Finite Sample Performance in Cointegration Analysis of Nonlinear Time Series with Long Memory

Afonso Gonçalves da Silva and Peter M. Robinson*

Department of Economics,
London School of Economics and Political Science,
Houghton Street,
London WC2A 2AE, UK

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6674

Discussion paper
No. EM/2006/501
April 2006

* Corresponding author. Tel.: +44 (0)20 7955 7516; fax: +44 (0)20 7955 6592.
E-mail: p.m.robinson@lse.ac.uk.

Abstract

Nonlinear functions of multivariate financial time series can exhibit long memory and fractional cointegration. However, tools for analysing these phenomena have principally been justified under assumptions that are invalid in this setting. Determination of asymptotic theory under more plausible assumptions can be complicated and lengthy. We discuss these issues and present a Monte Carlo study, showing that asymptotic theory should not necessarily be expected to provide a good approximation to finite-sample behaviour.

JEL Classification: C32
Keywords: Fractional cointegration; memory estimation; stochastic volatility.

1 Introduction

Fractional cointegration analysis is increasingly found to be a promising tool for dimensionality reduction in financial time series. On the one hand, series of asset returns may have little autocorrelation, whereas instantaneous nonlinear functions, such as squares, can exhibit evidence of long memory. Considering series on several assets, it is possible that there exists a linear combination of the nonlinear functions that has shorter memory. Then there is said to be fractional cointegration. Note that here, as implied by many stochastic volatility (SV) models, series are supposed to be stationary. By contrast, in analysing macroeconomic time series, levels are typically believed to be nonstationary with a unit root, and cointegration exists when there is a linear combination that is stationary (with short memory).

A variety of tools for analysing fractional cointegration in stationary series is becoming available. The main stress has been on “semiparametric” methods. These avoid full parameterisation of autocorrelation, in favour of a local power law for the spectral density around zero frequency. Estimates of memory parameters can be rendered inconsistent by misspecification of short memory properties. Moreover, when the cointegrating relation is expressed in regression form, with one of the observables on the left hand side, the other observables cannot plausibly be assumed orthogonal to the cointegrating errors. Thus (“full-band”) time domain procedures (in a stationary environment) such as least squares will inconsistently estimate the cointegrating vector. This leads to a focus on methods based on a vanishing neighbourhood of zero in the frequency domain, such that the number, m, of Fourier frequencies used increases with sample size n, but more slowly. An undesirable consequence of this semiparametric strategy is rates of convergence (in case of both memory parameters and cointegrating vector estimates) that are slower than would be possible in a fully parametric setting. However, parametric estimates of memory parameters and (due to the stationarity) cointegrating vectors can only converge at rate $n^{\frac{1}{2}}$ (there is no super-consistency), and the slower rates of the semiparametric methods (depending on m) may be acceptable when n is very large indeed, as is the case with many financial time series.

Asymptotic theory for the semiparametric estimates has been developed mainly under assumptions that are unfortunately implausible in this setting. Usually series have been assumed to be generated by linear filters of conditionally homoscedastic martingale differences. This is justified if, for example, series are Gaussian. Recall, however, that in financial series the long memory property, and the possibility of fractional cointegration, has tended to emerge only for certain nonlinear functions. It is possible (see e.g. Hurvich, Moulines, and Soulier, 2005) to specify SV models for which the log-squares transformation yields a linear representation, on which linear filter assumptions might be plausible. Note, however, that linear factor models for asset returns are widely used in the asset pricing literature. As discussed by Gonçalves da Silva and Robinson (2005), the presence of aditional additive errors, which seems realistic, would render this type of "linearisation" impossible. Furthermore, these results crucially hinge on particular parametric specifications for the SV model, which are proposed, at least in part, for reasons of technical convenience. As a result, linear-in-martingale-difference representations should not be assumed to necessarily hold for volatility measures. Models for them can be articulated, in terms of underlying independent and identically distributed (iid) sequences, say, but the nonlinearity makes derivation of asymptotic properties (already a delicate matter in the linear setting) extremely complicated and lengthy. Moreover, due to second order bias that affects some estimates, useful limit distribution theory is unavailable. As a result, relevant asymptotic theory is not well developed, in view of which Monte Carlo simulation here plays a rather larger role than the usual one of investigating relevance of asymptotic theory in finite samples.

In the following section we consider the modelling of cointegration of series that are generated by SV models. Section 3 discusses methods of estimating cointegrating coefficients (the stress being on relatively simple "single equation" methods), and also memory parameters. Section 4 presents Monte Carlo simulations.

2 Long memory, cointegration and stochastic volatility

Consider first a covariance stationary scalar process z_t, $t = 0, \pm 1, \ldots$, having spectral density $f_z(\lambda)$, $\lambda \in (-\pi, \pi]$. We say that z_t is a (fractionally integrated) $I(d)$ process, for $d \in (-\frac{1}{2}, \frac{1}{2})$, if

$$f_z(\lambda) \sim C\,|\lambda|^{-2d}, \qquad \text{as } \lambda \to 0, \tag{2.1}$$

for some $C \in (0, \infty)$, "$\sim$" meaning that the ratio of left- and right-hand sides tends to 1. We call d the "memory parameter" of z_t. An $I(0)$ process is said to have short memory, an $I(d)$ process for $d < 0$ is said to have negative memory, and an $I(d)$ process for $d > 0$

is said to have long memory. We will focus on cases $d \geq 0$.

Now consider a $p \times 1$ column vector $Z_t = (z_{1t}, \ldots, z_{pt})'$, such that z_{it} is $I(d_i)$, $d_i \in [0, \frac{1}{2})$, $i = 1, \ldots, p > 1$. In general it is supposed that there is cross-correlation between the z_{it} but it is not necessary at present to discuss the nature of this, except to note that by the Schwarz inequality the cross-spectral density at frequency λ between z_{it} and z_{jt} has modulus of order no greater than $|\lambda|^{-d_i - d_j}$ as $\lambda \to 0$. Now suppose that there exists an unknown nonzero $p \times 1$ vector α (the "cointegrating vector") such that the unobservable process $u_t = \alpha' Z_t$ is an $I(d_u)$ process, for $d_u < \min_i d_i$. Then Z_t is said to be fractionally cointegrated. Notice that if $p = 2$ a necessary condition for fractional cointegration is that $d_1 = d_2$. Alternative definitions of fractional cointegration are reviewed by Robinson and Yajima (2002), who also discuss the possibility of existence of two or more cointegrating relations, and methods for estimating the number of these.

It is desirable to reconcile these properties of long memory and fractional cointegration with a more fundamental modelling of Z_t, which is plausible in financial series. Consider a jointly strictly stationary $q \times 1$ vector process η_t, for $q \geq p$, such that

$$z_{it} = g_i(\eta_t), \qquad i = 1, \ldots, p, \tag{2.2}$$

where the g_i are nonlinear functions. As analysed in this kind of general setting by Robinson (2001), if at least one element of η_t has long memory, then, for given i, z_{it} may have long memory, though the existence of long memory, and the actual value of d_i, depends on the nature of g_i as well as memory parameters of elements of η_t. In view of the nonlinearity, theoretical analysis is greatly facilitated if η_t is Gaussian but it is not necessary to stress this possibility here.

It may be possible, further, to infer the cointegrating relation for Z_t from an underlying structural relation for η_t. We consider perhaps the simplest case. We take $p = 2$, $q = 4$, write $\eta_t = (\eta_{1t}, \eta_{2t}, \eta_{3t}, \eta_{4t})'$, and assume it is Gaussian. Suppose that the $\{\eta_{it}\}$ are mutually independent processes, that for $i = 1, 2, 3$ the η_{it} are iid with zero mean and variance σ_i^2, and that η_{4t} is an $I(d_4)$ process, for $d_4 > 0$. Suppose that we observe sequences x_t, y_t, generated by

$$x_t = \theta_1 \zeta_t + \eta_{1t}, \tag{2.3}$$

$$y_t = \theta_2 \zeta_t + \eta_{2t}, \tag{2.4}$$

where $\theta_1, \theta_2 \neq 0$ and

$$\zeta_t = \eta_{3t} h(\eta_{4t}), \tag{2.5}$$

where h is a possibly nonlinear function, with $E\{h(\eta_{4t})^2\} < \infty$.

This setup can be interpreted as a factor model for asset returns, x_t and y_t, where ζ_t is the (unobservable) market return, and θ_1, θ_2 are the market risk exposures of x_t and y_t, respectively. Since memory properties (of volatilities, in this case) are invariant to temporal aggregation (see Chambers, 1998), (2.3)-(2.5) should be a reasonable model across all sampling frequencies. Now ζ_t is not an iid sequence but it is a square-integrable martingale difference, and thus uncorrelated, sequence, as therefore are x_t and y_t. Thus x_t and y_t exhibit an ideal property of asset returns, say. Because x_t and y_t are therefore $I(0)$ sequences, and all linear combinations of them are also $I(0)$, they are not cointegrated. However, we can deduce a cointegrating relation between the squares $z_{1t} = x_t^2$, $z_{2t} = y_t^2$. We have

$$\begin{aligned} z_{2t} &= (\theta_2 \zeta_t + \eta_{2t})^2 \\ &= \beta z_{1t} + u_t, \end{aligned} \tag{2.6}$$

where $\beta = \theta_2^2/\theta_1^2$ and

$$u_t = \eta_{2t}^2 + 2\theta_2 \eta_{2t} \zeta_t - 2\beta\theta_1 \eta_{1t} \zeta_t - \beta \eta_{1t}^2. \tag{2.7}$$

Clearly u_t has no autocorrelation, and is thus an $I(0)$ process. We have

$$z_{1t} = \theta_1^2 \zeta_t^2 + 2\theta_1 \eta_{1t} \zeta_t + \eta_{1t}^2. \tag{2.8}$$

The last two terms on the right are also $I(0)$. However for suitable h, the leading term $\theta_1^2 \zeta_t^2$ has long memory, and thence so has z_{1t}. For example if $\zeta_t = \eta_{3t}\eta_{4t}^2$, z_{1t} is $I(2d_4 - \frac{1}{2})$, or if $\zeta_t = \eta_{3t} e^{\eta_{4t}}$, z_{1t} is $I(d_4)$. In either case, z_{2t} has the same memory parameter as z_{1t}, and $Z_t = (z_{1t}, z_{2t})'$ is fractionally cointegrated, with cointegrating vector $\alpha = (-\beta, 1)'$. A similar conclusion is drawn if, even more simply, η_{1t} is missing from (2.3). Notice that ζ_t is generated by a SV model and plays the role of a common factor. Fractional cointegration can also arise if η_{1t} and/or η_{2t} are replaced by processes with SV (so that u_t can have long memory), as shown by Gonçalves da Silva and Robinson (2005).

Though (2.6) is expressed in the form of a regression model, it does not possess the classical properties. The unobservable sequence u_t actually has nonzero mean (as does z_{1t}), but this situation is rectified by introducing an intercept. More important, however, u_t is not orthogonal to the right hand side observable z_{1t}:

$$\operatorname{Cov}(z_{1t}, u_t) = -2\beta\sigma_1^2 \left\{\sigma_1^2 + 2E(\zeta_t^2)\right\} < 0, \tag{2.9}$$

taking $\theta_1 = 1$ with no loss of generality. For general p, after rewriting $\alpha' Z_t = u_t$ in regression form, then even in the absence of an underlying structure like (2.3), (2.4) there is no reason to suppose that orthogonality between cointegrating errors and right-hand side regressors obtains, especially as the designation of left-hand variable is arbitrary.

3 Estimation of cointegrating vector and memory parameters

Assuming the p-th element of α is non-zero, adopting an arbitrary normalization, and designating z_{pt} as left-hand side variable, we rewrite the cointegrating relation $\alpha' Z_t = u_t$ as

$$Y_t = \beta' X_t + u_t, \tag{3.1}$$

where $Y_t = z_{pt}$, $X_t = (z_{1t}, \ldots, z_{p-1,t})'$ and β is a $(p-1) \times 1$ vector. It is desired to estimate the unknown $\beta = (\beta_1, \ldots, \beta_{p-1})'$, on the basis of observables Z_t, $t = 1, \ldots, n$.

The most obvious estimate of β is ordinary least squares (OLS) with intercept correction (bearing in mind that u_t may have non-zero mean, as the discussion of the previous section suggests). This is

$$\hat{\beta}_O = \left\{\sum_{t=1}^{n} (X_t - \bar{X}) X_t'\right\}^{-1} \sum_{t=1}^{n} (X_t - \bar{X}) Y_t, \tag{3.2}$$

where $\bar{X} = n^{-1}\Sigma_{t=1}^{n} X_t$. However, the correlation envisaged between u_t and X_t makes $\hat{\beta}_O$ inconsistent for β, bearing in mind also the stationarity of Z_t; this outcome differs from the familiar one in which Z_t has a unit root and u_t is $I(0)$, where the asymptotic dominance of sums of squares of u_t by those of X_t overwhelms the simultaneous equation bias, leading to n-consistency of $\hat{\beta}_O$.

A consistent estimate of β was proposed by Robinson (1994). For a vector sequence

a_t, define the discrete Fourier transform

$$w_a(\lambda) = (2\pi n)^{-\frac{1}{2}} \sum_{t=1}^{n} a_t e^{it\lambda}, \tag{3.3}$$

and for a vector sequence b_t, possibly the same as a_t, define the (cross-) periodogram matrix

$$I_{ab}(\lambda) = w_a(\lambda) w_b'(-\lambda). \tag{3.4}$$

For a sequence $m = m(n)$ such that

$$m \leq \frac{n}{2}, \tag{3.5}$$

$$\frac{1}{m} + \frac{m}{n} \to 0, \qquad \text{as } n \to \infty, \tag{3.6}$$

define the narrow-band least squares estimate of β,

$$\hat{\beta}_{NB} = \left(\sum_{j=1}^{m} \operatorname{Re}\{I_{XX}(\lambda_j)\} \right)^{-1} \sum_{j=1}^{m} \operatorname{Re}\{I_{XY}(\lambda_j)\}, \tag{3.7}$$

where $\lambda_j = 2\pi j/n$ and $\operatorname{Re}(\cdot)$ is the real part operator. Note that omission of the frequency $\lambda_0 = 0$ corresponds to a sample mean correction like that in (3.2), while if in contrast to (3.5), (3.6), $m = n - 1$ we have $\hat{\beta}_{NB} = \hat{\beta}_O$. However, the condition (3.6) is crucial to the consistency of $\hat{\beta}_{NB}$. The basic intuition for the consistency is as follows. By the Cauchy inequality, for $i = 1, \ldots, p-1$,

$$\left| \sum_{j=1}^{m} \operatorname{Re}\{I_{z_i u}(\lambda_j)\} \right| \leq \left\{ \sum_{j=1}^{m} I_{z_i z_i}(\lambda_j) \sum_{j=1}^{m} I_{uu}(\lambda_j) \right\}^{\frac{1}{2}}, \tag{3.8}$$

and under suitable conditions this is

$$O_p\left(n \left\{ \int_0^{\lambda_m} \lambda^{-2d_i} d\lambda \int_0^{\lambda_m} \lambda^{-2d_u} d\lambda \right\}^{\frac{1}{2}} \right) = O_p\left(n \left(\frac{m}{n} \right)^{1-d_i-d_u} \right). \tag{3.9}$$

On the other hand, under suitable conditions, for $\Lambda_m = \operatorname{diag}\left\{\lambda_m^{d_1}, \ldots, \lambda_m^{d_{p-1}}\right\}$,

$$\frac{1}{m} \Lambda_m \sum_{j=1}^{m} \operatorname{Re}\{I_{XX}(\lambda_j)\} \Lambda_m \to_p \Omega, \tag{3.10}$$

where Ω is a constant positive definite matrix. It follows that

$$\hat{\beta}_{NB,i} - \beta_i = O_p\left(\left(\frac{m}{n}\right)^{d_i - d_u}\right), \qquad i = 1, \ldots, p-1, \tag{3.11}$$

where β_i and $\hat{\beta}_{NB,i}$ are the i-th elements of β and $\hat{\beta}_{NB}$. Since cointegration entails $d_u < d_i$, $i = 1, \ldots, p-1$, $\hat{\beta}_{NB}$ is thus consistent for β. The key is the domination, near zero, of the spectral density of u_t by the spectral densities of $z_{1t}, \ldots, z_{p-1,t}$.

Consistency of $\hat{\beta}_{NB}$ was first shown by Robinson (1994) in case $p = 2$, and then, with the rate in (3.11), by Robinson and Marinucci (2003) for general p. The conditions they imposed to deduce the crucial properties (3.9) and (3.10) were that Z_t is generated by a linear moving average in conditionally homoscedastic martingale differences. As previously noted, this is inconsistent with our SV setup, such as illustrated in the previous section, albeit similar to one for log squared returns for a certain SV model (see e.g. Deo and Hurvich, 2001) and a multiplicative set-up in place of the additive one, typified in (2.3), (2.4). However, Gonçalves da Silva and Robinson (2005) have established (3.11) for $p = 2$ under a somewhat more general set-up than that described in connection with (2.3) and (2.4). The proof is exceedingly lengthy, however, requiring Hermite approximations to the first and second moments of the periogram, and things do not generalise immediately to the case $p > 2$, where one must also consider the asymptotic behaviour of the cross-periodogram to establish (3.10).

The estimate $\hat{\beta}_{NB}$ is desirably computationally simple, and the exclusion of high-frequency contributions makes this estimate robust to contamination by short run dynamics, such as those introduced by microstructure noise. It has been applied in fractional cointegration analyses of implied and realised volatility by Christensen and Nielsen (2004), Bandi and Perron (2004).

In general the rate in (3.11) is sharp, and indeed under additional conditions it seems that, for each i, $(n/m)^{d_i - d_u}(\hat{\beta}_{NBi} - \beta_i)$ converges in distribution not to a non-degenerate random variable, but to a constant. This is due to the presumed coherence between X_t and u_t around zero frequency. Without such coherence, asymptotic normality and a faster rate of convergence is possible. Christensen and Nielsen (2004) supposed that the cross-spectral density between z_{it} and u_t is $o(|\lambda|^{-d_i - d_u})$, as $\lambda \to 0$, rather than having real part behaving precisely like $|\lambda|^{-d_i - d_u}$. Assuming also that $d_i + d_u < \frac{1}{2}$, $i = 1, \ldots, p-1$, they deduced that $m^{\frac{1}{2}}(m/n)^{d_u}\Lambda_m^{-1}(\hat{\beta}_{NB} - \beta)$ is asymptotically multivariate normal; they assumed Z_t is linear in homoscedastic martingale differences, as in Robinson (1994), Robinson and Marinucci (2003).

Though the model constructed in Section 2, (2.6) based on (2.3)-(2.5) and $z_{1t} = X_t = x_t^2$, $z_{2t} = Y_t = y_t^2$, does not satisfy the linearity assumption of Christensen and Nielsen (2004), it does satisfy a lack-of-coherence assumption that corresponds to theirs. It is easily seen that $\text{Cov}(z_s, u_t) = 0$ if $s \neq t$, so in view of (2.8), the cross-spectral density of z_{1t}, u_t is finite and constant, and $o(|\lambda|^{-\delta})$, where $\delta > 0$ represents the memory parameter of z_{1t}. (In the cases discussed after (2.8), the possibilities that $\delta = d_4$ and $\delta = 2d_4 - \frac{1}{2}$ emerged.)

Violation of orthogonality represents an important way in which (3.1) disobeys classical regression conditions, but it is not the only one. Though the simple set-up with $p = 2$ analysed in the previous section ensured that u_t has no autocorrelation (see (2.7)), more generally u_t can be not only autocorrelated but even have long memory, as indicated by Gonçalves da Silva and Robinson (2005). In the absence of simultaneous equations bias, a suitable weighted frequency domain estimate will be more efficient. In (3.1) with short memory u_t orthogonal to X_t, Hannan (1963) showed that weighting inversely with respect to a nonparametric estimate of f_u can achieve the same asymptotic efficiency as generalised least squares based on a correctly specified parametric model for f_u. Hidalgo and Robinson (2002) extended this finding to long memory u_t, with unknown d_u. However, the "full-band" estimates will incur similar simultaneous equations bias to $\hat{\beta}_O$. Nevertheless, it is worth considering whether some such weighting can improve on $\hat{\beta}_{NB}$, since f_u changes even over the interval $[\lambda_1, \lambda_m]$. Smith and Chen (1996) proposed the weighted narrow-band estimate

$$\hat{\beta}_{WNB} = \tilde{\beta}(\hat{d}_u), \tag{3.12}$$

where

$$\tilde{\beta}(d) = \left(\sum_{j=1}^{m} \lambda_j^{2d} \operatorname{Re}\{I_{XX}(\lambda_j)\} \right)^{-1} \sum_{j=1}^{m} \lambda_j^{2d} \operatorname{Re}\{I_{XY}(\lambda_j)\}, \tag{3.13}$$

and $\hat{d}_u$ is a consistent estimate of d_u (see below). Note that $\tilde{\beta}(0) = \hat{\beta}_{NB}$. Smith and Chen (1996) in fact proposed $\hat{\beta}_{WNB}$ in a more traditional regression setting, with u_t orthogonal to X_t, and did not establish any asymptotic properties. Recently, Nielsen (2005), under the same kind of incoherence-near-zero assumption as Christensen and Nielsen (2004), established that for given d, which satisfies a suitable constraint relative to d_u and the d_i, $m^{\frac{1}{2}}(m/n)^{d_u}\Lambda_m^{-1}(\tilde{\beta}(d) - \beta)$ is asymptotically normal. Nielsen (2005) also discussed the relative efficiency of $\tilde{\beta}(d)$ and $\hat{\beta}_{NB}$, noting some circumstances in which

$\tilde{\beta}(d)$ can be the more efficient even when $d \neq d_u$.

However d_u is clearly an optimal choice of d, and given that d_u is unknown it is natural to focus on $\hat{\beta}_{WNB}$ which, like $\hat{\beta}_{NB}$, should still be consistent in the presence of coherence between u_t and X_t, violating Nielsen's (2005) condition. We have, say,

$$\left| \sum_{j=1}^{m} \lambda_j^{2\hat{d}_u} \operatorname{Re}\{I_{z_i u}(\lambda_j)\} \right| \leq \left\{ \sum_{j=1}^{m} \lambda_j^{2\hat{d}_u} I_{z_i z_i}(\lambda_j) \sum_{j=1}^{m} \lambda_j^{2\hat{d}_u} I_{uu}(\lambda_j) \right\}^{\frac{1}{2}}. \tag{3.14}$$

We assume (as in Robinson, 1994)

$$(\log n)(\hat{d}_u - d_u) \to_p 0, \qquad \text{as } n \to \infty, \tag{3.15}$$

which can readily be justified in view of asymptotic theory for various memory parameter estimates. Then

$$\lambda_j^{2\hat{d}_u} = \lambda_j^{2d_u} \lambda_j^{2(\hat{d}_u - d_u)} \leq \lambda_j^{2d_u} n^{o((\log n)^{-1})} \leq \lambda_j^{2d_u} e^{o(1)} \leq 2\lambda_j^{2d_u}, \tag{3.16}$$

for n sufficiently large. It is then readily seen that (3.14) is

$$O_p\left(n \left\{ \int_0^{\lambda_m} \lambda^{2(d_u - d_i)} d\lambda \frac{m}{n} \right\}^{\frac{1}{2}} \right) = O_p\left(n \left(\frac{m}{n} \right)^{1 + d_u - d_i} \right). \tag{3.17}$$

Also under (3.15), and similar conditions to those giving (3.10), we can justify the step

$$\frac{1}{m} \left(\frac{m}{n} \right)^{-2d_u} \Lambda_m \sum_{j=1}^{m} \left(\lambda_j^{2\hat{d}_u} - \lambda_j^{2d_u} \right) \operatorname{Re}\{I_{XX}(\lambda_j)\} \Lambda_m \to_p 0, \tag{3.18}$$

and then that

$$\frac{1}{m} \left(\frac{m}{n} \right)^{-2d_u} \Lambda_m \sum_{j=1}^{m} \lambda_j^{2d_u} \operatorname{Re}\{I_{XX}(\lambda_j)\} \Lambda_m \tag{3.19}$$

converges in probability to a constant positive definite matrix.

Notice that in the model (2.6) derived from (2.3)-(2.5), $d_u = 0$ so we expect no improvement of $\hat{\beta}_{WNB}$ over $\hat{\beta}_{NB}$. However, we can extend this model to allow at the same time $d_u > 0$, and incoherence at frequency zero between regressors and errors.

At least for linear processes, bias and autocorrelation can be corrected simultaneously by more elaborate methods. These are based on a full system of p equations that expresses also the long memory properties of the z_{it}, $i = 1, \ldots, p-1$, and lead to estimates of β which depend not only on $\hat{d}_u$, but also on estimates of the d_i, $i = 1, \ldots, p-1$. Such estimates of β are developed by Hualde and Robinson (2004); they are asymptotically normal (centered at β) with the same rate as described for $\hat{\beta}_{NB}$ and $\hat{\beta}_{WNB}$ under the incoherence-near-zero assumption, but without imposing that. We focus in our numerical study in the following section only on the "single-equation" estimates (based on (3.1)) we have discussed above, this is partly due to their computational simplicity, but also because incoherence-near-zero can often be justified in a factor model context, as discussed above, whence $\hat{\beta}_{NB}$ and $\hat{\beta}_{WNB}$ enjoy a reasonably fast rate of convergence.

Even if simple estimates of β are used, there may be interest in estimation of the d_i, as well as in estimation of d_u, as is required for $\hat{\beta}_{WNB}$. In particular, such estimates are useful in determining the existence and extent of cointegration, as described by Robinson and Yajima (2002). In this multivariate setting, efficiency gains are possible by estimating memory parameters jointly, especially if prior equality constraints are placed on the d_i. However, joint estimates have principally been developed under the assumption of no cointegration, and if there is cointegration they are liable to be inconsistent. Thus we describe some leading "semiparametric" estimates. We introduce a generic univariate stationary process v_t which can represent any of the z_{it}, or, where estimation of d_u is concerned, residuals $y_t - \tilde{\beta}' X_t$, such that $\tilde{\beta}$ represents one of our consistent estimates of β.

Denote by d the unknown memory parameter of v_t. Geweke and Porter-Hudak (1983) proposed a log-periodogram estimate, a simplified version of which, due to Robinson (1995a), is

$$\hat{d}_{LP} = \left\{ \sum_{j=1}^{m} \left(\ln j - \frac{1}{m} \sum_{i=1}^{m} \ln i \right)^2 \right\}^{-1} \sum_{j=1}^{m} \left(\ln j - \frac{1}{m} \sum_{i=1}^{m} \ln i \right) \ln I_{vv}(\lambda_j). \tag{3.20}$$

Assuming that m satisfies at least (3.5) and (3.6), Robinson (1995a), Hurvich, Deo, and Brodsky (1998) showed that

$$m^{\frac{1}{2}}(\hat{d}_{LP} - d) \to_d N(0, \pi^2/6), \qquad \text{as } n \to \infty. \tag{3.21}$$

An efficiency improvement is possible, for the same m sequence, via the local Whittle estimate (Künsch, 1987),

$$\hat{d}_{LW} = \arg\min_{d\in\mathcal{D}} \left\{ \ln\left(\sum_{j=1}^{m} j^{2d} I_{vv}(\lambda_j)\right) - d\frac{1}{m}\sum_{j=1}^{m}\ln j \right\}, \tag{3.22}$$

where $\mathcal{D}$ is a compact subset of $(-\frac{1}{2}, \frac{1}{2})$. This was shown by Robinson (1995b) to satisfy

$$m^{\frac{1}{2}}(\hat{d}_{LW} - d) \to_d N(0, \frac{1}{4}), \qquad \text{as } n \to \infty. \tag{3.23}$$

Note that the conditions imposed to deduce (3.21) and (3.23) do not cover the SV setup described in the previous section, but see e.g. Deo and Hurvich (2001). Various modifications, in particular bias corrections, have been introduced. Hurvich, Moulines, and Soulier (2005) allow for a more refined approximation to $f_v(\lambda)$ than $C\lambda^{-2d}$, there is something of a signal-plus-noise character to the model (2.3), (2.4), so we might consider the estimate

$$(\hat{d}_{MLW}, \hat{\theta}) = \arg\min_{(d,\theta)\in\mathcal{D}\times\Theta} \left\{ \ln\left(\sum_{j=1}^{m} \frac{I_{vv}(\lambda_j)}{j^{-2d}+\theta}\right) + \frac{1}{m}\sum_{j=1}^{m}\ln(j^{-2d}+\theta) \right\}, \tag{3.24}$$

where Θ is a compact subset of the positive real line. Hurvich, Moulines, and Soulier (2005) justify asymptotic normality of $\hat{d}_{MLW}$, but with a different asymptotic variance from that in (3.23).

4 Simulations

We now present a Monte Carlo study of finite-sample performance. For linear processes, Robinson and Marinucci (2003) reported simulation experiments of NBLS with $I(1)$ observables and $I(0)$ cointegrating errors, while Marinucci and Robinson (2001) explored different cases of fractional cointegration with nonstationary observables and stationary errors. Bandi and Perron (2004) examined NBLS for the regression between realised and implied volatility, generating the data from a discretised continuous time SV model. Gonçalves da Silva and Robinson (2005) reported experiments of NBLS in a SV framework similar to ours. We present results for two settings, one linear, and the other generalising (2.2)-(2.6). Under the linear model, we generate (see (2.3), (2.4))

$$z_{1t} = \zeta_t + \delta_t, \tag{4.1}$$

$$z_{2t} = \beta\zeta_t + \varepsilon_t, \tag{4.2}$$

where we use the abbreviated notation $\zeta_t = \eta_{1t}$, $\delta_t = \eta_{2t}$, $\varepsilon_t = \eta_{3t}$, and for $i = 1, 2, 3$, $\{\eta_{it}\}$ is a zero mean Gaussian ARFIMA$(0, d_i, 0)$ process with variance σ_i^2. In the nonlinear case, we use (see (2.3)-(2.6))

$$z_{1t} = (\zeta_t + \delta_t)^2, \tag{4.3}$$

$$z_{2t} = (\theta_2\zeta_t + \varepsilon_t)^2, \tag{4.4}$$

where $\zeta_t = \xi_{1t}h(\eta_{1t})$, $\delta_t = \xi_{2t}h(\eta_{2t})$, $\varepsilon_t = \xi_{3t}h(\eta_{3t})$, and for $i = 1, 2, 3$, $\{\xi_{it}\}$ is an independent standard Gaussian sequence, and $\{\eta_{it}\}$ a zero mean Gaussian ARFIMA$(0, d_i, 0)$ with variance σ_i^2. In both models, the basic processes $\{\xi_{it}\}$ and $\{\eta_{it}\}$, $i = 1, 2, 3$, are all generated independently of each other, and we will denote the variances of ζ_t, δ_t, ε_t by σ_ζ^2, σ_δ^2, σ_ε^2 respectively.

Under each model, we employ 1,000 replications of series of length $n = 2048$ and estimate β by narrow-band regressions of z_{2t} on z_{1t}, where $\beta = \theta_2^2$ in the nonlinear setting. Note that both models can be written as (2.6), with

$$u_t = \varepsilon_t - \beta\delta_t \tag{4.5}$$

in the linear setting and

$$u_t = \varepsilon_t^2 - \beta\delta_t^2 + 2\zeta_t\left(\theta_2\varepsilon_t - \beta\delta_t\right) \tag{4.6}$$

in the nonlinear setting. We present bias, standard deviation (SD) and root mean squared error (RMSE) of $\tilde{\beta}(d)$ given by (3.13), for various values of d, both fixed and estimated. All are evaluated at the bandwidth, m^*, that minimises RMSE.

Asymptotic theory

We first examine the performance of Nielsen's (2005) asymptotic theory under the linear model, when δ_t is absent in (4.1). We set $\beta = 1$, $d_1 = 0.4$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$

d	Asy. SD	MC SD	Ratio
0.10	0.0176	0.0213	1.211
0.15	0.0155	0.0203	1.307
0.20	0.0152	0.0201	1.323
0.25	0.0154	0.0204	1.328
0.30	0.0157	0.0209	1.332
0.35	0.0161	0.0216	1.337
0.40	0.0166	0.0223	1.341
0.45	0.0171	0.0230	1.344

Table 1: Asymptotic and Monte Carlo SD of WNBLS, for varying d; linear setting with δ_t absent.

and $\sigma_\varepsilon^2 = 2$. This simulation is comparable to his model A, although we focus on full-band estimates, i.e. $m = n/2$. (Given the independence between u_t and z_{1t}, this choice dominates any other value of m.) Table 1 reports asymptotic (Asy.) and Monte Carlo (MC) SD for different values of d. Monte Carlo bias is negligible in this setting and therefore omitted. Note that Nielsen's (2005) theory requires

$$(2d_1 + 2d_3 - 1)/4 < d \leq d_3, \tag{4.7}$$

which in this case is equivalent to $0.05 < d \leq 0.2$, but we compute his asymptotic SD also for $d > 0.2$. Here we find that Monte Carlo SD is almost always over 30% larger than the asymptotic one, so the asymptotic theory is not a good approximation even when $n = 2048$.

More realistically, a complete factor model such as (4.1), (4.2) allows the explanatory variable z_{1t} to include an idiosyncratic component, δ_t. The discrepancy between z_{1t} and the ideal explanatory variable, ζ_t, can be interpreted as a case of measurement error (ME), causing z_{1t} to be correlated with u_t. While still compatible with Nielsen's (2005) assumptions, this would increase the Monte Carlo SD even further without changing the asymptotic one (as long as $d_2 < d_3$), thereby widening the gap between them. Figure 1 plots the theoretical and Monte Carlo SD of $\tilde{\beta}(d)$ relative to that of $\tilde{\beta}(d_3)$, for different values of d. Although the asymptotic and Monte Carlo levels in Table 1 substantially differ, their ratios across d are comparable, and $d = d_3$ is the optimal choice in both.

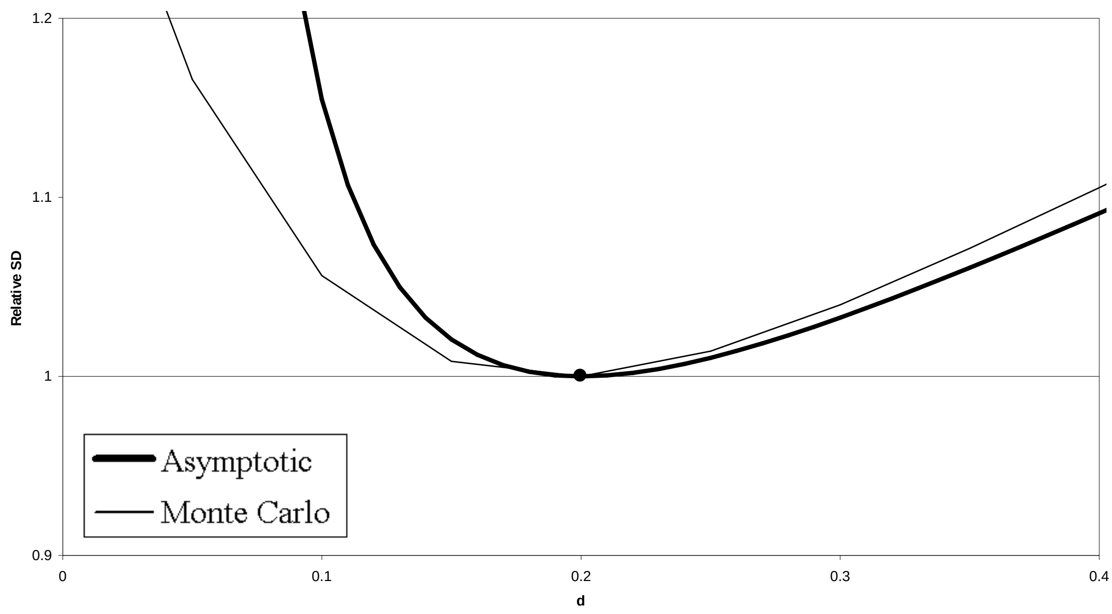

Figure 1: Asymptotic and Monte Carlo relative SD of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d; linear setting with δ_t absent.

Variation in measurement error

We present results for different types of ME, namely: no ME, i.e. δ_t absent in (4.1) or (4.3); antipersistent ME ($d_2 = -0.2$); iid ME ($d_2 = 0$); and long memory ME ($d_2 = 0.2$). In the nonlinear model, the antipersistent case would still generate $I(0)$ ME in (4.3) and is therefore omitted. In both settings we use $\beta = 1$, $d_1 = 0.4$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting.

Table 2 reports Monte Carlo optimal bandwidth, bias and RMSE, under the linear setting, for various regression estimates of β: unweighted NBLS, $\tilde{\beta}(0)$; the theoretically optimal but infeasible weighted estimate, $\tilde{\beta}(d_3)$; and feasible versions of it, $\tilde{\beta}(\hat{d}_3)$, where $\hat{d}_3$ is a consistent estimate of d_3. In these cases, d_3 is estimated using LP (3.20), LW (3.22), or MLW (3.24) based on the regression residuals from a first step unweighted NBLS regression; the same m is used in the first and second steps. Due to the modified

$\tilde{\beta}$	δ_t absent			$d_2 = -0.2$			$d_2 = 0$			$d_2 = 0.2$		
	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
NBLS	1024	-0.0005	0.0273	81	-0.0279	0.0642	25	-0.0470	0.0897	12	-0.1301	0.1752
True d_3	1024	-0.0001	0.0201	53	-0.0283	0.0652	23	-0.0555	0.0933	10	-0.1326	0.1789
LP	1024	-0.0001	0.0209	53	-0.0297	0.0651	23	-0.0524	0.0928	10	-0.1322	0.1799
LW	1024	-0.0001	0.0204	53	-0.0296	0.0650	23	-0.0524	0.0930	10	-0.1321	0.1800
MLW	1024	0.0001	0.0205	53	-0.0301	0.0650	23	-0.0539	0.0937	10	-0.1339	0.1807

Table 2: Monte Carlo bias and RMSE of regression estimates, for different types of measurement error; linear setting.

$\widehat{d}_3$	m	δ_t absent		$d_2 = -0.2$		$d_2 = 0$		$d_2 = 0.2$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
LP	80	-0.0020	0.0806	-0.0485	0.0934	-0.0574	0.0993	-0.0070	0.0821
LW	80	-0.0072	0.0628	-0.0519	0.0821	-0.0613	0.0892	-0.0108	0.0675
MLW	200	0.0491	0.1002	0.0569	0.1464	0.0184	0.1177	0.0418	0.0908

Table 3: Monte Carlo bias and RMSE of residual memory estimates, for different types of measurement error; linear setting.

spectral approximation in (3.24), when using MLW we compute WNBLS as

$$\tilde{\beta}(\hat{d},\hat{\theta}) = \left(\sum_{j=1}^{m} \frac{\operatorname{Re}\{I_{XX}(\lambda_j)\}}{j^{-2\hat{d}}+\hat{\theta}} \right)^{-1} \sum_{j=1}^{m} \frac{\operatorname{Re}\{I_{XY}(\lambda_j)\}}{j^{-2\hat{d}}+\hat{\theta}}, \tag{4.8}$$

instead of (3.13). Table 3 reports bias and RMSE for these preliminary estimates of d_3.

In the model without ME all regression estimates have, as expected, virtually no bias and perform best in the full-band case. Here, $\tilde{\beta}(d_3)$ clearly exhibits an efficiency gain over $\tilde{\beta}(0)$, which is equivalent to OLS (3.2). However, as progressively more persistent ME is introduced, both estimates have increasing bias, and the RMSE of $\tilde{\beta}(d_3)$ grows much faster than that of $\tilde{\beta}(0)$. Indeed, in the presence of ME, simple NBLS always outperforms the weighted estimate. Here and throughout all experiments, estimates are biased towards zero, due to the negative correlation between z_{1t} and u_t caused by ME. The feasible versions of WNBLS seem to closely match the infeasible one in both RMSE and bias, in many cases even appearing slightly better. This behaviour arises because whenever ME is present, the optimal weighting is actually obtained for $d < d_3$, so the negative bias of LP and LW, seen in Table 3, can actually work to their advantage. Although MLW actually displays positive bias, the weights in (4.8) do not depend on $\hat{d}_2$

$\tilde{\beta}$	m^*	δ_t absent Bias	RMSE	m^*	$d_2 = 0$ Bias	RMSE	m^*	$d_2 = 0.2$ Bias	RMSE
NBLS	973	-0.0042	0.0840	8	-0.1495	0.2717	8	-0.1944	0.3210
True d_3	973	-0.0042	0.0855	8	-0.1589	0.2829	8	-0.2020	0.3290
LP	973	-0.0042	0.0840	8	-0.1516	0.2753	8	-0.1969	0.3243
LW	973	-0.0043	0.0842	8	-0.1514	0.2747	8	-0.1966	0.3237
MLW	973	-0.0043	0.0847	8	-0.1532	0.2776	8	-0.1987	0.3266

Table 4: Monte Carlo bias and RMSE of regression estimates, for different types of measurement error; nonlinear setting.

alone but also on $\hat{\theta}$ in (3.24), allowing it to still outperform the infeasible estimate for $d_2 = -0.2$. The optimal bandwidths for each estimate are lower the more persistent the ME is, since frequencies closer to zero become more contaminated with the correlation between z_{1t} and u_t.

Table 3 shows that both LP and LW perform relatively well throughout. The small biases are insufficient for the bias reduction properties of MLW to make a difference; in fact, this estimate displays larger bias than LP and LW in three of the four cases. As expected, the much lower SD of LW makes it the best in RMSE. Although some of the bias can be attributed to estimation error, most of it surely comes from the "signal-plus-noise" nature of the residuals, as seen in (4.5). When δ_t is absent or when δ_t has the same memory as ε_t, LP and LW are essentially unbiased, while for $d_2 = -0.2, 0$ some bias is present.

Tables 4 and 5 present results for the nonlinear setting. Here it can be seen that the weighted estimate is always outperformed by NBLS, with ME causing much more significant bias. Even in the absence of ME, the optimal bandwidth is slightly below the full-band case, possibly as a consequence of u_t being orthogonal to but not independent of z_{1t}, as can be seen by setting $\delta_t = 0$ in (4.6). All feasible weighted estimates outperform the infeasible one, which can again be explained by the negative biases found in Table 5. Biases are stronger here than in the linear setting, partly because of the estimation error and the nonlinear setting, but also because of the signal-plus-noise structure. Note that in this setting the $I(0)$ noise in (4.6) does not vanish even if δ_t is absent. For both LP and LW, bias is the main component of RMSE. Therefore, the bias reduction provided by MLW allows it to dominate the other estimates in the presence of ME. Again, the inferior performance of the weighted estimate relative to simple NBLS demonstrates that $d = d_3$ is not the optimal choice in this setting.

$\widehat{d}_3$	m	δ_t absent		$d_2 = 0$		$d_2 = 0.2$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
LP	80	-0.1512	0.1704	-0.1827	0.2016	-0.1640	0.1861
LW	80	-0.1562	0.1690	-0.1847	0.1969	-0.1666	0.1810
MLW	200	-0.0411	0.1827	-0.0986	0.1840	-0.0697	0.1788

Table 5: Monte Carlo bias and RMSE of residual memory estimates, for different types of measurement error; nonlinear setting.

Naturally, if d_3 is no longer the optimal choice for d, the usefulness of estimating it from the data can be questioned. This is verified in Figures 2 and 3, which show the RMSE of $\tilde{\beta}(d)$ relative to that of $\tilde{\beta}(d_3)$, for different values of d, in the linear and nonlinear settings. Only in the linear case without ME is $d = d_3$ optimal; in all other cases, the optimal value is smaller, and it is reduced the more persistent the ME is. In the nonlinear case the optimal values for d are always negative, and in a region excluded by (4.7). It should also be noted that, in the absence of information on the optimal d, NBLS should be chosen over $\tilde{\beta}(d_3)$ (or its feasible versions). Tables 6 and 7 report optimal bandwidth, bias and RMSE for $\tilde{\beta}(d)$, with $d = 0$, 0.2 and the values of d that minimise RMSE in each case (indicated in bold-face), in the linear and nonlinear settings. The degradation in performance with more persistent ME can still be seen here, and bias is often slightly smaller for the optimal d. However, the variation in bias across d is relatively small, and most of the variation in RMSE can be explained by variations in SD.

The minimization of RMSE at values different from $d = d_3$ is surprising since it does not conform to the asymptotic theory. A frequency domain generalised least squares approach will weigh the contribution of each frequency by the inverse of their approximate SD, thereby "whitening" the observations. A possible explanation for the discrepancy lies in the approximation error in (2.1), which in the limit theory is made irrelevant by assuming enough smoothness in the spectral density, but can play a major role in finite samples. The whitening approach will give low weight to the frequencies closer to zero, where variance is higher but (2.1) is a more accurate approximation, and will boost the impact of more distant frequencies where the approximation is not so accurate. Another relevant factor is the coherence between z_{1t} and u_t, here generated by δ_t, which is the leading source of bias. Being of smaller order than the spectral pole, it will be irrelevant asymptotically, but this also means higher frequencies are more contaminated than

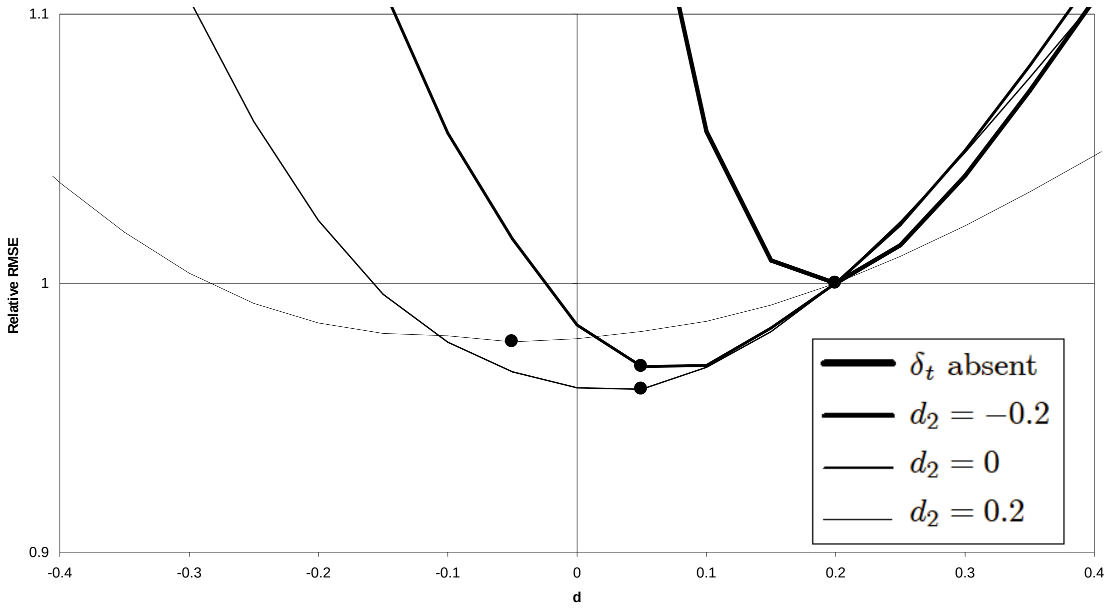

Figure 2: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and d_2; linear setting.

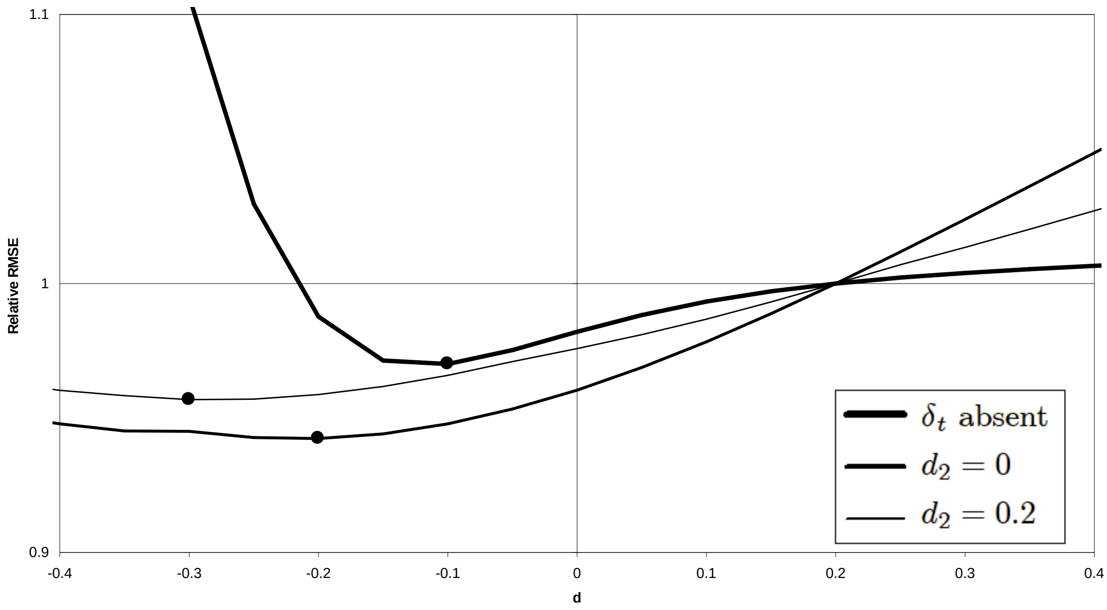

Figure 3: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and d_2; nonlinear setting.

d	δ_t absent			$d_2 = -0.2$			$d_2 = 0$			$d_2 = 0.2$		
	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.05	1024	-0.0005	0.0331	67	-0.0303	0.0663	33	-0.0519	0.0902	**12**	**-0.1269**	**0.1750**
0.00	1024	-0.0005	0.0273	55	-0.0279	0.0642	25	-0.0470	0.0897	12	-0.1301	0.1752
0.05	1024	-0.0004	0.0235	**52**	**-0.0294**	**0.0632**	**25**	**-0.0502**	**0.0896**	11	-0.1298	0.1757
0.20	**1024**	**-0.0001**	**0.0201**	39	-0.0283	0.0652	23	-0.0555	0.0933	10	-0.1326	0.1789

Table 6: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and different types of measurement error; linear setting. The minimum RMSE choice of d is indicated in bold-face.

d	δ_t absent			$d_2 = 0$			$d_2 = 0.2$		
	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	1022	-0.0025	0.0946	14	-0.1476	0.2674	**14**	**-0.1901**	**0.3148**
-0.20	976	-0.0033	0.0845	**8**	**-0.1386**	**0.2666**	14	-0.1963	0.3154
-0.10	**973**	**-0.0039**	**0.0830**	8	-0.1441	0.2681	14	-0.2022	0.3178
0.00	973	-0.0042	0.0840	8	-0.1495	0.2717	8	-0.1944	0.3210
0.20	973	-0.0042	0.0855	8	-0.1589	0.2829	8	-0.2020	0.3290

Table 7: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and different types of measurement error; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

lower ones. Again, decreasing the weight of the lowest frequencies is likely to worsen the estimation. Both these factors lead to an optimal d that will tend to be lower than d_2; in some circumstances they can outweigh the heteroskedasticity in the periodogram, and the optimal d will be negative, as can be seen in Figures 2 and 3 and Tables 6 and 7.

Variation in sample size

Failure of asymptotic theory to provide a good approximation in finite samples is further explored by changing the sample size. Figures 4 and 5 and Tables 8 and 9 present similar results to Figures 2 and 3 and Tables 6 and 7, for $n = 512$, 2048, 8192. We set $\beta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and use $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. In both the linear and nonlinear settings, the optimal value for d increases with n, but not dramatically. Even for $n = 8192$, the optimal d is not only below d_3, but also outside the parameter range in (4.7). For all

n	512			2048			8192		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
0.00	**15**	**-0.0950**	**0.1555**	25	-0.0470	0.0897	61	-0.0271	0.0532
0.05	13	-0.0905	0.1560	**25**	**-0.0502**	**0.0896**	**61**	**-0.0296**	**0.0525**
0.20	12	-0.0977	0.1601	23	-0.0555	0.0933	41	-0.0265	0.0538

Table 8: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and n; linear setting. The minimum RMSE choice of d is indicated in bold-face.

n	512			2048			8192		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	**15**	**-0.3060**	**0.4520**	14	-0.1476	0.2674	16	-0.0583	0.1419
-0.20	15	-0.3145	0.4528	**8**	**-0.1386**	**0.2666**	13	-0.0586	0.1396
-0.05	15	-0.3261	0.4572	8	-0.1468	0.2697	**11**	**-0.0594**	**0.1389**
0.00	15	-0.3296	0.4591	8	-0.1495	0.2717	11	-0.0608	0.1392
0.20	8	-0.3096	0.4669	8	-0.1589	0.2829	8	-0.0555	0.1424

Table 9: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and n; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

values of d, there is a strong improvement in both bias and RMSE as n increases. While in the linear case the optimal bandwidth for each d increases with n, in the nonlinear setting it is often higher for $n = 512$ than for $n = 8192$. Bandwidths for $n = 2048$ are the lowest of the three sample sizes, suggesting a "U-shaped" bandwidth profile that will continue diverging to infinity as the theory requires.

Figures 6 and 7 illustrate the distributional properties of NBLS by plotting kernel density estimates for varying n, under the linear and nonlinear setting. Density estimates are computed for a sequence of $s = 50,000$ NBLS estimates b_i, $i = 1, \ldots, s$, using

$$\widehat{f}(b) = \frac{1}{sh} \sum_{i=1}^{s} \phi\left(\frac{b_i - b}{h}\right), \tag{4.9}$$

where $\phi(\cdot)$ is the standard Gaussian density function and the bandwidth h is chosen using (3.31) of Silverman (1986),

$$h = 0.9 s^{-1/5} \min(SD, IQR/1.34), \tag{4.10}$$

where SD and IQR are the sample standard deviation and interquartile range of the

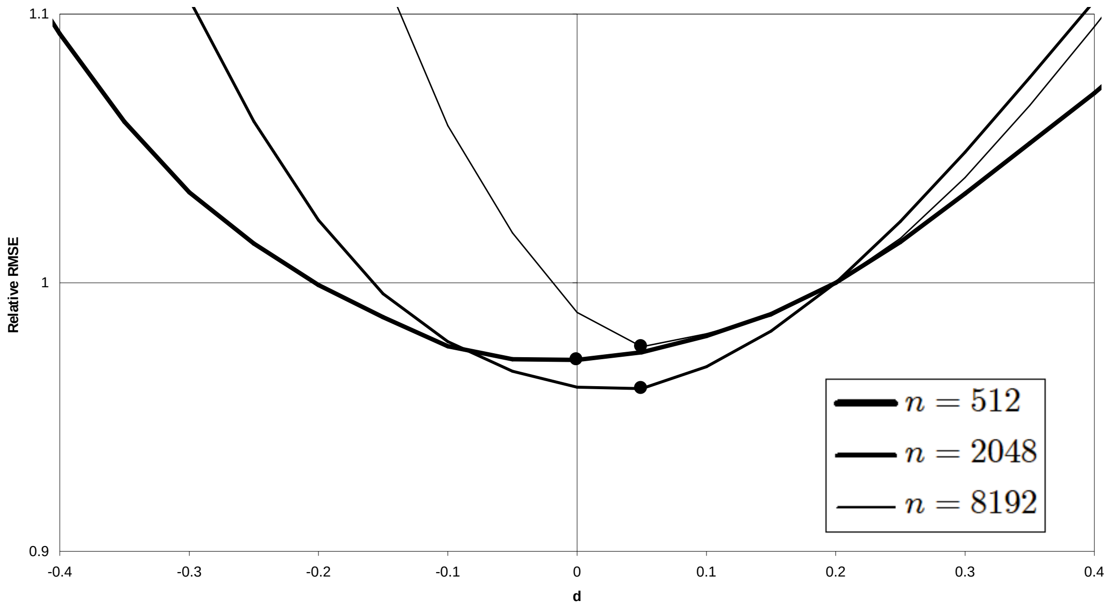

Figure 4: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and n; linear setting.

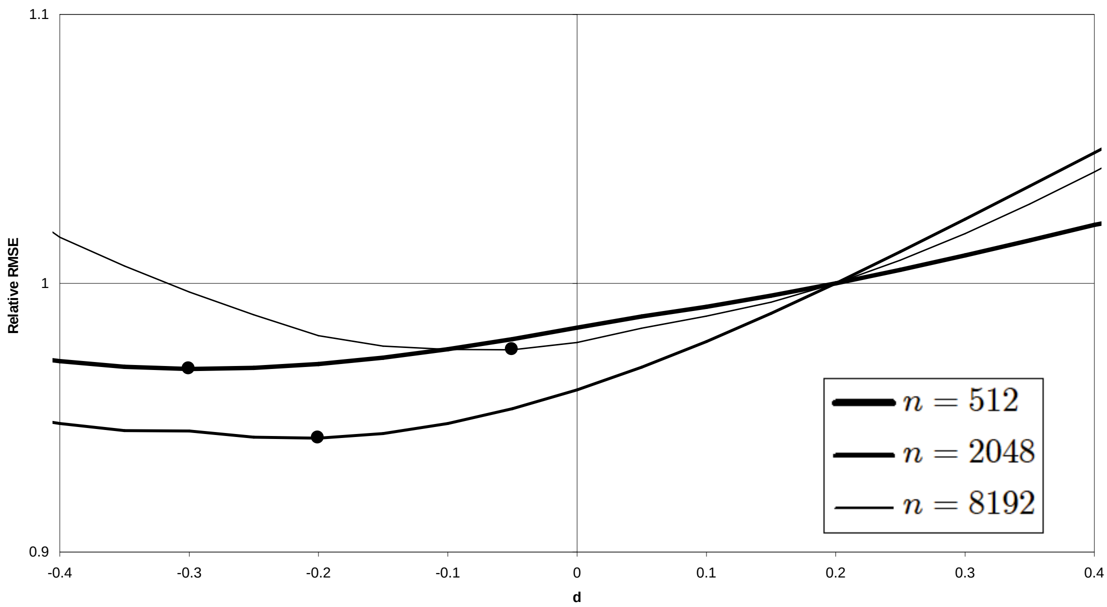

Figure 5: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and n; nonlinear setting.

b_i. Estimates for other values of d yield very similar shapes and are thus omitted, but available from the authors upon request. However, unlike for other values of d, NBLS is not covered by (4.7). Still, in the linear case all curves in Figure 6 seem to be fairly close in shape to that of a normal density. On the contrary, densities in Figure 7 are all highly skewed to the left, even for $n = 8192$, suggesting that the asymptotic distribution under the nonlinear setting might not be normal. In both settings, bias and SD seem to be decaying at the same rate, which is natural given our minimum RMSE bandwidth choice.

Variation in the signal-to-noise ratio

Figures 8 through 11 and Tables 10 through 13 can be interpreted in the same way as Figures 2 and 3 and Tables 6 and 7, for the linear and nonlinear settings, where we first change the variance of the ME, then the variance of the signal. In both experiments we start with $\beta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. The variance of the ME in the first experiment is then set to $\sigma_\delta^2 = 1/2$, 2, 8, by varying σ_2^2 in the linear setting, and by using $h_k(x) = k\exp(x)$, with $k = 1/2$, 1, 2, as the volatility function for δ_t, while keeping σ_2^2 constant, in the nonlinear setting. The resulting sequences δ_t are consequently the same, up to a multiplicative factor, for each value of σ_δ^2. In the second experiment, the variance of the signal is changed by choosing σ_1^2 so that $\sigma_\zeta^2 = 2$, 4, 8.

These parameters affect the accuracy of the estimates by influencing the relative variance of z_{1t} and u_t in (4.2), which can be interpreted as a signal-to-noise ratio, and the covariance between z_{1t} and u_t, which can be seen in (4.5) and (4.6) to depend crucially on δ_t; this was derived in (2.9), for a different setting.

Figures 8 and 9 and Tables 10 and 11 show that both m^* and the optimal d decrease rather heavily as σ_δ^2 increases, especially in the nonlinear setting. For large values of σ_δ^2, the common component in z_{1t} and u_t becomes very important, influencing even frequencies relatively close to zero. As a result, both the bandwidth and the weights should adjust so that only the lowest frequencies (where the spectral pole still dominates) have significant influence. Tables 10 and 11 display a strong degradation in both bias and RMSE, caused by the increased coherence between regressor and residuals.

While increasing σ_δ^2 influences both u_t and z_{1t}, scaling up the common component in both, increasing the cointegrating parameter β boosts the weight of the common

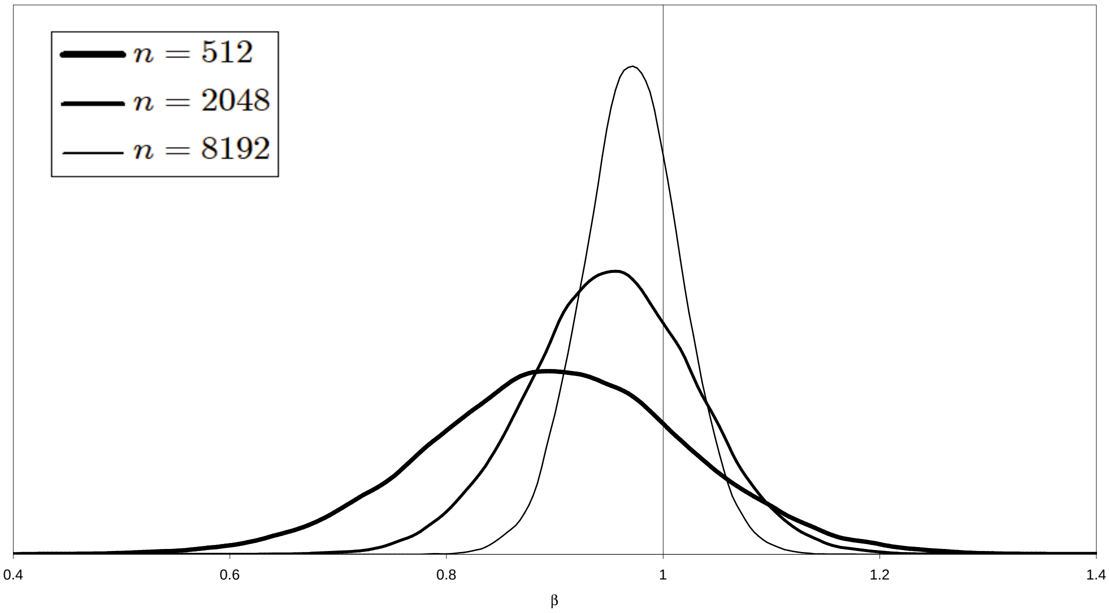

Figure 6: Kernel density estimates of NBLS for varying n; linear setting.

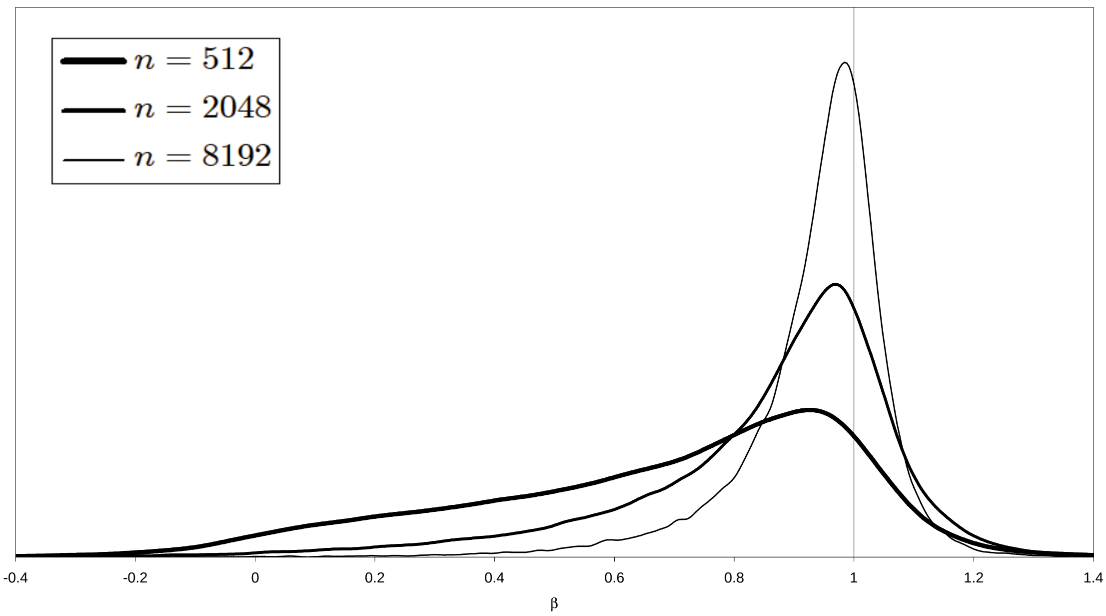

Figure 7: Kernel density estimates of NBLS for varying n; nonlinear setting.

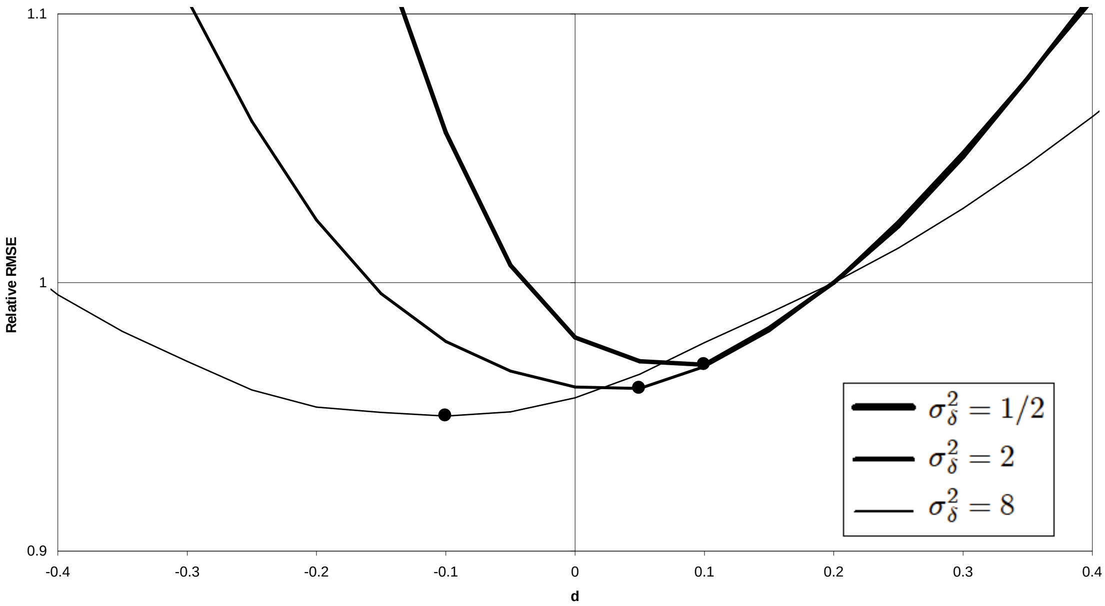

Figure 8: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and σ_δ^2; linear setting.

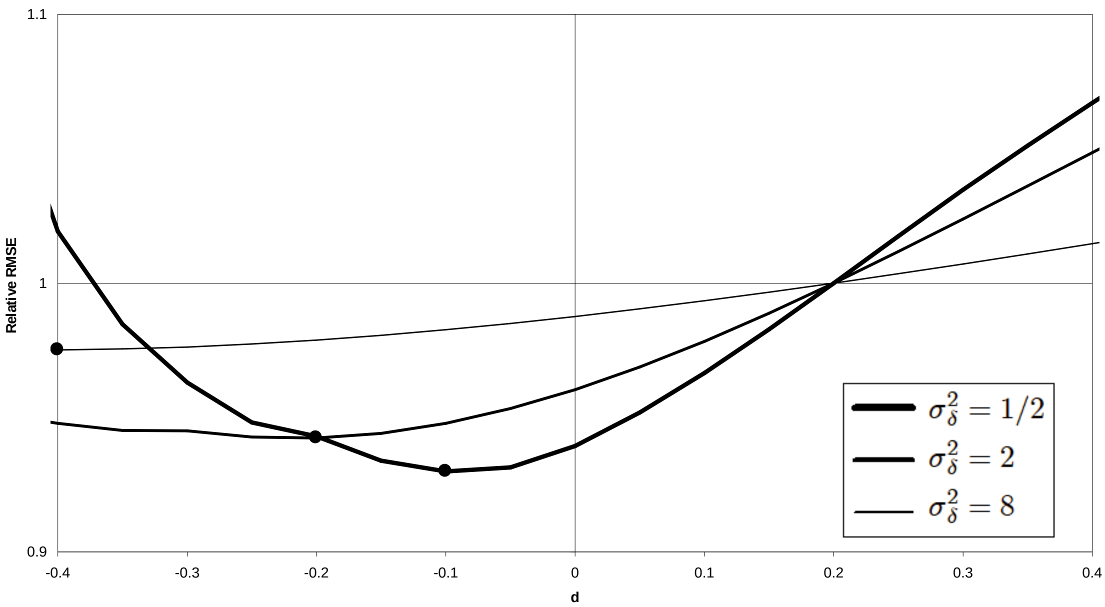

Figure 9: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and σ_δ^2; nonlinear setting.

σ_δ^2		1/2			2			8	
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.10	142	-0.0297	0.0629	39	-0.0524	0.0913	**10**	**-0.0897**	**0.1516**
0.00	112	-0.0329	0.0584	25	-0.0470	0.0897	10	-0.0969	0.1527
0.05	82	-0.0301	0.0579	**25**	**-0.0502**	**0.0896**	10	-0.1005	0.1541
0.10	**81**	**-0.0326**	**0.0578**	25	-0.0533	0.0904	10	-0.1041	0.1560
0.20	55	-0.0280	0.0596	23	-0.0555	0.0933	7	-0.0893	0.1595

Table 10: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and σ_δ^2; linear setting. The minimum RMSE choice of d is indicated in bold-face.

σ_δ^2		1/2			2			8	
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.40	1022	-0.0477	0.1333	18	-0.1454	0.2682	**4**	**-0.4607**	**0.5858**
-0.20	98	-0.0444	0.1233	**8**	**-0.1386**	**0.2666**	4	-0.4703	0.5880
-0.10	**68**	**-0.0447**	**0.1216**	8	-0.1441	0.2681	4	-0.4751	0.5903
0.00	66	-0.0491	0.1228	8	-0.1495	0.2717	4	-0.4798	0.5933
0.20	63	-0.0555	0.1308	8	-0.1589	0.2829	4	-0.4889	0.6007

Table 11: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and σ_δ^2; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

component in u_t alone, keeping z_{1t} constant. Still, this provokes a comparable increase in correlation, causing very similar effects to those reported for σ_δ^2. Monte Carlo results for this case are omitted but available upon request.

Figures 10 and 11 and Tables 12 and 13 display the effect of the strength of the signal ζ_t. In the linear case, this scales up the signal in z_{1t} without affecting u_t. In the nonlinear case, both are affected, but since the SV model used generates heavily leptokurtic processes (implying that the variance of ζ_t^2 is the major contribution to the variance of z_{1t}) and ζ_t only affects u_t through a white noise component (thus having a bounded contribution to the spectrum around the zero frequency), the impact on u_t will be minimal compared to that on z_{1t}. In both models, increasing σ_ζ^2 will have the double effect of increasing the variance of z_{1t}, thereby making the observables more correlated at all frequencies, and scaling up the spectral pole caused by the memory in η_{1t}, improving the local signal-to-noise ratio. While both effects will have a clearly positive influence on the accuracy of the estimates, as seen in Tables 12 and 13, the effect on m^* and on the optimal d is not clear, as even frequencies distant from zero become less contaminated by the dependence between z_{1t} and u_t. As a result, Figures 10 and 11 show very little

σ^2_ζ	2			4			8		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.05	**25**	**-0.0843**	**0.1380**	33	-0.0519	0.0902	42	-0.0310	0.0585
0.00	23	-0.0852	0.1382	25	-0.0470	0.0897	**40**	**-0.0327**	**0.0579**
0.05	22	-0.0883	0.1392	**25**	**-0.0502**	**0.0896**	39	-0.0346	0.0580
0.20	19	-0.0946	0.1455	23	-0.0555	0.0933	25	-0.0305	0.0603

Table 12: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and σ^2_ζ; linear setting. The minimum RMSE choice of d is indicated in bold-face.

σ^2_ζ	2			4			8		
d	m^*	Bias	RMSE	m^*	Bias	RMSE	m^*	Bias	RMSE
-0.30	**8**	**-0.3350**	**0.4722**	14	-0.1476	0.2674	18	-0.0702	0.1696
-0.20	8	-0.3453	0.4733	**8**	**-0.1386**	**0.2666**	8	-0.0628	0.1695
-0.15	8	-0.3504	0.4750	8	-0.1414	0.2671	**8**	**-0.0642**	**0.1691**
0.00	8	-0.3649	0.4833	8	-0.1495	0.2717	8	-0.0680	0.1705
0.20	7	-0.3714	0.4980	8	-0.1589	0.2829	8	-0.0726	0.1764

Table 13: Monte Carlo bias and RMSE of $\tilde{\beta}(d)$, for varying d and σ^2_ζ; nonlinear setting. The minimum RMSE choice of d is indicated in bold-face.

variation on relative RMSE with σ^2_ζ.

Distributional properties of residual memory estimates

While the previous experiments show that estimates of residual memory are not necessarily useful for choosing d in (3.13), they might still be relevant for other purposes, namely to verify if a cointegrating relationship exists at all. The use of the LP and LW estimates is well established by now, and their finite-sample properties have been examined in various settings (see e.g. Robinson and Henry, 1999; Nielsen and Frederiksen, 2005). In finite samples, LW is generally found to have bias of similar magnitude but lower variance than LP, to conform with (3.21) and (3.23). However, the recent MLW estimate has not yet been directly compared to LW. The findings of Hurvich, Moulines, and Soulier (2005), Hurvich and Ray (2003), and Table 5, indicate that, even for moderate sample sizes, MLW can successfully reduce bias in the presence of a "signal-plus-noise" structure, but at the cost of a substantially higher SE than LW. We now present a short comparison of finite-sample distributional properties of LW and MLW in the context of residual memory estimation, for $n = 512, 2048, 8192$. Residuals are

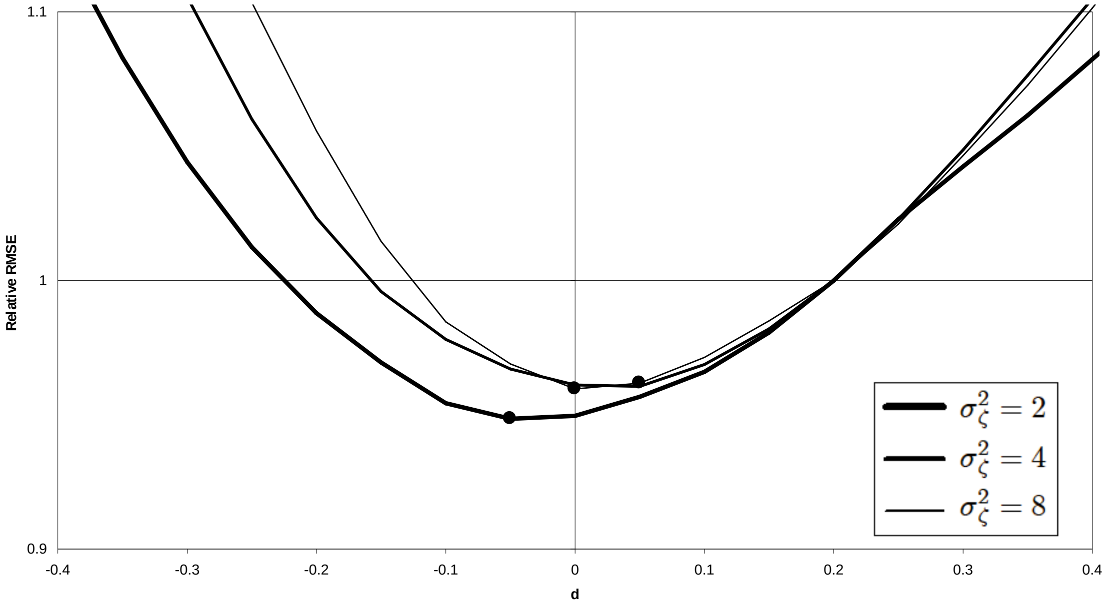

Figure 10: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and σ_ζ^2; linear setting.

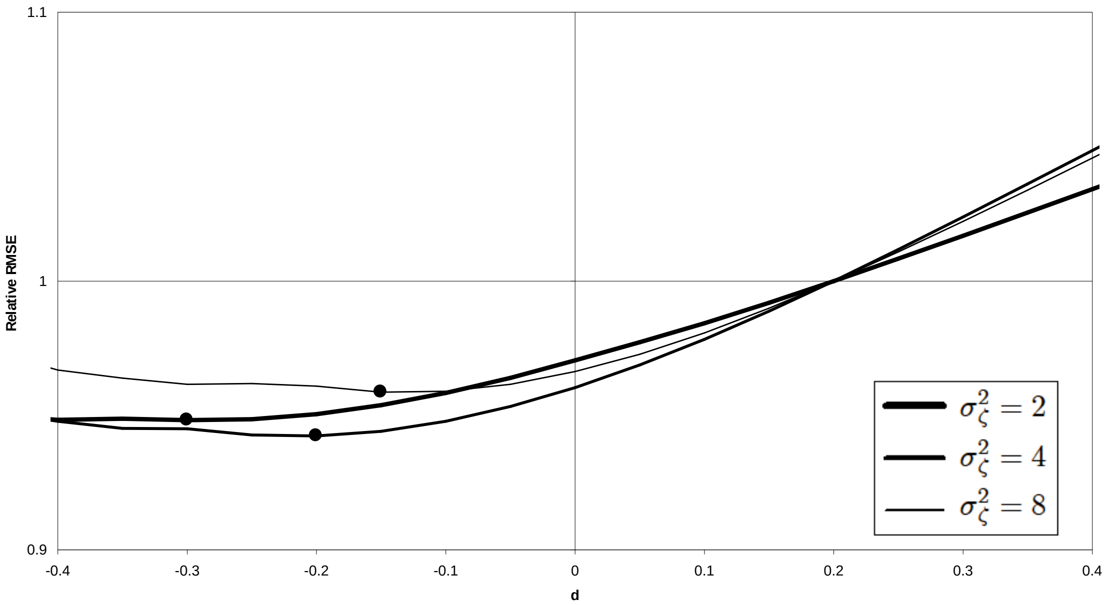

Figure 11: Relative RMSE of $\tilde{\beta}(d)$ versus $\tilde{\beta}(d_3)$, for varying d and σ_ζ^2; nonlinear setting.

	Linear setting			Nonlinear setting		
n	512	2048	8192	512	2048	8192
LW	140	270	360	240	320	300
MLW	240	940	3720	240	680	4010

Table 14: Approximate minimum RMSE bandwidths of $\hat{d}_{LW}$ and $\hat{d}_{MLW}$, for varying n; linear and nonlinear settings.

obtained from $s = 1,000$ replications of NBLS regression in the linear and nonlinear settings, with $\beta = 1$, $d_1 = 0.4$, $d_2 = 0$, $d_3 = 0.2$, $\sigma_\zeta^2 = 4$, $\sigma_\varepsilon^2 = \sigma_\delta^2 = 2$, and $h(x) = \exp(x)$ as the volatility function for the nonlinear setting. The minimum RMSE bandwidths reported in Tables 8 and 9 are used in this step. Then, LW and MLW estimates are constructed from the residuals for a grid of bandwidths (from 10 to $n/2$, with increments of 10), allowing us to approximately locate the minimum RMSE bandwidth for each memory estimate. Figures 12 through 15 show kernel density estimates (see (4.9), (4.10)) of LW and MLW, under the linear and nonlinear settings, using the approximately optimal bandwidths given in Table 14.

Table 14 shows that while LW works best with a narrow-band approach, MLW has optimal bandwidth rather close to $n/2$. This is possible because, unlike LW, MLW corrects for the presence of iid noise, and thus its spectral approximation is relatively accurate throughout all frequencies considered. However, for higher frequencies to be informative, the absence of short memory dynamics is crucial; the inclusion of, say, ARMA dynamics in any of the $\{\eta_{it}\}$ would undoubtedly require MLW bandwidths to be much lower.

All curves in Figures 12 and 13 suggest that the finite-sample density of LW is fairly close in shape to that of a normal density, but heavily biased downwards. While in the linear setting both bias and SD are substantially reduced when n increases, estimation in the nonlinear one seems surprisingly insensitive to sample size; even for $n = 8192$ the mean is much closer to 0 than to 0.2. Figures 14 and 15 highlight a potential problem of MLW in finite samples. In several cases, the distribution of MLW is bimodal, with peaks close to 0 and 1/2, the boundaries of the parameter space. In the nonlinear setting, this behaviour is apparent even for $n = 8192$, with a small mode close to the true parameter value being barely distinguishable. Performance in the linear setting is more encouraging: for $n = 8192$, the "boundary" modes disappear and are replaced by an essentially unbiased unimodal density. Still, it is worth noting that the SD in this case

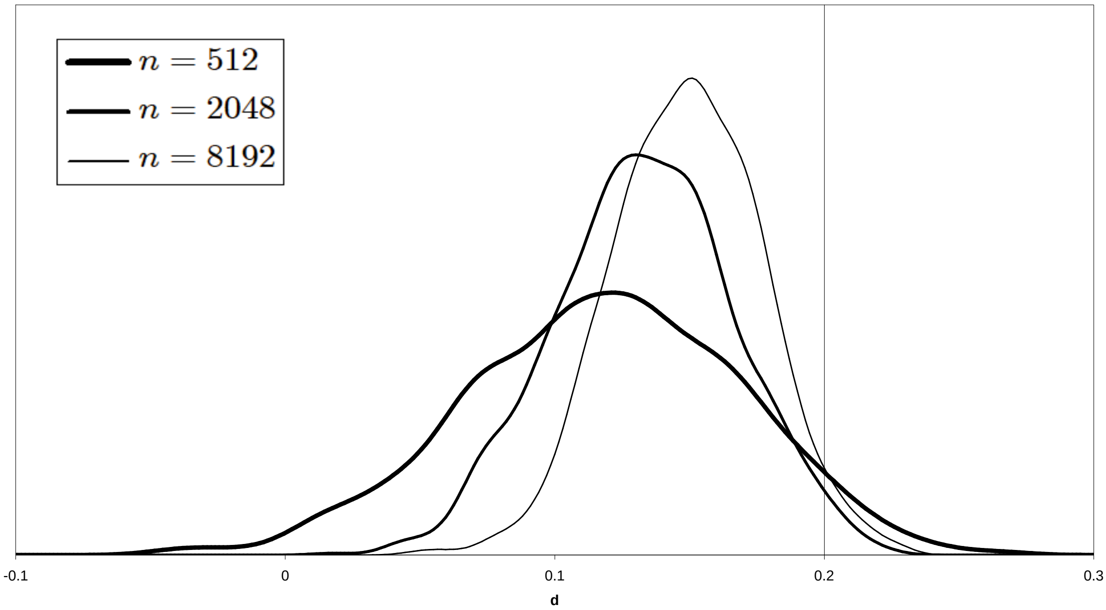

Figure 12: Kernel density estimates of LW for varying n; linear setting.

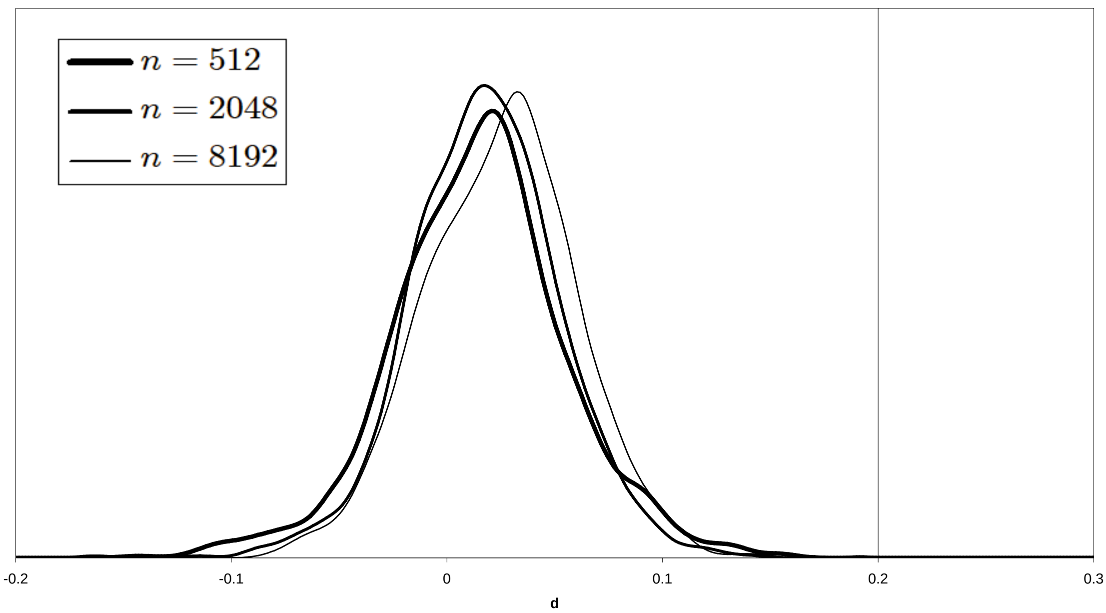

Figure 13: Kernel density estimates of LW for varying n; nonlinear setting.

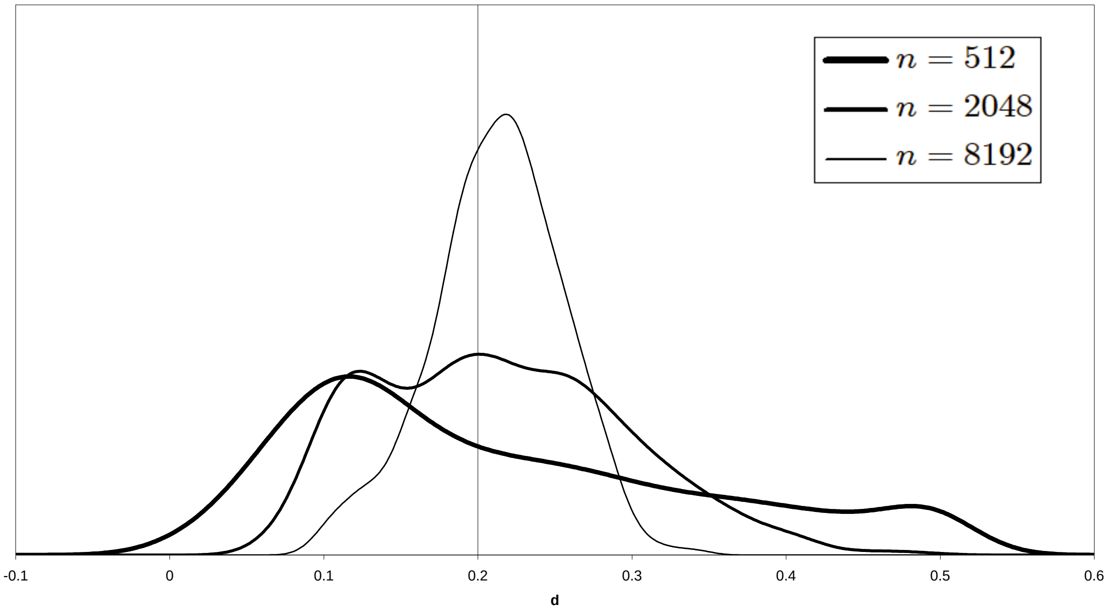

Figure 14: Kernel density estimates of MLW for varying n; linear setting.

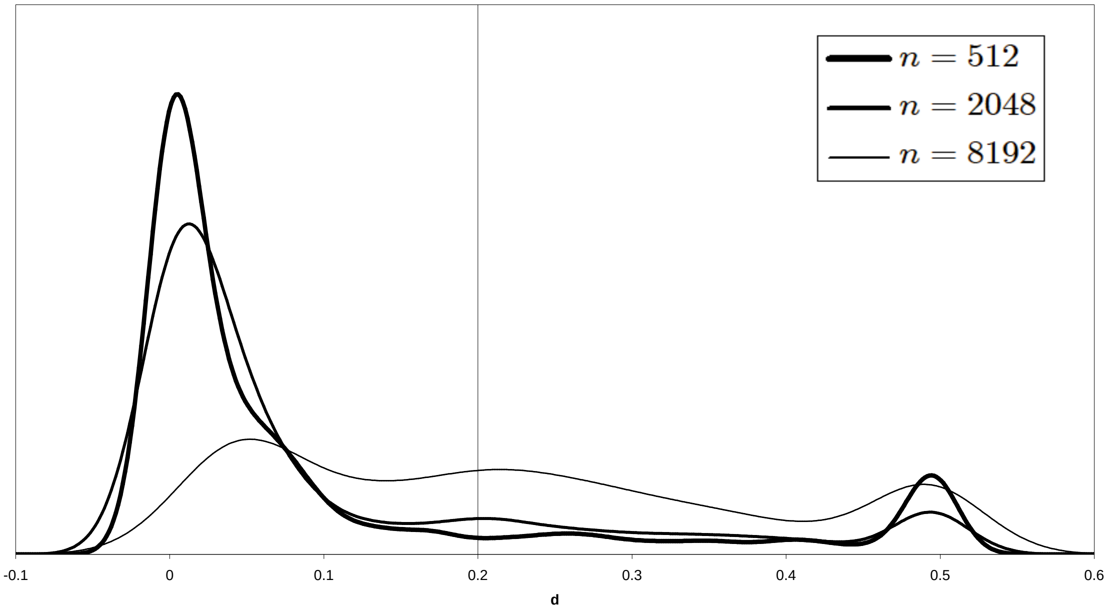

Figure 15: Kernel density estimates of MLW for varying n; nonlinear setting.

is roughly twice that of LW, and that the tails of the density are still moderately asymmetric. The findings of bimodality and higher SD in MLW are maintained in alternative (unreported) experiment designs, suggesting that they are linked to the additive noise structure itself, not to first step estimation error or nonlinearity. Estimation error in the first step regression actually contaminates the true errors (4.5), (4.6) with a higher memory component (in this case, of memory $d_1 = 0.4$), which should induce a positive contribution to both bias (thereby reducing the LW bias) and SD.

Concluding remarks

The results presented indicate that asymptotic theory should not necessarily be expected to provide a good approximation to finite-sample performance.

We first showed that, even in a standard setting, where error and regressor are independent Gaussian processes, Monte Carlo SD deviates substantially from its asymptotic counterpart. While in this setting $d = d_3$ is the optimal choice for WNBLS, further results demonstrate that the introduction of nonlinearity or ME makes this choice suboptimal, and indeed dominated by simple NBLS. Furthermore, the nonlinear setting always yields a negative optimal d, even in the absence of ME. Although optimal bandwidths somewhat vary, they appear to be lower than those implied by commonly used feasible rules. For instance, Nielsen (2005) uses $m = [n^{0.4}]$ and $m = [n^{0.5}]$, yielding $m = 21,\ 45$ for $n = 2048$, which would be clearly too high for most of the nonlinear settings considered. While in the linear setting the RMSE profiles seem to be relatively sensitive to the choice of d, in the nonlinear one a wide range of values for d perform comparably; this is possibly a consequence of the lower bandwidths used. The optimal choice of d seems to be sensitive to most parameters in the model, so a feasible rule would undoubtedly require preliminary estimation of these.

All the finite sample results were generated under assumptions which might not be realistic in practice, such as Gaussianity and independence of the underlying, unobservable processes, and the absence of short memory dynamics. These assumptions constitute a best-case scenario, and relaxing them might well widen the gap between theoretical predictions and finite sample performance. More elaborate methods, such as those of Hualde and Robinson (2004), exhibit more desirable asymptotic properties under conditions that are in some sense weak, though it is not clear to what extent they can be justified when the linearity assumptions underlying them are relaxed. Heavy dependence on preliminary estimates may also hamper their finite sample performance.

A brief comparison of residual memory estimates was also presented. It seems that, while MLW is found to dominate LW in RMSE for large enough n, due to the large negative bias of the latter, it displays high dispersion and bimodality, which can be especially misleading in cointegration analysis, where the focus is often on the difference between memory estimates obtained from observables and residuals. On the contrary, LW, being biased downwards in both cases, might yield more accurate inference on the existence and degree of fractional cointegration. Evaluation of these issues is left for future research.

Acknowledgements

Both authors' research was supported by ESRC grant R000239936. Gonçalves da Silva's research was also supported by FCT grant SFRH/BD/4783/2001. We are grateful for the helpful comments of two referees.

References

Bandi, F. M., and B. Perron (2004): "Long memory and the relation between implied and realized volatility," Economics Working Paper Archive at WUSTL, Econometrics.

Chambers, M. J. (1998): "Long memory and aggregation in macroeconomic time series," *International Economic Review*, 39(4), 1053–1072.

Christensen, B. J., and M. Ø. Nielsen (2004): "Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting," forthcoming in *Journal of Econometrics*.

Deo, R. S., and C. M. Hurvich (2001): "On the log-periodogram regression estimator of the memory parameter in long memory stochastic volatility models," *Econometric Theory*, 17(4), 686–710.

Geweke, J., and S. Porter-Hudak (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221–238.

Gonçalves da Silva, A., and P. M. Robinson (2005): "Fractional cointegration in stochastic volatility models," Preprint, London School of Economics.

Hannan, E. J. (1963): "Regression for time series," in *Time series analysis*, ed. by M. Rosenblatt, pp. 17–77. Wiley, New York.

Hidalgo, J., and P. M. Robinson (2002): "Adapting to unknown disturbance autocorrelation in regression with long memory," *Econometrica*, 20(4), 1545–1581.

Hualde, J., and P. M. Robinson (2004): "Semiparametric estimation of fractional cointegration," Preprint, London School of Economics.

Hurvich, C. M., R. Deo, and J. Brodsky (1998): "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series," *Journal of Time Series Analysis*, 19(1), 19–46.

Hurvich, C. M., E. Moulines, and P. Soulier (2005): "Estimating long memory in volatility," *Econometrica*, 73(4), 1283–1328.

Hurvich, C. M., and B. K. Ray (2003): "The Local Whittle estimator of long-memory stochastic volatility," *Journal of Financial Econometrics*, 1(3), 445–470.

Künsch, H. R. (1987): "Statistical aspects of self-similar processes," in *Proceedings of the first world congress of the Bernoulli Society*, ed. by Y. Prohorov, and V. Sazonov, vol. 1, pp. 67–75, Utrecht. VNU Science Press.

Marinucci, D., and P. M. Robinson (2001): "Semiparametric fractional cointegration analysis," *Journal of Econometrics*, 105(1), 225–247.

Nielsen, M. Ø. (2005): "Semiparametric estimation in time-series regression with long-range dependence," *Journal of Time Series Analysis*, 26(2), 279–304.

Nielsen, M. Ø., and P. H. Frederiksen (2005): "Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration," forthcoming in *Econometric Reviews*.

Robinson, P. M. (1994): "Semiparametric analysis of long-memory time series," *Annals of Statistics*, 22(1), 515–539.

——— (1995a): "Log-periodogram regression of time series with long range dependence," *Annals of Statistics*, 23(3), 1048–1072.

——— (1995b): "Gaussian semiparametric estimation of long range dependence," *Annals of Statistics*, 23(5), 1630–1661.

——— (2001): "The memory of stochastic volatility models," *Journal of Econometrics*, 101(2), 195–218.

Robinson, P. M., and M. Henry (1999): "Long and short memory conditional heteroscedasticity in estimating the memory parameter of levels," *Econometric Theory*, 15(3), 299–336.

Robinson, P. M., and D. Marinucci (2003): "Semiparametric frequency domain analysis of fractional cointegration," in *Time series with long memory*, ed. by P. M. Robinson, Advanced Texts in Econometrics, chap. 14, pp. 334–373. Oxford University Press, Oxford.

Robinson, P. M., and Y. Yajima (2002): "Determination of cointegrating rank in fractional systems," *Journal of Econometrics*, 106(2), 217–241.

Silverman, B. W. (1986): *Density estimation for statistics and data analysis*, vol. 26 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

Smith, R. L., and F.-L. Chen (1996): "Regression in long-memory time series," in *Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis in Memory of E.J. Hannan*, ed. by P. M. Robinson, and M. Rosenblatt, vol. 115 of *Lecture Notes in Statistics*, pp. 378–391, New York. Springer.