REFINED INFERENCE ON LONG MEMORY IN REALIZED VOLATILITY

By

Offer Lieberman and Peter C. B. Phillips

January 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1549

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Refined Inference on Long Memory in Realized Volatility

By

Offer Lieberman[1]

Technion—Israel Institute of Technology

and
Peter C.B. Phillips
Cowles Foundation for Research in Economics
Yale University

This version: December 12, 2005

[1]Address for correspondence: Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel. E-mail: offerl@ie.technion.ac.il

Abstract

There is an emerging consensus in empirical finance that realized volatility series typically display long range dependence with a memory parameter (d) around 0.4 (Andersen *et. al.* (2001), Martens *et. al.* (2004)). The present paper provides some analytical explanations for this evidence and shows how recent results in Lieberman and Phillips (2004a, 2004b) can be used to refine statistical inference about d with little computational effort. In contrast to standard asymptotic normal theory now used in the literature which has an $O\left(n^{-1/2}\right)$ error rate on error rejection probabilities, the asymptotic approximation used here has an error rate of $o\left(n^{-1/2}\right)$. The new formula is independent of unknown parameters, is simple to calculate and highly user-friendly. The method is applied to test whether the reported long memory parameter estimates of Andersen *et. al.* (2001) and Martens *et. al.* (2004) differ significantly from the lower boundary ($d = 0.5$) of nonstationary long memory.

Keywords and Phrases: ARFIMA; Edgeworth expansion; Fourier integral expansion; Fractional differencing; Improved inference; Long memory; Pivotal statistic; Realized volatility; Singularity.

JEL Classification: C 13, C22

1 Introduction

It is well known that by sampling intraday returns sufficiently frequently, the integrated volatility of the process can be very well approximated by the realized volatility (henceforth, RV), the latter defined as the sum of squared returns over a specified period (usually a day). See, for instance, Merton (1980), Andersen *et. al.* (2001) and Barndorff-Nielsen and Shephard (2002). This result holds, as in Andersen *et. al.* (2001), under the simple assumption that the returns propagate as semimartingales, processes which take the form of the sum of a local martingale and a predictable component with finite variation. Barndorff-Nielsen and Shephard (2002) have, in fact, demonstrated that for stochastic volatility models the integrated volatility can be recovered exactly from the entire path of the process, at least in the absence of microstructure noise. The implication of this result is that the actual volatility can be estimated quite accurately by simple summation of the squared intraday returns, such as those sampled at 5- or 30- minute frequencies.

Accurate measurement and forecasting of volatility are of great importance in the financial analysis and practice, be it for asset pricing, risk management or asset allocation. For this reason, the accuracy, nonparametric generality and practical convenience of the RV estimator has caused an understandable excitement in the literature lately, with applications to stock indices, exchange rates, futures and more. See, for instance, Martens *et. al.* (2004) and the references therein.

One of the core issues in the literature is the optimal sampling frequency for RV measurement. There is a trade off between accuracy, in terms of variance reduction, and microstructure bias. As the sampling frequency increases, microstructure noise becomes progressively more dominant in the RV estimate, to the extent that volatility estimates based on sampling every few seconds can overestimate the true volatility by a factor of two or more (Zhang *et. al.* (2005b)). Thus, while it may seem somewhat counterintuitive to use less frequently observed data in RV calculations, there is some consensus in the literature that use of a 5- to 30- minute sampling interval is optimal and effectively reduces bias (from microstructure noise) while limiting variance increases (Andersen and Bollerslev (1998), Ait Sahalia *et. al.* (2005)). Zhang *et. al.* (2005a) discussed five alternative estimators, suggesting subsampling, averaging and bias correction over two time scales as the ideal estimator.

Recent empirical studies indicate that one of the stylized facts of realized volatility series is that they display evidence of long memory with a fractional difference parameter d of around 0.4. See, among others, Andersen *et. al.* (2001) and Martens *et. al.* (2004). The former used the log periodogram (LP) regression estimator of Geweke and Porter Hudak (1983) and log-variance plots to estimate d. The long memory feature of realized volatility is perhaps not so surprising, given that RV is an increasing process constituted from squared returns and the latter are well known to manifest long range dependence. For modeling long memory, the ARFIMA (p,d,q)

model has been extensively employed. See, among others, Li (2002), Andersen *et. al.* (2003), Pong *et. al.* (2004), Martens and Zein (2004). Bandi and Perron (2001) considered spectral methods in estimation, whereas for forecasting, Deo *et. al.* (2004) suggested a long memory stochastic volatility model. In addition to the methods in the aforementioned references, there is a large array of techniques to choose from, almost every imaginable type of approach now being used: graphical, heuristic, nonparametric semiparametric and parametric. An early overview of some of these methods is given in Beran (1994), but since then many new techniques have been developed, some of which are designed to allow for data with nonstationary as well as stationary long memory (Phillips, 1999; Shimotsu and Phillips, 2005; Abadir, Distaso and Giraitis, 2005).

Of course, the literature on long memory processes has a long history and substantially predates that of RV, with applications in many fields such as hydrology, where studies began with Hurst (1951), economics, finance, physics, internet traffic and more.

At the crux of most estimation methods lies the fact that the spectral density, $f(\lambda)$, of a stationary long memory process with a parameter $d \in (0, 1/2)$ asymptotes at the origin, behaving like

$$C\lambda^{-2d} \text{ as } \lambda \to 0+, \tag{1}$$

where C is a finite and positive constant and λ is frequency, so that low frequency behavior is a dominant characteristic of the series.. For this reason, the slope of a graph of the log-periodogram, viewed as an estimate of $\log f(\lambda)$, against $\log \lambda$, for small enough λ-values, provides a preliminary indication of the value of d. Of course, this feature of $f(\lambda)$ near the origin motivated the LP regression estimator as a semiparametric procedure, its asymptotic properties being worked out by Robinson (1995a). Other semiparametric procedures include the local Whittle estimator (Künsch, 1987, and Robinson, 1995b) and the exact local Whittle estimator (Shimotsu and Phillips, 2005), which is consistent for all values of d.

Classical methods for the estimation of d include the maximum likelihood estimator (MLE) and the full band Whittle estimator. Under certain conditions, the $\sqrt{n}$-normalized and mean subtracted MLE and Whittle estimators of d are asymptotically $N(0, 6/\pi^2)$. The conditions for narrow band LP regression and Whittle estimation are weaker because the behavior of the spectrum $f(\lambda)$ in only an immediate neighborhood of the origin is used in developing the estimates and their asymptotic properties. To clarify, suppose that the true process is a stationary and invertible ARFIMA(p,d,q) model with spectrum

$$f(\lambda) = C(\lambda) \left|1 - e^{i\lambda}\right|^{-2d},$$

where

$$C(\lambda) = \frac{\sigma^2 \left|a\left(e^{i\lambda}\right)\right|^2}{2\pi \left|b\left(e^{i\lambda}\right)\right|^2},$$

and $a(\cdot)$ and $b(\cdot)$ are polynomials with all roots lying outside the unit circle. Now,

$$C(0) = \frac{\sigma^2 |a(1)|^2}{2\pi |b(1)|^2} = C,$$

where C is as in (1) and does not depend on λ. However, for λ sufficiently far from zero, $C(\lambda)$ may fluctuate substantially and have local peaks in the short memory spectrum away from the origin. In a certain sense, therefore, ignoring behavior of the spectrum over a wider band of larger λ values is tantamount to treating the process as an ARFIMA (0,d,0) process *locally*, because such a process *has* spectral density $f(\lambda) = C\left|1 - e^{i\lambda}\right|^{-2d}$, defined on $[-\pi, \pi]$, with C not depending on λ. The same is true for any estimator that merely uses the localizing feature (1).

The Whittle and exact maximum likelihood estimators are broadband estimators, taking account of the entire spectrum. For this reason, the conditions for the asymptotic $N(0, 6/\pi^2)$ are stronger than those assumed for LP and other semiparametric estimators based only on narrow band conditions like the local Whittle and exact local Whittle estimators. Specifically, what is required is that $f(\lambda) = C\left|1 - e^{i\lambda}\right|^{-2d}$, over the full band $[-\pi, \pi]$ and not only near the origin. In other words, for $\sqrt{n}$ convergence and the limiting $N(0, 6/\pi^2)$ distribution to apply, the process needs to follow an ARFIMA $(0, d, 0)$ model.

The asymptotic $N(0, 6/\pi^2)$ result is appealing in its simplicity and the fact that it does not involve unknown parameters. However, as with many asymptotic results, in finite samples the use of the asymptotic normal tables may result in inaccurate confidence intervals, p-values and rejection probabilities. In studying these issues, Lieberman and Phillips (2004a, 2004b) refined the limit theory by deriving second-order expansions for the distributions of the exact and Whittle MLE's which are uniform and second-order pivotal. Instead of the usual $O\left(n^{-1/2}\right)$ error rate for the asymptotic distribution, the higher order result has an error of reduced order $o\left(n^{-1/2}\right)$ and the pivotal characteristic (or independence of unknown parameters) of the second-order expansion makes it an attractive option in applications.

In Section 2, we provide analytical explanations for the apparent long memory property of RV series. In Section 3, we review the relevant results in Lieberman and Phillips (2004a, 2004b). As far as we know, this is the first attempt to apply Edgeworth expansions to do statistical inference on the long memory parameter with RV data, although Zhang *et. al.* (2005b) derived Edgeworth expansions for the RV estimator itself. Practical aspects of the expansion are discussed in Section 5. The usefulness of our results is demonstrated in Section 5. Section 6 concludes.

2 Autocovariance structure of the RV estimator

In this section we provide analytical explanations for the evidence of long memory in RV data. To fix ideas, denote by $\varepsilon = (\varepsilon_1, ..., \varepsilon_T)'$ the entire vector of intraday

returns and assume for simplicity that it is stationary $N(\mu 1, \Sigma)$, where 1 is a $T \times 1$ vector of 1's and Σ is positive definite. The assumption of Gaussianity is generally inessential for our arguments to carry through and is only made to simplify the notation. Denote by $\gamma_\varepsilon(s)$, $(s = 0, 1, 2, ...)$, the autocovariances of $\{\varepsilon_t\}_{t=1}^T$. Let $Y_s = \sum_{t=(s-1)N+1}^{sN} \varepsilon_t^2$, $(s = 1, \ldots, n)$, with $n = T/N$. In words, Y_s is the RV estimate of day s, in each day there are N records of high frequency returns and the sample is over n days. The autocovariance of order j of Y is given by

$$\begin{aligned}\gamma_Y(j) &= E(Y_1 Y_{1+j}) - E^2(Y_1) \\ &= \sum_{t=1}^{N} \sum_{l=jN+1}^{(j+1)N} E\left(\varepsilon_1^2 \varepsilon_{1+j}^2\right) - \left(E\left(\sum_{t=1}^{N} \varepsilon_t^2\right)\right)^2.\end{aligned}$$

To obtain the expectation, let e_t be a canonical vector with unity in the t-th position and zero's elsewhere. If $x \sim N(a, \Sigma)$ and if A and B are fixed, symmetric $T \times T$ matrices, then we know that (e.g., Searle, 1971)

$$\begin{aligned}E(x'Ax) &= \mathrm{tr}(A\Sigma) + a'Aa \\ Cov(x'Ax, x'Bx) &= 2\mathrm{tr}(A\Sigma B\Sigma) + 4a'A\Sigma Ba.\end{aligned}$$

Noting that $\varepsilon_s^2 = \varepsilon' e_s e_s' \varepsilon$, $(s = 1, ..., T)$, and using these results, we obtain

$$\begin{aligned}E\left(\varepsilon_1^2 \varepsilon_{1+j}^2\right) &= Cov\left(\varepsilon' e_1 e_1' \varepsilon, \varepsilon' e_{1+j} e_{1+j}' \varepsilon\right) + E^2\left(\varepsilon' e_1 e_1' \varepsilon\right) \\ &= 2\mathrm{tr}\left(e_1 e_1' \Sigma e_{1+j} e_{1+j}' \Sigma\right) + 4\mu^2 1' e_1 e_1' \Sigma e_{1+j} e_{1+j}' 1 \\ &\quad + \left(\mathrm{tr}\left(e_1 e_1' \Sigma\right) + \mu^2 1' e_1 e_1' 1\right)^2 \\ &= 2\gamma_\varepsilon^2(j) + 4\mu^2 \gamma_\varepsilon(j) + \left(\gamma_\varepsilon(0) + \mu^2\right)^2.\end{aligned}$$

Hence,

$$\begin{aligned}\gamma_Y(j) &= \sum_{t=1}^{N} \sum_{l=jN+1}^{(j+1)N} \left(2\gamma_\varepsilon^2(t-l) + 4\mu^2 \gamma_\varepsilon(t-l) + \left(\gamma_\varepsilon(0) + \mu^2\right)^2\right) - \left(N\left(\gamma_\varepsilon(0) + \mu^2\right)\right)^2 \\ &= 2N \sum_{h=-(N-1)}^{N-1} \left(1 - \frac{|h|}{N}\right)\left(\gamma_\varepsilon^2(jN+h) + 2\mu^2 \gamma_\varepsilon(jN+h)\right).\end{aligned}$$

The behavior of $\gamma_Y(j)$ for large j depends on those of $\gamma_\varepsilon(s)$ and whether or not N is fixed.

Next, we analyze the variance of $\bar{Y}_n = n^{-1}\sum_{s=1}^n Y_i$. It is well known that the variance of the sample mean of a long memory process behaves as n^{2d-1} (see, for instance, Beran 1994, equation (4.13)). Here, $Var\left(\bar{Y}_n\right) = n^{-2} var\left(\sum_{t=1}^T \varepsilon_t^2\right)$. If it simply holds that as a process, $\{\varepsilon_t^2\}_{t=1}^T$ are long memory with a memory parameter d and N is independent of T, then

$$Var\left(\bar{Y}_n\right) \leq K n^{-2} T^{2d+1} = K N^{2d+1} n^{2d-1}$$

for some $0 < K < \infty$, so that $\{Y_s\}_{s=1}^n$ is a long memory process with a memory parameter d. However, $\{Y_s\}_{s=1}^n$ can be long memory even if the $\{\varepsilon_t^2\}_{t=1}^T$ are short memory. To see this, notice that

$$\begin{aligned} Var\left(\bar{Y}_n\right) &= \frac{1}{n^2} var\left(\sum_{t=1}^{T} \varepsilon_t^2\right) \\ &= \frac{2T}{n^2} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \left(\gamma_\varepsilon^2(h) + 2\mu^2 \gamma_\varepsilon(h)\right) \\ &= \frac{2N}{n} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \left(\gamma_\varepsilon^2(h) + 2\mu^2 \gamma_\varepsilon(h)\right). \end{aligned}$$

If it holds that

$$\lim_{T\to\infty} \sum_{h=-(T-1)}^{T-1} |h| \left|\gamma_\varepsilon(h)\right| = M < \infty, \tag{2}$$

then

$$\frac{n}{2N} Var\left(\bar{Y}_n\right) \to_{T\to\infty} 2\pi \left(f_Z(0) + 2\mu^2 f_\varepsilon(0)\right) < \infty, \tag{3}$$

where $f_Z(0)$ and $f_\varepsilon(0)$ are the spectral densities of a random variable Z with autocovariances $\gamma_Z(s) = \gamma_\varepsilon^2(s)$, $(s = 0, 1, 2, ...)$, and of ε, respectively, evaluated at zero. The bound on the right hand side of (3) is a consequence of (2). All stationary and invertible ARMA(p,q) models satisfy (2). If N is fixed, then $Var\left(\bar{Y}_n\right) = O\left(n^{-1}\right)$, which characterizes a short memory process. However, if N is chosen to satisfy $N = O\left(n^{2d}\right)$, then even though the intraday increments $\{\varepsilon_t\}_{t=1}^T$ are short memory satisfying (2), this sampling scheme renders $\{Y_t\}_{t=1}^n$ to be long memory.

While in particular research one usually sets N apriori to be independent of n, in a sample which has a large N relative to T, the series $\{Y_t\}_{t=1}^n$ may exhibit long memory characteristics, even though in actual fact it is not. We emphasize that the preceding arguments are sufficient for the generation of a long memory time series. They are not necessary.

3 A simple formula for refined inference on d

Denote by $\hat{d}_n$ either the Whittle or ordinary MLE of d and by d_0 the true value of d, which is assumed to lie in $(0, 1/2)$. Let $\hat{\delta}_n = \sqrt{n}\left(\hat{d}_n - d_0\right)$. For an ARFIMA(0,d,0) model with unknown mean and variance, Lieberman and Phillips (2004b) showed that the second order distribution function of $\hat{\delta}_n$, evaluated at $x/\sqrt{\kappa_{n,1,1}}$, is

$$\tilde{H}_{\hat{\delta}_n}^{(1),A}(x/\sqrt{\kappa_{n,1,1}}) = \Phi(x) + \frac{\zeta(3)}{\sqrt{n}\zeta^{3/2}(2)} \phi(x) \left\{x^2 + 2\right\}, \tag{4}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cdf and pdf respectively, $\zeta(\cdot)$ is the Riemann-zeta function and $\kappa_{n,1,1}$ is the variance of the score function. For instance, for the Whittle score,

$$\kappa_{n,1,1} = \frac{1}{2n} tr\left(M_n \dot{T}^d_{W,n} M_n T^d_n\right)^2,$$

where $M_n = I_n - n^{-1}11'$, I_n is an identity matrix of order n, T^d_n is an $n \times n$ Toeplitz matrix with elements

$$T^d_n(j,k) = \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \left|1 - e^{-i\lambda}\right|^{-2d} e^{i(j-k)\lambda} d\lambda,$$

and $\dot{T}^d_{W,n}$ is the first-order derivative with respect to d of the matrix $T^d_{W,n}$ with elements

$$T^d_{W,n}(j,k) = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} \left|1 - e^{-i\lambda}\right|^{2d} e^{i(j-k)\lambda} d\lambda.$$

Note that for both estimators it holds that

$$\kappa_{n,1,1} = \frac{\pi^2}{6} + O\left(n^{-1+2\epsilon}\right), \forall \epsilon > 0,$$

which is consistent with the fact that the asymptotic variance of the $\sqrt{n}$–normalized estimators is $6/\pi^2$.

The main feature of (4) is that it does not depend on unknown parameters, thereby making it attractive in applications. The expansion is uniform and valid in the sense that

$$\sup_{x\in R}\sup_{d\in D^*}\left|\Pr_{d_0}\left(\hat{\delta}_n \leq x/\sqrt{\kappa_{n,1,1}}\right) - \tilde{H}^{(1),A}_{\hat{\delta}_n}(x/\sqrt{\kappa_{n,1,1}})\right| = o\left(n^{-1/2}\right), \tag{5}$$

where D^* is any closed subset of $(0, 1/2)$. The improvement over the asymptotic $N(0, 6/\pi^2)$ lies in the fact that the error of the approximation is $o\left(n^{-1/2}\right)$ rather than $O\left(n^{-1/2}\right)$.

We emphasize that the developments leading to (4) and (5) hold strictly only under the assumption that the process is Gaussian ARFIMA $(0, d, 0)$ with unknown mean and variance. However, the application of the approximation can be extended to higher order ARFIMA (p, d, q) models and to other processes in a local way as follows. Suppose that the underlying process $\{X_t\}$ has spectral density

$$f_X(\lambda) = \left|1 - e^{i\lambda}\right|^{-2d} f_u(\lambda),$$

where $f_u(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_u(h) e^{-ih\lambda}$ and $\sum_{h=-\infty}^{\infty} |h| |\gamma_u(h)| < \infty$, which includes all stationary and invertible ARFIMA (p, d, q) models. Then, since $\left|1 - e^{i\lambda}\right|^2 =$

$4\sin^2\left(\frac{\lambda}{2}\right)$, we have

$$\begin{aligned} f_X(\lambda) &= \left|1-e^{i\lambda}\right|^{-2d} f_u(\lambda) = \lambda^{-2d} f_u(\lambda)\left[\frac{\sin^2(\lambda/2)}{(\lambda/2)^2}\right]^{-d} \\ &= \lambda^{-2d}\left\{f_u(0)+O(\lambda)\right\}, \text{ as } \lambda \to 0,. \end{aligned}$$

so that (1) holds. The k'th autocovariance of X_t is therefore

$$\begin{aligned} \gamma_x(k) &= \int_{-\pi}^{\pi} e^{i\lambda k} f_u(\lambda)\left|1-e^{i\lambda}\right|^{-2d} d\lambda \\ &= 2\,\mathrm{Re}\left\{\int_0^{\pi} e^{i\lambda k} f_u(\lambda)\left|1-e^{i\lambda}\right|^{-2d} d\lambda\right\} \\ &= 2\,\mathrm{Re}\left\{\int_0^{\pi} e^{i\lambda k}\lambda^{-2d} g(\lambda)\, d\lambda\right\}, \end{aligned} \tag{6}$$

where $g(\lambda) = f_u(\lambda)\left[\frac{\sin^2(\lambda/2)}{(\lambda/2)^2}\right]^{-d}$ is continuously differentiable over $[0,\pi]$. The Fourier integral in (6) has a critical point at $\lambda = 0$ and may be expanded as $k \to \infty$ by standard methods for Fourier integral asymptotic expansions (e.g., Bleistein and Handelsman (1975) p. 91), giving

$$\begin{aligned} \int_0^{\pi} e^{i\lambda k}\lambda^{-2d} g(\lambda)\, d\lambda &= \Gamma(1-2d)\, e^{\pi i(1-2d)/2} g(0)\, k^{-(1-2d)}\left\{1+O\left(k^{-1}\right)\right\} \\ &= \Gamma(1-2d)\, e^{\pi i/2} e^{-\pi d i} f_u(0)\, k^{-(1-2d)}\left\{1+O\left(k^{-1}\right)\right\} \\ &= i\Gamma(1-2d)\, e^{-\pi d i} f_u(0)\, k^{-(1-2d)}\left\{1+O\left(k^{-1}\right)\right\}. \end{aligned}$$

Hence

$$\gamma_x(k) = \frac{2\Gamma(1-2d)\sin(\pi d)\, f_u(0)}{k^{1-2d}}\left\{1+O\left(k^{-1}\right)\right\}, \tag{7}$$

so that the autocovariances decay according to the power law $O\left(1/k^{1-2d}\right)$. At least to the first order, these autocovariances correspond to those of an ARFIMA $(0,d,0)$ model with error variance G_0. The latter have the explicit form (e.g., Brockwell and Davis, 1991, page 468)

$$\gamma_x(k) = \frac{(-1)^k\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}G_0 = \frac{(-1)^k\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)}G_0$$

and, using the reflection formula $\Gamma(1-k-d) = \pi/\left\{\Gamma(k+d)\sin(\pi(k+d))\right\}$ and asymptotic expansion $\Gamma(k+d)/\Gamma(k-d+1) = k^{-(1-2d)}\left\{1+O\left(k^{-1}\right)\right\}$, we get

$$\begin{aligned} \gamma_x(k) &= \frac{(-1)^k\Gamma(1-2d)\sin(\pi(k+d))\Gamma(k+d)}{\pi\Gamma(k-d+1)}G_0 \\ &\sim \frac{\Gamma(1-2d)\sin(\pi d)}{\pi k^{1-2d}}G_0\left\{1+O\left(\frac{1}{k}\right)\right\}, \end{aligned}$$

which corresponds to (7) when $f_u(0) = G_0/2\pi$.

This approximation provides some justification for the use of our formula outside the strict ARFIMA $(0, d, 0)$ setting, at least when the error spectrum $f_u(\lambda)$ is flat in some neighborhood of the origin.

4 Practical aspects of the approximation

Frequently, the analysis of RV data is conducted in two stages. In the first stage, the series is analyzed for its degree of integration d and differenced (or fractionally differenced) by this degree. The second stage seeks to model the differenced series. See Anderson *et. al.* (2001) and Martens *et. al.* (2004). For this reason, it is important to obtain good estimates and perform accurate inference on d at the first stage. Our higher order asymptotic formula (4) can be used for this purpose and this section explains some of the practical aspects of its implementation.

For a one-sided hypothesis, consider the results summarized in Table 1 and note that $\zeta(2) = 1.64493$ and that $\zeta(3) = 1.20206$. It is clear that the difference between the normal approximation and the Edgeworth expansion is substantial, even for $n = 1000$.

For a two-sided hypothesis of the form $H_0 : d = d_0$ vs. $H_1 : d \neq d_0$, the normal approximation as well as the Edgeworth expansion yield $P\left(\left|\hat{\delta}_n\right| \leq x/\sqrt{\kappa_{n,1,1}}\right) = 2\Phi(x) - 1$. However, our result (5) implies that the error is $o\left(n^{-1/2}\right)$, rather than $O\left(n^{-1/2}\right)$. This refinement is an outcome of the correction term of the expansion (4) being an even function in x.

With the normal approximation, the usual asymptotic 95% confidence interval is $\hat{d}_n \pm 1.96\sqrt{6}/(\pi\sqrt{n})$. This confidence interval is symmetric and equal–tailed. With the same $x = 1.96$ and $n = 1000$, say, the upper tail based on the Edgeworth expansion equals 0.019 whereas the lower tail equals 0.031, so that the Edgeworth expansion confidence interval is not an equal–tailed one. This is due to the skewness of the distribution expansion. To obtain an equal–tailed confidence interval with the Edgeworth expansion, we need to find x_1 and x_2 such that $\tilde{H}^{(1),A}_{\hat{\delta}_n}(x_1/\sqrt{\kappa_{n,1,1}}) = 1 - \tilde{H}^{(1),A}_{\hat{\delta}_n}(x_2/\sqrt{\kappa_{n,1,1}})$. With $n = 1000$, for example, $\tilde{H}^{(1),A}_{\hat{\delta}_n}(1.852/\sqrt{\kappa_{n,1,1}}) = 0.975$ and $\tilde{H}^{(1),A}_{\hat{\delta}_n}(-2.061/\sqrt{\kappa_{n,1,1}}) = 0.025$.

5 Empirical Examples

Martens *et. al.* (2004) analyzed the S&P 500 series over the period January 3, 1994 until December 29, 2000. Their RV series are based on 5-minute intraday and 30-minute intranight returns. It appears from their Figure 2 that the log RV series is approximately Gaussian. An ARFIMA model for that series yielded a fitted

memory parameter of $\hat{d}_n = 0.471$. Four other modeling schemes resulted in a fitted $\hat{d}_n$ between 0.363 and 0.495. See Table 3 of their paper.

To demonstrate how our method can be used, suppose that the hypothesis of interest is $H_0 : d = 0.5$ vs. $H_1 : d < 0.5$[2]. In this case, there are $n = 1767$ observations (see p 4 of Martens *et. al.* (2004)) so the standardized statistic $\sqrt{1767/6}\pi\,(0.471 - 0.5) = -1.5635$. In this case, $\Phi\,(-1.5635) = 0.059$ and the Edgeworth expansion (4) gives a 0.068 probability. The difference 0.009 is due to the correction term of the expansion. Hence, with both the normal and Edgeworth approximations we cannot reject the null of nonstationary long memory, with a higher p-value assigned by the Edgeworth expansion.

Similar evidence of long memory is given in Table 3 of Andersen *et. al.* (2001) who analyzed DM/$ and Yen/$ series. They used LP regression with an optimal narrow band frequency setting of $m = [n^{4/5}]$ frequencies (Hurvich *et. al.* (1998)) and reported what they termed a "typical value" of 0.4 - see p 52 of their paper. These estimates coincide with the results of Granger *et. al.* (2000) for daily absolute and square returns series and the intraday DM/$ returns considered by Andersen and Bollerslev (1997). It is clear that the estimates are statistically significant from zero. With 2449 observations, $\sqrt{2449/6}\pi\,(0.4 - 0.5) = -6.347$, so the hypothesis $H_0 : d = 0.5$ can be firmly rejected by both the normal and Edgeworth approximations. The latter, in fact, gives a 6.5×10^{-10} p-value.

6 Remarks

One of the stylized facts emerging about realized volatility is that such series display stationary long memory with memory parameter d around 0.4. It is important to draw accurate inference on d and to assist in this process the present paper proposes the use of second–order refinements by Lieberman and Phillips (2004a, 2004b) of conventional asymptotic formulae. An advantage of the procedure is that the statistic is second order pivotal, so that the distributional expansion depends only on known constants. While these refinements are strictly valid only for ARFIMA$(0, d, 0)$ processes, the paper shows that they are locally valid for the general ARFIMA(p, d, q) case since the lag k autocovariance of a general ARFIMA(p, d, q) process has leading term given by the lag k autocovariance of an ARFIMA$(0, d, 0)$ process, with an error of the order k^{-1}.

[2] Our results only hold under stationarity, so that technically speaking the null hypothesis should be $H_0 : d = 0.5 - \varepsilon$, for some small positive ε.

Table 1: Normal approximation and Edgeworth expansion[3]

	Normal		Edgeworth	
x	$\Phi(x)$	$n = 100$	$n = 1000$	$n = 10000$
1.645	0.95	0.978	0.959	0.953
1.96	0.975	0.994	0.981	0.977
2.326	0.99	1.001	0.994	0.991

[3]The Edgeworth expansion is given in (2).

REFERENCES

Abadir, K., Distaso, W., Giraitis, L. (2005). Local Whittle estimation, fully extended for nonstationarity. University of York Discussion Paper 05/16.

Ait-Sahalia, Y., Mykland, P. A., Zhang, L. (2005). How often to sample a continuous–time process in a presence of market microstructure noise. *Review of Financial Studies* 18:351–416.

Andersen, G. A., Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: uncovering the long–run in high frequency returns. *Journal of Finance* 52:975–1005.

Andersen, G. A., Bollerslev, T. (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39(4): 885–905.

Andersen, G. A., Bollerslev, T., Diebold, F. X., Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96(453):42–55.

Andersen, G. A., Bollerslev, T., Diebold, F. X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

Bandi, F. M., Perron, B. (2001). Long memory and the relation between implied and realized volatility. C.R.D.E. discussion paper.

Barndorff-Nielsen, O. E., Shephard, N. (2002). Economic analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B* 64(2):253–280.

Beran, J. (1994). *Statistics for Long-Memory Processes*. London: Chapman-Hall.

Bleistein, N., Handelsman, R. A. (1975). *Asymptotic Expansions of Integrals.* Holt, Reinehart and Winston.

Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods,* 2nd edn, Springer–Verlag, New York.

Deo, R., Hurvich, C., Lu, Y. (2004). Forecasting realized volatility using a long memory stochastic volatility model: estimation prediction and seasonal adjustment. Mimeo.

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

Granger, C., Scot, W. J., Spear, A., Ding, Z. (2000). Stylized facts on the temporal and distributional properties of absolute returns: an update. In Chan, W., Li, W. K., Tong, H. (eds.), *Statistics and Finance: an Interface.*

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineering* 116:770–808.

Hurvich, C. M., Deo, R., Brodsky, J. (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* 19:19–46.

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society (*Yu Prokhorov and V. V. Sazanov eds.) 1:67–74. VNU Science Press, Utrecht.

Li, K. (2002). Long memory versus option-implied volatility predictions. *Journal of Derivatives* 9:9–25.

Lieberman, O., Phillips, P. C. B. (2004a). Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory* 20(3):464–484.

Lieberman, O., Phillips, P. C. B. (2004b). Expansions for approximate maximum likelihood estimators of the fractional difference parameter. Accepted to *The Econometrics Journal.*

Martens, M., Zein, J. (2004). Predicting financial volatility: high frequency time series forecasts vis a vis implied volatility. *Journal of Future Markets* 24(11):1005–1028.

Martnes, M., Van Dijk, D., De Pooter, M. (2004). Modeling and forecasting S&P 500 volatility: long memory, structural breaks and nonlinearity. Tinbergen Institute Discussion Paper 2004-067/4.

Merton, R. C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8:323–361.

Phillips, P. C. B. (1999). "Discrete Fourier Transforms of Fractional Processes", *Cowles Foundation Discussion Paper,* No. 1243, Yale University.

Pong, S., Shackleton, M. B., Taylor, S. J., Xu, X. (2004). Forecasting currency volatility: a comparison of implied volatilities and AR(FI)MA models. *Journal of Banking and Finance* 28(10):2541–2563.

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

Robinson, P. M. (1995b). Gaussian semiparametric estimation of long memory. *Annals of Statistics* 23:1630–1661.

Searle, R. (1971). *Linear Models.* Wiley, New York.

Shimotsu, K., Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics* 33(4):1890–1933.

Zhang, L., Mykland, P. A., Ait-Sahalia, Y. (2005a). A tale of two time scales: determining integrated volatility with noisy high–frequency data. *Journal of the American Statistical Association* 100(472):1394–1411.

Zhang, L., Mykland, P. A., Ait-Sahalia, Y. (2005b). Edgeworth expansions for realized volatility and related estimators. Mimeo.