure is fast and ensures that, at least, the variances and covariances are diminishing as t increases at the correct rate.

To illustrate the performance of the interpolation procedure, Figure 3 plots, for the case $d = 0.4$ and $n = 150$, the first 4 columns of $\boldsymbol{V}_n$ by exact calculation (solid lines) and also by interpolation from $p = 50$ (dashed lines). The differences are apparently negligible. This is the largest n for which this direct comparison is possible, but our simulation results suggest the method also works well in cases up to $n = 1000$. Table 3 shows the theoretical standard deviations of the random variables $X(1)$ and $X^*(1)$, with the same quantities estimated by Monte Carlo from samples of size $n = 1000$ for comparison. The table indicates that the proposed simulation strategy replicates the distribution very accurately, in general. Only for the extreme negative

values of d_1 does the approximation prove poor, the approach to the asymptote as $n \to \infty$ appearing to be very slow in this region. However, note that this phenomenon effects the type I and type II models equally.

	Type I		Type II	
d	Theoretical	Monte Carlo	Theoretical	Monte Carlo
0.4	1.389	1.383	0.840	0.842
0.2	0.997	0.993	0.920	0.917
0	1	1.0085	1	1.0085
−0.2	1.176	1.167	1.109	1.104
−0.4	1.877	1.76	1.501	1.41

Table 3: Standard Deviations of Type I and II Processes. Monte Carlo estimates for n=1000, from 10,000 replications

Now consider the application of this method to general fractional processes. In the case of the ARFIMA$(0, d, 0)$, as was used in the construction of Table 3, the autocovariance formula is taken from (3.7). The sequence $\{b_j\}$ is easily constructed by the recursion $b_j = b_{j-1}(j + d - 1)/j$ for $j > 0$ with $b_0 = 1$. It would be possible to extend the method directly to the ARFIMA(p, d, q)

$$\phi(L)\Delta^d x_t = \theta(L)u_t, \quad u_t \sim \text{iid}(0, \sigma^2) \tag{4.5}$$

by taking the required covariance formulae from Sowell (1992). In practice, however, there is little need for this elaboration. To see why, note that defining $z_t = \phi(L)x_t$ we may write

$$\begin{aligned} z_t &= z_t^* + z_t^{**} \\ &= \Delta^{-d}\theta(L)u_t 1(t \geq 1) + \Delta^{-d}\theta(L)u_t 1(t \leq 0). \end{aligned}$$

The first term can be simulated in the usual manner as a partial sum from zero initial values, whereas the second term is well approximated, by $\sum_{j=0}^{\infty} b_{t+j}v_{-j}$ where $v_t \sim \text{iid}(0, \theta(1)^2\sigma^2)$, and $\{b_j\}$ is obtained by the recursion just described. The autoregressive component is now easily added, given initial values $z_{1-p}, \ldots, z_0$, by the recursion

$$x_t = z_t - \sum_{j=1}^{p} \phi_j x_{t-j}.$$

4.2 The Multivariate Case

To generalize this method to generate vectors of two or more type I processes, say $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{mt})$ for any $m > 1$, we need to write the model in final (Wold) form as

$$\boldsymbol{x}_t = \sum_{j=0}^{\infty} \boldsymbol{B}_j \boldsymbol{u}_{t-j},$$

where the $\boldsymbol{B}_j$ $(m \times m)$ are matrices of lag coefficients, and $\{\boldsymbol{u}_t\}$ $(m \times 1)$ is the vector of shocks with covariance matrix $\boldsymbol{\Sigma}$. The autocovariance matrices accordingly take the form

$$\boldsymbol{\Gamma}(k) = E(\boldsymbol{x}_0 \boldsymbol{x}'_{0-k}) = \sum_{j=0}^{\infty} \boldsymbol{B}_j \boldsymbol{\Sigma} \boldsymbol{B}'_{j+k}.$$

It easily follows by the preceding arguments that

$$E(\boldsymbol{x}_t^{**}\boldsymbol{x}_s^{**\prime}) = \sum_{j=0}^{\infty} \boldsymbol{B}_{j+t}\boldsymbol{\Sigma}\boldsymbol{B}_{j+s}' \\ = \boldsymbol{\Gamma}(s-t) - \sum_{j=0}^{t-1} \boldsymbol{B}_j\boldsymbol{\Sigma}\boldsymbol{B}_{j+s-t}'$$

for $t \leq s$, and take the transpose of this matrix for the case $t \geq s$.

Accordingly, stack the components $\boldsymbol{x}_1^{**}, \ldots, \boldsymbol{x}_m^{**}$ into a vector $\boldsymbol{x}^{**} = (\boldsymbol{x}_1^{**\prime}, \ldots, \boldsymbol{x}_m^{**\prime})'$ $(mn \times 1)$ having covariance matrix

$$E(\boldsymbol{x}^{**}\boldsymbol{x}^{**\prime}) = \begin{bmatrix} \boldsymbol{C}_{11,n} & \cdots & \boldsymbol{C}_{1m,n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{C}_{m1,n} & \cdots & \boldsymbol{C}_{mm,n} \end{bmatrix} = \boldsymbol{C}_n.$$

Letting $\boldsymbol{b}_{k,j}'$ represent the kth row of $\boldsymbol{B}_j$, note that the cross-covariance matrices $\boldsymbol{C}_{kh,n}$ for $k, h = 1, \ldots, m$ have elements of the form

$$E(x_{kt}^{**}x_{hs}^{**}) = \sum_{j=0}^{\infty} \boldsymbol{b}_{k,j+t}'\boldsymbol{\Sigma}\boldsymbol{b}_{h,j+s} \\ = \gamma_{kh}(s-t) - \sum_{j=0}^{t-1} \boldsymbol{b}_{k,j}'\boldsymbol{\Sigma}\boldsymbol{b}_{h,j+s-t}.$$

for the cases $s \geq t$, and with $E(x_{kt}^{**}x_{hs}^{**}) = E(x_{ks}^{**}x_{ht}^{**})$ for the cases $t \geq s$, such that $\boldsymbol{C}_{kh,n} = \boldsymbol{C}_{hk,n}'$.

The decomposition (4.4) can now be computed as before, for this stacked matrix, to yield $\boldsymbol{V}_n = (\boldsymbol{V}_{1n}', \ldots, \boldsymbol{V}_{mn}')'$. The blocks $\boldsymbol{V}_{jn}$ $(n \times s)$ for $j = 1, \ldots, m$ are used to generate replications of each process, from the formula $\boldsymbol{x}_j^{**} = \boldsymbol{V}_{jn}\boldsymbol{z}$ where, in this case, as before, $\boldsymbol{z}$ is a standard normal drawing of conformable dimension. Given that we are limited by $mn \leq 150$, this method has to be modified by the extrapolation step described above, for cases with $n > [150/m]$. Hence, large-dimensional systems potentially entail an additional compromise in terms of approximation error, relative to the univariate case. However, for the reasons stated above we would not expect this to be a critical issue for most purposes; thus, the case $m = 3$ and $n = 150$ will yield an approximation comparable to that illustrated in Figure 3.

For the case where $\boldsymbol{B}_j = \text{diag}(b_{1j}, \ldots, b_{mj})$ and $b_{kj} = \Gamma(j + d_k)/(\Gamma(d_k)\Gamma(j+1))$, the following generalization of (3.7) provides the cross-autocovariances. Without loss of generality, consider the bivariate case, as follows.

Theorem 4.2 *For x_{1t} and x_{2t} defined by (1.4) with respect to i.i.d. shock processes u_{1t} and u_{2t} with covariance $E(u_{1t}u_{h2}) = \sigma_{12}$,*

$$E(x_{10}x_{2,-k}) = \sigma_{12}\frac{\sin \pi d_1}{\pi}\frac{\Gamma(1-d_1-d_2)\,\Gamma(d_1+k)}{\Gamma(1-d_2+k)}.$$

Note that this formula yields (3.7) in the case $x_{1t} = x_{2t}$. Extending the procedure to the simulation of vector ARFIMA systems is a simple matter of replacing $\boldsymbol{\Sigma}$ by $\boldsymbol{\Theta}(1)\boldsymbol{\Sigma}\boldsymbol{\Theta}(1)'$ to cope with a vector moving average contribution $\boldsymbol{\Theta}(L)$ $(m \times m)$, and then applying the autoregressive recursion to the augmented series in the manner described in the previous section.

5 Estimation of Type I ARFIMA Models

Compare the stationary fractional noise model

$$(1-L)^d Y_t = u_t,\ t = 1, \ldots, n \tag{5.1}$$

where $\{u_t\}_{-\infty}^{\infty}$ is i.i.d.$(0, \sigma^2)$ and $|d| < \frac{1}{2}$ with its feasible counterpart

$$(1-L)^d Y_t^* = u_t^*,\ t = 1, \ldots, n \tag{5.2}$$

where u_t^* is defined by (1.8) and Y_t^* is defined by the equation. In other words, if the sequence $\{a_j\}$ repesents the coefficients in the expansion of $(1-L)^d$,

$$\begin{aligned} Y_1^* &= u_1 \\ Y_2^* &= u_2 - a_1 Y_1^* \\ &\cdots \\ Y_n^* &= u_T - a_1 Y_{t-1}^* - \cdots - a_{n-1} Y_1^*. \end{aligned}$$

In the standard time domain estimation framework, we will normally maximize the likelihood implied by (5.2), although using the data $Y_1, \ldots, Y_n$, generated by (5.1) by hypothesis.

Write

$$\Upsilon_t(L; d) = \sum_{j=0}^{t-1} a_j L^j$$

to represent the truncation of the expansion of $(1-L)^{-d}$ at the tth term. This operator is sometimes written $(1-L)_+^{-d}$, but we choose in this context to emphasize the dependence on t. Note that

$$\Upsilon_t(L; -d) = \Upsilon_t(L; d)^{-1}$$

follows immediately from matching terms in the identity $(1-L)^d(1-L)^{-d} = 1$. With this notation, we can write the solution of (5.2) as

$$Y_t^* = (1-L)^{-d} u_t^* = \Upsilon_t(L; -d) u_t.$$

However, notice that the solution of (5.1) has the approximate form

$$\begin{aligned} Y_t &= (1-L)^{-d} u_t \\ &\approx \Upsilon_t(L; -d) u_t + \boldsymbol{v}_t(d, \sigma)' \boldsymbol{z} \end{aligned}$$

where $\boldsymbol{v}_t(d, \sigma)'$ is row t of the $n \times s$ matrix defined by (4.4), and $\boldsymbol{z}$ $(s \times 1)$ is a standard normal vector. Therefore consider the approximate form of (5.1) taking the form

$$\begin{aligned} \Upsilon_t(L; d) Y_t &= \Upsilon_t(L; d) \boldsymbol{v}_t(d, \sigma)' \boldsymbol{z} + u_t \\ &= \boldsymbol{v}_t^*(d, \sigma)' \boldsymbol{z} + u_t \end{aligned} \tag{5.3}$$

where the second equality defines $\boldsymbol{v}_t^*$. The vectors $\boldsymbol{v}_t^*(d, \sigma)$ can be computed, given values for d and σ, and the elements of $\boldsymbol{z}$ can be treated as s additional unknown parameters. Therefore, the true model (5.1) can be estimated, in principle, by inserting the 'regressors' $\boldsymbol{v}_t^*$ into the equation and estimating the parameters $(d, \sigma, \boldsymbol{z})$ jointly, by conditional maximum likelihood.

The same technique is straightforwardly extended to estimating the ARFIMA(p, d, q) model, with the form

$$\phi(L)(1-L)^d (Y_t - \alpha) = \theta(L) u_t,\ t = 1 + \max(p, q), \ldots, n$$

where $\alpha = E(Y_t)$. The approximate model in this case takes the form

$$\phi(L)\Upsilon_t(L;d)(Y_t - \alpha) = \boldsymbol{v}_t^*(d, \sigma|\theta(1)|)'\boldsymbol{z} + \theta(L)u_t,\ t = 1 + \max(p,q), \ldots, n. \tag{5.4}$$

Notice that in this case the variance of the presample shocks must be calculated as $\sigma^2\theta(1)^2$, and hence the $\boldsymbol{v}_t^*$ depend additionally on the moving average parameters. Be careful to distinguish between this model and that having the form

$$\phi(L)\Upsilon_t(L;d)Y_t = \mu + \boldsymbol{v}_t^*(d, \sigma|\theta(1)|)'\boldsymbol{z} + \theta(L)u_t.$$

This latter model has a solution of the form

$$Y_t = \frac{\mu}{\phi(1)}\Upsilon_t(1;-d) + Y_t^*$$

where Y_t^* is a zero-mean ARFIMA and $\Upsilon_t(1;-d) = O(t^d)$, and hence contains a deterministic fractional trend.

Note that this modification of the conditional ML estimator is of small order and irrelevant to the asymptotic distribution. Since $\boldsymbol{v}_t^* \to \boldsymbol{0}$ as $t \to \infty$, the estimator of $\boldsymbol{z}$ is not consistent. However, under the distribution conditional on the presample realization of the process, the omission of the terms in $\boldsymbol{z}$ is a potential source of finite-sample estimation bias. Including these terms implies a bias-efficiency trade-off depending on sample size, and whether it is desirable on balance is a question needing to be considered in context. The method is best understood by contrasting the exact maximum likelihood estimator (MLE) with the conditional MLE. The former estimator was derived by Sowell (1992) for the Gaussian case, while the latter is equivalent to least squares with presample values set to 0. The question we pose is whether introducing the extra parameters allows a finite sample correction comparable to that provided by doing exact rather than conditional ML.

With these issues in mind, we considered the well-known series for annual minima of the Nile, as studied by Hurst (1951) and reproduced in Beran (1994). This series of 663 annual observations (622–1284AD) appears as a stationary process, having a sample mean of 1148.16. The time plot is reproduced (in mean deviation form) in Figure 4.

The natural linear representation of such a process is (5.4) where α represents the unconditional mean. The fact that the true α is unknown is a complicating factor for our analysis, which to date has implicity considered zero mean processes. Ideally we should like to fit α econometrically, in the context of the type I model. However, preliminary attempts revealed a very substantial loss of efficiency. The difficulty of fitting the mean of fractional models is a well known problem, documented for example by Cheung and Diebold (1994). There proves to be too little information in this sample to allow α and $\boldsymbol{z}$ to be estimated jointly, so for the purposes of the exercise we subtract off the sample mean at the outset. For the centred series, α is fixed at 0.

A second important question is the choice of s, the number of elements of $\boldsymbol{z}$ to be fitted. The elements of $\boldsymbol{v}_t^*$ depend on the magnitude of d but, beyond the first element, get very rapidly small from the outset, even when d is large (see Figure 3). A practical limit for s of at most one or two emerges from this and other cases examined.

In Table 1, we report estimates for the cases $s = 0, 1$ and 2, the first of these corresponding to the usual type II model. In view of the leptokurtic shock distribution evident from Figure 4, we also opted to maximize the Student t likelihood, which allows the degrees of freedom of the distribution to be estimated as an additional parameter. The columns headed MLE show for comparison the Sowell (1992) exact Gaussian maximium likelihood estimator. This is the Ox implementation ARFIMA 1.04, due to Doornik and Ooms (2006). The fact that the available implementations do not allow for non-Gaussian disturbances is one advantage of our approximate

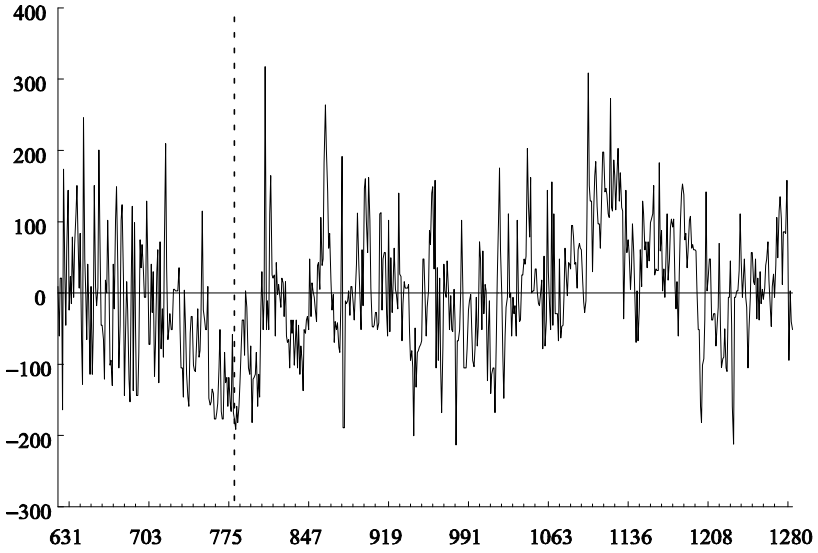

Figure 4: Annual Nile minima (mean deviations)

	622–1284AD				784–1284AD			
	$s=0$	$s=1$	$s=2$	MLE	$s=0$	$s=1$	$s=2$	MLE
ARFIMA d	0.4182 (0.0316)	0.4187 (0.0315)	0.4185 (0.0310)	0.3932 (0.0299)	0.4504 (0.0383)	0.4398 (0.0377)	0.4289 (0.0315)	0.4374 (0.0336)
type I Frac.,Z_1	–	−0.465 (0.516)	−0.908 (0.679)	–	–	−0.9841 (0.672)	−0.5301 (0.554)	–
type I Frac., Z_2	–	–	1.894 (1.771)	–	–	–	−3.485 (1.842)	–
Shock SD	70.547 (2.946)	70.665 (3.004)	70.865 (3.075)	69.90	66.981 (3.757)	66.958 (3.891)	66.542 (3.825)	65.37
Student t DF	2.345 (0.245)	2.314 (0.239)	2.273 (0.234)	–	2.1248 (0.214)	2.088 (0.206)	2.1248 (0.214)	–
Log-likelihood	−3738	−3737	−3737	−3757	−2786	−2783	−2782	−2806
Residual $Q(12)$	7.6426	7.524	7.027	–	5.250	5.686	5.897	–

Table 4: Annual Nile minima: ARFIMA(0,d,0) estimated by Student t conditional ML (robust standard errors in parentheses) and Sowell (1992) exact ML.

method over exact ML. ARFIMA(0,d,0) models are fitted, and the residual Box-Pierce Q statistics indicate that these models account adequately for the autocorrelation in the series.

The first four columns of the table show the estimates for the complete sample of 633 years. It is apparent from the time plot that the initial observations are quite close to the mean of the series. Presample components happen to cancel out here, and have a small net influence on the initial observations. In other words, the 'type II' assumption that the pre-sample shocks are zero is not too implausible at this date. However, moving forward in time to the late 700s places us in the middle of a prolonged dry period. Observe that the Nile's flow was substantially lower than average, in every year except one, between 758AD and 806AD. Of course, it is climatic variations of this type that give rise to the 'long memory' characterization of the series. If our sample had happened to start in (say) the year 784AD, instead of 622AD, the pre-sample shocks would have been relatively influential, and the 'type II' assumption correspondingly inadequate to account for them.

Columns 4-6 of the table show the results of estimating the model from the observations from 784AD onwards (marked with the dotted line in Figure 7). Note the substantial difference between the 'type I' and 'type II' estimates in this case. If we take as a benchmark the estimate of the memory parameter d for the whole period (0.418), note that in the shorter sample the conventional type II model ($s = 0$) appears to overstate d significantly. Also, fitting the type I components applies a much more substantial correction than before. The estimate 0.429, while still a little larger than the full-sample benchmark, is a great deal closer to it than the estimate 0.450 obtained from the 'type II' model.

The estimates of the $\boldsymbol{z}$ components are evidently inefficient, especially when two are fitted. Thus, since we know that these coefficients are standard normal drawings, the estimate of -3.48 is clearly excessive, a result that can be understood as due to a trading-off of two highly collinear components. However, it is also clear that neglecting the presample shocks can in certain circumstances induce bias with respect to the conditional distribution. The ability to correct for these effects may in some circumstances provide a useful addition to the modeller's armoury.

6 Conclusion

In this paper, we have considered the issue of modelling fractionally integrated processes for econometric applications. Since inference in these models will generally depend on teaming an invariance principle with a scheme for numerical simulation of the assumed asymptotic distribution, it is of some importance to make an appropriate choice of data generation process. We show that simulating the more natural type I representation of fractional Brownian motion can be achieved with as little computational cost as the type II model often used in practice, although conventional simulation methods work poorly. Our firm recommendation to practitioners is to use type I simulations wherever this difference is likely to be crucial, unless there are particular reasons for doing otherwise.

We note the existence of important exceptions to this rule, such as the unit root test against fractional alternatives proposed by Dolado, Gonzalo and Mayoral (2002). Here, the statistic is computed using the fractional difference of the observed series, where since this is naturally truncated to the observation period, the induced asymptotic distribution is of type II by construction. Hence the tables reported by these authors for this case of the null hypothesis are correct. However, they also propose, although do not analyse in any detail, a test for the null hypothesis of a fractional process with parameter d_0 against an alternative d_1. For these cases, the tables would need to be generated according to the assumed type of the observed data, and the test outcomes could depend on this decision in a crucial manner. We would recommend the methods proposed here in such a case.

A Appendix: Proofs

Proof of Proposition 2.1 We derive this expectation as the limit of the expression

$$\frac{1}{n^{1+d_1+d_2}}\sum_{t=1}^{n-1}\sum_{s=1}^{t} Ex_{1s}^{*}x_{2,t+1}^{*}$$

where $x_{pt}^{*}=\sum_{j=0}^{t-1} b_{pj}u_{p,t-j}$ for $p=1,2$, and u_{1t} and u_{2t} are i.i.d. with $E(u_{1t}u_{2s})=\sigma_{12}$ if $t=s$, and 0 otherwise. Note that

$$\sum_{s=1}^{t} x_{1s}^{*}=\sum_{s=0}^{t-1}\left(\sum_{k=0}^{s} b_{1k}\right)u_{1,t-s}$$

and hence

$$E\sum_{s=1}^{t} x_{1s}^{*}x_{2,t+1}^{*}=\sigma_{12}\sum_{s=0}^{t-1}\left(\sum_{k=0}^{s} b_{1k}\right)b_{2,s+1}.$$

Applying Stirling's approximation formula, note that

$$\sum_{k=0}^{s} b_{1k}\sim\frac{1}{\Gamma(d_1)}\int_0^s \xi^{d_1-1}d\xi=\frac{s^{d_1}}{\Gamma(d_1+1)},$$

where '$\sim$' denotes that the ratio of the two sides converges to 1 (see Davidson and de Jong 2000, Lemma 3.1). Hence, by a similar argument

$$\begin{aligned}\frac{1}{n^{1+d_1+d_2}}\sum_{t=1}^{n-1}\sum_{s=1}^{t} Ex_{1s}^{*}x_{2,t+1}^{*}&=\frac{\sigma_{12}}{n^{1+d_1+d_2}}\sum_{t=1}^{n-1}\sum_{s=0}^{t-1}\left(\sum_{k=0}^{s} b_{1k}\right)b_{2,s+1}\\&\sim\frac{\sigma_{12}d_2}{\Gamma(d_1+1)\Gamma(d_2+1)}\int_0^1\int_0^\tau \zeta^{d_1+d_2-1}d\zeta d\tau\end{aligned}$$

and the stated result follows directly. ∎

Proof of Theorem 4.1
Note that $X_n^{**}(r)$, $0\le r\le 1$ is Gaussian with covariance structure converging to that of the limit process X^{**}, by construction. It therefore remains to show that the sequence is uniformly tight, which we demonstrate by establishing the criterion of Theorem 15.6 of Billingsley (1968). In the present case, this is easily shown to be implied by

$$E(X_n^{**}(r+\delta)-X_n^{**}(r))^2\le C\delta^{2\alpha}$$

for $\alpha>\frac{1}{2}$ and all $0\le r\le 1-\delta$, and $C<\infty$ represents a generic positive constant. However,

$$X_n^{**}(r+\delta)-X_n^{**}(r)=\frac{1}{n^{1/2+d}}\sum_{t=[nr]+1}^{[n(r+\delta)]} x_t^{**}.$$

It follows from (4.3) that for $k\ge 0$,

$$E(x_t^{**}x_{t+k}^{**})=O(t^{2d-1})$$

Hence, the proof is completed by noting that

$$E(X_n^{**}(r+\delta)-X_n^{**}(r))^2\le C\frac{(n\delta)^2(nr)^{2d-1}}{n^{1+2d}}$$

$$= C\delta^2 r^{2d-1}. \qquad \blacksquare$$

Proof of Theorem 4.2

To compute the cross-covariance use the harmonizable representation:

$$\begin{aligned}\gamma_{12}(k) = E(x_{1t}x_{2,t-k}) &= \frac{\sigma_{12}}{2\pi}\int_{-\pi}^{\pi}\left(1-e^{-i\lambda}\right)^{-d_1}e^{it\lambda}\left(1-e^{i\lambda}\right)^{-d_2}e^{-i(t-k)\lambda}d\lambda \\ &= \frac{\sigma_{12}}{2\pi}\int_{-\pi}^{\pi}\left(1-e^{-i\lambda}\right)^{-d_1}\left(1-e^{i\lambda}\right)^{-d_2}e^{ik\lambda}d\lambda. \qquad \text{(A-1)}\end{aligned}$$

Denoting the integrand in (A-1) by $F(\lambda)$ observe that

$$\int_{-\pi}^{\pi}F(\lambda)\,d\lambda = \int_{0}^{\pi}\left[F(\lambda)+\overline{F(\lambda)}\right]d\lambda \qquad \text{(A-2)}$$

where the upper bar denotes complex conjugate. Further, using

$$\begin{aligned}1-e^{\mp i\lambda} &= \pm e^{\mp i\lambda/2}\left(e^{i\lambda/2}-e^{-i\lambda/2}\right) = \pm 2ie^{\mp i\lambda/2}\sin\frac{\lambda}{2} \\ &= 2e^{\pm i(\pi-\lambda)/2}\sin\frac{\lambda}{2}\end{aligned}$$

and noting that $\sin(\lambda/2)$ is non-negative for $0\le\lambda\le\pi$, rewrite the integral in (A-2) as

$$\begin{aligned}&\int_0^{\pi}\left[F(\lambda)+\overline{F(\lambda)}\right]d\lambda \\ &= \int_0^{\pi}\left[\left(2e^{i(\pi-\lambda)/2}\sin\frac{\lambda}{2}\right)^{-d_1}\left(2e^{-i(\pi-\lambda)/2}\sin\frac{\lambda}{2}\right)^{-d_2}e^{ik\lambda}\right. \\ &\qquad\left.+\left(2e^{-i(\pi-\lambda)/2}\sin\frac{\lambda}{2}\right)^{-d_1}\left(2e^{i(\pi-\lambda)/2}\sin\frac{\lambda}{2}\right)^{-d_2}e^{-ik\lambda}\right]d\lambda \\ &= 2^{-d_1-d_2}\int_0^{\pi}\sin^{-d_1-d_2}\frac{\lambda}{2}\left[e^{i[-(d_1-d_2)\pi/2+(d_1-d_2+2k)\lambda/2]}+e^{-i[-(d_1-d_2)\pi/2+(d_1-d_2+2k)\lambda/2]}\right]d\lambda \\ &= 2^{1-d_1-d_2}\int_0^{\pi}\sin^{-d_1-d_2}\frac{\lambda}{2}\cos\left[-(d_1-d_2)\pi/2+(d_1-d_2+2k)\lambda/2\right]d\lambda. \qquad \text{(A-3)}\end{aligned}$$

The integral in (A-3) can be transformed, using the change of variable $x=(\pi-\lambda)/2$, into

$$\begin{aligned}&\int_0^{\pi}\sin^{-d_1-d_2}\frac{\lambda}{2}\cos\left[-(d_1-d_2)\pi/2+(d_1-d_2+2k)\lambda/2\right]d\lambda \\ &\qquad = 2\int_0^{\pi/2}\cos^{-d_1-d_2}x\cos\left[-(d_1-d_2)\pi/2+(d_1-d_2+2k)(\pi/2-x)\right]dx \\ &\qquad = 2\int_0^{\pi/2}\cos^{-d_1-d_2}x\cos\left[\pi k-(d_1-d_2+2k)x\right]dx \\ &\qquad = (-1)^k 2\int_0^{\pi/2}\cos^{-d_1-d_2}x\cos(d_1-d_2+2k)x\,dx.\end{aligned}$$

Using Relation 3.631.9 of Gradshteyn and Ryzhik (2000) and the properties of beta and gamma functions,

$$\int_0^{\pi/2}\cos^{-d_1-d_2}x\cos(d_1-d_2+2k)x\,dx = \frac{2^{-(1-d_1-d_2)}\pi}{(1-d_1-d_2)\,B(1-d_2+k,1-d_1-k)}$$

$$= \frac{2^{-(1-d_1-d_2)}\pi}{1-d_1-d_2}\frac{\Gamma(2-d_1-d_2)}{\Gamma(1-d_2+k)\Gamma(1-d_1-k)}$$
$$= \frac{\Gamma(1-d_1-d_2)}{2^{1-d_1-d_2}}\frac{\Gamma(d_1+k)}{\Gamma(1-d_2+k)}\sin\pi(d_1+k)$$
$$= (-1)^k\frac{\Gamma(1-d_1-d_2)}{2^{1-d_1-d_2}}\frac{\Gamma(d_1+k)}{\Gamma(1-d_2+k)}\sin\pi d_1.$$

Finally,

$$\gamma_{12}(k) = \sigma_{12}\frac{\sin\pi d_1}{\pi}\frac{\Gamma(1-d_1-d_2)\Gamma(d_1+k)}{\Gamma(1-d_2+k)}. \qquad \blacksquare$$

B Appendix: Distributions of Fractional Brownian Functionals

The following simulations of familiar statistics associated with nonstationary regression analysis are each based on 100,000 Monte Carlo replications. Although these statistics have an established role in hypothesis testing, we are abstracting here from any specific testing problem. The object is solely to investigate how far these representative fractional Brownian functionals differ, under the alternative definitions.

In the following formulae, the expression on the left of the "$\approx$" symbol is evaluated from data in each case and the expression on the right is the random variable whose distribution we seek to estimate. The model in (1.4) with independent Gaussian(0,1) shocks is used to generate the data with a sample size $n = 1000$ in each case. For the "type I" model, $x_t = x_t^* + x_t^{**}$ where x_t^* is generated as in (4.1), and x_t^{**} is constructed as detailed in Section 4. These kernel densities are plotted with solid lines in the figures. The corresponding expressions for the "type II" case are obtained by simply setting $x_t = x_t^*$ and $X = X^*$ throughout. These densities are plotted with broken lines in the figures.

1. *Dickey-Fuller statistics.*

$$n\hat{\phi} = n\frac{\sum_{t=1}^{n-1} S_t x_{t+1}}{\sum_{t=1}^{n-1} S_t^2} \approx \frac{\int_0^1 X dX}{\int_0^1 X^2 ds}$$

and

$$n\hat{\phi}_\mu = n\frac{\sum_{t=1}^{n-1}\left(S_t-\bar{S}\right)x_{+1}}{\sum_{t=1}^{n-1}\left(S_t-\bar{S}\right)^2} \approx \frac{\int_0^1 X dX - X(1)\int_0^1 X ds}{\int_0^1 X^2 ds - \left(\int_0^1 X ds\right)^2}.$$

These are the normalized coefficients of the regression of x_{t+1} on $S_t = \sum_{s=1}^t x_s$, with and without intercept. Note that these statistics are $O_p(1)$ for $d \geq 0$. The corresponding Dickey-Fuller t statistics diverge at the rate $O_p(n^d)$ in the same case (see Davidson 2006). Table 5 shows some leading quantiles, and Figure 5 plots the kernel densities.

2. *Bivariate Stochastic Integrals*

$$\frac{\sum_{t=1}^n S_{1t}x_{2t}}{n^{1+d_1+d_2}} \approx \int_0^1 X_1 dX_2,$$

and

$$\frac{\sum_{t=1}^n (S_{1t}-\bar{S}_1)x_{2t}}{n^{1+d_1+d_2}} \approx \int_0^1 X_1 dX_2 - X_2(1)\int_0^1 X_1 ds,$$

	$P(\leq)$	0.01	0.05	0.1	0.9	0.95	0.99
no intercept	Type I	−0.12	0.04	0.28	2.39	2.88	4.17
	Type II	−0.24	−0.04	0.16	2.71	3.30	4.68
with intercept	Type I	−4.51	−2.75	−2.05	1.97	2.67	4.25
	Type II	−4.02	−2.45	−1.78	2.47	3.15	4.94

Table 5: Quantiles of the Dickey-Fuller statistics, $d = 0.4$.

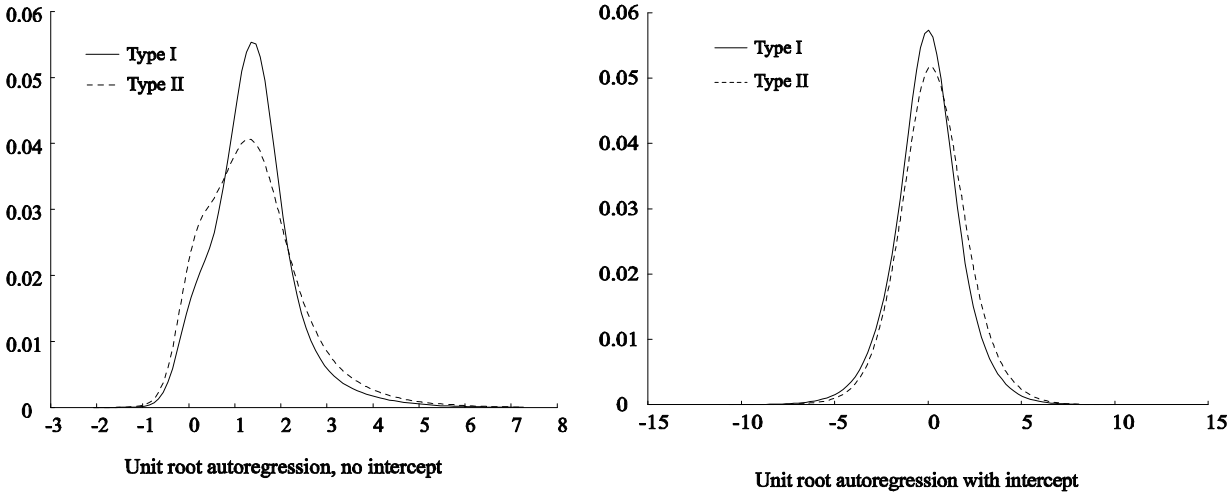

Figure 5: Simulation of unit root autoregression: d = 0.4, 1000 observations, 100,000 replications

where $S_{1t} = \sum_{s=1}^{t} x_{1s}$ and the pair $\{x_{1t}, x_{2t}\}$ are fractional noise processes where $\{u_{1t}, u_{2t}\}$ are Gaussian(0,1) and contemporaneously correlated with correlation coefficient 0.5. Figure 6 shows kernel densities for the cases where either X_2 is fBM with $d = 0.4$ and X_1 is a regular Brownian motion (and hence the difference in the distributions depends wholly X_2) or both processes are fBM with the same d of 0.4.

3. *Fractional cointegrating regression 't statistics'*

$$\frac{n^{1/2-d_2}\sum_{t=1}^{n} S_{1t}x_{2t}}{\sqrt{\sum_{t=1}^{n} S_{1t}^2 \sum_{t=1}^{n} x_{2t}^2 - \left(\sum_{t=1}^{n} S_{1t}x_{2t}\right)^2}} \approx \frac{\int_0^1 X_1 dX_2}{\sigma_2 \sqrt{\int_0^1 X_1^2 ds}}$$

and

$$\frac{n^{1/2-d_2}\sum_{t=1}^{n} (S_{1t}-\bar{S}_1)x_{2t}}{\sqrt{\sum_{t=1}^{n} (S_{1t}-\bar{S}_1)^2 \sum_{t=1}^{n} x_{2t}^2 - \left(\sum_{t=1}^{n} (S_{1t}-\bar{S}_1)x_{2t}\right)^2}} \approx \frac{\int_0^1 X_1 dX_2 - X_2(1)\int_0^1 X_1 ds}{\sigma_2 \sqrt{\int_0^1 X_1^2 - \left(\int_0^1 X_1\right)^2}}$$

where $S_{1t} = \sum_{s=1}^{t} x_{1s}$, $\sigma_2^2 = \operatorname{plim} n^{-1} \sum_{t=1}^{n} x_{2t}^2$. In these expressions the stochastic integrals from 2. appear in the numerator. These statistics are normalized to be $O_p(1)$ using the facts that $\sum_{t=1}^{n} S_{1t}x_{2t} = O_p(n^{1+d_1+d_2})$ and $\sum_{t=1}^{n} S_{1t}^2 = O(n^{2+2d_1})$. Quantiles of the distributions are given in Table 6 and the kernel densities are plotted in Figure 7.

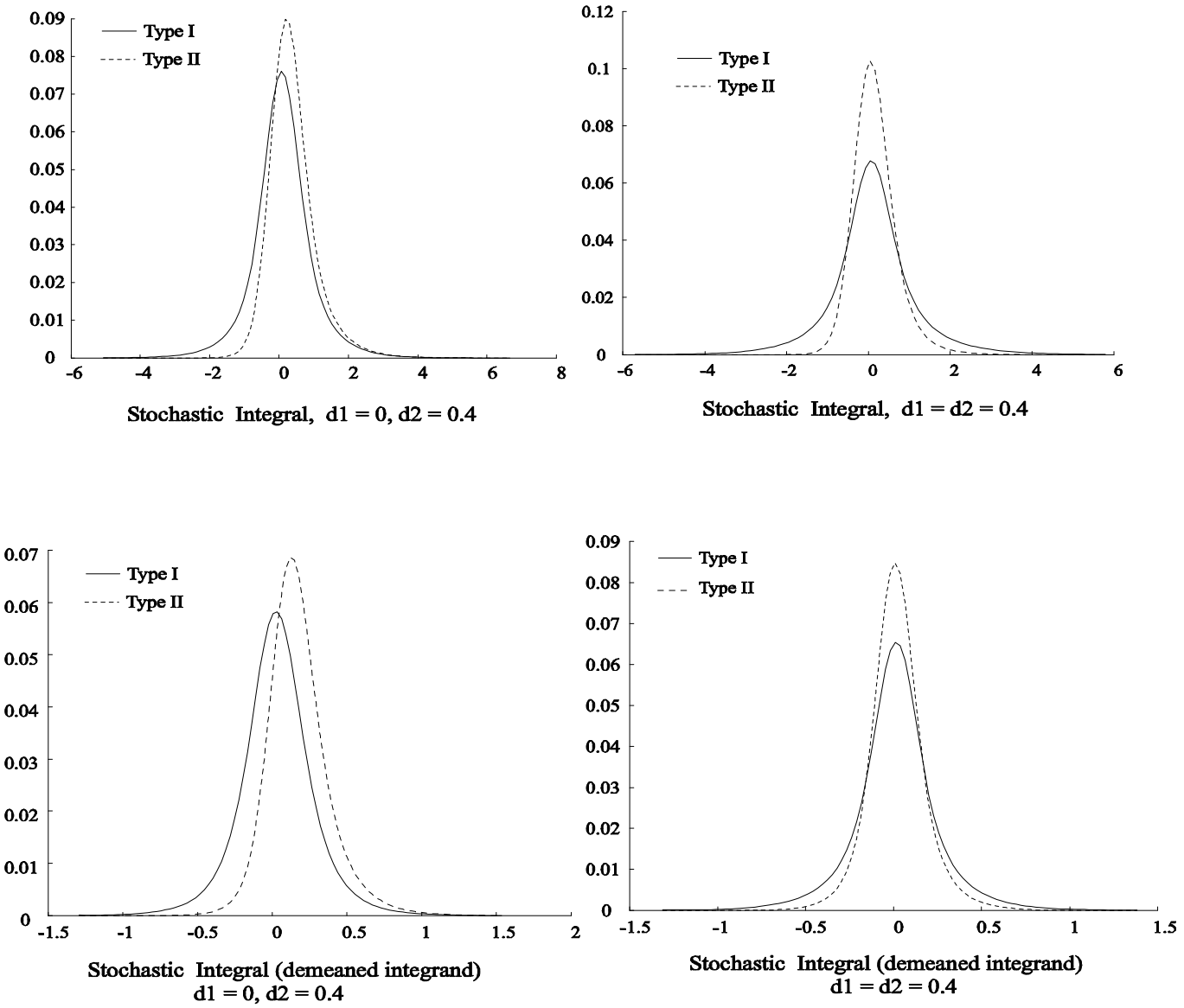

Figure 6: Simulations of a bivariate distribution with correlation 0.5. Integrand has parameter d1, integrator has parameter d2. 1000 observations, 100,000 replications.

		$P(\leq)$	0.01	0.05	0.1	0.9	0.95	0.99
no intercept	$d_1 = 0$	Type I	−1.711	−1.135	−0.879	0.977	1.297	1.810
		Type II	−0.672	−0.322	−0.172	1.125	1.375	1.774
	$d_1 = 0.4$	Type I	−1.913	−1.353	−1.033	1.206	1.526	2.086
		Type II	−0.986	−0.643	−0.446	0.977	1.173	1.566
with intercept	$d_1 = 0$	Type I	−0.868	−0.570	−0.437	0.523	0.689	0.954
		Type II	−0.381	−0.175	−0.056	0.770	0.888	1.124
	$d_1 = 0.4$	Type I	−0.885	−0.623	−0.460	0.487	0.650	0.912
		Type II	−0.778	−0.550	−0.387	0.525	0.655	0.916

Table 6: Quantiles of the cointegrating regression "t statistics"

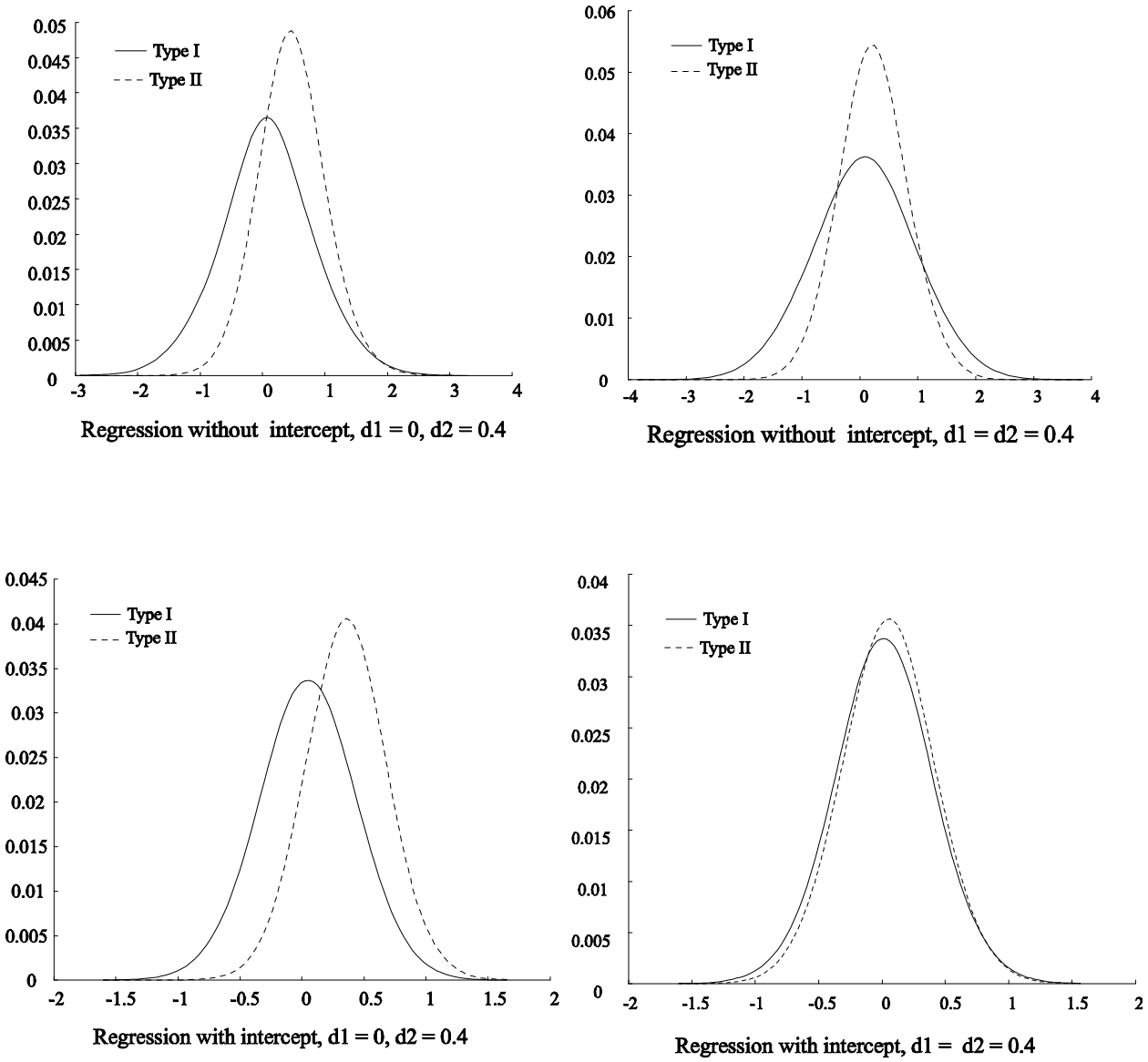

Figure 7: Simulations of regression t-ratios. 1000 observations, 100,000 replications.

Acknowlegements

The first draft of this paper was prepared for the 'Long Memory for Francesc Marmol' Conference, Madrid, May 4-5 2006. We thank Ulrich Mueller and Søren Johansen for helpful comments and discussion.

References

Abry, P. and F. Sellan (1996) The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and Fast Implementation. *Applied and Computational Harmonic Analysis* 3, 377-383.

Beran, J. (1994) *Statistics for Long Memory Processes*. New York: Chapman and Hall.

Billingsley, P (1968) *Convergence of Probability Measures*, John Wiley and Sons

Caporale, G.-M. and L. Gil-Alana (2008) Modelling the US, UK and Japanese unemployment rates: Fractional integration and structural breaks. *Computational Statistics & Data Analysis* 52 (11), 4998-5013.

Cheung. Y.-W. and 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

Coakley, J,.J. Dollery and N. Kellard (2008) The role of long memory in hedging effectiveness, *Computational Statistics & Data Analysis* 52 (6) 3075-3082

Davidson, J. and R. M. de Jong (2000) The functional central limit theorem and convergence to stochastic integrals II: fractionally integrated processes. *Econometric Theory* 16, 5, 643–666.

Davidson, J. (2008) *Time Series Modelling 4.27*, at http://www.timeseriesmodelling.com

Davidson, J. (2006) Alternative bootstrap procedures for testing cointegration in fractionally integrated processes. *Journal of Econometrics* 133 (2), 741-777.

Davidson, J. and P. Sibbertsen (2005) Generating schemes for long memory processes: regimes, aggregation and linearity *Journal of Econometrics* 128, 253-282

Davidson, J. and N. Hashimzade (2008) Alternative frequency and time domain versions of fractional Brownian motion. *Econometric Theory* 24(1), 256-293.

Davidson, J. and N. Hashimzade (2009) Representation and weak convergence of stochastic integrals with fractional integrator processes. *Econometric Theory*, forthcoming.

Davies, R. B. and D. S. Harte (1987) Tests for Hurst effect. *Biometrika* 74, 95-102

Dolado, J., J. Gonzalo and L. Mayoral (2002) A fractional Dickey Fuller test for unit roots, *Econometrica* 70, 1963-2006

Doornik, J. A. (2006) *Ox: an Object-Oriented Matrix Programming Language*. Timberlake Consultants Ltd.

Doornik, L. A. and Ooms (2006) A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.04 for Ox, at http://www.doornik.com.

Enriquez, N. (2004) A simple construction of the fractional Brownian motion, *Stochastic Processes and their Applications* 109, 203-223.

Gradshteyn and Ryzhik (2000) *Tables of Integrals, Series, and Products* (6th Edn), eds A. Jeffrey and D. Zwillinger. Academic Press.

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

Granger, C.W.J. and Roselyne Joyeux (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 1, 15-29.

Haldrup, N. and M. Ø. Nielsen (2007) Estimation of fractional integration in the presence of data noise. C*omputational Statistics & Data Analysis* 51 (6), 3100-3114

Johansen, S and M. Ø. Nielsen (2008) Likelihood inference for a nonstationary fractional autoregressive model, CREATES working paper, University of Aarhus

Mandelbrot, B. B. and J. W. van Ness (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 4, 422–437.

Marinucci, D. and P. M. Robinson (1999) Alternative forms of fractional Brownian motion. *Journal of Statistical Inference and Planning* 80, 111–122.

Marinucci, D. and P. M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, pp.103-120.

Meyer, Y., F. Sellan and M. S. Taqqu (1999) Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion. *Journal of Fourier Analysis and Applications* 5 (5), 465-494.

Pipiras, V., (2004) On the usefulness of wavelet-based simulation of fractional Brownian motion, Working paper, University of North Carolina at Chapel Hill.

Pipiras, V. (2005) Wavelet-based simulation of fractional Brownian motion revisited. *Applied Computational and Harmonic Analysis* 19, 49–60

Sowell, F (1992) Maximum likelihood estimation of stationary fractionally integrated time series models. *Journal of Econometrics* 53, 165-188