The Local Whittle Estimator of Long Memory Stochastic Volatility

Clifford M. Hurvich
Department of Statistics and Operations Research
New York University
44 W. 4^{th} Street
New York NY 10012 USA
churvich@stern.nyu.edu

Bonnie K. Ray*
Department of Mathematical Sciences
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598 USA
bonnier@us.ibm.com

Abstract

We propose a new semiparametric estimator of the degree of persistence in volatility for long memory stochastic volatility (LMSV) models. The estimator uses the periodogram of the log squared returns in a local Whittle criterion which explicitly accounts for the noise term in the LMSV model. Finite-sample and asymptotic standard errors for the estimator are provided. An extensive simulation study reveals that the local Whittle estimator is much less biased and that the finite-sample standard errors yield more accurate confidence intervals than the widely-used GPH estimator. The estimator is also found to be robust against possible leverage effects. In an empirical analysis of the daily Deutsche Mark/US Dollar exchange rate, the new estimator indicates stronger persistence in volatility than the GPH estimator, provided that a large number of frequencies is used.
Key Words: long-range dependence; nonlinearity; semiparametric estimation

1. INTRODUCTION

Long memory in volatility of financial returns has received considerable attention in recent years. See, *e.g.* Ding, Granger and Engle (1993), de Lima and Crato (1993), Baillie, Bollerslev and Mikkelsen (1996), Andersen and Bollerslev (1997a,b), Comte and Renault (1998), Lobato and Savin (1998), Lobato and Robinson (1998), Ray and Tsay (2000), Lobato and Velasco (2000), Andersen, Bollerslev, Diebold and Labys (2001), Robinson (2001), and Wright (2002). A widely-used methodology for determining the degree of persistence in volatility, parameterized

*Corresponding author

by d, is to estimate d semiparametrically using log periodogram regression based on squared or absolute returns. The log periodogram regression estimator, $\widehat{d}_{GPH}$, was originally proposed by Geweke and Porter-Hudak (1983), in a non-volatility context. Properties of this estimator for stationary Gaussian processes, which are linear and hence free of volatility clustering, were derived by Robinson (1995a) and Hurvich, Deo and Brodsky (1998). In this case, $\widehat{d}_{GPH}$ is consistent and asymptotically normal under certain regularity conditions. The GPH method is practically appealing, as it is may be computed using simple linear regression.

To model observed persistence in volatility of financial returns, the long memory stochastic volatility (LMSV) model was introduced independently by Breidt, Crato and de Lima (1998) and Harvey (1998). For a comprehensive discussion of stochastic volatility models, including various long memory specifications such as LMSV, see Ghysels, Harvey, and Renault (1996) and the references therein. The series of logarithms of squared values of an LMSV process is modeled as a long-range dependent process plus added noise (See Section 2). However, Deo and Hurvich (2001) show that $\widehat{d}_{GPH}$ based on log squared returns in the LMSV model suffers from a potentially severe negative bias which does not arise in the Gaussian case, and which depends on d, becoming worse as d goes to zero. Deo and Hurvich (2001) is, to the best of our knowledge, the first paper to derive theoretical properties for any semiparametric estimator of d in the context of volatility.

In this paper, we propose a new semiparametric estimator of d in the LMSV context, designed with a view towards bias reduction in comparison with $\widehat{d}_{GPH}$. The new estimator, $\widehat{d}_{LWN}$, is a local Whittle estimator which explicitly accounts for the noise term in the LMSV model. This noise term introduces a certain degree of roughness, which is determined by d, in the short memory component of the spectral density in a neighborhood of zero frequency. The estimator $\widehat{d}_{LWN}$ is implicitly defined, and may be computed using a two-dimensional nonlinear optimization algorithm.

1.1 Analysis of transformed returns

We focus in this paper on estimators of d for series of log squared returns. This choice of transformation seems to be justified empirically; Ding, Granger and Engle (1993) observed that autocorrelations of absolute returns raised to the power c were typically maximized by taking c close to 1. Deo and Hurvich (2003) have proposed an explanation in terms of outlier effects for the fact that absolute and squared returns typically have smaller sample autocorrelations than log squared returns. An analogous phenomenon presumably holds for the degree of persistence implied by periodograms. Indeed, Wright (2002) has shown using simulations under both LMSV and ARCH-type models that periodogram-based semiparametric estimators of d are less negatively biased if log squared returns are used, instead of absolute or squared returns.

1.2 Using GPH to assess persistence in volatility

Even using log squared returns for analysis, however, the GPH estimator of persistence in volatility in LMSV models still suffers from a potentially severe negative bias. This bias, which is given explicitly in Theorem 1 of Deo and Hurvich (2001), implies a slow rate of convergence for $\widehat{d}_{GPH}$. In general, in order to guarantee that $\sqrt{m}(\widehat{d}_{GPH} - d)$ will be asymptotically normal with zero mean, m must grow more slowly than $n^{4d/(4d+1)}$, where n is the sample size and m is the number of frequencies used in the regression. For example, if $d = .25$, then m must grow more slowly than $n^{1/2}$, while if $d = .1$, m must grow more slowly than $n^{2/7}$. In no situation with $0 < d < 1/2$ can m grow faster than $n^{2/3}$.

Now, when $d = 0$ in the LMSV model, Hurvich and Soulier (2002) have shown that $\sqrt{m}(\widehat{d}_{GPH} - d)$ is asymptotically normal with mean zero and variance $\pi^2/24$, as long as m grows more slowly than $n^{4/5}$. Thus, an asymptotically valid test for long memory in volatility is to reject the null hypothesis of $d = 0$ in favor of $d > 0$ if the test statistic $\widehat{d}_{GPH}/\sqrt{\pi^2/(24m)}$ is greater than the $1 - \alpha$ quantile of the standard normal distribution, where α is the desired significance level. This would seem to suggest that $\widehat{d}_{GPH}$ is satisfactory for assessing the existence of persistence in volatility.

Nevertheless, the fact that the bias in $\widehat{d}_{GPH}$ depends on d makes statistical inference based on $\widehat{d}_{GPH}$ difficult, if not impossible, in general. Indeed, even if we knew that $d > 0$, we could not construct an asymptotically valid confidence interval for d based on $\widehat{d}_{GPH}$ without an *a priori*, strictly positive lower bound for d. Such a bound, which would seldom if ever be available in practice, would be needed to prevent the practitioner from selecting too large a value of m, and thereby invalidating the confidence interval by introducing excessive bias in $\widehat{d}_{GPH}$. Thus, a better estimator of long memory in volatility is desirable.

1.3 Outline of paper

Here, we investigate the properties of the local Whittle estimator, $\widehat{d}_{LWN}$ compared to $\widehat{d}_{GPH}$ in practice. We also compare the proposed method to the local polynomial GPH estimator, $\widehat{d}_{LP-GPH}$ of Andrews and Guggenberger (2003), which reduces the bias of GPH for sufficiently regular linear processes. We present extensive simulation studies comparing the performance of $\widehat{d}_{GPH}$, $\widehat{d}_{LP-GPH}$ and $\widehat{d}_{LWN}$. The simulations reinforce the fact that $\widehat{d}_{GPH}$ can be extremely negatively biased. This is of considerable practical relevance, since it suggests, in conjunction with our data analysis, that many of the published data analyses may be understating the strength of the true persistence in volatility. The local polynomial GPH estimator is slightly less biased, but at the cost of increased variability. We find that $\widehat{d}_{LWN}$ has much less bias than $\widehat{d}_{GPH}$, and its variance inflation compared with $\widehat{d}_{GPH}$ is not unreasonably large. Thus, $\widehat{d}_{LWN}$ seems to hold great promise for estimating persistence in volatility. The theoretical properties of $\hat{d}_{LWN}$ have been studied by Hurvich, Moulines and Soulier (2002). We summarize here the most relevant aspects of that theory, including an expression for the asymptotic variance of $\hat{d}_{LWN}$, which depends on d. We also provide a feasible, finite-sample expression for the variance of $\hat{d}_{LWN}$. The accuracy of these approximations, as well as resulting confidence intervals, is assessed in our simulation study, which also explores the robustness of $\hat{d}_{LWN}$ in the presence of leverage effects. Finally, we present an empirical analysis of the daily Dollar/Deutsche Mark exchange rate, and find a higher degree of persistence in volatility than suggested by the GPH estimator when a large number of frequencies is used.

2. ESTIMATION OF d IN THE LMSV MODEL

The LMSV model for returns $\{r_t\}$ as defined by Breidt, Crato and de Lima (1998) takes form

$$r_t = \eta \exp(Y_t/2) e_t \tag{1}$$

where $\eta > 0$ is a scale parameter, $\{e_t\}$ are independent identically distributed (*i.i.d.*) shocks, not necessarily Gaussian, with zero mean and unit variance , and $\{Y_t\}$ is a zero-mean stationary Gaussian process, independent of $\{e_t\}$, with spectral density

$$f_Y(x) = x^{-2d} f_Y^*(x), \tag{2}$$

where f_Y^* is an even, positive, continuous function on $[-\pi, \pi]$ and d is the memory parameter, $0 \le d < 1/2$. The function f_Y^* may be thought of as the spectral density of the short-memory component of $\{Y_t\}$. We assume hereafter that $d > 0$. The assumption that $\{Y_t\}$ is independent of $\{e_t\}$ precludes the possibility of leverage effects. We will consider relaxing this assumption in Section 4.3. Under the LMSV model, the logarithms of the squared returns, $X_t = \log(r_t^2)$, may be expressed as

$$X_t = Y_t + Z_t, \tag{3}$$

where $\{Z_t\} = \{\log e_t^2 + \log \eta^2\}$ is *i.i.d.* with variance $\sigma_Z^2 < \infty$.

The assumptions given above for the LMSV model imply that the spectral density of X_t may be written as

$$f_X(x) = f_Y(x) + \sigma_Z^2/(2\pi). \tag{4}$$

The LMSV model described above can be generalized in various ways. The $\{Y_t\}$ series can be non-Gaussian, subject to the regularity conditions described below. Additionally, the log squared returns can be nonstationary, with memory parameter $d \in (1/2, 1]$. In this nonstationary case, we define the model by $r_t = \eta \exp(U_t/2) e_t$ where $U_t = \sum_{s=1}^t Y_s$ and $f_Y(x) = x^{-2(d-1)} f_Y^*(x)$, so that here $\{Y_t\}$ has memory parameter $d_Y \in (-1/2, 0]$. Since $\{U_t\}$ is nonstationary, it does not have a spectral density, but it does have a pseudo spectral density given by $|1 - e^{ix}|^{-2} f_Y(x)$. This pseudo spectral density plays a similar role to that of the ordinary spectral density in determining the properties of the periodogram when $d > 1/2$. See, e.g., Solo (1992), Hurvich and Ray (1995), Velasco (1999).

Overall, then, our generalized model is

$$r_t = \begin{cases} \eta \exp(Y_t/2) e_t \ , & d \in (0, 1/2) \\ \eta \exp(\sum_{s=1}^t Y_s/2) e_t \ , & d \in (1/2, 1) \end{cases}$$

such that $\{Y_t\}$ is independent of the *i.i.d.* process $\{e_t\}$, where $\{Y_t\}$ is stationary and invertible with spectral density $f_Y(x) = x^{-2d_Y} f_Y^*(x)$, $d_Y \in (-1/2, 1/2)$, and

$$d_Y = \begin{cases} d \text{ if } d \in (0, 1/2) \\ d - 1 \text{ if } d \in (1/2, 1) \end{cases} .$$

The log squared return series $\{\log r_t^2\}$ is given by

$$X_t = \begin{cases} Y_t + Z_t \text{ if } d \in (0, 1/2) \\ \sum_{s=1}^t Y_s + Z_t \text{ if } d \in (1/2, 1) \end{cases} .$$

In both cases, $\{Z_t\} = \{\log e_t^2 + \log \eta^2\}$ is an *i.i.d.* process with finite variance, independent of $\{Y_t\}$.

2.1 The GPH Estimator

Define the periodogram of the observations $X_1, \cdots, X_n$ at the k^{th} Fourier frequency $x_k = 2\pi k/n$ by

$$I_{n,k}^X = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{itx_k}\right|^2 .$$

The GPH estimator of d using the first m Fourier frequencies may be written as

$$\widehat{d}_{GPH} = -\frac{1}{2S_{ww}}\sum_{k=1}^{m} a_k \log I_{n,k}^X,$$

where $a_k = W_k - \overline{W}$, $W_k = \log|2\sin(x_k/2)|$, $\overline{W} = m^{-1}\sum_{k=1}^{m} W_k$ and $S_{ww} = \sum_{k=1}^{m} a_k^2$. Note that the intuition behind the GPH estimator in the standard Gaussian case is the linear relation at low frequencies between the logarithm of the spectral density of a long memory process and the logarithm of the corresponding frequencies, as can be seen from (2). The $\{Z_t\}$ process in (3) may be viewed as an additive noise term which corrupts this linear relationship and impairs our ability to estimate the memory parameter in the signal process $\{Y_t\}$.

2.2 The Local Polynomial GPH Estimator, $\widehat{d}_{LP-GPH}$

Andrews and Guggenberger (2003) proposed a local polynomial GPH estimator of long memory. We will consider the simplest version here, in which the estimator $\hat{d}_{LP-GPH}$ is defined as the coefficient of $-2\log x_k$ in an ordinary least squares regression of $\log I_{n,k}^X$ on a constant, $-2\log x_k$ and x_k^2, for $k = 1, \ldots, m$. For a Gaussian (and therefore linear) process such that the spectral density of the short memory component is sufficiently smooth, specifically, smooth of order $s \geq 1$ at zero frequency, the optimal rate of convergence of mean squared error (MSE) of $\hat{d}_{LP-GPH}$ is proportional to $n^{-2\phi/(2\phi+1)}$ where $\phi = min\{s, 4\}$. Unfortunately, in the context of the LMSV model, we have $s = 2d$ (see Equation (6) below), presumably leading to an optimal mean squared error proportional to $n^{-4d/(4d+1)}$. This rate is identical to the rate attained by GPH in the LMSV context as given by Deo and Hurvich (2001), and is inferior to the optimal rate of $n^{-4/5+\epsilon}$ attained by the MSE of $\hat{d}_{LWN}$, as will be shown in Section 3 below. Nevertheless, for completeness we include $\hat{d}_{LP-GPH}$ in our comparative Monte Carlo study in Section 4.

2.3 The Local Whittle with Noise Estimator, $\widehat{d}_{LWN}$

We assume in this section that

$$f_Y^*(x) = f_Y^*(0) + Cx^2 + R(x), \tag{5}$$

where $R(x) = o(x^2)$ as $x \to 0$. This assumption holds for most short-memory models in current use, including all stationary invertible ARMA models (see Robinson 1995a), and exponential models (see Bloomfield, 1973). To avoid a conflict of notation, in this and the next section we denote the true value of the memory parameter by d_0. Then from Equations (2), (4) and (5), we can write

$$f_X(x) = \frac{\sigma_Z^2}{2\pi}\left[1 + \frac{2\pi f_Y^*(0)}{\sigma_Z^2} x^{-2d_0}\right] + O(x^{2-2d_0}) \ . \tag{6}$$

Stationarity is implicitly assumed in writing (6), but an argument based on pseudo-spectral densities shows that (6) holds even in the nonstationary case. See Hurvich and Ray (1995), or Hurvich, Moulines and Soulier (2002).

Since the final $O(x^{2-2d_0})$ term is negligible with respect to the other terms in (6) for x close to 0, it seems reasonable to try locally fitting a model of form

$$g_\theta(x) = b_0(1 + b_1 x^{-2d}) \tag{7}$$

in a neighborhood of zero frequency, where $\theta = (b_0, b_1, d)'$ is the vector of parameters. Model (7) explicitly accounts for the noise term in (3).

For local fitting of model (7), we propose to minimize the local Whittle criterion

$$L(\theta) = \sum_{j=1}^{m}\left[\log\ g_\theta(x_j) + \frac{I_{n,j}^X}{g_\theta(x_j)}\right] \ , \tag{8}$$

where the minimization is carried out in a compact set $\Theta \subset \mathbb{R}^+ \times \mathbb{R}^+ \times (0, 0.75)$, and m is a positive integer such that $1/m + m/n \to 0$ as $n \to \infty$. We assume that θ_0 is an interior point of Θ, where $\theta_0 = [\sigma_Z^2/(2\pi)\,,\, 2\pi f_Y^*(0)/\sigma_Z^2\,,\, d_0]'$ is the vector of true parameters.

The parameter b_0 can be concentrated out of (8), so minimizing $L(\theta)$ is equivalent to finding (b_1, d) to minimize

$$\tilde{L}(b_1, d) = \sum_{j=1}^{m}\left[\log\ \tilde{g}_{\tilde{\theta}}(x_j) + \frac{I_{n,j}^X}{\tilde{g}_{\tilde{\theta}}(x_j)}\right] \ , \tag{9}$$

where $\tilde{\theta} = (b_1, d)'$,

$$\tilde{g}_{\tilde{\theta}}(x_j) = b_0^{\tilde{\theta}}(1 + b_1 x_j^{-2d}) \ , \tag{10}$$

and

$$b_0^{\tilde{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\frac{I_{n,j}^X}{1 + b_1 x_j^{-2d}} \ . \tag{11}$$

The vector of estimated parameters is $\widehat{\theta} = (\widehat{b}_0\,,\, \widehat{b}_1\,,\, \widehat{d}_{LWN})'$, where $\widehat{b}_1, \widehat{d}_{LWN}$ minimize $\tilde{L}$, and $\widehat{b}_0 = b_0^{(\widehat{b}_1, \widehat{d}_{LWN})'}$. Here, the minimization is carried out in a compact set $\Theta \subset \mathbb{R}^+ \times (0, 0.75)$.

In the discussion above, it was implicitly assumed that the minimizer of $\tilde{L}$ occurs at an interior point of Θ. In this case, the estimators $\hat{b}_1$ and $\hat{d}_{LWN}$ satisfy the so-called first order conditions (FOC), that is, the partial derivatives of $\tilde{L}$ are zero at $(b_1, d) = (\hat{b}_1, \hat{d}_{LWN})$. In fact, we need to slightly modify the definition of $\widehat{d}_{LWN}$ to account for possible solutions to (7) on the boundary.

If the global minimizer of $\tilde{L}$ occurs at a boundary point of Θ, then, although there may be several interior points which satisfy the FOC, none of these local optima corresponds to a global optimum, and we define our estimator as follows. (1) If there are no solutions to the FOC, we use the global optimum (boundary point) as our estimator. (2) If there are any solutions to the FOC, then our estimator is defined to be that solution which is closest in the sense of ordinary Euclidean distance to the global optimum (boundary point).

It should be noted that the above algorithm implies that a local optimum will be chosen over the global optimum when the latter is a boundary point. The reason for this choice is to facilitate the development of theory, as suggested by Andrews and Sun (2001). The context for the suggestion of Andrews and Sun (2001) was a local polynomial Whittle estimator of long memory, in a non-volatility context. There, as here, the estimator involves minimization of a multidimensional criterion function, and the individual components of the estimator converge at different rates.

3 PROPERTIES OF $\widehat{d}_{LWN}$

The asymptotic properties of $\hat{d}_{LWN}$ and other related estimators are derived in Hurvich, Moulines and Soulier (2002). We present here the result for $\hat{d}_{LWN}$ under simplified assumptions.

We assume that $\{Y_t\}$ has an infinite order moving average representation

$$Y_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j} \ , \tag{12}$$

where $\{\epsilon_t\}$ is a zero-mean white noise process, not necessarily Gaussian, with $Var[\epsilon_t] = \sigma_\epsilon^2$, and $\sum_{j=0}^{\infty} a_j^2 < \infty$. Note that $\{\epsilon_t\}$ is independent of $\{Z_t\}$. We lose no generality in assuming that $\{Y_t\}$ has zero mean, since the estimators considered in this paper are all functions of the periodogram at nonzero Fourier frequencies. In the nonstationary case, the assumption that $\{Y_t\}$ has mean zero ensures that $\{X_t\}$ is free of linear trends.

Define $a(x) = \sum_{j=0}^{\infty} a_j e^{\mathrm{i}jx}$. The spectral density of the process $\{Y_t\}$ is then $f_Y(x) = |a(x)|^2 \sigma_\epsilon^2/(2\pi)$, and we assume that it can be expressed as

$$f_Y(x) = x^{-2d_Y} f_Y^*(x), \tag{13}$$

with $d_Y \in (-1/2, 1/2)$.

To present our theoretical results, we require the following definition.

Definition 1. *For $\alpha \in (0, \pi]$, $\beta > 0$ and $0 < \mu < \infty$, $\mathcal{F}_0(\alpha, \beta, \mu)$ is the set of functions g defined on $[-\pi, \pi]$ satisfying $\int_{-\pi}^{\pi} |g(x)| dx \leq \mu$ and for all $x \in [-\alpha, \alpha]$,*

$$|g(x)| \leq \mu |x|^{\beta}. \tag{14}$$

We also require the following assumption, which was made in Robinson (1995b) as well.

$\{\epsilon_t\}$ is a martingale difference sequence such that for all t, $\mathbb{E}[\epsilon_t^4] := \mu_4 < \infty$ and $\mathbb{E}[\epsilon_t^2 \mid \epsilon_s, s < t] = 1$ almost surely.

Theorem 1. *Let $\{Y_t\}$ have a moving average representation (12) with respect to a white noise $\{\epsilon_t\}$ which satisfies (**A1**) and such that the function $a(x) = \sum_{j=0}^{\infty} a_j e^{ijx}$ can be expressed as $a(x) = x^{-d_Y} a^*(x)$, where $(a^*(0)^{-1}a^*(x) - 1) \in \mathcal{F}_0(\alpha, \beta, \mu)$ for some $\beta > 2d_0$, $\alpha > 0$ and $\mu > 0$. Assume that $d_0 \in (0, .75)$. If m is a non decreasing sequence of integers such that*

$$\lim_{n\to\infty} (m^{-4d_0-1} n^{4d_0} + n^{-2\beta} m^{2\beta+1} \log^2(m)) = 0, \tag{15}$$

then $m^{1/2}(\hat{d}_{LWN} - d_0)$ is asymptotically Gaussian with zero mean and variance $(1+2d_0)^2/(16{d_0}^2)$.

Thus, if $\beta = 2$ (as is most commonly assumed) and we use $m = n^{4/5-2\epsilon}$ for some small ϵ, then $\hat{d}_{LWN}$ is $n^{2/5-\epsilon}$-consistent, i.e., the same rate of convergence enjoyed by Robinson's (1995b) Gaussian semiparametric estimator in the linear case. The first term in (15) imposes a lower bound on the allowable value of m, requiring that m tend to ∞ faster than $n^{4d_0/(4d_0+1)}$. Thus, for example, if $d_0 = .4$ and $\beta = 2$ then m must tend to ∞ faster than $n^{8/13} \approx n^{.62}$ and slower than $n^{4/5}$ in order for Theorem 1 to be valid.

Note that the asymptotic variance of $\hat{d}_{LWN}$ in Theorem 1 depends only on d_0, and is a decreasing function of d_0. Unfortunately, unless the noise to signal ratio (nsr) is quite small, this asymptotic variance may not accurately reflect the actual variance, even in the relatively large sample sizes considered in this paper. An alternative approach is to construct a finite-sample approximation to the variance. Here, we proceed heuristically and omit some long algebraic calculations, but we note that our approximation is asymptotically equivalent to the result in Theorem 1, and its improvement over the asymptotic expression

$$\mathrm{var}(\hat{d}_{LWN}) \approx (1/m)(1 + 2d_0)^2/(16{d_0}^2)$$

is documented in our simulation study.

Standard likelihood theory suggests that the covariance matrix of $(\hat{b}_1, \hat{d}_{LWN})'$ is well approximated by the inverse of the Fisher information matrix,

$$Cov[\nabla\tilde{L}(b_{1,0}, d_0)] \quad ,$$

where $b_{1,0}$ is the signal to noise ratio, $b_{1,0} = 2\pi f_Y^*(0)/\sigma_Z^2$, and ∇ is the gradient, i.e., the vector of partial derivatives with respect to b_1 and d. Treating the periodogram values $I_{n,j}^X$ as independently distributed as $g_{\theta_0}(x_j)(1/2)\chi_2^2$, and ignoring a multiplying constant which converges in probability to 1, we obtain after a long calculation the approximation $Cov[\nabla\tilde{L}(b_{1,0}, d_0)] \approx M$, where

$$\begin{aligned}
M_{11} &= \sum_{k=1}^{m} \frac{1}{(x_k^{-2d_0} + b_{1,0}^{-1})^2} - \frac{1}{m}\left(\sum_{k=1}^{m} \frac{1}{x_k^{-2d_0} + b_{1,0}^{-1}}\right)^2 \\
M_{12} &= -2\sum_{k=1}^{m} \frac{\log x_k\, x_k^{-2d_0}}{(x_k^{-2d_0} + b_{1,0}^{-1})^2} + \left(\frac{2}{m}\sum_{k=1}^{m} \frac{\log x_k\, x_k^{-2d_0}}{x_k^{-2d_0} + b_{1,0}^{-1}}\right)\left(\sum_{j=1}^{m} \frac{1}{x_j^{-2d_0} + b_{1,0}^{-1}}\right) \\
M_{21} &= M_{12} \\
M_{22} &= 4\sum_{k=1}^{m} \left(\frac{\log x_k\, x_k^{-2d_0}}{x_k^{-2d_0} + b_{1,0}^{-1}}\right)^2 - \frac{4}{m}\left(\sum_{k=1}^{m} \frac{\log x_k\, x_k^{-2d_0}}{x_k^{-2d_0} + b_{1,0}^{-1}}\right)^2 \quad .
\end{aligned} \tag{16}$$

Thus, we may approximate $Var(\hat{d}_{LWN})$ by $A_{2,2}$, the (2,2) entry of $A = M^{-1}$. The use of $A_{2,2}$ is not feasible in practice, since d_0 and $b_{1,0}$ are not known. We can, however, use the feasible version $\hat{A}_{2,2}$ where d_0 and $b_{1,0}$ are replaced by $\hat{d}_{LWN}$ and $\hat{b}_1$ in (16).

In the next section, we compare the performance of $\widehat{d}_{LWN}$ relative to that of $\widehat{d}_{GPH}$ and $\widehat{d}_{LP-GPH}$ and assess the accuracy of the asymptotic and finite-sample expressions for $Var(\widehat{d}_{LWN})$ using simulation.

4 SIMULATION RESULTS

4.1 Assessment of Empirical Bias and Variance for $\widehat{d}_{LWN}$

We simulated logarithms of squared LMSV processes by first simulating Gaussian ARFIMA(p, d, q) data. The PACF method of Hosking (1984) was used to generate data from a Gaussian ARFIMA$(0, d, 0)$ process. An ARMA(p, q) filter was then applied to give ARFIMA(p, d, q) data. An independent sequence of logarithms of squared standard normal random variates was added to the ARFIMA data to produce a series of logarithms of a squared LMSV-ARFIMA(p, d, q) process. One thousand realizations were generated for each value of $n = (1000, 5000, 10000)$, and for each of two values of the noise to signal ratio, $nsr = b_{1,0}^{-1}$. Since we take the $\{e_t\}$ to be standard normal, we have $\sigma_Z^2 = \pi^2/2$. The values $nsr = 5$ and $nsr = 10$ were chosen to correspond to the large nsr values observed in other empirical studies of LMSV models in finance (e.g., Breidt, Crato, and de Lima, 1998) and to see how the estimates of d are influenced by nsr in practice. For each realization, the $\widehat{d}_{GPH}$, $\widehat{d}_{LP-GPH}$ and $\widehat{d}_{LWN}$ estimators were evaluated for $m = ([n^{.4}], [n^{.5}], [n^{.6}], [n^{.7}], [n^{.8}])$. We investigated the LMSV-ARFIMA$(0, d, 0)$ model for values of $d = 0.3, 0.4, 0.45, 0.49$. These values were chosen based on previous findings of relatively strong persistence in financial time series (e.g. Lobato and Savin, 1999; Ray and Tsay, 2000). We also investigated the influence of ARMA components on the estimates by considering three LMSV-ARFIMA models having nonzero ARMA terms, that of an LMSV-ARFIMA$(1, d, 0)$ model with $d = 0.4$ and $\phi = 0.5, 0.8$ where ϕ is the autoregressive parameter in the ARFIMA$(1, d, 0)$ model, that is, $(1 - B)^d(1 - \phi B)y_t = \eta_t$ with $\{\eta_t\}$ $i.i.d$ normal random variates having standard deviation such that the specified nsr is obtained, and that of an LMSV-ARFIMA$(0, d, 1)$ model with $d = 0.4$ and $\theta = -0.8$, where θ is the moving-average parameter in the ARFIMA$(0, d, 1)$ model, that is, $(1 - B)^d y_t = (1 - \theta B)\eta_t$. The $\widehat{d}_{LWN}$ estimator was obtained by numerical optimization of (9) as a function of d and b_1. The value of d was constrained to lie in the range $[0.01, .75]$, while $\log(b_1)$ was constrained to the region $[-8, 20]$.

All computations were performed using Microsoft Visual FORTRAN in conjunction with the IMSL numerical libraries running on a 1 GHz Pentium 3 processor with 512 MB RAM in a Windows 2000 operating environment. The IMSL function DBCONF with default control parameters was used for optimization. The initial value used in computing $\widehat{d}_{LWN}$ for a given m was the $\widehat{d}_{GPH}$ estimator based on the same value of m. To find solutions to the FOC when the global optimum was obtained at a boundary point, we divided Θ into 16 equal-sized, non-overlapping rectangular regions. For each of these regions, (9) was optimized using DBCONF with starting value given by the midpoint of the region. Any interior solutions obtained by DBCONF were assumed to be solutions to the FOC. Complete simulation results for all sample sizes and nsr values for each specified ARFIMA(p, d, q) model required approximately 15 hours of CPU time.

Figure 1 presents representative results for the LMSV-ARFIMA$(0, 0.4, 0)$ model graphically, in the form of boxplots, for the $nsr = 5$ case, while Table 1 provides detailed numerical results in the LMSV-ARFIMA$(0, 0.4, 0)$ case for each nsr value when $n = 5000$. Tables and figures corresponding to additional simulation results are available for download from `http://www.stern.nyu.edu/sor/research/wp03.html`.

We start by discussing the results for the LMSV-ARFIMA$(0, d, 0)$ processes. Overall, in most situations studied, $\widehat{d}_{LWN}$ has a smaller root mean squared error (RMSE) than either $\widehat{d}_{GPH}$ or $\widehat{d}_{LP-GPH}$. As m increases for given values of n, nsr and d, the RMSE for $\widehat{d}_{LWN}$ typically decreases, while the RMSE for $\widehat{d}_{GPH}$ and $\widehat{d}_{LP-GPH}$ is typically a convex function of m. The minimum RMSE with respect to m for a given situation is typically smaller for $\widehat{d}_{LWN}$ than for $\widehat{d}_{GPH}$ or $\widehat{d}_{LP-GPH}$.

The bias of $\widehat{d}_{LWN}$ is uniformly small, while the biases of $\widehat{d}_{GPH}$ and $\widehat{d}_{LP-GPH}$ become increasingly negative as either m or nsr is increased. This is in agreement with the theoretical results of Deo and Hurvich (2001). Even for samples of size $n = 10000$, the bias of $\widehat{d}_{GPH}$ may be quite severe. For example, for the LMSV-ARFIMA$(0, 0.40, 0)$ process with $n = 10000$, $m = [n^{.8}]$, $nsr = 10$, the bias in $\widehat{d}_{GPH}$ is -0.262, rendering the estimate nearly useless. The bias in $\widehat{d}_{LP-GPH}$, although smaller, is still -0.182. See Table 1.

The standard errors of both $\widehat{d}_{GPH}$ and $\widehat{d}_{LWN}$ decrease as m or n is increased, holding everything else fixed. Consistent with theory, the standard error of $\widehat{d}_{GPH}$ is often smaller than that of the corresponding $\widehat{d}_{LWN}$. For a given n, m, d, the standard error for $\widehat{d}_{GPH}$ is insensitive to nsr while the standard error for $\widehat{d}_{LWN}$ increases as nsr increases. Thus, for large nsr, the standard error for $\widehat{d}_{LWN}$ can become dramatically larger than the standard error for $\widehat{d}_{GPH}$ (except when m is small). However, this inflation in standard error for $\widehat{d}_{LWN}$ is usually not enough to offset the inflation in bias for $\widehat{d}_{GPH}$, so that $\widehat{d}_{LWN}$ typically has the smaller RMSE. The boxplots illustrate very nicely the trade-off between bias and variance, clearly showing the superiority of $\widehat{d}_{LWN}$ when m is large.

Additional simulation results over a range of d values show that as d is increased, holding everything else fixed, the standard error for $\widehat{d}_{LWN}$ goes down, while that for $\widehat{d}_{GPH}$ remains stable. Furthermore, as d is increased, the bias for $\widehat{d}_{LWN}$ remains stable, while negative bias for $\widehat{d}_{GPH}$ becomes more severe. These findings are consistent with the theoretical results of Theorem 1 for $\widehat{d}_{LWN}$ and those of Deo and Hurvich (2001) for $\widehat{d}_{GPH}$, showing strong superiority of $\widehat{d}_{LWN}$ to $\widehat{d}_{GPH}$ in terms of RMSE when d is large.

Boxplots of estimates for the LMSV-ARFIMA$(1, d, 0)$ model with $\phi = 0.8$ and $n = 5000$ are shown in Figure 2. For this model, $\widehat{d}_{GPH}$ appears less biased than it was when the autoregressive parameter was absent. This can be explained by noting that the presence of the autoregressive parameter tends to increase the expected value of $\widehat{d}_{GPH}$, and thereby results in a less negatively biased estimator. Nevertheless, in almost all situations studied for this model, $\widehat{d}_{LWN}$ has a smaller RMSE than $\widehat{d}_{GPH}$. This is true despite the strong positive short-range correlation induced by the autoregressive parameter $\phi = 0.8$. Similar results were found for the other ARMA component models considered.

Overall, our simulation results suggest that $\widehat{d}_{LWN}$ is preferable to $\widehat{d}_{GPH}$ since the latter estimator may suffer from a very strong negative bias due to the noise term in the LMSV model, while the former estimator suffers from no such bias.

4.2 Assessment of Approximate Variance Expression for $\widehat{d}_{LWN}$

According to the asymptotic theory given in Theorem 1, the variance of $\widehat{d}_{LWN}$ does not depend on nsr. Our simulations appear to be at least somewhat at odds with that theory, as seen from the above discussion. The first two rows of the boxes labeled Asymptotic in Table 2 show the average and median standard errors across replications obtained using the asymptotic expression $(1+2d)/(4dm^{1/2})$ evaluated using $\hat{d}_{LWN}$ for the ARFIMA$(0, 0.4, 0)$ model with $n = 5000$, while the third row gives the asymptotic value computed using the true value of d. The mean values are much larger than the values obtained using the asymptotic expression with the true value of d, especially when m is small. We attribute this to a few outlying values of $\widehat{d}_{LWN}$, as can be seen from the boxplots in Figure 1. Although the median value for the standard errors based on estimated d values is close to that based on the true value of d, the values typically do not match closely the standard errors observed in the simulations, which increase as nsr increases (see rows labeled Simulation in Table 2). Thus, for the sample sizes typically encountered in practice, the asymptotic expression does not seem to provide a reliable approximation to the actual standard error of $\widehat{d}_{LWN}$.

We also explored whether $A_{2,2}$, the $(2, 2)$ entry of M^{-1}, provides a better approximation, where the entries of M are given by (16). Note that $A_{2,2}$ depends not only on d, but also on b_1. A feasible version can be computed by substituting estimates of the unknown parameters in the expression for $A_{2,2}$. The first two rows of the boxes labeled Hessian in Table 2 give the mean and median values of the standard errors computing using (16) with LWN-estimated parameter values, while the third row gives the value obtained when the true parameter values are used in (16). Again we see that the mean value of the standard errors computed using estimated parameter values can be extremely large. This is true in particular when $n = 1000$ and also when n is larger but m is small. This is due to large variations in the estimated nsr values used in the computation of (16). Large sample sizes and large values of m, i.e. $m = [n^{.7}], [n^{.8}]$ are needed to accurately estimate nsr. When this is the case, both the mean and median values are very close to the values observed in the simulations.

We also compared the empirical 90% and 95% coverage obtained for Gaussian-based confidence intervals on d constructed using the estimated standard errors based on the asymptotic formula, the formula of (16) with LWN-estimated parameters, and the formula of (16) with known parameters. For completeness, these coverages were compared to those obtained from the GPH estimator with variance $\pi^2/(24S_{ww})$. Table 3 shows a representative result of these comparisons for the ARFIMA$(0, 0.4, 0)$ case when $n = 5000$. The values in parentheses denote the median lengths of the constructed intervals. The LWN-based confidence intervals provide close to nominal coverage when d is estimated using a large number of Fourier frequencies and the interval is constructed using the standard errors computed from (16) with estimated parameter values. The GPH-based confidence intervals, in contrast, provide very poor coverage. These results indicate that reliable determination of the degree of persistence in an LMSV model can be made using the Local Whittle method.

4.3 Robustness of $\widehat{d}_{LWN}$ Against Leverage Effects

Our assumption that $\{e_t\}$ and $\{Y_t\}$ are independent in model (1) implies that there is no leverage effect, whereby the conditional variance of the return for the next period would respond asymmetrically to a current return of a given magnitude, according to its sign. This may be viewed as a drawback as financial data sets often do present evidence of a leverage effect. (See Jacquier, Polson and Rossi 1999). We therefore now consider replacing the model (1) by

$$r_t = \eta \exp(Y_{t-1}/2) e_t \quad , \quad d \in (0, 1/2) \quad , \tag{17}$$

where the series $\{Y_t\}$ is given by (12), so that $\{Y_t\}$ has a one-sided linear representation with respect to a series of potentially non-Gaussian shocks $\{\epsilon_t\}$, which we assume here to be *i.i.d.* with zero mean and finite fourth moment. We assume further that e_t and ϵ_u are independent if $t \neq u$, but that e_t and ϵ_t may be correlated. We will use the notation $\rho = Corr(e_t, \epsilon_t)$. A negative value of ρ would induce the type of leverage effect observed in practice, whereby a current negative return is associated with an increase in future volatility. An important reason for using Y_{t-1} rather than Y_t in (17) (see, e.g., Shephard 1996, *pp.* 22, 37 and Ghysels, Harvey and Renault 1996) is that if $\rho \neq 0$, the series $\{r_t\}$ in (1) would not be a Martingale difference sequence, whereas $\{r_t\}$ in (17) will be a Martingale difference sequence, and this is a desirable property for financial returns.

We study here the effect that a nonzero value of ρ in model (17) has on the properties of $\hat{d}_{LWN}$. The fact that $\hat{d}_{LWN}$ is computed from the series $\{\log r_t^2\}$ induces some robustness of $\hat{d}_{LWN}$ against nonzero ρ. In particular, as long as the joint distribution of (e_t, ϵ_t) is symmetric around $(0, 0)$ (as would happen, for example, if (e_t, ϵ_t) were bivariate normal or bivariate t), we obtain $\rho_{z,\epsilon} = Corr(\log e_t^2, \epsilon_t) = 0$. Then the decomposition (6) continues to hold, and it can be shown that the asymptotic properties for $\hat{d}_{LWN}$ are completely unchanged. If, on the other hand, $\rho_{z,\epsilon} \neq 0$, then there will be an additional term of order x^{-d_0} in (6) which is not accounted for by $\hat{d}_{LWN}$. This would be expected to inflate the asymptotic bias of $\hat{d}_{LWN}$. One way to remove the bias would be to consider a modification of $\hat{d}_{LWN}$ in which the additional term is directly fitted. Theoretical properties of the modified estimator were developed in Hurvich, Moulines and Soulier (2002). Here, we remain focused on the unmodified estimator $\hat{d}_{LWN}$ and evaluate its robustness to misspecified correlation structure.

Our simulations here were carried out for model (17) with an $ARFIMA(0, d, 0)$ model for $\{Y_t\}$ having the representation (12), and two different specifications for the distribution of (e_t, ϵ_t). Under the first specification, we take (e_t, ϵ_t) to be independent bivariate normal with zero mean and covariance matrix

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

and $\rho = -0.6$. Under the second specification, we take the $\epsilon_t \overset{iid}{\sim} \delta(1/2)(\chi_2^2 - 2)$, and

$$e_t = (\gamma^2 \delta^2 + 2)^{-1/2} [\gamma \epsilon_t + (\chi_1^2 - 1)] \quad ,$$

where the χ_1^2 are $i.i.d$ and independent of the χ_2^2 random variables. For this specification, we have $\text{var}(\epsilon_t) = \delta^2$, $\text{var}(e_t) = 1$, $Corr(e_t, \epsilon_t) = \rho = \gamma\delta(\gamma^2\delta^2 + 2)^{-1/2}$. We chose $\gamma = -0.875$, $\delta = \sqrt{1.496}$ to correspond to $\rho = -0.60$. From simulations, we then calculate that the noise to signal ratio is $Var[\log e_t^2]/Var[\epsilon_t] = 5.00$, and $\rho_{z,\epsilon} = 0.493$.

Figure 3 shows boxplots of the results for the LMSV-ARFIMA$(0, 0.4, 0)$ model with $n = 5000$ in the case of Gaussian and non-Gaussian correlated e_t and ϵ_t, generated as described above, while Tables 4 and 5 provide detailed numerical results concerning the bias, standard deviation, and RMSE for these models. As expected from the above discussion, the results for the Gaussian correlated errors are basically unchanged from those shown in Figure 1. In the non-Gaussian case, the unaccounted for correlation results in a negative bias in $\hat{d}_{LWN}$ for large m, while the already negative bias observed for $\hat{d}_{GPH}$ is increased even more. An understanding of this phenomenon is similar to that underlying the negative bias observed for $\hat{d}_{GPH}$ in the standard LMSV model, *i.e.*, the correlation results in an additional additive term that is unaccounted for by $\hat{d}_{LWN}$, thereby resulting in model misspecification. The simulation standard deviation in the correlated non-Gaussian model is slightly smaller than that observed in the standard case, although the increased bias results in larger RMSE. For $m = [n^\alpha], \alpha = 0.6, 0.7, 0.8$, $\hat{d}_{LWN}$ still provides better performance than either $\hat{d}_{GPH}$ or $\hat{d}_{LP-GPH}$ for LMSV processes with leverage.

5 ANALYSIS OF CURRENCY EXCHANGE RATES

We consider a data set previously analyzed in Li, Deo and Hurvich (2000) consisting of daily returns on the Deutsche Mark / US Dollar exchange rate, from Jan 2 1985 to May 12 1998, $n = 3485$. Several of the returns r_t were zero. Adjusted log squared returns were constructed, using the method of Fuller (1996), computing

$$X_t = log(r_t^2 + \kappa) - \frac{\kappa}{r_t^2 + \kappa} ,$$

where $\kappa = \tau(n^{-1} \sum r_t^2)$ and $\tau = 0.02$. Time series plots of the returns series and volatility series are shown in Figure 4, while Figure 5 shows the sample autocorrelation function for the volatility series. The volatilities of DM/\$ exchange rates exhibit the apparently changing mean levels characteristic of long-range dependent processes. The sample ACF values, although small, are positive even at large lags.

Table 6 presents the $\widehat{d}_{GPH}$ and $\widehat{d}_{LWN}$ estimators for various values of m. The $\widehat{d}_{GPH}$ values decrease as m increases, a pattern which is consistent with the theoretical fact that the bias in $\widehat{d}_{GPH}$ becomes strongly negative for large values of m. On the other hand, the $\widehat{d}_{LWN}$ values increase with m, reaching 0.556 for $m = [n^{0.8}]$. For each given value of m, except for $m = [n^{0.5}]$, $\widehat{d}_{LWN}$ exceeds the corresponding value of $\widehat{d}_{GPH}$.

To gain some insight on the proper choice of m for $\widehat{d}_{LWN}$ in this exchange rate dataset, we carried out some additional simulations, using a fully parametric LMSV-ARFIMA$(1, d, 0)$ model fitted to the periodogram of $\{X_t\}$ at all Fourier frequencies using the Whittle likelihood. This model was found to fit well according to diagnostic tests performed in Li, Deo and Hurvich (2000). The fitted model has spectral density

$$f_X(x) = f_Y(x) + f_Z(x) = \frac{|2\,\sin(x/2)|^{-2\widehat{d}}\,\widehat{\sigma}_\eta^2}{2\pi|1 - \widehat{\phi}\,\exp(-ix)|^2} + \widehat{\sigma}_Z^2/(2\pi) ,$$

with $\widehat{d} = 0.4086$, $\widehat{\phi} = -0.1556$, $\widehat{\sigma}_\eta = 0.8452$, and $\widehat{\sigma}_Z = 2.4652$. The simulations were done by generating data from this model, using a Gaussian $\{Y_t\}$ process and a noise process given by $Z_t = \log e_t^2$ where e_t are *i.i.d.* with a $t(3)$ distribution. The value of the degrees of freedom for e_t was chosen so that the standard error for Z_t nearly matches the estimated value, $\widehat{\sigma}_Z = 2.4652$. Note that the asymptotic results of Theorem 1 are not dependent on a Gaussian assumption for the multiplicative noise in the LMSV-ARFIMA model.

Table 7 gives the bias and RMSE of $\widehat{d}_{LWN}$ based on one hundred simulated realizations. It is seen that the bias is stable with respect to m, and is quite small, while the RMSE decreases uniformly in m. Overall, $m = [n^{0.8}]$ would appear to be the best choice for this data set, leading to $\widehat{d}_{LWN} = 0.556$. It is notable that this value is so large that it lies outside the range of d values corresponding to a weakly stationary process. The estimated nsr for this series is 23.89. Using (16) with $\hat{d} = 0.556$ and $\hat{b}_1 = 1/23.89$, we obtain an estimated standard error of 0.095. A corresponding confidence interval for d includes values in both the stationary and non-stationary range.

6 SUMMARY

We have investigated the efficacy of a modified Local Whittle method for semiparametrically estimating the degree of long memory in an LMSV process. Our simulation study has focused on the weakly stationary case, $d < 0.5$, with and without leverage effects. The LWN estimator clearly dominates existing methods, such as GPH and the local polynomial GPH method of Andrews and Guggenberger (2003), in the presence of noisy observations. Reliable estimates of standard errors can be obtained using a finite-sample approximation to the asymptotic variance of the modified Local Whittle estimator.

Acknowledgements

We wish to thank the Editor and an anonymous Referee for their helpful remarks and suggestions. We also wish to thank Rohit Deo for informative discussions.

References

[1] Andersen, T. and Bollerslev, T. (1997a). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance* **LII,** 975–1005.

[2] Andersen, T. and Bollerslev, T. (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4,** 115–158.

[3] Andersen, T., Bollerslev, T., Diebold, F. and Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association* **96,** 42–55.

[4] Andrews, D. and Guggenberger, P. (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica* **71,** 675–712.

[5] Andrews, D. and Sun, Y. (2001). Local polynomial Whittle estimation of long-range dependence. Cowles Foundation Discussion Paper 1269. (http://cowles.econ.yale.edu/P/cd/d12b/d1269.pdf)

[6] Baillie, R. Bollerslev, T. and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **74,** 3–30.

[7] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60,** 217–226.

[8] Breidt, F. J., Crato, N., and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* **83,** 325–348.

[9] Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* **8,** 291–323.

[10] de Lima, P. and Crato, N. (1993). Long range dependence in the conditional variance of stock returns. Proceedings of the Business and Economics Statistics Section, Joint Statistical Meetings.

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

[12] Deo, R. S. and Hurvich, C. M. (2003). Estimation of long memory in volatility. In: Taqqu, M., Oppenheim, G., Doukhan, P. (Eds.), Theory and Applications of Long Range Dependence, 313–324. Birkhauser, Boston.

[13] Ding, Z. Granger, C. W. J. and Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* **1,** 83–106.

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

[15] Ghysels, E., Harvey, A. C. and Renault, E. (1996). Stochastic Volatility. In: *Handbook of Statistics*, Vol. 14 (Statistical Methods in Finance), 119–191. North-Holland, Amsterdam.

[16] Harvey, A. C. (1998). Long memory in stochastic volatility. In: Knight, J., Satchell, S. (Eds.), *Forecasting volatility in financial markets.* Butterworth-Heinemann, London.

[17] Hosking, J. R. M. (1984). Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* **20,** 1898–1908.

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

[19] Hurvich, C. M. Moulines, E. and Soulier, P. (2002). Estimating long memory in volatility. New York University Leonard N. Stern School of Business Working Paper SOR–2002–2.

[20] Hurvich, C.M. and Ray, B.K. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of Time Series Analysis*, **16,** (1995), 17–41.

[21] Hurvich, C. M. and Soulier, P. (2002). Testing for long memory in volatility. *Econometric Theory* **18,** 1291–1308.

[22] Jacquier, E., Polson, N. G. and Rossi, P. E. (1999). Stochastic volatility: univariate and multivariate extensions. Scientific Series, CIRANO, Montréal.

[23] Li, K. Deo, R. S. and Hurvich, C. M. (2000). On estimation, diagnostic testing and smoothing of long memory stochastic volatility models. Solomon Center (New York University Stern School of Business) Working Paper Number 2000–27.

[24] Lobato I. N. and Robinson, P. M. (1998). A nonparametric test for I(0). *Rev. Econom. Stud.* **65,** 475–495.

[25] Lobato, I. N. and Savin, N. E. (1998). Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics* **16,** 261–283.

[26] Lobato, I. N. and Velasco, C. (2000). Long memory in stock-market trading volume. *Journal of Business and Economic Statistics* **18,** 410–427.

[27] Ray, B. and Tsay, R. (2000). Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics* **18,** 254–262.

[28] Robinson, P. (1994). Semiparametric analysis of long memory time series. *Annals of Statistics* **2**2, 515–539.

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

[30] Robinson, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* **24,** 1630–1661.

[31] Robinson, P. M. (2001). The memory of stochastic volatility models. *Journal of Econometrics* **101,** 195–218.

[32] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinckley, D.V., Barndorff-Nielsen, O.E. (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*. Chapman & Hall, New York.

[33] Solo, V. Intrinsic random functions and the paradox of 1/f noise. *SIAM J. Appl. Math.* **52** (1992), 270–291.

[34] Velasco, C. Gaussian semiparametric estimation of non-stationary time series *Journal of Time Series Analysis* **20** (1999), 87–127.

[35] Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* **20,** 87–127.

[36] Wright, J. H. (2002). Log-periodogram estimation of long memory volatility dependencies with conditionally heavy tailed returns. *Econometric Reviews* **21,** 397–417.

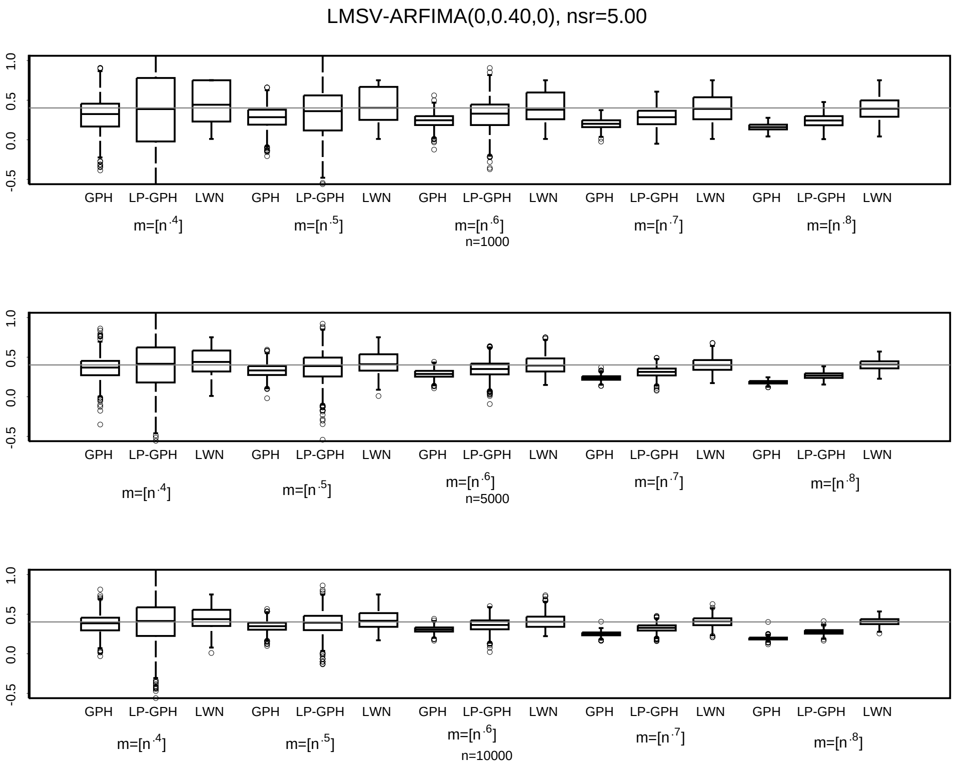

Figure 1: Box-plots of $\hat{d}_{GPH}$, $\hat{d}_{LP-GPH}$, and $\hat{d}_{LWN}$ for the LMSV-ARFIMA$(0, 0.4, 0)$ model with $nsr = 5$. Estimates were obtained using $m = [n^{\alpha}]$ Fourier frequencies, where $\alpha = 0.4, 0.5, 0.6, 0.7, 0.8$. The solid line indicates the true value of $d = 0.4$.

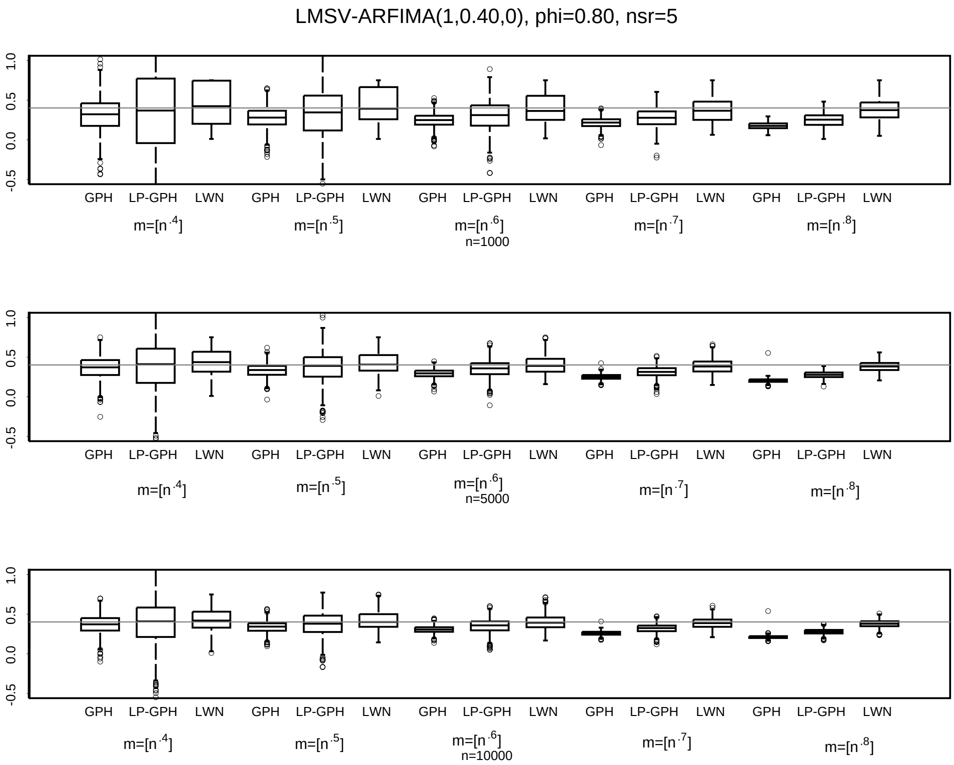

Figure 2: Box-plots of $\hat{d}_{GPH}$, $\hat{d}_{LP-GPH}$, and $\hat{d}_{LWN}$ for the LMSV-ARFIMA$(1, 0.4, 0)$ model with $\phi = 0.8$ and $nsr = 5$. Estimates were obtained using $m = [n^{\alpha}]$ Fourier frequencies, where $\alpha = 0.4, 0.5, 0.6, 0.7, 0.8$. The solid line indicates the true value of $d = 0.4$.

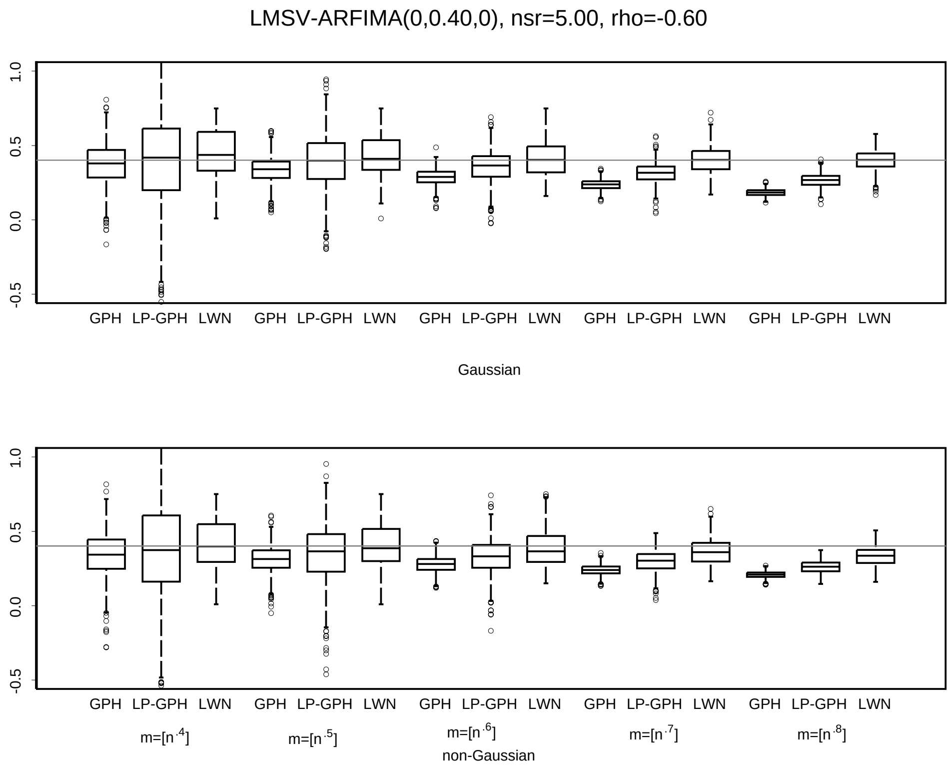

Figure 3: Box-plots of $\hat{d}_{GPH}$, $\hat{d}_{LP-GPH}$, and $\hat{d}_{LWN}$ for the LMSV-ARFIMA$(0, 0.4, 0)$ model with $nsr = 5$, $n = 5000$ having Gaussian or non-Gaussian correlated errors ϵ_t and e_t. Estimates were obtained using $m = [n^{\alpha}]$ Fourier frequencies, where $\alpha = 0.4, 0.5, 0.6, 0.7, 0.8$. The solid line indicates the true value of $d = 0.4$.

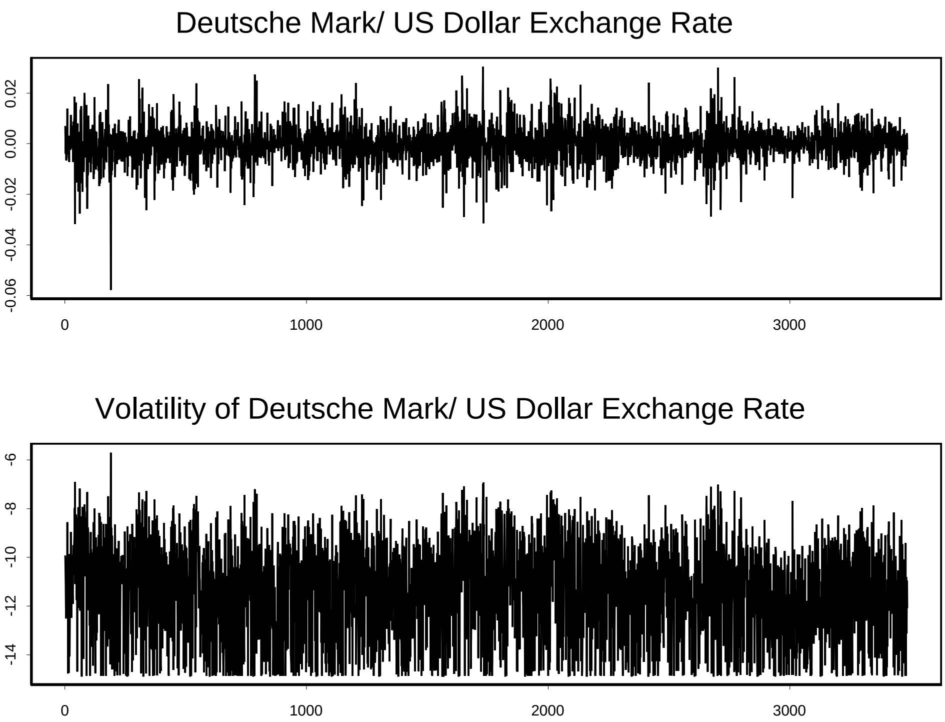

Figure 4: Top plot: Deutsche Mark/ US Dollar exchange rate from Jan 2, 1985 to May 12, 1998. Bottom plot: Volatility series for Deutsche Mark/ US Dollar exchange rate constructed using adjusted log squared returns.

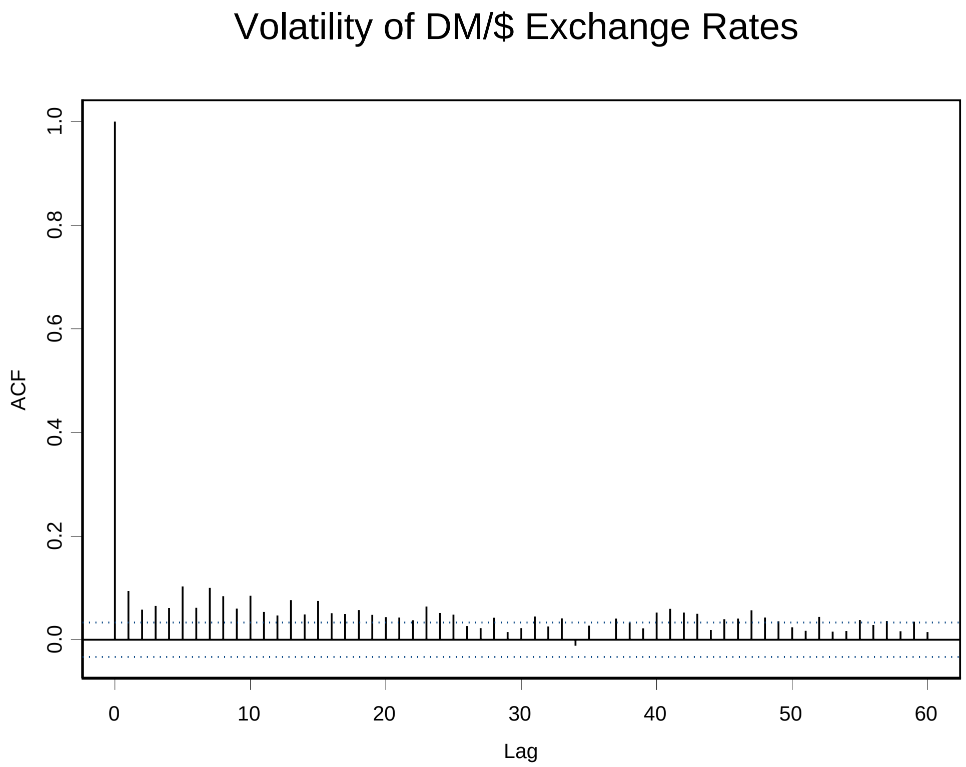

Figure 5: Sample ACF for volatility series of Deutsche Mark/ US Dollar exchange rates