On the detection and estimation of long memory in stochastic volatility [1]

F. Jay Breidt
Iowa State University
Nuno Crato
Stevens Institute of Technology
Pedro de Lima
The Johns Hopkins University

Abstract

Recent studies have suggested that stock markets' volatility has a type of long-range dependence that is not appropriately described by the usual Generalized Autoregressive Conditional Heteroskedastic (GARCH) and Exponential GARCH (EGARCH) models. In this paper, different models for describing this long-range dependence are examined and the properties of a Long-Memory Stochastic Volatility (LMSV) model, constructed by incorporating an Autoregressive Fractionally Integrated Moving Average (ARFIMA) process in a stochastic volatility scheme, are discussed. Strongly consistent estimators for the parameters of this LMSV model are obtained by maximizing the spectral likelihood. The distribution of the estimators is analyzed by means of a Monte Carlo study. The LMSV is applied to daily stock market returns providing an improved description of the volatility behavior. In order to assess the empirical relevance of this approach, tests for long-memory volatility are described and applied to an extensive set of stock market series, presenting substantial evidence for the existence of long-memory volatility.

Keywords

EGARCH, Fractional ARIMA, GARCH, LMSV, long-memory tests, persistence, spectral likelihood estimators, stochastic variance.

[1]The authors had the privilege of presenting this paper at the conference in honor of Professor Carl Christ held at the Johns Hopkins University in April 1995. We are thankful to Patrick Asea and participants at the conference, as well as to Francis Diebold, Thomas Epps, and Roger Pinkham for helpful comments and suggestions which led to improvements of this paper. Any remaining insufficiencies are our own.

1 Introduction

The recent empirical investigation of conditional variance models has suggested that financial markets' volatility may display a type of long-range persistence commonly called long memory. This type of persistence, as we will show, cannot be appropriately modeled by the now classic autoregressive conditional heteroskedastic (ARCH) models and their common variations, such as the generalized ARCH (GARCH) and the exponential GARCH (EGARCH). The same limitation applies to the emerging class of stochastic volatility (SV) models in their standard formulation. See, e.g., Bollerslev, Chou and Kroner (1992) for a review of ARCH and GARCH-type models, and Taylor (1994) for a recent review of SV models.

The evidence for the existence of some type of persistence in financial data can be traced back to the unit-root findings in GARCH and EGARCH estimated models. In fact, many GARCH and EGARCH applications involving high-frequency financial data have indicated the presence of an approximate unit root in the univariate representation for the volatility. For the standard GARCH(1,1) model, it is often found that the sum of the coefficients in the conditional variance equation is close to one. The case where this sum is actually one is known as the Integrated GARCH (IGARCH) model. For the EGARCH model, it is often found that the autoregressive polynomial in the logs of the volatility has a root close to the unit circle. The case where a unit root actually exists is known as the Integrated EGARCH (IEGARCH) model.

These findings show that financial markets' volatility displays persistent features, but since both GARCH and the EGARCH are short-memory models, the only way they have to reproduce persistence is by approximating a unit root. Direct evidence that this persistence in stock markets' volatility can be characterized as long memory has been presented in recent research, as we now describe briefly.

Ding, Granger and Engle (1993) discussed the decay of the autocorrelations of fractional moments of returns series. Using the Standard & Poor's 500 daily closing price index from January 3, 1928 to August 30, 1991, they computed the returns and constructed series of fractional moments. They found very slowly decaying autocorrelations for these series and proposed a new model on the fractional moments, the Asymmetric Power-ARCH. However, this model is also finitely parameterized as an ARMA and does not reproduce long-memory persistent features.

de Lima and Crato (1993) applied some long-memory tests to the squared residuals of filtered U.S. stock returns indexes and rejected the null of short memory for the high-frequency (daily) series.

Bollerslev and Mikkelsen (1996) found slowly decaying autocorrelations for the absolute returns of the Standard & Poor's 500 index. They proposed Fractionally Integrated GARCH (FIGARCH) models as a generalization that nests GARCH and

IGARCH and is the analogue for conditional variances of long-memory fractionally integrated models for conditional means. In a similar fashion, they discussed Fractionally Integrated EGARCH (FIEGARCH)

In light of these recent findings, this paper discusses the modeling, estimation, and testing for long memory in the context of volatility models. The paper has two main contributions.

First, we describe, characterize, and apply tests for the existence of long memory in the volatilities. In this context, the power and size of these tests is assessed for simulated series typical of financial applications.

Second, we describe, characterize and apply a feasible frequency domain likelihood estimator for the parameters in a Long Memory Stochastic Volatility (LMSV) model. We suggest and study the LMSV model, as in an earlier version of this paper (Breidt, Crato and de Lima 1994). This work was directly motivated by the empirical results of de Lima and Crato (1993). The LMSV model is constructed by incorporating an Autoregressive Fractionally Integrated Moving Average (ARFIMA) process in a stochastic volatility scheme. Harvey (1993) independently proposed a stochastic volatility model driven by fractional noise and applied it to exchange rate series, obtaining smoothed estimates of the underlying volatilities.

The LMSV model has significant advantages. It is is well-defined in the mean square sense and so many of its stochastic features are easy to establish. It has well-known counterparts in models for level series and so it inherits most of their statistical properties.

The plan for the rest of this paper is as follows. In Section 2 we discuss fractional GARCH and EGARCH models and describe the Long-Memory Stochastic Volatility model. In Section 3 we present further empirical evidence on the relevance of this model. We test the short memory null for proxies of the conditional variances in an extensive set of U.S. stock return indexes. In Section 4 we discuss a Whittle-type estimator for the LMSV model parameters obtained by maximizing the spectral approximation to the Gaussian likelihood. We present finite-sample simulation evidence about the properties of the estimators and, as an example, we study the daily returns for the value-weighted CRSP market index. In Section 5 we conclude. Proofs are found in the appendices.

2 Models for persistent volatility

In this paper, we follow Brockwell and Davis (1991) by saying that a weakly stationary process has *short memory* when its autocorrelation function (ACF), say $\rho(h)$, is geometrically bounded

$$|\rho(h)| \leq Cr^{|h|} \text{ for some } C > 0,\ 0 < r < 1.$$

In contrast to a short-memory process with a geometrically decaying ACF, a weakly stationary process has *long memory* if its ACF $\rho(\cdot)$ has a hyperbolic decay,

$$\rho(h) \sim Ch^{2d-1} \quad \text{as} \quad h \to \infty,$$

where $C \neq 0$ and $d < 0.5$ (e.g., Brockwell and Davis 1991, section 13.2). Alternatively, we can say the process has long memory if its spectrum $f(\lambda)$ has the asymptotic decay

$$f(\lambda) \sim C|\lambda|^{-2d} \quad \text{as} \quad \lambda \to 0 \quad \text{with} \quad d \neq 0. \tag{1}$$

If, in addition, $d > 0$, then the autocorrelations are not absolutely summable, $\sum |\rho(h)| = \infty$, and the spectrum diverges at zero, $f(\lambda) \uparrow \infty$ as $\lambda \to 0$. In this case we will say that the process is *persistent*. For a discussion of alternative long-memory characterizations see sections 2.2 and 2.3 of Baillie (1996).

2.1 GARCH and EGARCH models

Following Engle (1982), Bollerslev (1986), and Nelson (1991), let the prediction error y_t satisfy

$$y_t = \sigma_t \xi_t,$$

where $\{\xi_t\}$ is independent and identically distributed (iid) with mean zero and variance one, and σ_t^2 is the variance of y_t given information at time $t-1$. Among the most successful specifications for the conditional variance σ_t^2 are the GARCH and EGARCH models. A GARCH specification is given by

$$\sigma_t^2 = \omega + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 + \sum_{j=1}^{p} a_j y_{t-j}^2, \tag{2}$$

where $\omega > 0$. Constraints on $\{b_j\}$ and $\{a_j\}$ are discussed below. More compactly, we can write equation (2) as

$$b(B)\sigma_t^2 = \omega + a(B) y_t^2.$$

where B is the backshift operator ($B^j v_t = v_{t-j}$, $j = 0, \pm 1, \pm 2, \ldots$), $b(z) = 1 - b_1 z - \cdots - b_q z^q$ and $a(z) = a_1 z + \cdots + a_p z^p$.

As an alternative to the GARCH specification, Nelson (1991) proposed the Exponential GARCH (EGARCH) model

$$\log \sigma_t^2 = \mu_t + \sum_{j=0}^{\infty} \psi_j g(\xi_{t-j-1}), \quad \psi_0 \equiv 1, \tag{3}$$

where no restriction is needed for the signs of the coefficients. The function $g(\cdot)$ may be chosen to allow for asymmetric changes, depending on the sign of ξ_t.

It is known (Bollerslev 1986) that the GARCH model can also be written in an ARMA($\max\{p,q\}, q$) form, with the process $\{y_t^2\}$ being driven by the noise $\nu_t = y_t^2 - \sigma_t^2$. From this representation it is clear that the autocorrelation function for $\{y_t^2\}$ has a short-memory geometric decay.

The EGARCH models have the general representation in (3), but they are also usually parameterized with weights $\{\psi_j\}$ corresponding to an ARMA(p,q). Thus, the the usual EGARCH specification can be written

$$\phi(B)\left(\log\sigma_t^2 - \mu_t\right) = \theta(B)g(\xi_{t-j-1}),$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$ for $|z| \leq 1$ is an autoregressive polynomial, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ is a moving average polynomial and $\theta(z)$ has no roots in common with $\phi(z)$.

The empirical evidence previously discussed points in the direction of long memory, both in the squared process $\{y_t^2\}$ and in the process of log squares $\{\log y_t^2\}$. This contrasts with the usual short-memory formulations of GARCH and EGARCH models. We will look for the formulation of models with persistent properties.

2.2 Long-memory GARCH

In order to accommodate the findings of long memory, a sensible approach is to generalize GARCH models by using fractional differences, along lines earlier suggested by Robinson (1991, p. 82). The fractional differencing operator is defined through the expansion

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} B^j,$$

from which Baillie, Bollerslev, and Mikkelsen (1996) formulated the fractionally integrated GARCH (FIGARCH) model

$$(1-B)^d b(B)(\sigma_t^2 - \mu) = a(B)(y_t^2 - \mu).$$

Baillie, Bollerslev, and Mikkelsen (1996) suggested quasi maximum-likelihood estimation methods for this model.

In order to have a well-defined process, the parameters $\{a_j\}$, $\{b_j\}$, and d are constrained so that the coefficients ψ_j in the representation

$$\sigma_t^2 = \mu + (1-B)^{-d} a(B) b^{-1}(B) y_t^2 = \mu + \sum_{j=0}^{\infty} \psi_j \left(y_{t-1-j}^2 - \mu\right), \qquad (4)$$

are all nonnegative, as otherwise we will have $\sigma_t^2 < 0$ with positive probability. This implies that the parameters $\{a_j\}$ and $\{b_j\}$ are constrained as in the standard GARCH models. This also implies that the parameter d is constrained to be positive, and so $\sum_{j=0}^{\infty} \psi_j = \infty$. But then the sum of all coefficients is bigger than one, hence $\{y_t\}$ is not covariance stationary as it follows from a now standard result in Bollerslev (1986). Consequently, the autocovariance function (ACVF) of the process $\{y_t^2\}$ is not defined,[2] the series $\sum_{j=0}^{\infty} \psi_j(y_{t-1-j}^2 - \mu)$ is not defined in L^2, and the use of spectral and time-domain autocorrelation methods is not justifiable in a standard way. In addition, initializing the quasi-likelihood, which is usually done with unconditional moments of out-of-sample σ_t^2, can be problematical, although Baillie, Bollerslev, and Mikkelsen (1996) reported good results for the quasi maximum-likelihood estimation method.

As an alternative, persistence can be modeled in the log squares with a long-memory specification of an EGARCH model, as Nelson (1991, p. 352) noted and Bollerslev and Mikkelsen (1996) formulated explicitly.

A fractionally integrated EGARCH is a model of the form

$$\log \sigma_t^2 = \mu_t + \theta(B)\phi(B)^{-1}(1-B)^{-d} g(\xi_{t-1}), \tag{5}$$

where $\phi(z)$ and $\theta(z)$ are defined above. This generalization of EGARCH with fractional noise gives a strictly stationary and ergodic process. The condition for the covariance stationarity of $\{\log \sigma_t^2 - \mu_t\}$ is $\sum_{j=0}^{\infty} \psi_j^2 < 1$ (Theorem 2.1 of Nelson 1991), which is met for a parameter $d < 1/2$.

EGARCH models have the convenient feature that the coefficients in the moving average expansion (5) are not restricted to be positive. However, asymptotic results about the estimators have proved extremely hard to obtain, even when $d = 0$.

2.3 LMSV

A different approach, based on stochastic volatility (SV) models like those discussed by Melino and Turnbull (1990) and Harvey, Ruiz and Shephard (1994), has many appealing features as we now discuss.

The Stochastic Volatility model is defined by

$$y_t = \sigma_t \xi_t, \quad \sigma_t = \sigma \exp(v_t/2), \tag{6}$$

where $\{v_t\}$ is independent of $\{\xi_t\}$, $\{\xi_t\}$ is independent and identically distributed (iid) with mean zero and variance one, and $\{v_t\}$ is an ARMA model.

[2] However, this model displays the important property of having a bounded cumulative impulse-response function for any $d < 1$ as Baillie, Bollerslev, and Mikkelsen (1996) were able to show.

The Long Memory Stochastic Volatility (LMSV) model we now introduce is defined by (6) with $\{v_t\}$ being a stationary long-memory process.

Restricting attention to Gaussian $\{v_t\}$, it follows that $\{y_t\}$ is both covariance and strictly stationary. Denote by $\gamma(\cdot)$ the ACVF of $\{v_t\}$. The covariance structure of y_t is obtained from properties of the lognormal distribution:

$$\mathrm{E}\,[y_t] = 0, \;\; \mathrm{Var}\,(y_t) = \exp\{\gamma(0)/2\}\sigma^2, \;\; \text{and} \;\; \mathrm{Cov}\,(y_t, y_{t+h}) = 0 \text{ for } h \neq 0,$$

so that $\{y_t\}$ is a white noise sequence. In fact, $\{y_t\}$ is a martingale difference, a property inherited from $\{\xi_t\}$. An appealing property of this model in terms of its empirical relevance is the excess kurtosis displayed by y_t, which is

$$\frac{\mathrm{E}\,[y_t^4]}{\mathrm{E}\,[y_t^2]^2} - 3 = 3\,(\exp\{\gamma(0)\} - 1)$$

whenever the driving noise $\{\xi_t\}$ is Gaussian.

The process $\{y_t^2\}$ is also both covariance and strictly stationary. Moments of y_t^2 are again obtained from properties of the lognormal distribution:

$$\begin{aligned}
\mathrm{E}\left[y_t^2\right] &= \exp\{\gamma(0)/2\}\sigma^2, \\
\mathrm{Var}\left(y_t^2\right) &= \sigma^4\left[\left\{1 + \mathrm{Var}\left(\xi_t^2\right)\right\}\exp\{2\gamma(0)\} - \exp\{\gamma(0)\}\right], \\
\text{and } \mathrm{Cov}\left(y_t^2, y_{t+h}^2\right) &= \sigma^4\left[\exp\{\gamma(0) + \gamma(h)\} - \exp\{\gamma(0)\}\right], \text{ for } h \neq 0.
\end{aligned}$$

The series is simple to analyze after transforming to the stationary process

$$\begin{aligned}
x_t &= \log y_t^2 \\
&= \log\sigma^2 + \mathrm{E}\left[\log\xi_t^2\right] + v_t + \left(\log\xi_t^2 - \mathrm{E}\left[\log\xi_t^2\right]\right) \\
&= \mu + v_t + \epsilon_t,
\end{aligned}$$

where $\{\epsilon_t\}$ is iid with mean zero and variance σ_ϵ^2. For example, if ξ_t is standard normal, then $\log\xi_t^2$ is distributed as the log of a χ_1^2 random variable, $\mathrm{E}\,[\log\xi_t^2] = -1.27$ and $\sigma_\epsilon^2 = \pi^2/2$ (Wishart 1947).

The process $\{x_t\}$ is thus a long-memory Gaussian signal plus an iid non-Gaussian noise, with $\mathrm{E}\,[x_t] = \mu$ and

$$\gamma_x(h) = \mathrm{Cov}\,(x_t, x_{t+h}) = \gamma(h) + \sigma_\epsilon^2 I_{\{h=0\}}, \tag{7}$$

where $I_{\{h=0\}}$ is one if $h = 0$ and zero otherwise. It turns out that the ACVF of the process $\{\log y_t^2\}$ is the same as that of a fractionally integrated EGARCH model whenever $\delta_2 = 0$ (see Appendix 2).

A simple long-memory model for $\{v_t\}$ is the fractionally integrated Gaussian noise defined as the unique stationary solution of the difference equations

$$(1-B)^d v_t = \eta_t, \quad \{\eta_t\} \text{ iid } \mathrm{N}(0, \sigma_\eta^2), \tag{8}$$

where $d \in (-0.5, 0.5)$. The spectral density, ACVF, and ACF of $\{v_t\}$, denoted by $f(\cdot)$, $\gamma(\cdot)$, and $\rho(\cdot)$, respectively, are given by

$$\begin{aligned} f(\lambda) &= \frac{\sigma_\eta^2}{2\pi}\left|1-e^{-i\lambda}\right|^{-2d}, \quad -\pi \le \lambda \le \pi, \\ \gamma(0) &= \sigma_\eta^2 \Gamma(1-2d)/\Gamma^2(1-d), \\ \rho(h) &= \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} \quad h = 1, 2, \ldots \end{aligned}$$

(e.g., Brockwell and Davis 1991, p. 522).

More generally, $\{v_t\}$ can be modeled as an ARFIMA(p,d,q), defined as the unique stationary solution of the difference equations

$$(1-B)^d \phi(B) v_t = \theta(B)\eta_t, \quad \{\eta_t\} \text{ iid } \mathrm{N}(0, \sigma_\eta^2). \tag{9}$$

The spectral density of $\{x_t\}$, denoted by $f_\beta(\cdot)$, is then given by

$$f_\beta(\lambda) = \frac{\sigma_\eta^2\left|\theta\left(e^{-i\lambda}\right)\right|^2}{2\pi\left|1-e^{-i\lambda}\right|^{2d}\left|\phi\left(e^{-i\lambda}\right)\right|^2} + \frac{\sigma_\epsilon^2}{2\pi}, \quad -\pi \le \lambda \le \pi, \tag{10}$$

where $\beta = (d, \sigma_\eta^2, \sigma_\epsilon^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$.

3 Evidence of long memory in volatility

In order to check for the appropriateness of a long-memory component in the stochastic volatility model, we formally test for the existence of long memory in the volatilities of stock markets' series. This will be achieved by analyzing two traditional volatility proxies, namely the squared series and the logarithm of the squared series.

3.1 Testing for long memory in volatility

There are several ways of testing for long memory, ranging from fully parametric to nonparametric approaches. The present paper uses both a semiparametric and a nonparametric test.

The first test is constructed by regressing the logarithm of the periodogram at low frequencies on a function of the frequencies; the expected slope is dependent on the long-memory parameter d, as can be seen from equation (1). This method was introduced by Geweke and Porter-Hudak (1983) and developed by Robinson (1993).

Geweke and Porter-Hudak suggested the use of only the first ordinates of the periodogram, up to m_U, say, and argued that the resulting regression estimator for d could capture the long-memory behavior without being "contaminated" by the short-memory behavior of the process. Robinson suggested an additional truncation of the very first ordinates, up to m_L, say, in order to avoid biases. However, no clear rule exists about the choice of either m_U or m_L, and we therefore adhere to the common practice of experimenting with a few different values. To test the null hypothesis of short memory against long-memory alternatives, we perform the usual t-test for the hypothesis that $d = 0$ against $d \neq 0$. The standard deviation is obtained from the output of the regression.

It should be stressed that we will apply this regression as a test of short memory without assuming any particular form of long-memory alternatives. The asymptotics in equation (1) define long-memory processes.

The second statistic used in this paper is the *normalized rescaled range*, the R/S statistic (see, e.g., Beran 1994). The *adjusted range* R is defined as

$$R(n) = \max_{1 \leq k \leq n} \left\{ \sum_{i=1}^{k} X_i - k\bar{X} \right\} - \min_{1 \leq k \leq n} \left\{ \sum_{i=1}^{k} X_i - k\bar{X} \right\},$$

where $\bar{X}$ represents the sample mean. The *normalization factor* S can be defined as the square root of a consistent estimator for the variance, given by

$$S^2(n,q) = \sum_{j=-q}^{q} w_q(j)\hat{\gamma}(j),$$

with $\hat{\gamma}(j)$ representing the usual estimators for the autocovariances. The weights $w_q(j)$ we used are those from the Bartlett window. The R/S statistic is

$$Q(n,q) = \frac{R(n)}{S(n,q)},$$

and when $q = 0$ we have the classical R/S statistic of Hurst. The so-called Hurst exponent J is estimated as

$$\hat{J}(n,q) = \frac{\log Q(n,q)}{\log n}.$$

If short memory only is present, then $\hat{J}(n,q)$ converges to 1/2. If persistent long memory is present, then $\hat{J}(n,q)$ converges to a value larger than 1/2 (see, e.g., Mandelbrot and Taqqu 1979).

If a process satisfies a set of regularity conditions, including the existence of moments of order $4+\delta$, with $\delta > 0$, Lo (1991) shows that under the short memory null the statistic $V = n^{-1/2}Q(n,q)$ converges weakly to the range of the Brownian bridge on the unit interval. The distribution function for this range, say F_V, is

$$F_V(v) = \sum_{k=-\infty}^{\infty} (1 - 4v^2k^2)e^{-2v^2k^2}.$$

If a short-memory process does not have finite second-order moments then the classical Hurst estimate $\hat{J}(n,0)$ still converges to $1/2$, as discussed in Mandelbrot and Taqqu (1979). Therefore, the estimate $\hat{J}(n,0)$ can still provide an indication of long memory. However, no distribution theory is available in this case.

In the absence of clear rules for the choice of q, we experimented with a few values. First, we used $q = 0$, corresponding to the classical estimate. Second, we used $q = q^*$, chosen by Andrews' (1991) data-dependent formula as in Lo (1991, p. 1302). Finally, we tried $q = 200$ in an attempt to yield statistics which are more robust against short-memory effects.

3.2 Finite sample performance of the long-memory tests

In this section we consider the finite sample performance of the spectral regression test and the R/S analysis under both short and long memory. The generated processes were long memory ($d \neq 0$) and short memory ($d = 0$) stochastic volatility models, defined in equations (6) and (9) above. Here, we focus on detecting long memory in the log-squared observations; results for the squared observations are qualitatively similar. An analogous Monte Carlo study for long memory in the levels series is reported by Cheung (1993).

In designing this simulation experiment, we chose as a short memory benchmark the first-order autoregressive stochastic volatility (ARSV) model given by equations (6) and (9) with $p = 1$, $d = 0$ and $q = 0$. This model has been studied extensively; see Jacquier, Polson, and Rossi (1994) and the references therein. We chose four ARSV parameter settings from Table 4 of Jacquier, Polson, and Rossi (1994). To make the LMSV results comparable, we chose for each ARSV model an ARFIMA$(0,d,0)$ LMSV model which matched the ARSV parameterizations in two ways: first, the ratio

$$\mathrm{Var}(\sigma_t^2)/\mathrm{E}^2[\sigma_t^2]$$

is the same for both models (implying that the excess kurtosis of y_t is the same for both models), and second, the lag-one autocorrelation of v_t is the same for both models. Under these parameterizations, the job of distinguishing long and short memory is quite challenging. Finally, we considered an ARFIMA$(1, d, 0)$ LMSV model similar to the one fitted to the value-weighted CRSP data in Section 4.3. All processes were simulated with ξ_t and η_t Gaussian.

Simulation means and standard deviations over 1000 simulated realizations of each model are given for the spectral regression test in Table 1. Also tabled is the proportion of rejections of the short-memory null hypothesis $d = 0$, using the standard t-test with nominal significance level 0.05. Some conclusions from the results reported in Table 1 are:

- Under the short-memory null, the size of the test is not far from nominal if the upper truncation is taken to be less than $[n^{0.5}]$; $[n^{0.45}]$ seems to be an all-around good choice. For this sample size, at least, larger upper truncations have little value: they distort the size under short memory and bias the point estimates under all models considered.

- The spectral regression test has good power against all the long memory models considered. Power is lower for the third and fourth LMSV models, since these models have a weaker long memory signal (i.e., a smaller value of $\mathrm{Var}(\sigma_t^2)/\mathrm{E}^2[\sigma_t^2]$) than the first and second LMSV models.

- The point estimates of d under long memory have large negative biases, which increase with m_U, reflecting the contamination by short-memory effects. Even with this downward bias, estimates of d under long memory are clearly different from those under short memory.

——INSERT Table 1 AROUND HERE——

Simulation means and standard deviations for the R/S analysis over 1000 simulated realizations of each model are also given in Table 1. Some general conclusions from the results in Table 1 are:

- Long and short memory are quite distinguishable.

- The classical Hurst exponent is substantially larger than 1/2 under the short memory models we have considered.

- Andrews' data-dependent formula for choice of q goes a long way toward reducing the bias of the classical Hurst exponent, though $\hat{J}(n, q^*)$ is still above 1/2 on average.

- The Hurst exponents estimated with values of $q = 200$ provide some robustness against even highly-correlated short memory.

The overall conclusions from these tables are that the spectral regression tests and the R/S analyses can be useful indicators of long memory in stochastic volatility, but as with any asymptotic tests, they should be interpreted with caution. We recommend that additional diagnostics, in particular the shape of the estimated autocorrelation function, be used to help assess the usefulness of LMSV in any particular application.

3.3 Empirical evidence

The tests for long memory were performed over several market indexes' daily returns. The data and the designations used in the tables are as follows.

From the Center for Research in Security Prices (CRSP) tapes we used series starting on the first trading day of July 1962 and ending on the last trading day of July 1989. We computed returns for both the equally-weighted and the value-weighted data, here denoted ECRSP and VCRSP, respectively.

Using the same raw data, we also constructed the excess returns series based on the monthly Treasury bill returns. We followed the usual simplification of assuming the riskless returns were constant within each month and subtracted these latter returns from the ones in the stock market indexes.

We have also used the long series constructed by Schwert (1990), complemented with the more recent CRSP value-weighted index. This series, here denoted SCHWERT, spans from the first trading day of February 1885 to the last trading day of 1990.

In each case, in order to whiten the series of returns, we followed the usual practice of first removing any apparent correlation in the data, namely the day-of-the-week and the month-of-the-year effects, by applying standard filters.

For each of the series we applied the long-memory tests over the squared returns and the logarithms of the squared returns.

In the first three columns of Table 2 we show the results of the spectral regression tests. We immediately note that in almost all cases and all series the tests are highly significant, even when the high-frequency cut-off is severe ($u = 0.40$). Interestingly, the memory of the volatilities is reduced when the excess returns are computed. In the case of the equally-weighted index the tests are less significant. In some cases, they do not reject the null of sole existence of short memory in the volatilities. We should note, however, that the equally-weighted indexes are economically much less sensible as representatives of the overall financial markets' activity than the value-weighted ones. Finally, it is interesting to note that the log squared series reveal the existence of a much more significant long-memory component.

——INSERT Table 2 AROUND HERE——

In the last three columns of Table 2 we show the estimates $\hat{J}(n,q)$ and the p-values for the statistic $V = n^{-1/2}Q(n,q)$ for all the series described. All estimates point in the direction of persistent long memory. The J estimates computed with Andrews' (1991) data-dependent formula are highly significant for all series but the squared excess returns of the equally-weighted index. When the number of lags q increases, the significance of all statistics is reduced, as it is natural to expect in persistent processes. Even so, most of the computations show J estimates significantly larger than 1/2.

These tests can be questioned on the grounds that long data sets may display nonstationarity in the variances and that we may be detecting nonstationarity instead of long memory. In particular, the evidence points in this direction for the Schwert long indexes, since some d estimates are larger than 1/2. Diebold (1986) and Lamoureux and Lastrapes (1990), among others, have interpreted the findings of persistence in volatility as the outcome of shifts in the unconditional variances.

We complemented the results with tests for shorter series, starting in the beginning of 1978 and ending in September 1987. These shorter series avoid the crashes of 1976 and 1987 and display a period known for its rather stable volatility.

The same tests, reproduced in the second part of Table 2, still reveal long memory in the conditional variances, with the exception of the already noted equally-weighted index. This fact is quite significant and suggests that long-memory models provide an alternative to nonstationarity for volatility modeling.

4 Estimation for LMSV

The exact likelihood of the parameter vector β given $(y_1, \ldots, y_n)$ involves an n-dimensional integral and is very difficult to evaluate. Jacquier, Polson, and Rossi (1994) have developed a Markov chain simulation methodology for likelihood-based inference in an autoregressive stochastic volatility model (ARSV). Their algorithm, a cyclic independence Metropolis chain, requires specification of prior distributions on all parameters and relies heavily on the special Markovian structure of pure autoregressive processes. Other simulation-based estimation methods for the first-order ARSV exist, see, e.g., Danielsson (1994), but it is not clear whether they apply to more general stochastic volatility models. All these methods are very computationally intensive; simpler estimation strategies will be considered here since the LMSV model is much more complicated than ARSV.

Other methods for estimation from SV models have been proposed. A method of moments (MM) estimator, which avoids the problem of evaluating the likelihood function, was suggested by Taylor (1986) and Melino and Turnbull (1990). While easy to implement, MM estimators for parameters in the ARSV model have a number of disadvantages. The MM method seems relatively ineficient when some kind of persistence in the autocorrelations is present, as it is the case of nearly non-stationary AR models (see Jacquier, Polson, and Rossi, 1994, and Anderson and Sorenson, 1994, for a discussion). Moreover, the choice of appropriate moments can be problematic.

Though the process $\{x_t\}$ is non-Gaussian, a reasonable estimation procedure is to maximize the quasi-likelihood, or likelihood computed as if $\{x_t\}$ was Gaussian with ACVF $\gamma_x(h)$. See Nelson (1988) and Harvey, Ruiz, and Shephard (1994) for discussion of QML in the context of short-memory stochastic volatility models. In the context of ARFIMA models, exact computation of the quasi-likelihood is possible (e.g., Sowell 1992) but it presents convergence problems, especially for long time series, and it is extremely slow. A version of this method is conceivable for LMSV models. However, the computational problems are likely to be amplified.

We suggest a spectral-domain estimator. This is a computationally simple method for which we provide an asymptotic characterization.

4.1 The spectral likelihood estimator

A simple alternative to maximizing the time-domain Gaussian likelihood is to maximize its frequency-domain representation, as discussed in a long-memory context by Fox and Taqqu (1986), Dahlhaus (1989) and Giraitis and Surgailis (1990). The simulation results of Cheung and Diebold (1994) suggest that spectral likelihood estimators have efficiency comparable to exact QML estimators when the process has an unknown mean. The following result gives the strong consistency of estimators obtained by minimizing minus the logarithm of the spectral likelihood function,

$$\mathcal{L}_n(\beta) = 2\pi n^{-1} \sum_{k=1}^{[n/2]} \left\{ \log f_\beta(\omega_k) + \frac{I_n(\omega_k)}{f_\beta(\omega_k)} \right\}, \tag{11}$$

where $[\cdot]$ denotes the integer part, $\omega_k = 2\pi k n^{-1}$ is the kth Fourier frequency, and

$$I_n(\omega_k) = \frac{1}{2\pi n} \left(\sum_{t=1}^{n} x_t \cos \omega_k t \right)^2 + \frac{1}{2\pi n} \left(\sum_{t=1}^{n} x_t \sin \omega_k t \right)^2$$

is the kth normalized periodogram ordinate. For a general justification of the method see, e.g., Beran (1994, chapter 6).

Theorem 1 *Assume that the parameter vector*

$$\beta = (d, \sigma_\eta^2, \sigma_\epsilon^2, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$$

is an element of the compact parameter space Θ and assume that $f_{\beta_1}(\omega) \equiv f_{\beta_2}(\omega)$ for all ω in $[0, \pi]$ implies that $\beta_1 = \beta_2$, where $f_\beta(\cdot)$ is defined in (10). Let $\hat{\beta}_n$ minimize (11) over Θ and let β_0 denote the vector of true parameter values. Then $\hat{\beta}_n \to \beta_0$ almost surely.

The proof is given in Appendix A1.

Remarks: 1. The proof follows Dahlhaus (1989) in avoiding the special parameterization of Fox and Taqqu (1986). Dahlhaus' (1989) result is not directly applicable to our case because his explicit assumptions include Gaussianity and his objective function is an integral version of (11). For the non-Gaussian case, we verify Dahlhaus' remark (p. 1753) that his results extend to the function (11).
2. The component $|1 - e^{-i\lambda}|^{-2d} = (\sqrt{2 - 2\cos\lambda})^{-2d}$ of $f_\beta(\lambda)$ introduces in the likelihood a term proportional to

$$d \sum_{k=1}^{[n/2]} \log(2 - 2\cos\omega_k) 2\pi n^{-1}; \tag{12}$$

the corresponding integral is improper, but converges to zero (see Appendix A1). In the course of the proof, we show that the effect on the estimators of dropping the term (12) is negligible.
3. The identifiability condition in Theorem 1 is met if σ_ϵ^2 is known from an assumed distribution for ξ_t; for example, $\xi_t \sim$N$(0, 1)$ implies $\sigma_\epsilon^2 = \pi^2/2$. If σ_ϵ^2 is not known, the model is identifiable only if the ARFIMA component is not white noise; that is, if $\phi_p \neq 0$ for some p, $\theta_q \neq 0$ for some q, or $d \neq 0$.

4.2 Finite sample properties of the spectral likelihood estimator

This subsection presents a simulation study of the finite sample properties of the maximum likelihood spectral estimator previously proposed. In this experiment we consider two different sample sizes (n = 1024 and n = 4096) and three classes of LMSV models given, respectively, by ARFIMA(0,d,0), ARFIMA(1,d,0), and ARFIMA(1,d,1). Within each class of models, several combinations for the parameters of these models are considered—see Table 3. The variance of the iid innovations in the ARFIMA component was set to one and so was the variance of the noise component.

——INSERT Table 3 AROUND HERE——

All the results reported in this section are obtained from 1000 realizations of each model. Table 3 presents simulation means and standard deviations for the parameter estimates. Figure 1 presents boxplots for some of the models considered in the simulation. The cases considered in this figure are representative of the overall results.

——INSERT Figure 1 AROUND HERE——

Some general conclusions from the table and the boxplots are:

- Maximum likelihood estimation in the spectral domain performed well for relatively large samples, such as those found in the high frequency financial markets' data.
- The biases were relatively small and decreased uniformly from $n = 1024$ to $n = 4096$. The increase in the sample size also reduced significantly the dispersion of the results.
- The boxplots in Figure 1 also show that some less positive aspects of the results obtained for $n = 1024$ tended to be smoothed out for $n = 4096$, namely, the asymmetry of the distribution of the estimates. An extreme case was the LMSV model with an ARFIMA(0,-0.4,0) component.
- The maximum likelihood spectral estimator provided less biased and more precise parameter estimates in processes in which the fractional parameter d was positive. This includes both the estimate of d and the estimates of the other parameters in the model.
- The performance of the maximum likelihood spectral estimator in small samples might be less than ideal as illustrated by the boxplots of the smaller sample size. Moreover, some very large outliers occurred when $n = 1024$.
- The procedure had some difficulties in estimating the moving average term, even when the number of observations was 4096 (although the magnitude of the problem decreased for the larger sample size).

Given the overall good performance of the estimator when $n = 4096$, these sampling experiments indicate that maximum likelihood spectral estimation of LMSV models may be a very effective method for the type of financial applications that have suggested this line of research.

Moreover, this maximum likelihood estimator is easy to implement. Convergence for a LMSV model with an ARFIMA(0,d,0) component and $n = 4096$ was typically attained in less than 20 iterations and less than 4 seconds of CPU time on a Pentium 100Mhz. [3]

[3] In these simulations we used the set of routines for maximum likelihood estimation provided by the GAUSS programming language. The algorithm used is the derivative-based procedure of Broyden, Fletcher, Goldfarb, and Shanno, as described in the GAUSS manual. Analytical derivatives were provided. The code is available from the authors upon request.

4.3 Modeling volatility of stock returns

Nelson (1991) introduced the EGARCH model using as an example the daily returns for the value-weighted market index from the CRSP tapes for July 1962–December 1987. He selected an ARMA(2, 1) model for $\log \sigma_t^2$ and found the largest estimated AR root to be 0.99962, suggesting substantial persistence.

For comparison, we fitted a model with long-memory stochastic volatility to the log squares of the VCRSP series described in Section 3.3 above. This log squared series, denoted $\{x_t\}$, consists of $n = 6801$ mean-corrected observations modeled as

$$x_t = v_t + \epsilon_t$$

where $\{\epsilon_t\}$ is iid $(0, \sigma_\epsilon^2)$ independent of $\{v_t\}$ and $\{v_t\}$ is the ARFIMA$(1, d, 0)$,

$$(1 - B)^d (1 - \phi B) v_t = \eta_t,$$

with $\{\eta_t\}$ iid $(0, \sigma_\eta^2)$.

The spectral likelihood for the x_t's was formed as in equation (11) replacing(12) with zero, which we have found useful though this needs further investigation. The resulting likelihood was maximized with respect to the parameters σ_η^2, d, ϕ and σ_ϵ^2 yielding the estimates $\hat{\sigma}_\eta^2 = 0.00318$, $\hat{d} = 0.444$, $\hat{\phi} = 0.932$ and $\hat{\sigma}_\epsilon^2 = 5.238$.

——INSERT Figure 2 AROUND HERE——

Figure 2 shows the empirical and fitted autocorrelations for the series $\{x_t\}$. The empirical autocorrelations show a slow decay, remaining non-negligible for hundreds of lags. The ACF of the fitted LMSV model was derived from the ARFIMA(1,d,0) formulae in Hosking (1981) and adjusted for the bias due to the existence of long memory as in Theorem 5 of Hosking (1993). In Figure 2, the bias-adjusted ACF for LMSV accurately reflects the slow decay of the empirical ACF.

We also fitted short-memory GARCH and EGARCH models as well as an IGARCH model to the same VCRSP series. In order to compare the properties of fitted GARCH, IGARCH, EGARCH, and LMSV models with the observations, we computed the autocorrelations of the fitted models and plotted them against the sample autocorrelations of the series. The order of the models was selected by SIC.

The fitted GARCH and IGARCH were very similar, as often happens in practice: SIC selected a GARCH(1,2) and an IGARCH(1,2). The fitted GARCH had parameter estimates $\hat{a}_1 = 0.923$, $\hat{b}_1 = 0.143$, and $\hat{b}_2 = -0.067$. Their sum is 0.999. The fitted IGARCH had parameter estimates $\hat{a} = 0.923$, $\hat{b}_1 = 0.145$, and $\hat{b}_2 = -0.068$. Using the GARCH parameter estimates, we simulated 1,000 GARCH realizations, each of length $n = 6801$, and computed the sample ACF of the log squares for each realization. The same simulations were done for IGARCH. The average of the

GARCH sample ACF's, plotted in Figure 2 and labeled "GARCH/IGARCH", is almost indistinguishable from the average of the IGARCH sample ACF's (not plotted). The GARCH models, nearly integrated or integrated, seem "too persistent" to model these data.

The SIC criterion selected an EGARCH(2,0). The fitted EGARCH had parameter estimates $\hat{\delta}_1 = 0.0185$, $\hat{\delta}_2 = 0.200$, $\hat{\phi}_1 = 0.577$ and $\hat{\phi}_2 = 0.359$. The ACF for the log squares corresponding to the fitted EGARCH model was obtained theoretically through the formulae derived in Appendix A2. The short-memory EGARCH model clearly fails to reflect the slow decay of the empirical ACF.

5 Conclusions

Empirical evidence suggests that the recent interest in long-memory conditional variance models for stock market indexes is well-founded; both a nonparametric and semiparametric test found evidence of long memory in variance proxies for many series. A simulation exercise shows these tests are able to distinguish long from short memory in the volatilities.

The Long Memory Stochastic Volatility (LMSV) model is an analytically tractable model of this persistence in the conditional variances. The LMSV is easily fitted and analyzed using standard tools for weakly stationary processes. In particular, the LMSV model is built from the widely-used ARFIMA class of long-memory time series models, so many of its properties are well-understood.

The spectral likelihood estimator proposed for this model is strongly consistent and finite-sample simulation results show it has reasonable properties for series of the length usually found in financial data.

An example with a long series of stock prices shows that short-memory models are unable to reproduce more than the short-term structure of the autocorrelations. In contrast, a parsimonious LMSV model fit to the data is able to reproduce closely the empirical autocorrelation structure of the conditional volatilities.

These results are encouraging and suggest some lines of future research. We believe it will be interesting to investigate further the empirical relevance of the LMSV model, namely on its relevance to estimating and forecasting the volatilities and to pricing derivatives. We also believe it will be useful to compare properties of the LMSV with other models of persistence in the volatilities.

Appendices

A1. Proof of strong consistency for spectral likelihood estimators.

Let $\hat{\beta}_n$ minimize (11) and let

$$\mathcal{L}(\beta) = 2\int_0^{\pi} \left\{ \log f_\beta(\omega) + \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)} \right\} d\omega,$$

where β_0 denotes the vector of true parameter values. Then

$$\begin{aligned} |\mathcal{L}_n(\beta) - \mathcal{L}(\beta)| &\leq \left| 2\pi n^{-1} \sum_{k=1}^{[n/2]} g_\beta(\omega_k) - 2\int_0^{\pi} g_\beta(\omega)\, d\omega \right| \\ &\quad + \left| 2\pi n^{-1} \sum_{k=1}^{[n/2]} d \log(2 - 2\cos\omega_k) - 2\int_0^{\pi} d\log(2 - 2\cos\omega)\, d\omega \right| \\ &\quad + \left| 2\pi n^{-1} \sum_{k=1}^{[n/2]} \frac{I_n(\omega_k)}{f_\beta(\omega_k)} - 2\int_0^{\pi} \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)}\, d\omega \right| \\ &= M_{1n}(\beta) + M_{2n}(\beta) + M_{3n}(\beta), \end{aligned}$$

where

$$g_\beta(\lambda) = \log\left\{ \frac{\sigma_\eta^2 \left|\theta\left(e^{-i\lambda}\right)\right|^2 + \sigma_\epsilon^2 \left|\phi\left(e^{-i\lambda}\right)\right|^2 \left|1 - e^{-i\lambda}\right|^{2d}}{2\pi \left|\phi\left(e^{-i\lambda}\right)\right|^2} \right\}.$$

Now $M_{1n}(\beta)$ converges to zero uniformly in β by Riemann integrability of $g_\beta(\omega)$, continuity in β of the integral and compactness of Θ. Given $\delta > 0$, $M_{1n}(\beta)$ can be bounded above and below by the upper Riemann sum plus δ and the lower Riemann sum minus δ for a partition $\mathcal{P}_n$ of $[0, \pi]$, where $\mathcal{P}_n$ contains the nth-order Fourier frequencies and $\mathcal{P}_n \subset \mathcal{P}_{n+1} \subset \cdots$. For each β, these bounds converge to zero$\pm\delta$ monotonically, and so uniform convergence in β follows by Dini's theorem.

Next, $M_{2n}(\beta)$ can be bounded uniformly in β by

$$0.5 \left| 2\pi n^{-1} \sum_{k=1}^{[n/2]} \log(2 - 2\cos\omega_k) - 2\int_0^{\pi} \log(2 - 2\cos\omega)\, d\omega \right|,$$

which converges to zero since

$$\int_0^{\pi} \log(2 - 2\cos\omega)\, d\omega = 0.$$

Finally, $M_{3n}(\beta)$ can be shown to converge almost surely (a.s.) to zero uniformly in β by modifying Lemma 1 of Hannan (1973) (see also Lemma 1 of Fox and Taqqu (1986) and Dahlhaus (1989)). First, $1/f_\beta(\omega)$ satisfies the continuity condition of Hannan (1973) and so the Césaro sum of its Fourier series converges uniformly in (ω, β) for $\beta \in \Theta$. Second, the process $\{x_t\}$ is ergodic since $\{v_t\}$ is a linear process with iid innovations and square-summable coefficients (e.g., Hannan 1970, p. 204) and $\{\epsilon_t\}$ is iid, independent of $\{v_t\}$. From these two facts, Lemma 1 of Hannan (1973) follows.

Hence,

$$\sup_{\beta\in\Theta} |\mathcal{L}_n(\beta) - \mathcal{L}(\beta)| \to 0 \text{ a.s.}$$

Since $-\log x \geq 1 - x$, with equality holding if and only if $x = 1$,

$$\begin{aligned}
\mathcal{L}(\beta) &= 2\int_0^\pi \left\{-\log \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)} + \log f_{\beta_0}(\omega) + \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)}\right\} d\omega \\
&\geq 2\int_0^\pi \left\{1 - \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)} + \log f_{\beta_0}(\omega) + \frac{f_{\beta_0}(\omega)}{f_\beta(\omega)}\right\} d\omega \\
&= 2\int_0^\pi \left\{\log f_{\beta_0}(\omega) + \frac{f_{\beta_0}(\omega)}{f_{\beta_0}(\omega)}\right\} d\omega \\
&= \mathcal{L}(\beta_0),
\end{aligned}$$

and so (using the identifiability condition) β_0 uniquely minimizes $\mathcal{L}(\beta)$. Thus

$$\mathcal{L}_n\left(\hat{\beta}_n\right) \leq \mathcal{L}_n(\beta_0) \text{ and } \mathcal{L}(\beta_0) \leq \mathcal{L}\left(\hat{\beta}_n\right),$$

which implies that $\mathcal{L}\left(\hat{\beta}_n\right) \to \mathcal{L}(\beta_0)$ a.s. and therefore also $\hat{\beta}_n \to \beta_0$ a.s. by compactness of Θ. $\Box$.

A2. Autocovariance function of log squares under EGARCH.

Under an EGARCH model for $\{y_t\}$, the ACVF for the series $\{x_t\} = \{\log y_t^2 - \mu_t\}$, where $\{\mu_t\}$ are the deterministic volatility changes in equation (3), can be computed as follows

$$\begin{aligned}
\mathrm{Cov}(x_t, x_{t+h}) &= \mathrm{Cov}\left(\sum_{j=0}^\infty \psi_j g(\xi_{t-j-1}) + \log \xi_t^2, \sum_{j=0}^\infty \psi_j g(\xi_{t+h-j-1}) + \log \xi_{t+h}^2\right) \\
&= \mathrm{Var}\{g(\xi_t)\}\,\gamma(h) + \psi_{h-1}\mathrm{E}\left[g(\xi_t)\log\xi_t^2\right] + \mathrm{Var}\left(\log\xi_t^2\right) I_{\{h=0\}},
\end{aligned}$$

where $\gamma(h)$ is the autocovariance function

$$\gamma(h) = \sum_{j=0}^{\infty} \psi_j \psi_{j+h}$$

and $\psi_{-1} := 0$. If, as it was originally suggested by Nelson (1991), the function $g(\cdot)$ is chosen to be

$$g(\xi_t) = \delta_1 \xi_t + \delta_2 \left(|\xi_t| - \mathrm{E}|\xi_t| \right),$$

then we have

$$\mathrm{Var}\left\{ g(\xi_t) \right\} = \delta_1^2 + \delta_2^2 \left(1 - \mathrm{E}^2 |\xi_t| \right).$$

For Gaussian ξ_t, $\mathrm{E}|\xi_t| = \sqrt{2/\pi}$, $\mathrm{Var}\left(\log \xi_t^2\right) = \pi^2/2$ and

$$\mathrm{E}\left[g(\xi_t) \log \xi_t^2 \right] = \frac{2\delta_2}{\sqrt{2\pi}} \left(\log 2 - \kappa + 1.27 \right),$$

where $\kappa \simeq 0.577216$ is Euler's constant. Thus, if $\delta_2 = 0$, the Gaussian EGARCH ACVF has the same form as the SV ACVF in (7).

References

Andrews, D.W.Q. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Anderson, T.G. and Sørenson, B.E. (1994). GMM and QML asymptotic standard deviations in stochastic volatility models: A response to Ruiz. Working Paper 189, Dept. of Finance, Kellog School, Northwestern University.

Baillie, R.T. (1996). Long-memory processes and fractional integration in econometrics. *Journal of Econometrics*, forthcoming.

Baillie, R.T., Bollerslev, T., and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, forthcoming.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992). ARCH modeling in finance, *Journal of Econometrics* **52**, 5–59.

Bollerslev, T. and Mikkelsen, H.O.A. (1996). Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, forthcoming.

Breidt, F.J., de Lima, P., and Crato, N. (1994). Modeling long memory stochastic volatility. Working Papers in Economics No. 323, Johns Hopkins University.

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

Cheung, Y.W. (1993). Tests for fractional integration: a Monte Carlo investigation. *J. Time Series Analysis* **14**, 331–345.

Cheung, Y.W. and Diebold, F.X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *J. Econometrics* **62**, 301–316.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Annals of Statistics* **17**, 1749–1766.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics* **64**, 375–400.

de Lima, P.J.F. and Crato, N. (1993). Long-range dependence in the conditional variance of stock returns. August 1993 Joint Statistical Meetings, San Francisco. *Proceedings of the Business and Economic Statistics Section.*

Diebold, F.X. (1986). On modeling the persistence of conditional variances. *Econometric Reviews* **5**, 51–56.

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

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14**, 517–532.

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

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Th. Rel. Fields* **86**, 87–104

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

Hannan, E.J. (1973). The asymptotic theory of linear time-series models. *J. Applied Probability* **10**, 130–145.

Hannan, E.J. (1970). *Multiple Time Series,* John Wiley, New York.

Harvey, A.C. (1993). Long memory in stochastic volatility. Mimeo. LSE.

Harvey, A.C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**, 247–264.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika* **68**, 165–176.

Hosking, J.R.M. (1995). Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series. Forthcoming in *J. of Econometrics*

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* **12**, 371–417.

Lo, A.W. (1991). Long-term memory in stock market prices. *Econometrica* **59**, 1279-1313.

Mandelbrot, B.B. and Taqqu, M. (1979). Robust R/S analysis of long-run serial correlation, *Proceedings of the 42nd Session of the International Statistical Institute*, International Statistical Institute.

Melino, A. and and Turnbull, S.M. (1990). Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 239–265.

Nelson, D.B. (1988). Time series behavior of stock market volatility and returns. *Unpubl. PhD diss.,* Massachusetts Institute of Technology, Econ. Dept.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**, 347–370.

Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**, 67–84.

Robinson, P.M. (1993). Log-periodogram regression of time series with long range dependence. Mimeo. London School of Economics.

Schwert, G.W. (1990). Indexes of U.S. stock prices from 1802 to 1987. *Journal of Business* **63**, 399-426.

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

Taylor, S. (1986). *Modelling Financial Time Series*, Wiley, New York.

Taylor, S. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance* **4**, 183–204.

Wishart, J. (1947). The cumulants of the z and of the logarithmic χ^2 and t distributions. *Biometrika* **34**, 170–178.

Table 1: FINITE SAMPLE PERFORMANCE OF THE SPECTRAL REGRESSION TESTS AND THE R/S ANALYSIS UNDER STOCHASTIC VOLATILITY

(Simulation means and standard deviations (in parentheses) over 1000 replications of estimated d parameters and of Hurst exponents $\hat{J}(n,q)$. Also reported is the proportion of rejections of the short-memory null hypotheses $d = 0$ and $J = 1/2$ using two-sided tests with nominal significance level 0.05. The d parameters were estimated by spectral regression using the periodogram of the log squares. Indices of the Fourier frequencies used in the regression have a lower truncation at $m_L = [n^{0.1}]$ and different upper truncations $m_U = [n^u]$ with $u = 0.45$, 0.50, and 0.55. Hurst exponents $\hat{J}(n,q)$ were estimated from log squares with $q = 0$, $q = q^*$, which is the value chosen by Andrew's data-dependent formula, and $q = 200$. Sample size is $n = 6144$.)

Model	$\hat{d}_{u=0.45}$	$\hat{d}_{u=0.50}$	$\hat{d}_{u=0.55}$	$\hat{J}(n,0)$	$\hat{J}(n,q^*)$	$\hat{J}(n,200)$
ARSV: $\phi = 0.9$, $\sigma_\eta^2 = 0.45$	0.031	0.061	0.112	0.627	0.549	0.522
	(0.118)	(0.089)	(0.070)	(0.025)	(0.025)	(0.022)
rejection proportion	0.057	0.110	0.388		0.190	0.024
LMSV: $d = 0.47$, $\sigma_\eta^2 = 0.37$	0.423	0.392	0.356	0.707	0.669	0.562
	(0.120)	(0.087)	(0.067)	(0.039)	(0.032)	(0.023)
rejection proportion	0.923	0.992	0.999		0.997	0.428
ARSV: $\phi = 0.95$, $\sigma_\eta^2 = 0.23$	0.111	0.189	0.278	0.661	0.571	0.522
	(0.115)	(0.088)	(0.069)	(0.027)	(0.026)	(0.023)
rejection proportion	0.145	0.562	0.978		0.507	0.030
LMSV: $d = 0.49$, $\sigma_\eta^2 = 0.19$	0.384	0.348	0.307	0.688	0.667	0.564
	(0.121)	(0.089)	(0.070)	(0.039)	(0.034)	(0.023)
rejection proportion	0.885	0.965	0.988		0.995	0.438
ARSV: $\phi = 0.9$, $\sigma_\eta^2 = 0.13$	0.026	0.049	0.085	0.585	0.556	0.523
	(0.119)	(0.085)	(0.067)	(0.026)	(0.025)	(0.021)
rejection proportion	0.054	0.083	0.238		0.285	0.025
LMSV: $d = 0.47$, $\sigma_\eta^2 = 0.11$	0.302	0.263	0.224	0.651	0.643	0.562
	(0.121)	(0.089)	(0.068)	(0.039)	(0.036)	(0.024)
rejection proportion	0.704	0.832	0.907		0.967	0.399
ARSV: $\phi = 0.95$, $\sigma_\eta^2 = 0.07$	0.092	0.157	0.221	0.614	0.581	0.523
	(0.117)	(0.086)	(0.067)	(0.027)	(0.026)	(0.022)
rejection proportion	0.133	0.425	0.906		0.651	0.026
LMSV: $d = 0.49$, $\sigma_\eta^2 = 0.05$	0.255	0.212	0.176	0.629	0.626	0.560
	(0.120)	(0.091)	(0.069)	(0.038)	(0.037)	(0.023)
rejection proportion	0.587	0.665	0.746		0.929	0.362
LMSV: $d = 0.44$, $\sigma_\eta^2 = 0.003$	0.459	0.455	0.442	0.717	0.677	0.560
$\phi = 0.93$	(0.121)	(0.092)	(0.068)	(0.038)	(0.030)	(0.024)
rejection proportion	0.957	0.998	1.000		0.998	0.366

Table 2: RESULTS OF THE SPECTRAL TESTS AND THE R/S ANALYSIS

(The integration parameters d are estimated with a lower truncation at $m_L = [n^{0.1}]$ and different upper truncations $m_U = [n^u]$ with $u = 0.45$, 0.50, and 0.55. Hurst exponents $\hat{J}(n,q)$ are estimated with $q = 0$, $q = q^*$, which is the value chosen by Andrew's data-dependent formula, and $q = 200$. Unilateral test p-values for d and for $V = n^{-1/2}Q(n,q)$ are displayed within parentheses.)

Series	$\hat{d}_{u=0.45}$	$\hat{d}_{u=0.50}$	$\hat{d}_{u=0.55}$	$\hat{J}(n,0)$	$\hat{J}(n,q^*)$	$\hat{J}(n,200)$
VCRSP	0.295	0.314	0.435	0.740	0.671	0.567
(Jul62-Jul89)	(0.003)	(0.000)	(0.000)		(0.000)	(0.036)
lnVCRSP	0.382	0.342	0.365	0.732	0.696	0.575
(Jul62-Jul89)	(0.002)	(0.000)	(0.000)		(0.000)	(0.015)
ECRSP	0.333	0.218	0.295	0.696	0.619	0.538
(Jul62-Jul89)	(0.024)	(0.029)	(0.000)		(0.000)	(0.232)
lnECRSP	0.263	0.186	0.248	0.703	0.667	0.565
(Jul62-Jul89)	(0.020)	(0.018)	(0.000)		(0.000)	(0.044)
ExRt-VCRSP	0.075	0.011	0.153	0.618	0.566	0.517
(Jul62-Jul89)	(0.140)	(0.006)	(0.000)		(0.039)	(0.591)
lnExRt-VCRSP	0.446	0.391	0.347	0.746	0.703	0.583
(Jul62-Jul89)	(0.001)	(0.000)	(0.000)		(0.000)	(0.005)
ExRt-ECRSP	0.032	0.063	0.101	0.588	0.519	0.490
(Jul62-Jul89)	(0.316)	(0.099)	(0.003)		(0.562)	(0.908)
lnExRt-ECRSP	0.346	0.322	0.303	0.660	0.650	0.557
(Jul62-Jul89)	(0.005)	(0.001)	(0.000)		(0.000)	(0.061)
SCHWERT	0.781	0.482	0.407	0.742	0.667	0.560
(Feb1885-Dec1990)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
lnSCHWERT	0.582	0.540	0.501	0.736	0.684	0.593
(Feb1885-Dec1990)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)
VCRSP	0.399	0.432	0.409	0.667	0.652	0.546
(Jan78-Sep87)	(0.005)	(0.000)	(0.000)		(0.000)	(0.240)
lnVCRSP	0.437	0.386	0.384	0.653	0.653	0.550
(Jan78-Sep87)	(0.003)	(0.000)	(0.000)		(0.000)	(0.197)
ECRSP	-0.019	-0.072	-0.017	0.654	0.603	0.561
(Jan78-Sep87)	(0.460)	(0.287)	(0.427)		(0.002)	(0.106)
lnECRSP	0.165	0.168	0.142	0.660	0.650	0.557
(Jan78-Sep87)	(0.164)	(0.107)	(0.073)		(0.000)	(0.131)
ExRt-VCRSP	0.399	0.435	0.410	0.667	0.652	0.546
(Jan78-Sep87)	(0.005)	(0.000)	(0.000)		(0.000)	(0.238)
lnExRt-VCRSP	0.408	0.351	0.335	0.654	0.654	0.551
(Jan78-Sep87)	(0.011)	(0.002)	(0.000)		(0.000)	(0.189)
ExRt-ECRSP	-0.017	-0.071	-0.016	0.654	0.603	0.561
(Jan78-Sep87)	(0.464)	(0.228)	(0.121)		(0.002)	(0.106)
lnExRt-ECRSP	0.244	0.240	0.190	0.661	0.651	0.555
(Jan78-Sep87)	(0.082)	(0.047)	(0.032)		(0.000)	(0.149)

Abbreviations used in the table are as follows.
VCRSP: squared returns from the filtered value-weighted CRSP series
ECRSP: squared returns from the filtered equally-weighted CRSP series
ln: logarithms of the squared return series
ExRt: series of excess returns

Table 3: FINITE-SAMPLE RESULTS FOR THE SPECTRAL LIKELIHOOD ESTIMATOR

(For each model and set of parameters, 1000 replications were performed with length $n = 1024$ and $n = 4096$. The LMSV model parameters are given within parentheses. The values in the table represent the simulation means and standard deviations (in parentheses) for the estimated parameters.)

Parameters	ϕ		d		θ	
ϕ, d, θ	n=1024	n=4096	n=1024	n=4096	n=1024	n=4096
(0,-0.4,0)			-0.550	-0.419		
			(0.550)	(0.189)		
(0,-0.2,0)			-0.337	-0.223		
			(0.521)	(0.152)		
(0,0.0,0)			-0.0678	-0.0169		
			(0.296)	(0.107)		
(0,0.2,0)			0.189	0.196		
			(0.187)	(0.042)		
(0,0.4,0)			0.407	0.401		
			(0.086)	(0.036)		
(0.8,-0.4,0)	0.756	0.781	-0.369	-0.389		
	(0.169)	(0.114)	(0.234)	(0.172)		
(0.8,-0.2,0)	0.771	0.795	-0.211	-0.215		
	(0.157)	(0.081)	(0.207)	(0.129)		
(0.8,0.0,0)	0.773	0.797	-0.0213	-0.0142		
	(0.147)	(0.063)	(0.188)	(0.101)		
(0.8,0.2,0)	0.774	0.798	0.180	0.187		
	(0.158)	(0.057)	(0.191)	(0.096)		
(0.8,0.4,0)	0.774	0.797	0.381	0.394		
	(0.147)	(0.052)	(0.196)	(0.085)		
(0.4,-0.4,0)	0.366	0.391	-0.420	-0.400		
	(0.300)	(0.228)	(0.389)	(0.247)		
(0.4,-0.2,0)	0.398	0.435	-0.242	-0.255		
	(0.293)	(0.212)	(0.271)	(0.213)		
(0.4,0.0,0)	0.434	0.427	-0.0828	-0.0423		
	(0.279)	(0.169)	(0.233)	(0.145)		
(0.4,0.2,0)	0.425	0.403	0.142	0.191		
	(0.250)	(0.121)	(0.172)	(0.059)		
(0.4,0.4,0)	0.373	0.390	0.382	0.399		
	(0.240)	(0.112)	(0.129)	(0.046)		
(0.8,0.2,0.3)	0.788	0.800	0.161	0.186	0.319	0.134
	(0.128)	(0.050)	(0.175)	(0.083)	(0.521)	(0.378)
(0.8,0.2,-0.3)	0.908	0.839	0.167	0.183	0.308	0.144
	(0.251)	(0.134)	(0.172)	(0.093)	(0.530)	(0.368)
(0.4,0.2,0.3)	0.345	0.371	0.128	0.188	0.303	0.114
	(0.410)	(0.243)	(0.204)	(0.067)	(0.525)	(0.373)
(0.4,0.2,-0.3)	0.399	0.394	0.123	0.187	0.277	0.0864
	(0.374)	(0.229)	(0.210)	(0.060)	(0.555)	(0.400)
(0.8,0.4,0.3)	0.790	0.802	0.367	0.390	0.320	0.147
	(0.113)	(0.050)	(0.177)	(0.088)	(0.548)	(0.403)
(0.8,0.4,-0.3)	0.978	0.893	0.369	0.388	0.349	0.188
	(0.316)	(0.190)	(0.172)	(0.089)	(0.557)	(0.422)
(0.4,0.4,0.3)	0.302	0.365	0.362	0.393	0.371	0.177
	(0.419)	(0.220)	(0.151)	(0.050)	(0.551)	(0.426)
(0.4,0.4,-0.3)	0.395	0.374	0.348	0.394	0.297	0.147
	(0.352)	(0.221)	(0.172)	(0.051)	(0.587)	(0.427)

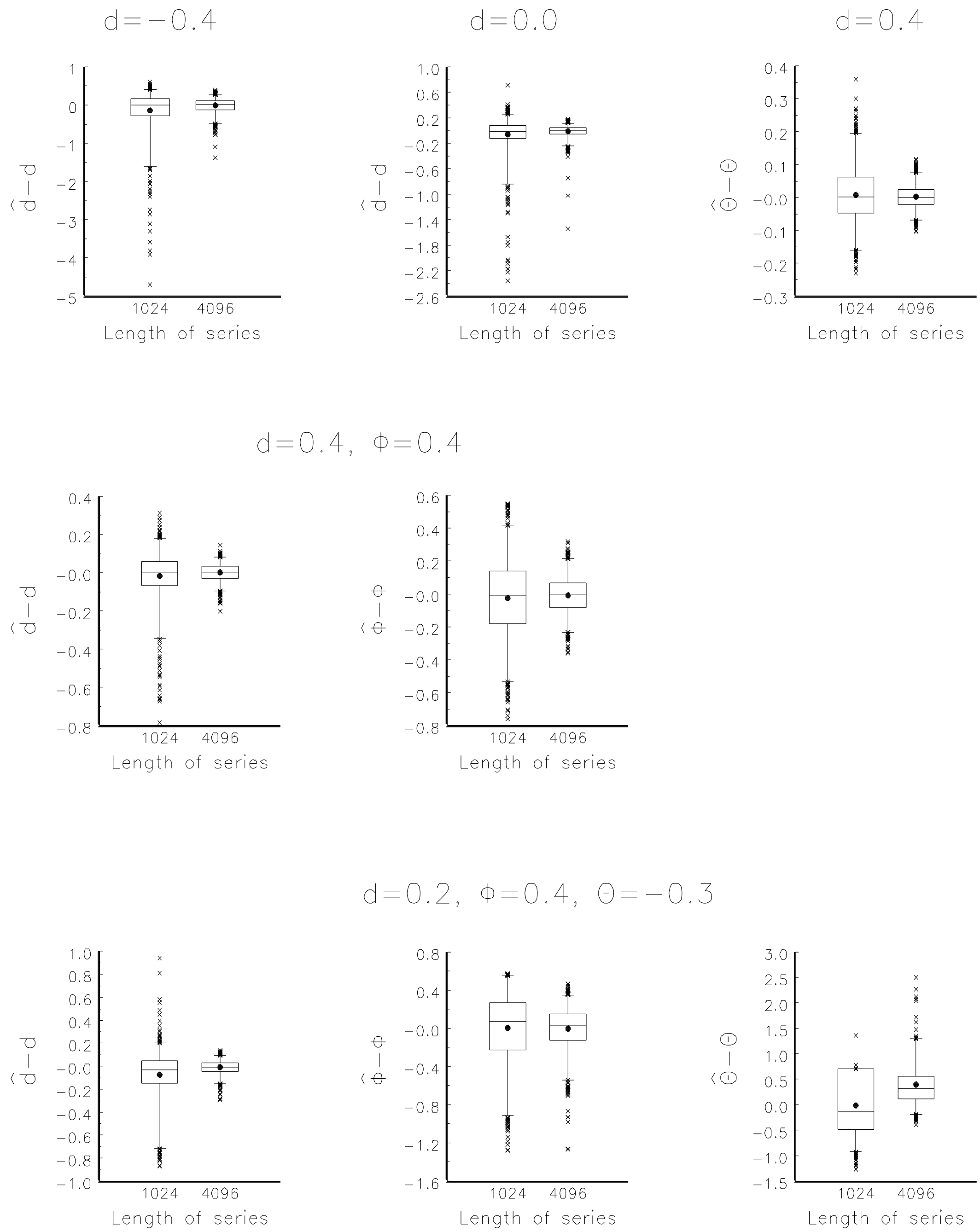

Figure 1: Boxplots represent the deviations of the estimated parameters from the true values for five different LMSV models. On the first row are three different cases of the ARFIMA(0,d,0). On the second row is an ARFIMA(1,d,0). On the third row is an ARFIMA(1,d,1). The horizontal line inside each box represents the corresponding sample median and the dot the sample average. Whiskers are drawn out to the empirical percentiles 2.5 and 97.5.

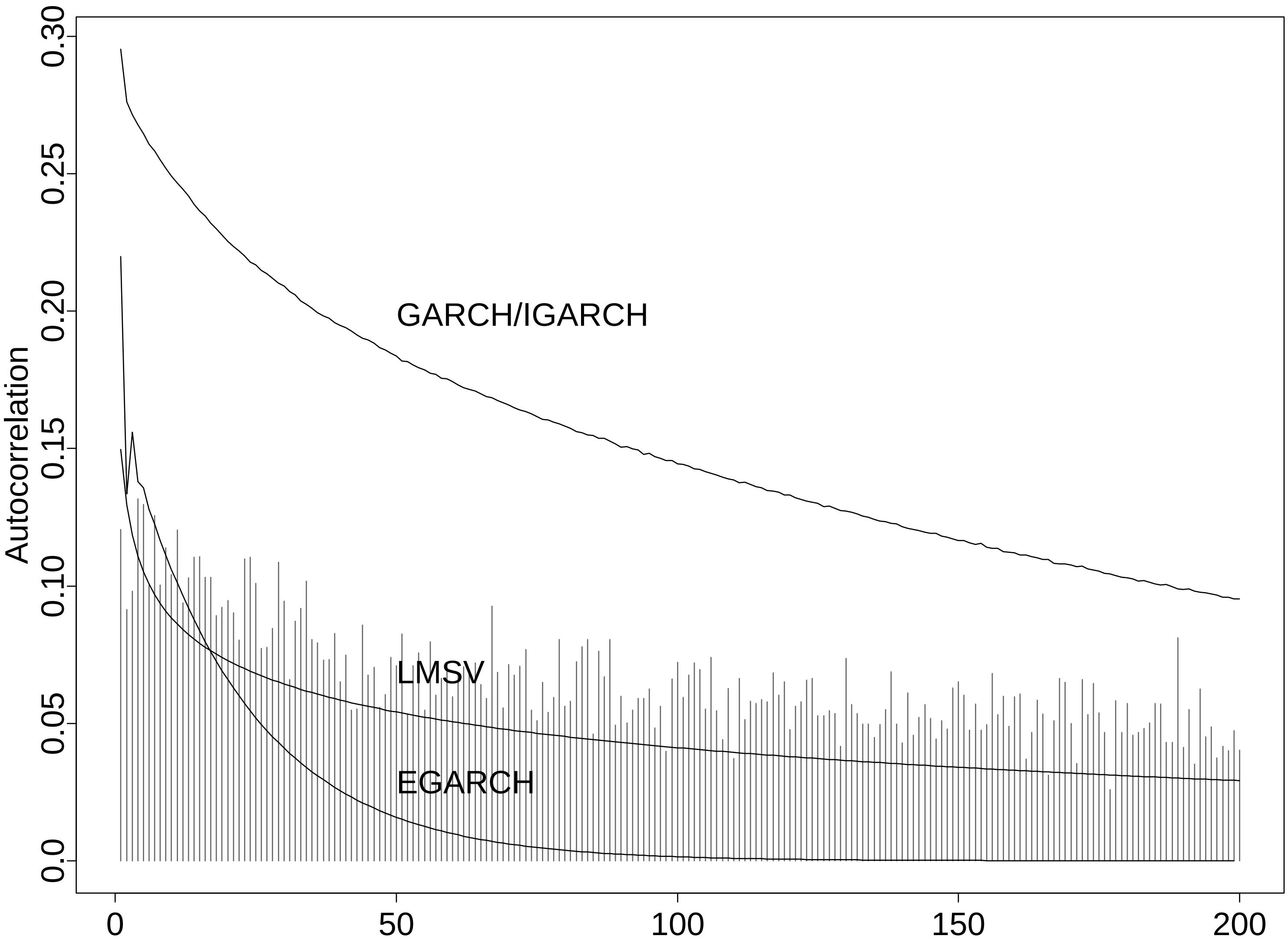

Figure 2: Empirical and fitted autocorrelation functions for the log squares VCRSP series.