

Monte Carlo Maximum Likelihood Estimation for Generalized Long-Memory Time Series Models

G. Mesters¹, S. J. Koopman^{2,3} and M. Ooms²

¹*Universitat Pompeu Fabra & Barcelona GSE, Barcelona, Spain, and Netherlands Institute for the Study of Crime and Law Enforcement*

²*Department of Econometrics, VU University Amsterdam, The Netherlands*

³*Tinbergen Institute, Amsterdam, The Netherlands, and CREATES, Aarhus University, Aarhus, Denmark*

An exact maximum likelihood method is developed for the estimation of parameters in a non-Gaussian nonlinear density function that depends on a latent Gaussian dynamic process with long-memory properties. Our method relies on the method of importance sampling and on a linear Gaussian approximating model from which the latent process can be simulated. Given the presence of a latent long-memory process, we require a modification of the importance sampling technique. In particular, the long-memory process needs to be approximated by a finite dynamic linear process. Two possible approximations are discussed and are compared with each other. We show that an autoregression obtained from minimizing mean squared prediction errors leads to an effective and feasible method. In our empirical study, we analyze ten daily log-return series from the S&P 500 stock index by univariate and multivariate long-memory stochastic volatility models. We compare the in-sample and out-of-sample performance of a number of models within the class of long-memory stochastic volatility models.

Keywords Fractional integration; Forecasting; Importance sampling; Kalman filter; Latent factors; Stochastic volatility.

JEL Classification C32.

1. INTRODUCTION

In this article, we develop a maximum likelihood estimation method for the class of generalized long-memory time series models that is proposed by Brockwell (2007). The long-memory stochastic volatility model as in Breidt et al. (1998) and Wright (1999), and the long-memory censored Gaussian model as in Brockwell and Chan

Address correspondence to Prof. Dr. S. J. Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; E-mail: s.j.koopman@vu.nl

(2006) belong to this class of models. The generalized long-memory model consists of a latent autoregressive fractionally integrated moving average (ARFIMA) process with Gaussian innovations and an arbitrary observation density that is conditional on the latent ARFIMA process. A further development in this article is the extension towards the simultaneous analysis of multiple time series which allows the treatment of generalized long-memory dynamic factor models.

The presence of long-memory in an observed time series becomes apparent when its autocovariance function decays slower than an exponential decay. The time series is then said to be subject to long-range dependence. Such time series appear in many fields including finance, meteorology, and computer science. The modeling of long-memory time series has received much interest since the seminal article of Mandelbrot (1969). Surveys on specification and parameter estimation for long-memory models are given by Robinson (1994) and Baillie (1996). A recent textbook treatment of theory and methods for long-range dependent data is given by Palma (2007). We consider the ARFIMA model with Gaussian innovations for the long-memory latent variable. This model was independently introduced by Granger, Joyeux (1980) and Hosking (1981). Beran (1994) discusses inference but also forecasting for ARFIMA models. In practice, parameter estimation for ARFIMA models is based on approximate maximum likelihood methods. Sowell (1992) has shown that exact maximum likelihood estimation is feasible via the direct calculation of the full autocovariance function and by means of prediction error decomposition or Durbin–Levinson algorithms, see Durbin (1960). Computational refinements of this method are proposed in Doornik and Ooms (2003).

The main motivation to consider the generalized class of latent long-memory models is its member, the long-memory stochastic volatility (LMSV) model, which we discuss in detail in Sections 5 and 6 below. Another example is the long-memory censored Gaussian model of Brockwell and Chan (2006). We refer to Deo et al. (2006) for more examples and discussion regarding the model class. Brockwell (2007) has developed a general Bayesian procedure based on the Markov chain Monte Carlo method for the estimation of the parameters in models of this class. We propose a maximum likelihood procedure based on importance sampling methods such as those developed by Shephard and Pitt (1997) and Durbin and Koopman (1997). The difficulties in estimation are two-fold. First, the latent Gaussian ARFIMA process is unobserved such that the likelihood function becomes an integral over all possible latent time paths. Second, the ARFIMA process cannot be written in state space form with a finite state vector. The importance sampling method evaluates the likelihood via Monte Carlo integration based on simulating latent paths from an adequate approximation of the model of interest. For this purpose, we develop a linear Gaussian state space model that approximates both the possibly non-Gaussian nonlinear features of the observations and the dynamic long-memory features of the model.

The LMSV model is designed for time series of daily returns that are subject to a stochastically time-varying volatility with long-memory dynamic properties. LMSV

models are first considered by Breidt et al. (1998) and Harvey (1998). They have adopted quasi-maximum likelihood procedures for parameter estimation; see also Wright (1999). Our key contribution is the development of an exact maximum likelihood method. Many other volatility models with long-memory features have been proposed in the context of autoregressive conditional heteroskedasticity models; see, for example, Ding et al. (1993). The volatility models involving realized volatility measures, that are computed from high-frequency data, also rely typically on long-memory dynamic processes; see, for example, Andersen et al. (2001), Barndorff-Nielsen and Shephard (2001), Corsi (2009), and Asai et al. (2012). More recently, models have appeared that combine equations for daily returns and realized volatility measures to extract long-memory volatility estimates simultaneously as in Koopman and Scharth (2013). In the empirical work for this article, we extract long-memory volatility only from daily returns which are modeled by the LMSV model. We show that exact maximum likelihood based on Monte Carlo integration methods are feasible for parameter estimation and for producing accurate volatility forecasts.

In our general framework, we can also consider vectors of time series that are subject to long-memory dynamics. When the number of long-memory processes (or factors) are limited to one or two, the methodology can still be carried out as an exact maximum likelihood estimation procedure. We will argue that the number of time series in the observation density is not relevant in this respect as the method remains exact. However, when the number of factors becomes larger, the numerical challenge becomes very high for an exact method, and we may need to resort to approximating methods. We explore the feasibility of our approach in detail.

The remainder of the article is organized as follows. In the next section, we present the generalized latent long-memory time series model. In Section 3, we describe the general procedure of importance sampling to evaluate the exact likelihood function. Our importance sampling method for models with latent long-memory time series processes is developed in Section 4. All developments are presented for the general multivariate framework. In Section 5, we show the effectiveness of our approach for univariate LMSV models. We present Monte Carlo evidence for the small-sample properties of our estimation procedure, and we provide an empirical illustration analyzing the volatility underlying the log-returns of ten constituents of the S&P 500 stock index. We study both the in-sample and the out-of-sample performance of the long memory stochastic volatility models. This illustration is extended in Section 6, where we examine a multivariate LMSV model. In Section 7, we summarize and present directions for further research.

2. MODELING FRAMEWORK

Consider a vector time series Y_t , for $t = 1, \dots, n$, that is subject to nonlinear, non-Gaussian and long-memory characteristics. A general modeling framework is provided by

the decomposition model

$$Y \sim p(Y|Z), \tag{1}$$

with $Y = (Y'_1, \dots, Y'_n)'$ and where $p(Y|Z)$ can be any density function for Y given the latent vector $Z = (Z'_1, \dots, Z'_n)'$, which is often referred to as the signal. In this article, we assume that the latent process for vector Z_t can be represented by a sum of linear Gaussian dynamic processes of which a selection can have long-memory properties. In particular, we have

$$Z_t = \mathcal{A}X_t + \mathcal{B}U_t, \tag{2}$$

where U_t is a vector of independent short-memory dynamic processes and X_t is a vector of independent long-memory processes. The matrices \mathcal{A} and \mathcal{B} have appropriate dimensions and can be regarded as fixed selection or weight matrices which may depend on an unknown coefficient vector. The dimensions of the vectors X_t and U_t can be determined for each given model. Although our proposed methodology can be used under more general conditions, for presentational purposes we assume that the short- and long-memory variables are modeled as linear dynamic processes with Gaussian innovations. In particular, we will assume that the i th element of X_t can be represented by the ARFIMA process as given by

$$\phi_i(B) (1 - B)^{d_i} X_{it} = \theta_i(B)\varepsilon_{it}, \quad \varepsilon_{it} \sim N(0, \sigma_i^2), \tag{3}$$

for given i , where B is the backshift operator for time index t with $B^m X_{it} = X_{i,t-m}$ for any integer m , the autoregressive $\phi_i(B)$ and moving average $\theta_i(B)$ are finite backshift polynomial functions, d_i is the fractional integration coefficient, and ε_{it} is a serially uncorrelated and normally distributed sequence with zero mean and variance σ_i^2 . The disturbances ε_{it} are mutually and serially uncorrelated at all time periods t and for all i . For the modeling of co-integrated vector long-memory processes, a number of representations exists; see Sela and Hurvich (2009). The evaluation of the implied covariances and determinants of these specifications typically require computationally intensive methods. By allowing a selection of elements of \mathcal{B} to depend on unknown parameters, we can allow for contemporaneous correlation between the long-memory processes. The backshift polynomials are given by

$$\phi_i(B) = 1 - \phi_{i,1}B - \dots - \phi_{i,p}B^p, \quad \theta_i(B) = 1 + \theta_{i,1}B + \dots + \theta_{i,q}B^q, \tag{4}$$

for known non-negative integer values p and q , unknown autoregressive coefficients $\phi_{i,j}$, and unknown moving average coefficients $\theta_{i,k}$ with $i = 1, \dots, p$ and $k = 1, \dots, q$, for each i , where p and q can be chosen differently for a different i . We assume that the roots of the polynomials $\phi_i(B)$ and $\theta_i(B)$ lie strictly outside the unit circle and that these polynomials

have no common roots for each i . The fractional integration part can be expressed as the binomial expansion given by

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

where the parameter d_i is a real valued constant in the range of $-1 < d_i < 0.5$. The assumptions ensure that for each i , the process X_{it} is stationary, invertible, and causal; see (Palma, 2007, Theorem 3.4) for a proof. The dynamic process (3) and its assumptions also apply to elements of U_t but with the additional assumption that each process has $d_i = 0$ in (3). We obtain the autoregressive moving average (ARMA) process

$$\phi_j^*(B) U_{jt} = \theta_j^*(B) \eta_{jt}, \quad \eta_{jt} \sim N(0, \sigma_j^2), \tag{5}$$

for given j , where $\phi_j^*(B)$ and $\theta_j^*(B)$ are defined as $\phi_j(B)$ and $\theta_j(B)$ in (4), respectively. The disturbance sequence η_{jt} is similarly defined as ε_{it} , and they are uncorrelated for all i and j . A particular element of U_t can represent a white noise sequence. It requires $\phi_j^*(B) = 1$ and $\theta_j^*(B) = 1$ in (5) to obtain $U_{jt} = \eta_{jt}$.

3. LIKELIHOOD EVALUATION AND SIGNAL EXTRACTION

When we observe a realization z of Z directly, likelihood evaluation can take place via the multiplicative representation of one-step ahead predictive densities. For the weak stationary Gaussian process Z_t , the predictive density is

$$Z_t | Z_{t-1}, \dots, Z_1 \sim N(\hat{Z}_t, V_t), \quad t = 1, \dots, n, \tag{6}$$

where

$$\hat{Z}_1 = 0, \quad \hat{Z}_t = \sum_{j=1}^{t-1} C_{t-1,j} Z_{t-j}, \tag{7}$$

with $V_t = E[(Z_t - \hat{Z}_t)(Z_t - \hat{Z}_t)' | Z_{t-1}, \dots, Z_1]$ and with the coefficient matrices $C_{i,j}$ determined by the Durbin–Levinson algorithm for a given autocovariance function of Z ; see (Durbin, 1960, Appendix 1). The density $p(Z)$ can then be expressed as

$$p(Z) = p(Z_1) \prod_{t=2}^n p(Z_t | Z_{t-1}, \dots, Z_1) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi|V_t|}} \exp[-(Z_t - \hat{Z}_t)' V_t^{-1} (Z_t - \hat{Z}_t)], \tag{8}$$

which can be computed for any realization $Z = z$. However, Z is not observed but is treated as a latent vector. We observe Y that is dependent of Z . Given the model

$Y \sim p(Y|Z)$ and $Z \sim p(Z)$, the density for Y is given by

$$p(Y) = \int p(Y, Z)dZ = \int p(Y|Z)p(Z)dZ. \tag{9}$$

For a realization y of Y , the likelihood function is defined as $\ell(\psi) = p(y)$ where ψ is a vector of fixed unknown coefficients and typically contains the ARFIMA parameters. An analytical expression for $p(Y)$ is in most cases not available because we consider $p(Y|Z)$ as nonlinear and/or non-Gaussian. Therefore, we rely on numerical methods. Given the potentially high dimension of the integral in (9), we rely on Monte Carlo methods for its evaluation. In particular, we generate M samples of Z , denoted by $\{z^{(1)}, \dots, z^{(M)}\}$, from $p(Z)$ and compute the average $M^{-1} \sum_{i=1}^M p(y|z^{(i)})$ to obtain a Monte Carlo estimate of the likelihood function. The estimate is however inefficient since most unconditional samples from $p(Z)$ will not resemble the observational process of Y . A more efficient approach is obtained by importance sampling, see Ripley (1987). It is based on an importance density $g(Z|Y)$ with the following properties: (i) $g(Z|Y) > 0$ whenever $p(Y, Z) > 0$, (ii) it is close in proportionality to $p(Y, Z)$, (iii) it is easy to sample from, and (iv) it is easy to compute. In practice, we therefore choose the importance density from the Gaussian family and adjust its mean and variance to get it close in proportionality to $p(Y, Z)$. The likelihood function is then based on

$$p(Y) = \int \frac{p(Y|Z)p(Z)}{g(Z|Y)}g(Z|Y)dZ = g(Y) \int \frac{p(Y|Z)}{g(Y|Z)}g(Z|Y)dZ, \tag{10}$$

since $p(Z) = g(Z)$ is a Gaussian density. Also, since $g(Y, Z) = g(Y|Z)p(Z)$ represents a Gaussian density, an analytic expression for $g(Y)$ is available and can be computed easily. A Monte Carlo estimate of the likelihood function is then given by

$$\hat{\ell}(\psi) = g(y)M^{-1} \sum_{i=1}^M \frac{p(y|z^{(i)})}{g(y|z^{(i)})}, \tag{11}$$

where the samples $\{z^{(1)}, \dots, z^{(M)}\}$ are drawn from the importance density $g(Z|Y)$ for the realisation $Y = y$. The quality of the Monte Carlo estimator (11) depends on how well $g(Z|Y)$ approximates $p(Y, Z)$. The choice of an appropriate importance density is taken on a case by case basis. We discuss the choice for our model in the next section.

For any choice of importance density, Kolmogorov’s strong law of large numbers implies that $\hat{\ell}(\psi) \rightarrow \ell(\psi)$ as $M \rightarrow \infty$. To guarantee a \sqrt{M} rate of convergence, we can rely on the Lindeberg–Levy central limit theorem for which a necessary condition is the existence of a variance for the importance weights $p(Y|Z) / g(Y|Z)$, for $Y = y$, see Geweke (1989). Diagnostic statistics for checking the existence of the variance of the importance weights can be based on the application of extreme value theory, see Monahan (2001) and Koopman et al. (2009). We will present a selection of these diagnostic statistics when we study our choice of importance densities for LMSV models in Section 5.

4. IMPORTANCE SAMPLING FOR LONG-MEMORY PROCESSES

Next we construct a linear Gaussian state space model for our importance density $g(Y, Z)$ which will satisfy properties (iii) and (iv). The effectiveness of the model depends on how well density $g(Z|Y)$ will approximate $p(Y, Z)$, for realisation $Y = y$. Once the importance density is established, a computationally efficient method is required to sample from $g(Z|y)$.

4.1. Basic Approximation

While keeping the long-memory properties of Z , we linearize the observation density $p(Y|Z)$ and contrast its mean and variance with those of the linear Gaussian density. In effect, we want to modify the mean and variance of the Gaussian density such that its mode is equal to the mode of the original observation density. Such a strategy is followed by Shephard and Pitt (1997) and Durbin and Koopman (2000) So (2003) and Jungbacker and Koopman (2007) argue that this strategy can be implemented by numerically maximizing $\log p(Z|Y) = \log p(Y|Z) + \log p(Z) - \log p(Y)$ with respect to Z . The instrumental basis is the linear Gaussian model

$$Y_t = c_t + Z_t + u_t, \quad u_t \sim N(0, D_t), \quad t = 1, \dots, n, \tag{12}$$

where c_t and D_t are known, the stochastic variables Z_t and u_s are mutually uncorrelated, and u_t is serially uncorrelated, for all time indices $t, s = 1, \dots, n$. It follows that

$$g(Y|Z) = \prod_{t=1}^n g(Y_t|Z_t), \quad g(Y_t|Z_t) \equiv N(c_t + Z_t, D_t). \tag{13}$$

The maximization of $\log p(Z|Y)$ with respect to Z can be carried out via the Newton–Raphson method and reduces to the following iterative procedure. At each step, we consider (12) where fixed mean vector c_t and variance matrix D_t are determined from the output of the previous iteration, for $t = 1, \dots, n$. We summarize the procedure as follows.

Algorithm A.

- (i) Choose a value z^* as a guess of Z .
- (ii) Given the set of two equations

$$\frac{\partial \log p(Y_t|Z_t)}{\partial Z_t} = \frac{\partial \log g(Y_t|Z_t)}{\partial Z_t}, \quad \frac{\partial^2 \log p(Y_t|Z_t)}{\partial Z_t \partial Z_t'} = \frac{\partial^2 \log g(Y_t|Z_t)}{\partial Z_t \partial Z_t'},$$

for $t = 1, \dots, n$, where $p(Y_t|Z_t)$ is the observation model and $g(Y_t|Z_t)$ is given by (13), we can deduct expressions for c_t and D_t as functions of Z , and compute $c_t = c_t^*$ and $D_t = D_t^*$ for $Z = z^*$.

- (iii) Compute $\hat{Z} = E_g(Z|Y)$ from the resulting model (12) with $c_t = c_t^*$ and $D_t = D_t^*$.
- (iv) Replace z^* by $z^* = \hat{Z}$.
- (v) Iterate between (ii), (iii), and (iv) until convergence.

The Algorithm A can be compared with the Gauss–Newton regression (GNR) method as described in Davidson and MacKinnon (2004) although Algorithm A is based on a second-order Taylor expansion. The computations can be carried out for any realization $Y = y$.

4.2. Long-memory Approximation

When $\mathscr{A} = 0$ in (2), the approximation method can be implemented as described in detail by Jungbacker and Koopman (2007). The short-memory process U_t is then formulated in a linear state space form and the computations in Algorithm A, specifically in step (iii), can be carried out using the Kalman filter and smoothing methods; see Anderson and Moore (1979) and Durbin and Koopman (2012, Chapter 4). However, in our general model with $\mathscr{A} \neq 0$, the long-memory process cannot be formulated in state space form with a finite state vector, see the discussion in Chan and Palma (1998). The approximating model $g(Y, Z)$ in (12) is linear and Gaussian nonetheless, and a standard lemma insists that

$$E_g(Z|Y) = E_g(Z) + Cov_g(Z, Y)Var_g(Y)^{-1}[Y - E_g(Y)].$$

When variance matrix $Var_g(Y)$ has a convenient structure, the computations can exploit the structure, and calculating $E_g(Z|Y)$ is still feasible. For example, a variance matrix with a Toeplitz structure as implied by the ARFIMA model (3) can rely on the computationally efficient Durbin–Levinson algorithm; see Sowell (1992) and Doornik and Ooms (2003). The model (12) implies however a variance matrix $Var_g(Y)$ that is equal to the sum of a Toeplitz and a block-diagonal matrix. A computationally efficient algorithm for computing $E_g(Z|Y)$ is unfortunately not available when $Var_g(Y)$ has this structure. The same arguments apply to sampling from $g(Z|Y)$ when computing (11), an appropriate algorithm is not available when Z is subject to long-memory dynamics.

We therefore need to introduce an additional approximation of formulating a short-memory dynamic model for a long-memory process. For this purpose, we propose the following two approximation methods.

4.2.1. Yule–Walker Approximation

Define the m th order backshift polynomial $\delta(B) = 1 - \delta_1 B - \dots - \delta_m B^m$. The autoregressive model of order m , the AR(m) model, is defined as (3) with $d_i = 0$, $\phi_i(B) = \delta(B)$ and $\theta_i(B) = 1$. We consider the AR(m) model with autoregressive polynomial $\delta(m)$

as an approximation to the ARFIMA model (3). The coefficients $\delta_1, \dots, \delta_m$ are set equal to the corresponding coefficients of the partial autocorrelation function (PACF) which are obtained from solving a sequence of m Yule–Walker equations. The Yule–Walker equations are based on the autocovariance function of the given model (2). The resulting coefficients have a minimum mean square prediction error property for a given finite order m . A convenient property of the PACF coefficients is their rapid convergence to zero as the backshift order increases. A threshold value for the decaying PACF coefficients can determine the order m . By a set of simulation exercises for a range of ARFIMA specifications, we have set the order fixed at $m = 10$, which appears adequate in most cases and leads to a computationally feasible method.

4.2.2. ARMA Model Approximation

The autocovariance function of the ARFIMA model (3) can also be approximated by a rational ARMA process as defined by (3) with $d_i = 0$. The approximation can be established by minimizing the mean square error directly, similar to Tiao and Tsay (1994), as follows. Both the ARFIMA and ARMA models can be represented as infinite moving averages with coefficients, say, $\theta_{\text{arfima},j}^*$ and $\theta_{\text{arma},j}^*$, respectively, for $j = 1, 2, \dots$. Given a set of ARFIMA model coefficients, we can obtain the ARMA coefficients by minimizing $\sum_{j=1}^{\infty} (\theta_{\text{arfima},j}^* - \theta_{\text{arma},j}^*)^2$. In practice, we truncate the infinite sum at 1,000. This minimization problem is nonlinear and need to be carried out for each ARFIMA specification and for each set of parameter values. A similar approach is discussed by Hsu and Breidt (2003) where they recommend an ARMA approximation based on the polynomials in (4) of orders $p = 3$ and $q = 2$.

4.3. Sampling from the Importance Density

Once the ARFIMA approximation is obtained, we can obtain the full approximating model in a similar as described by Algorithm A. This approach is summarized in Algorithm B.

Algorithm B.

- (i) Approximate the ARFIMA model for X_t in (2) using one of the methods described in Sections 4.2.1 and 4.2.2. Consider Z_t of (2) where X_t is modelled by its short-memory approximation.
- (ii) Carry out the steps of Algorithm A. Step (iii) of Algorithm A can be carried by Kalman filter and smoothing methods since we have a short-memory process Z_t .

The linear Gaussian approximating model obtained from Algorithm B is the result of two approximations: (a) the second-order Taylor expansion for treating the nonlinear

non-Gaussian observation equation, and (b) the short-memory approximation for the long-memory process X_t . This can be made explicit by having the expression in (10) changed to

$$p(Y) = g(Y) \int \frac{p(Y|Z)}{g(Y|Z)} \frac{g(Z|Y)}{g_s(Z|Y)} g_s(Z|Y) dZ = g_s(Y) \int \frac{p(Y|Z)}{g(Y|Z)} \frac{g(Z)}{g_s(Z)} g_s(Z|Y) dZ, \quad (14)$$

where $g_s(\cdot)$ refers to the approximating model (12) where the long-memory process X_t in Z_t of (2) is substituted by its short-memory approximation. The second equality in (14) follows since $g(Y|Z) = g_s(Y|Z)$. The ratio $g(Z) / g_s(Z)$ can be regarded as the error due to the short-memory approximation of the long-memory process X_t . For a given realization of Z , we can compute both $g(Z)$ and $g_s(Z)$ via the Durbin–Levinson algorithm.

For realization $Y = y$, the Monte Carlo estimate of the likelihood function is given by

$$\tilde{\ell}(\psi) = g_s(y) M^{-1} \sum_{i=1}^M \frac{p(y|z^{(i)})}{g(y|z^{(i)})} \frac{g(z^{(i)})}{g_s(z^{(i)})}, \quad (15)$$

where the samples $\{z^{(1)}, \dots, z^{(M)}\}$ are drawn from the importance density $g_s(Z|y)$ that is obtained from Algorithm B. Since $g_s(\cdot)$ is the approximating model (12) where Z_t has short-memory, we can represent Eqs. (12) and (2) as a linear Gaussian state space model. Hence, simulation from $g_s(Z|Y)$ can be based on the simulation smoother methods of Frühwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995), and Durbin and Koopman (2002). The convergence of the estimator $\tilde{\ell}(\psi) \rightarrow \ell(\psi)$ as $M \rightarrow \infty$ depends on the existence of the variance of the importance weights $w(Y, Z)$ as given by

$$w(Y, Z) = \frac{p(Y|Z)}{g(Y|Z)} \frac{g(Z)}{g_s(Z)}. \quad (16)$$

4.4. Parameter Estimation

Given a particular model specification for (1) and (2), together with a realization of the time series Y and a particular value for the parameter vector ψ , we can compute the Monte Carlo estimate of the likelihood function via (14). The method of maximum likelihood relies on the direct numerical optimization of (14) with respect to ψ . A change of the parameter vector ψ leads to a different value of likelihood function when the parameters are properly identified. The value of ψ that maximizes (14) is the Monte Carlo maximum likelihood estimate and can be found recursively. Quasi-Newton methods can be used effectively to maximize the estimate $\tilde{\ell}(\psi)$ with respect to the parameter vector ψ . In the simulation exercises and in the empirical studies below, we make use of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, see Nocedal and Wright (1999). However, the likelihood estimate (14) is subject to Monte Carlo error. A different set of random values leads to a numerically different value for the likelihood estimate. Hence, for the estimation of ψ , the same set of random values is used for each likelihood evaluation.

Elements of ψ are restricted to their stationary regions within the optimization algorithm, for example $-1 < d < 0.5$. Let $\tilde{\psi}$ denote the estimated parameters, obtained by maximizing $\tilde{\ell}(\psi)$. Standard errors for elements of estimates $\tilde{\psi}$ are computed by inverting the Hessian matrix at $\tilde{\ell}(\tilde{\psi})$ as

$$\tilde{\Sigma} = - \left\{ \frac{\partial^2 \log \tilde{\ell}(\tilde{\psi})}{\partial \psi \partial \psi} \right\}^{-1},$$

which is an asymptotic estimate of the variance matrix of $\tilde{\psi}$. It can be calculated numerically from values ψ around $\tilde{\psi}$. The standard errors of elements of ψ are given by the square root of the diagonal elements of $\tilde{\Sigma}$.

4.5. Signal Extraction

Given parameter vector ψ the location of a general function of latent vector Z , denoted by $h(Z)$ can be determined using importance sampling. For the original model, as specified by equations (1) and (2), it can be shown that

$$E_p(h(Z)|Y) = \int h(Z) \cdot p(Y, Z) dZ = \frac{\int h(Z) w(Y, Z) g_s(Z|Y) dZ}{\int w(Y, Z) g_s(Z|Y) dZ}, \tag{17}$$

where $w(Y, Z)$ is given by (16). The estimation of $E_p(h(Z)|Y)$ by importance sampling, for a realization $Y = y$, can be achieved by

$$h(\tilde{Z}) = \frac{\sum_{i=1}^M w_i h(z^{(i)})}{\sum_{i=1}^M w_i}, \tag{18}$$

where $w_i = w(Y = y, Z = z^{(i)})$ in (16) with the simulated value $z^{(i)}$ drawn from $g_s(Z|y)$.

5. UNIVARIATE LMSV

We illustrate the methods developed in the previous sections for the univariate long-memory stochastic volatility model. We provide a large Monte Carlo study to evaluate the finite-sample properties of the estimation procedure. Finally, we present an empirical study to ten daily log-return time series from constituents of the S&P 500 stock index. We evaluate both the in-sample and the out-of-sample performance of the long-memory stochastic volatility model.

5.1. Model Specification

Consider a time series of speculative asset log-returns y_t that is assumed to have constant zero mean and time-varying variance $\exp(X_t)$. The observations are sampled at daily

intervals. The general class of stochastic volatility models has a stochastically time-varying process for the log-variance X_t . Here we take X_t as a long-memory process and specifically treat the stochastic volatility for y_t as given by

$$y_t = \exp(X_t/2)\xi_t, \quad \xi_t \sim t(0, \sigma_\xi^2, \nu), \quad t = 1, \dots, n, \tag{19}$$

where X_t is modeled as the ARFIMA process (3) and $t(0, \sigma_\xi^2, \nu)$ denotes the Student's t density with mean zero, scaling σ_ξ and degrees of freedom ν . The assumption of a normal density for the disturbances ξ_t can be obtained by fixing the degrees of freedom parameter ν at some large value, say $\nu = 1,000$. The resulting models are referred to as the LMSV- t model, with Student's t disturbances, and the LMSV model, with Gaussian disturbances, which are obtained by fixing $\nu = 1,000$. The parameters of the models are collected in vectors ψ^t and ψ^g , respectively. The LMSV model is introduced by Breidt et al. (1998) and Harvey (1998). Their estimation methods for ψ are based on quasi-maximum likelihood methods. Wright (1999) also considers the LMSV model and proposes to estimate ψ via the general method of moments based on the estimated log-periodogram of Geweke and Porter-Hudak (1983). A comparison of the different estimation methods for the LMSV model is provided by Deo et al. (2006) where also an enhanced quasi-maximum likelihood method is proposed. Bayesian estimation methods for the LMSV model are considered by So (2002), Hsu and Breidt (2003), and Jensen (1983).

5.2. Simulation Design

In our simulation study we generate univariate observations y_t from the model (19) with the univariate ARFIMA process $X_t = X_{1,t}$ specified as in (3) with autoregressive polynomial of order $p = p_1 = 1$ and moving average polynomial of order $q = q_1 = 0$ in (4). We take the scale of the observations equal to unity, that is $\sigma_\xi^2 = 1$. In the Student's t case, we have $\psi^t = (d, \phi, \sigma, \nu)$, where $d = d_1$, $\phi = \phi_{1,1}$ and $\sigma^2 = \sigma_1^2$. In the Gaussian case, we obtain parameter vector $\psi^g = (d, \phi, \sigma)$ as $\nu = 1,000$ is fixed for this model.

Our Monte Carlo study is based on generating $n = 2,000$ observations from the model (19) with the log-variance X_t specified as the ARFIMA process (3). We have adopted eight different parameter vectors as given in Table 1. We have kept the scaling of the volatility σ similar at 0.2 in all sets of simulations since results are not much affected by different choices of the scaling parameter. The more interesting variations in parameter values are related to the intensity of the fractional integration parameter, $d = 0.1, 0.2, 0.3, 0.4$, in combination with either none or high stationary persistence, $\phi = 0.0$ or $\phi = 0.9$. The case where $\phi = 0.0$ corresponds to the ARFIMA model with $p = 0$ and $q = 0$, which has been used for describing log-variance X_t in Bollerslev and Jubinski (1999) and Ray and Tsay (2000). When simulating observations from the Student's t -density, we consider only $\nu = 10$ since other values for ν have shown to produce similar results.

5.3. Importance Sampling Diagnostics

To assess whether the use of importance sampling methods is effective for the Monte Carlo evaluation of the likelihood function, we discuss a set of diagnostics proposed by Koopman et al. (2009). These diagnostic statistics are based on testing the null hypothesis of the existence of a variance in a sequence of importance weights, see the discussion at the end of Section 3.

The diagnostic statistics are computed as follows. We simulate a time series y_t from model (19) of length $n = 2,000$ and with a particular value of ψ^l or ψ^g from Table 1. Next we estimate the parameters using the Monte Carlo maximum likelihood methods of Section 4.4. The parameter vector is then replaced by its resulting estimate, and we generate 100,000 importance sampling weights $w(Y, Z)$ in (16) with Y set equal to the simulated time series. For a given threshold w^{\min} , we only consider the weights that are larger than the threshold. These, say r , exceedence values x_1, \dots, x_r are assumed to come from the generalized Pareto distribution with logdensity function $f(a, b) = -\log b - (1 + a^{-1}) \log(1 + ab^{-1}x_i)$ for $i = 1, \dots, r$, where unknown parameters a and b determine the shape and scale of the density, respectively. For an appropriately chosen threshold and

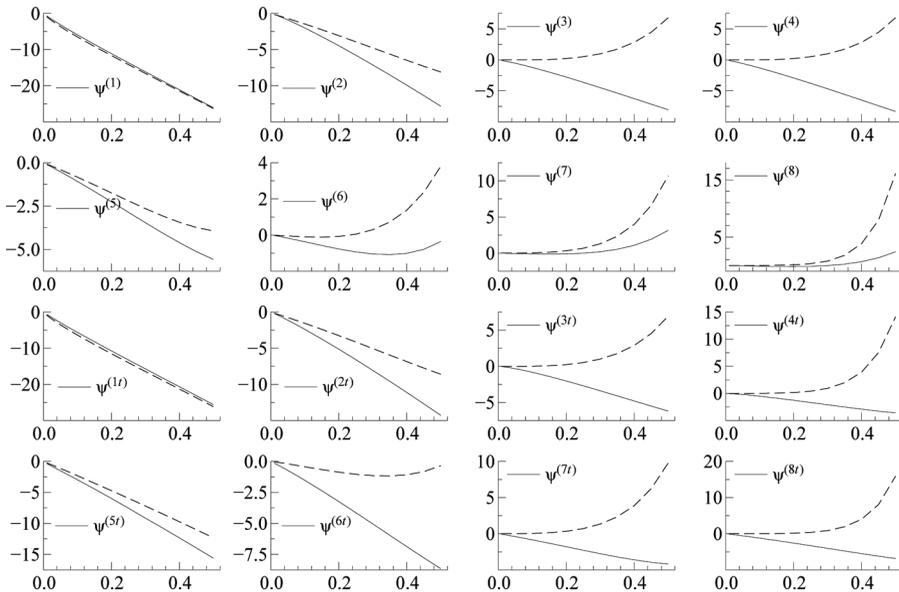


FIGURE 1 Importance sampling diagnostics for parameters $\psi^{(ig)}$ and $\psi^{(it)}$, for $i = 1, \dots, 8$, based on 100,000 simulations of weights $w(Y, Z)$ defined in (16). In each panel the solid line represents the estimated test statistics t_w based on the Yule-Walker approximation and the dashed line represents the test statistics from the ARMA approximation. The test statistics are computed for different thresholds w^{\min} , by procedures explained in Section 5.3. Thresholds are based on the number of exceedence values x_1, \dots, x_r included. We have taken $0.01 = r/100,000$, $0.025 = r/100,000$, $0.05 = r/100,000$, until $0.5 = r/100,000$.

TABLE 1
Parameter Vectors Used for Simulating Observations in the Monte Carlo Study

ψ^g	$\psi^{(1g)}$	$\psi^{(2g)}$	$\psi^{(3g)}$	$\psi^{(4g)}$	$\psi^{(5g)}$	$\psi^{(6g)}$	$\psi^{(7g)}$	$\psi^{(8g)}$
d	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
ϕ	—	—	—	—	0.9	0.9	0.9	0.9
σ	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
ψ^t	$\psi^{(1t)}$	$\psi^{(2t)}$	$\psi^{(3t)}$	$\psi^{(4t)}$	$\psi^{(5t)}$	$\psi^{(6t)}$	$\psi^{(7t)}$	$\psi^{(8t)}$
d	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
ϕ	—	—	—	—	0.9	0.9	0.9	0.9
σ	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
ν	10	10	10	10	10	10	10	10

when $a \leq 0.5$, the variance of the importance sampling weights exists. We estimate a and b by maximum likelihood, denoted by \hat{a} and \hat{b} , respectively, and compute the t -test statistic $t_w = \hat{b}^{-1} \sqrt{r/3}(\hat{a} - 0.5)$ for the null hypothesis $H_0 : a = 0.5$. As $n \rightarrow \infty$ and under the null hypothesis, the distribution of the test-statistic converges to the standard normal. We reject the null hypothesis when the statistic is positive and significantly different from zero, that is, when it is larger than 1.96 with 95% confidence.

Since the test statistics depend on the choice of the threshold w^{\min} , we compute the statistics for different threshold values. In Fig. 1, we report the test statistics based on the importance weights from the Yule–Walker and ARMA approximations, for the parameter values $\psi^{(it)}$ and $\psi^{(ig)}$ from Table 1, for $i = 1, \dots, 8$. In the ARMA approximation case, the test statistics are sufficiently large, especially when considering parameter vectors for which $d > 0.2$, to reject the null hypothesis. Many test statistics diverge exponentially to infinity as the threshold w^{\min} decreases, note that the portion of weights included than increases. We therefore regard this importance sampler as less reliable. When using the Yule–Walker approximation, the test statistics are overall smaller and in the majority of cases sufficiently small or negative. These results have been the motivation to opt for an importance density based on the Yule–Walker approximation. The Student’s t LMSV model produces overall much better statistics compared to the Gaussian model. The simulation results confirm earlier findings that the importance weights are more likely to have a variance when the serial dependence in the time series is weak, say for the combinations where $d + \phi < 1.1$. Furthermore, we present strong evidence that the importance weights have a variance for the LMSV model where the disturbances come from the Student’s t density in comparison to the Gaussian density.

5.4. Simulation Results

For the simulation study, we consider the LMSV- t model (19) with different parameter settings. We fix $\nu = 1,000$ to obtain the Gaussian LMSV model, while with $\nu > 2$ we

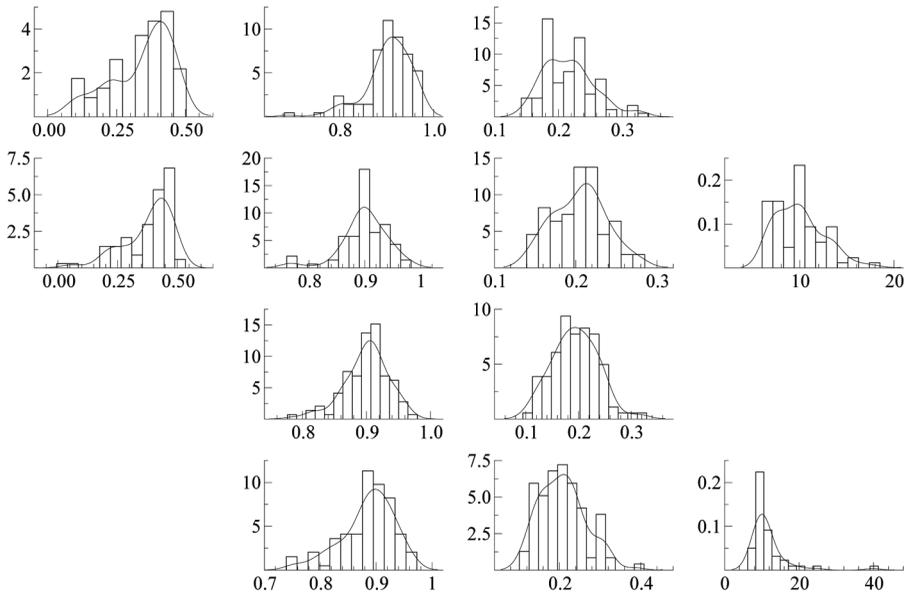


FIGURE 2 Sample densities of estimated parameters in the Monte Carlo study. We present a 4×4 matrix of density plots from a sample of 100 estimates of parameters in the LMSV- t model (19). The four columns are associated with parameters d , ϕ , σ^2 , and ν , respectively. The four rows are associated with the LMSV model with $\nu = 1,000$ (LMSV Gaussian), $\nu > 2$ (LMSV Student's t), $d = 0, \nu = 1,000$ (SV Gaussian) and $d = 0, \nu > 2$ (SV Student's t), respectively. The simulations are based on parameter vectors $\psi^{(8g)}$: $d = 0.4, \phi = 0.9, \sigma = 0.2$ and $\psi^{(8t)}$: $d = 0.4, \phi = 0.9, \sigma = 0.2, \nu = 10$. Note that ν is fixed for the Gaussian models and d is fixed for the short-memory models.

obtain the model with Student's t disturbances. For each parameter vector value from Table 1, we simulate 100 time series of length $n = 2,000$, and we estimate the parameter vector for each simulated time series, which is then treated as the observed time series. The estimation procedures are implemented as described in Section 4.4, and by using the Yule-Walker long-memory approximation with 10 lags as described in Section 4.2. In this way we obtain 100 estimates of vector ψ . For each element in ψ , we report the average estimation bias and standard deviation in Table 2. To make a comparison with the standard SV model, we also consider the settings $d = 0$ and $\nu = 1,000$ to obtain the Gaussian SV model and $d = 0$ and $\nu > 2$ to obtain the Student's t SV model. For the parameter vectors $\psi^{(8t)}$ and $\psi^{(8g)}$, we present the resulting sample histograms in Fig. 2. When the estimation method is successful, the average estimation bias should be close to zero, and the sample standard deviation should be relatively small. The results of our Monte Carlo study convincingly show that the estimation procedure is successful. All parameter estimates center around their "true" values for all different models and parameter values. The sample variation in the set of estimates appears to be smallest for the LMSV- t model.

TABLE 2

Descriptive Statistics of Estimated Parameters From the Simulation Study. For Each Parameter Vector the Average Estimation Bias and Standard Deviation (as Subscript) is Computed as Follows. We Sample 100 Time-series of Length $n = 2,000$ From the Univariate Long-memory Stochastic Volatility Model as Defined by Equations (19) and (3) for Different Parameter Vectors From Table 1. Each Simulated Time-series is Treated as Observed and Estimated Using Procedures From Section 4.4, the Yule–Walker Approximation with 10 Lags and $M = 400$ Importance Simulations. From the 100 Estimated Parameter Vectors we Present Their Average Bias, with Respect to the Corresponding Parameter Vector From Table 1, and Standard Deviation. This is Repeated for Each Parameter Vector considered

ψ	$\psi^{(1)}$	$\psi^{(2)}$	$\psi^{(3)}$	$\psi^{(4)}$
d	-0.039 _{0.153}	-0.018 _{0.111}	-0.019 _{0.098}	-0.033 _{0.107}
ϕ	—	—	—	—
σ	0.001 _{0.080}	-0.002 _{0.048}	0.008 _{0.032}	-0.011 _{0.064}
ψ	$\psi^{(5)}$	$\psi^{(6)}$	$\psi^{(7)}$	$\psi^{(8)}$
d	-0.018 _{0.125}	-0.022 _{0.145}	-0.036 _{0.109}	-0.032 _{0.106}
ϕ	-0.001 _{0.035}	-0.009 _{0.065}	0.000 _{0.045}	0.002 _{0.047}
σ	0.031 _{0.052}	0.024 _{0.051}	0.011 _{0.043}	0.017 _{0.042}
ψ	$\psi^{(1r)}$	$\psi^{(2r)}$	$\psi^{(3r)}$	$\psi^{(4r)}$
d	-0.031 _{0.123}	-0.004 _{0.135}	-0.016 _{0.111}	-0.012 _{0.101}
ϕ	—	—	—	—
σ	0.021 _{0.061}	0.008 _{0.072}	0.031 _{0.034}	0.001 _{0.059}
v	0.461 _{3.169}	0.215 _{4.126}	0.345 _{3.059}	0.826 _{2.180}
ψ	$\psi^{(5r)}$	$\psi^{(6r)}$	$\psi^{(7r)}$	$\psi^{(8r)}$
d	-0.009 _{0.102}	-0.012 _{0.122}	-0.029 _{0.102}	-0.024 _{0.098}
ϕ	0.003 _{0.047}	0.002 _{0.050}	0.004 _{0.040}	0.002 _{0.038}
σ	0.013 _{0.062}	-0.004 _{0.041}	0.005 _{0.042}	0.003 _{0.033}
v	1.226 _{4.543}	1.920 _{4.586}	0.850 _{2.981}	-0.150 _{2.485}

5.5. In-sample Empirical Evidence for Ten S&P 500 Stocks

The presence of long-memory in the volatility of the S&P 500 stock index is documented in Ding et al. (1993) and further analyzed for different components of the index by Ray and Tsay (2000). In our empirical study, we consider the top-ten constituents of this index. The included stocks are selected by their market capitalization as of 06-12-2010: the largest ten are listed in Table 3. For each stock, we create a sample of daily adjusted closing prices between 01-01-2004 and 06-12-2010. In Figs. 3a and 3b, we present, for $t = 1, \dots, n$, with $n = 1,745$; the adjusted daily closing prices P_t , the inflated daily differences of the log adjusted closing prices $y_t = 100 \log(P_t / P_{t-1})$, and the sample autocorrelation functions of $\log y_t^2$. The sample autocorrelation function of $\log y_t^2$ can be regarded as an indicator of the autocorrelation function of log-variance X_t , see the discussion in So and Kwok (2006). Most of the sample autocorrelation functions for $\log y_t^2$ show hyperbolic

TABLE 3
Top-ten Companies from the S&P 500 Index (by Market Capitalization, 06-12-2010)

Symbol	Constituent	GICS Sector
XOM	Exxon Mobil Corp	Energy
AAPL	Apple Inc	Information Technology
MSFT	Microsoft Corp	Information Technology
IBM	Intl. Business Machines Corp	Information Technology
GE	General Electric Corp	Industrials
PG	Procter & Gamble	Consumer Staples
JNJ	Johnson & Johnson	Health Care
CVX	Chevron Corp	Energy
T	AT&T	Technology
JPM	JP Morgan Chase	Financial

decays which may indicate a presence of long-memory in the volatility of the series. The series with autocorrelations close to zero may have been affected by outliers which can be due to unexpected returns.

Before the estimation procedure is started, we remove the sample mean $n^{-1} \sum_{t=1}^n y_t$ from the returns y_t to avoid taking logs of zero return values. Monte Carlo maximum likelihood estimation is carried as described in Section 3. Univariate estimation results, based on the Yule–Walker approximation and $M = 400$ simulations from the importance density for likelihood evaluation, are presented in Tables 4a and 4b. We consider the unrestricted LMSV- t model, the LMSV- t model with restriction $d = 0$ and the LMSV- t model with restriction $\phi = 0$. The three LMSV models are also considered with the Gaussian density, where we impose the additional restriction that $\nu = 1,000$.

For all time series the log-likelihood value is highest for the LMSV- t model where no restrictions are placed on d or ϕ . The well-known Akaike information criterion (AIC) confirms the superior in-sample performance of this model for most of the time series. All estimates of the fractional differencing parameter are significant and often around 0.45, confirming the presence of long-memory in the log volatility of the components of the S&P 500 stock index. In addition for most stocks, high levels of stationary persistence ϕ are estimated. The LMSV- t model estimates show the advantage of using fat tails to describe the distribution of the log-returns. Estimated degrees of freedom ν are often low indicating substantially larger tails compared to the Gaussian distribution. In Fig. 4 we present the estimates of the time-varying volatility paths $\exp(X_t)$ for the LMSV- t model estimates. The estimated time-varying volatilities are computed for the maximum likelihood estimate of the parameter vector, $\tilde{\psi}$, as presented in Tables 4a and 4b. The computations are given by (18) and based on $M = 400$ simulations from the importance density. Many of the estimated log-volatilities in Fig. 4 have similar patterns. We may therefore consider the reduction of the number of volatility components underlying the

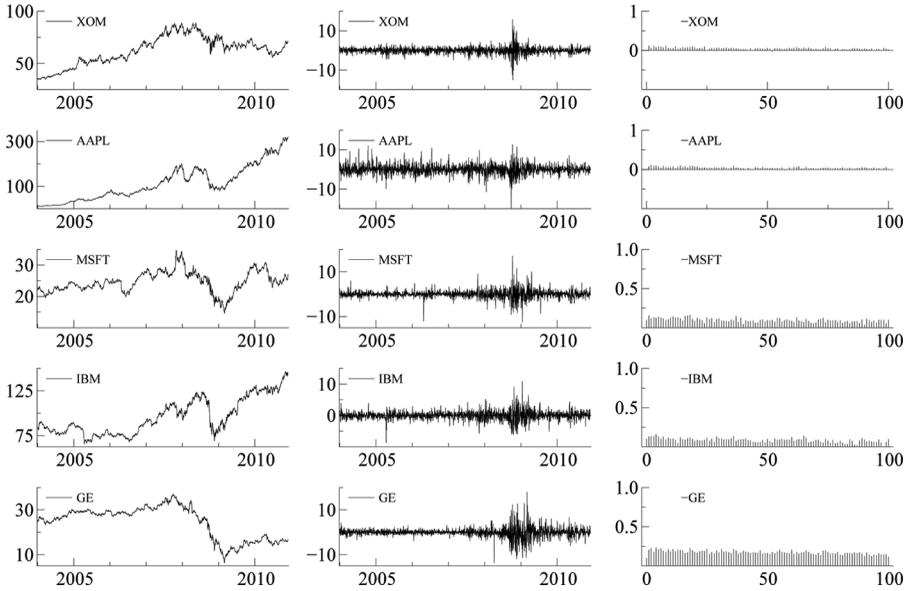


FIGURE 3a Data descriptives on constituents 1 to 5 by market cap of the S&P 500 stock index. Each series contains $n = 1,745$ trading days between 01-01-2004 and 06-12-2010. The columns show adjusted daily closing prices, P_t ; the daily differences of the log adjusted closing prices, y_t ; and the sample autocorrelation function of the squared daily differences of the log adjusted closing prices, $\log y_t^2$.

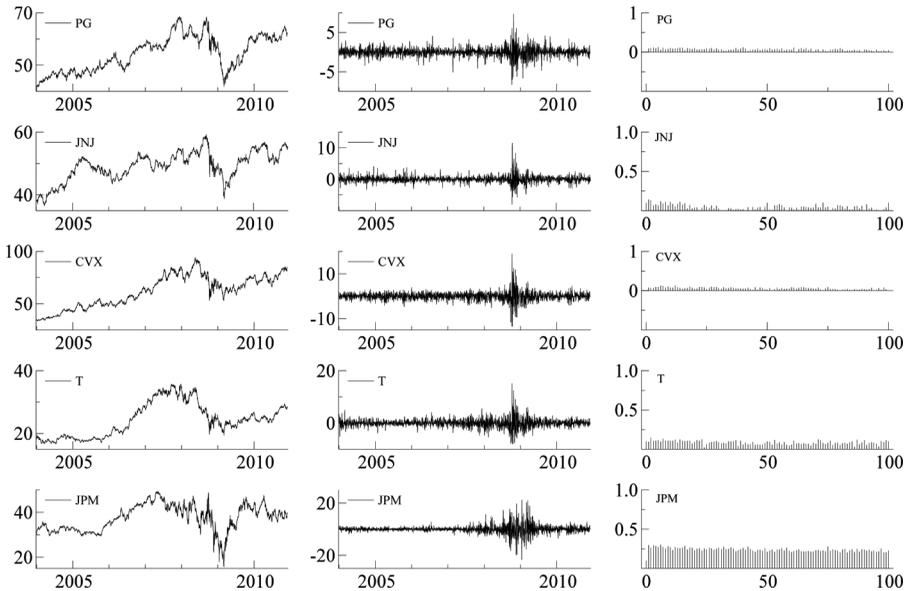


FIGURE 3b Data descriptives on constituents 6 to 10 by market cap of the S&P 500 stock index. Each series contains $n = 1,745$ trading days between 01-01-2004 and 06-12-2010. The columns show adjusted daily closing prices, P_t ; the daily differences of the log adjusted closing prices, y_t ; and the sample autocorrelation function of the squared daily differences of the log adjusted closing prices, $\log y_t^2$.

TABLE 4a

Estimated Parameters for Univariate LMSV and LMSV- t Models Using Log-returns From Constituents 1 to 5 By Market Cap of the S&P 500 Index, for Days From 01-01-2004 to 6-12-2010 ($n = 1,745$ Trading Days). The Standard Errors of the Estimates are Given as Subscripts. The Method of Monte Carlo Maximum Likelihood is Based on Importance Sampling Using the Yule-Walker Approximation with ten Lags and Using $M = 400$ Importance Simulations. The Akaike Information Criterion (AIC) is Computed as $AIC = 2\mathcal{P} - 2\log \tilde{\ell}(\tilde{\psi})$, where \mathcal{P} Denotes the Number of Elements in $\tilde{\psi}$. The last two columns report the maximized loglikelihood and AIC values; the bold values are the respective maximum and minimum values over the six different models

	d	ϕ	σ	ν	$\log \tilde{\ell}(\tilde{\psi})$	AIC
XOM						
LMSV- t	0.379 _{0.108}	0.942 _{0.029}	0.069 _{0.019}	20.183 _{9.740}	-4675.5	9359.0
LMSV- t ($\phi = 0$)	0.496 _{0.006}	—	0.505 _{0.032}	1887.96 _{4.296}	-4691.4	9388.8
LMSV- t ($d = 0$)	—	0.991 _{0.004}	0.130 _{0.019}	21.588 _{10.147}	-4677.3	9360.6
LMSV	0.371 _{0.111}	0.934 _{0.034}	0.079 _{0.021}	—	-4677.8	9361.6
LMSV ($\phi = 0$)	0.496 _{0.006}	—	0.505 _{0.032}	—	-4691.4	9386.8
LMSV ($d = 0$)	—	0.989 _{0.004}	0.141 _{0.020}	—	-4679.2	9362.4
AAPL						
LMSV- t	0.475 _{0.032}	0.943 _{0.026}	0.041 _{0.017}	7.019 _{1.227}	-5544.2	11096.4
LMSV- t ($\phi = 0$)	0.498 _{0.003}	—	0.594 _{0.031}	12.360 _{3.519}	-5567.5	11141.0
LMSV- t ($d = 0$)	—	0.998 _{0.001}	0.082 _{0.017}	6.880 _{1.157}	-5545.8	11097.6
LMSV	0.489 _{0.016}	0.664 _{0.167}	0.226 _{0.099}	—	-5560.6	11127.2
LMSV ($\phi = 0$)	0.498 _{0.003}	—	0.596 _{0.030}	—	-5572.8	11149.6
LMSV ($d = 0$)	—	0.996 _{0.002}	0.146 _{0.024}	—	-5567.4	11138.8
MSFT						
LMSV- t	0.428 _{0.085}	0.938 _{0.033}	0.062 _{0.019}	6.011 _{0.853}	-4626.7	9261.4
LMSV- t ($\phi = 0$)	0.495 _{0.006}	—	0.553 _{0.040}	10.100 _{2.896}	-4640.1	9286.2
LMSV- t ($d = 0$)	—	0.995 _{0.003}	0.115 _{0.021}	5.925 _{0.824}	-4627.7	9261.4
LMSV	0.490 _{0.013}	-0.040 _{0.157}	0.602 _{0.069}	—	-4643.6	9293.2
LMSV ($\phi = 0$)	0.492 _{0.010}	—	0.595 _{0.036}	—	-4644.4	9292.8
LMSV ($d = 0$)	—	0.978 _{0.007}	0.242 _{0.028}	—	-4656.9	9317.8
IBM						
LMSV- t	0.414 _{0.097}	0.894 _{0.061}	0.093 _{0.032}	9.481 _{2.391}	-4371.5	8750.9
LMSV- t ($\phi = 0$)	0.492 _{0.011}	—	0.531 _{0.037}	43.394 _{3.330}	-4378.2	8762.4
LMSV- t ($d = 0$)	—	0.989 _{0.005}	0.126 _{0.026}	8.822 _{1.946}	-4372.0	8750.0
LMSV	0.486 _{0.020}	0.152 _{0.334}	0.471 _{0.143}	—	-4378.5	8763.0
LMSV ($\phi = 0$)	0.490 _{0.013}	—	0.541 _{0.037}	—	-4378.9	8761.8
LMSV ($d = 0$)	—	0.972 _{0.009}	0.208 _{0.028}	—	-4382.4	8768.9
GE						
LMSV- t	0.455 _{0.065}	0.998 _{0.125}	0.157 _{0.042}	8.134 _{2.168}	-4662.0	9332.0
LMSV- t ($\phi = 0$)	0.497 _{0.004}	—	0.572 _{0.034}	21.164 _{5.176}	-4679.3	9364.6
LMSV- t ($d = 0$)	—	0.997 _{0.002}	0.098 _{0.017}	7.681 _{1.341}	-4662.3	9330.6
LMSV	0.493 _{0.010}	0.341 _{0.181}	0.415 _{0.097}	—	-4678.2	9362.4
LMSV ($\phi = 0$)	0.497 _{0.004}	—	0.581 _{0.032}	—	-4680.8	9365.7
LMSV ($d = 0$)	—	0.992 _{0.004}	0.164 _{0.027}	—	-4682.2	9368.4

daily price differences. In other words, the volatilities in the ten constituents of the S&P500 index may depend on a small number of common components. We investigate this further in Section 6.

TABLE 4b

Estimated Parameters for Univariate LMSV and LMSV- t Models Using Log-returns From Constituents 6 to 10 by Market Cap of the S&P 500 Index, for Days From 01-01-2004 to 6-12-2010 ($n = 1,745$ Trading Days). The Standard Errors of the Estimates are Given as Subscripts. The Method of Monte Carlo Maximum Likelihood is Based on Importance Sampling Using the Yule-Walker Approximation with Ten Lags and Using $M = 400$ Importance Simulations. The Akaike Information Criterion is Computed as $AIC = 2\mathcal{P} - 2 \log \tilde{\ell}(\tilde{\psi})$, where \mathcal{P} Denotes the Number of Elements in $\tilde{\psi}$. The last two columns report the maximized loglikelihood and AIC values; the bold values are the respective maximum and minimum values over the six different models

	d	ϕ	σ	ν	$\log \tilde{\ell}(\tilde{\psi})$	AIC
PG						
LMSV- t	0.442 _{0.068}	0.731 _{0.223}	0.193 _{0.130}	7.809 _{3.269}	-4069.5	8147.0
LMSV- t ($\phi = 0$)	0.476 _{0.028}	—	0.577 _{0.055}	18.832 _{1.756}	-4069.5	8145.0
LMSV- t ($d = 0$)	—	0.986 _{0.007}	0.128 _{0.032}	6.062 _{0.0989}	-4070.4	8146.8
LMSV	0.462 _{0.037}	0.008 _{0.201}	0.612 _{0.087}	—	-4073.8	8153.6
LMSV ($\phi = 0$)	0.461 _{0.038}	—	0.623 _{0.054}	—	-4070.6	8145.2
LMSV ($d = 0$)	—	0.947 _{0.014}	0.287 _{0.037}	—	-4085.9	8175.9
JNJ						
LMSV- t	0.462 _{0.050}	0.892 _{0.071}	0.076 _{0.038}	8.107 _{1.686}	-4368.3	8744.6
LMSV- t ($\phi = 0$)	0.494 _{0.009}	—	0.534 _{0.036}	29.192 _{8.126}	-4372.9	8751.8
LMSV- t ($d = 0$)	—	0.996 _{0.003}	0.091 _{0.022}	7.212 _{1.235}	-4368.5	8743.0
LMSV	0.484 _{0.020}	0.201 _{0.179}	0.464 _{0.081}	—	-4373.1	8752.2
LMSV ($\phi = 0$)	0.492 _{0.010}	—	0.546 _{0.035}	—	-4373.6	8751.1
LMSV ($d = 0$)	—	0.981 _{0.007}	0.199 _{0.028}	—	-4384.9	8773.8
CVX						
LMSV- t	0.425 _{0.003}	0.944 _{0.015}	0.054 _{0.013}	236.500 _{3.567}	-4774.0	9556.0
LMSV- t ($\phi = 0$)	0.498 _{0.004}	—	0.455 _{0.029}	529.130 _{4.145}	-4805.1	9616.2
LMSV- t ($d = 0$)	—	0.994 _{0.003}	0.110 _{0.015}	341.200 _{1.941}	-4776.5	9559.0
LMSV	0.409 _{0.090}	0.948 _{0.023}	0.053 _{0.014}	—	-4774.1	9554.2
LMSV ($\phi = 0$)	0.497 _{0.004}	—	0.455 _{0.029}	—	-4805.1	9614.2
LMSV ($d = 0$)	—	0.994 _{0.003}	0.110 _{0.015}	—	-4776.5	9557.0
T						
LMSV- t	0.343 _{0.098}	0.970 _{0.088}	0.047 _{0.054}	17.141 _{1.180}	-4476.4	8960.8
LMSV- t ($\phi = 0$)	0.495 _{0.007}	—	0.494 _{0.034}	911.448 _{21.378}	-4486.4	8978.8
LMSV- t ($d = 0$)	—	0.995 _{0.003}	0.097 _{0.018}	17.840 _{8.987}	-4477.0	8960.0
LMSV	0.405 _{0.136}	0.942 _{0.047}	0.060 _{0.020}	—	-4480.0	8966.0
LMSV ($\phi = 0$)	0.495 _{0.007}	—	0.494 _{0.034}	—	-4486.4	8976.8
LMSV ($d = 0$)	—	0.993 _{0.004}	0.113 _{0.019}	—	-4479.2	8962.4
JPM						
LMSV- t	0.465 _{0.070}	0.948 _{0.084}	0.056 _{0.049}	9.672 _{1.214}	-5037.2	10082.4
LMSV- t ($\phi = 0$)	0.498 _{0.003}	—	0.613 _{0.032}	29.559 _{3.582}	-5056.9	10119.8
LMSV- t ($d = 0$)	—	0.997 _{0.002}	0.118 _{0.019}	9.911 _{2.471}	-5039.0	10084.0
LMSV	0.488 _{0.017}	0.775 _{0.089}	0.173 _{0.057}	—	-5045.1	10096.2
LMSV ($\phi = 0$)	0.498 _{0.003}	—	0.612 _{0.031}	—	-5057.8	10119.6
LMSV ($d = 0$)	—	0.996 _{0.002}	0.151 _{0.023}	—	-5045.6	10095.2

5.6. Out-of-sample Empirical Evidence for Ten S&P 500 Stocks

We investigate the out-of-sample performance of the LMSV- t model next. In particular, we compare the volatility forecasts from the LMSV- t model to the forecasts from the

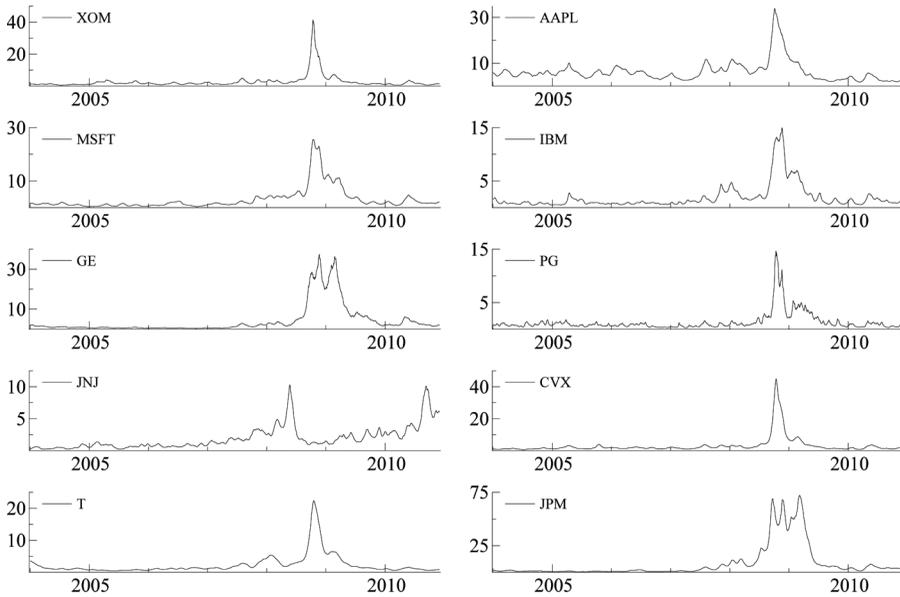


FIGURE 4 Estimated volatility paths $\exp(\hat{X}_t)$ of constituents 1 to 10 by market cap of the S&P 500 stock index. Log-variance X_t follows the ARFIMA process of Eq. (3) with $p = 1$ and $q = 0$. Log-returns y_t are modelled by Eq. (19) where ζ_t follows the Student's t distribution. We present an estimates for each trading day between 01-01-2004 and 6-12-2010, that is $n = 1745$. Estimates are computed as described in section 4.5 based on the estimated parameters from Tables 4a and 4b and $M = 400$ simulations from the importance density.

Gaussian LMSV model. We also include specifications that restrict either $d = 0$ or $\phi = 0$. The restriction $d = 0$ implies the short memory SV- t and SV models, whereas the restriction $\phi = 0$ implies a pure fractionally integrated process (ARFIMA(0, d ,0)) for the log-variance similar as in Ray and Tsay (2000). We consider a rolling forecast window where the first sample concerns all trading days from 01-01-2004 until 20-12-2007. This sample is used to estimate the model parameters and to compute the one-day-ahead forecast for 21-12-2007. The final window includes trading days from 14-12-2006 until 05-12-2010 and is used to compute the forecast for 06-12-2010. Each window includes 1,000 observations that we use for estimation. We obtain a total of 745 one-step-ahead forecasts.

The volatility forecasts for the LMSV- t models are given by

$$\hat{\sigma}_{n+1|n}^2 = \exp\left(\hat{X}_{n+1|n} + 0.5P_{n+1|n}\right), \tag{20}$$

where $\hat{X}_{n+1|n}$ and $P_{n+1|n}$ are computed by equation (18), with $h(z^{(i)}) = x_{n+1|n}^{(i)}$ and $h(z^{(i)}) = (x_{n+1|n}^{(i)})^2 - \hat{X}_{n+1|n}$, see Koopman et al. (2005). The predictions $x_{n+1|n}^{(i)}$ are computed by the Kalman filter for each simulation $i = 1, \dots, M$. We again use $M = 400$ draws from the importance density.

We use realized volatility as an accurate measure of daily volatility for each day. The realized volatility measures are denoted by $\tilde{\sigma}_n^2$ and are computed from 5 minute intradaily returns. For each forecast we compute the squared error loss given by

$$L_{n+1|n} = (\hat{\sigma}_{n+1|n}^2 - \tilde{\sigma}_{n+1}^2)^2. \tag{21}$$

Based on these losses we compare the forecasting performance of the different models using model confidence sets (MSC), as introduced in Hansen et al. (2011). We keep the notation used in Hansen et al. (2011) and refer to their article for a more detailed exposition.

The MSC is the set of models that contains the best performing models for a given level of confidence. By denoting $\mathcal{M}^0 = \mathcal{M}$ as the initial set of models considered, we define the MCS for this set of models as

$$\mathcal{M}^* \equiv \{l \in \mathcal{M}^0 : \mu^{lk} \leq 0 \text{ for all } k \in \mathcal{M}^0\}, \tag{22}$$

where $\mu^{lk} = E(d_{n+1|n}^{lk})$ is the difference in expected loss between models l and k , with $d_{n+1|n}^{lk} = L_{n+1|n}^l - L_{n+1|n}^k$. We have indexed the loss μ^{lk} by l and k to discriminate between the models l and k . For example, $L_{n+1|n}^l$ can denote the loss of the LMSV- t model, and $L_{n+1|n}^k$ can denote the loss of the LMSV model.

The objective of the MCS procedure is to have $\hat{\mathcal{M}}_{1-\alpha}^*$ as an asymptotically consistent estimator of \mathcal{M}^* with confidence $1 - \alpha$. This is accomplished by iterating between an equivalence test and an elimination rule until the equivalence test is not rejected. The equivalence test is used to test $\{H_{0,\mathcal{M}} : \mu^{lk} = 0 \text{ for all } l, k \in \mathcal{M} \subseteq \mathcal{M}^0\}$, and the elimination rule $e_{\mathcal{M}}$ identifies the model that should be removed from \mathcal{M} if H_0 is rejected.

To test $H_{0,\mathcal{M}}$ we construct t -test statistics based on the average over the 745 loss differences $d_{n+1|n}^{lk}$, which we denote by \bar{d}^{lk} . We consider the supremum type test statistic over all model combinations, that is

$$T_{\mathcal{M}} \equiv \max_{l,k \in \mathcal{M}} |t_{lk}|, \quad t_{lk} = \frac{\bar{d}^{lk}}{\sqrt{\widehat{Var}(\bar{d}^{lk})}}. \tag{23}$$

Whether or not the null of no difference in model performance is rejected depends only on the model with the greatest relative loss. The accompanying coherent elimination rule is given by $e_{\mathcal{M}} = \arg \max_{l \in \mathcal{M}} \sup_{k \in \mathcal{M}} t_{lk}$. The asymptotic distribution of this test statistic is nonstandard as it depends on nuisance parameters. However, the distribution can be computed using bootstrap methods which implicitly deal with the nuisance parameter problem, see for details Hansen (2005) and Hansen et al. (2011).

The results from the iterative MSC procedure are summarized as follows. Define $p(l)$ as the p -value of the l th model that gets eliminated if the null is rejected. The MSC p -value is given by $p_{MSC}(l) = \max_{k \leq l} p(k)$, which implies that the MSC p -value cannot be

TABLE 5

Results From Out-of-sample Forecasting Study. We Present the Mean Squared Losses and the Accompanying MSC p -values For all 10 Stocks. The One-step-ahead Forecasts are Computed Based on a Rolling Window With 1,000 Observations, Where the First Window is From 01-01-2004 Until 20-12-2007. We Computed 745 Out-of-sample Forecasts Which Were Evaluated Against Realized Volatility to Compute the Mean Squared Error Loss. The MSC is Determined as Discussed in Section 5.6. The * and ** indicates that the model is in the 90% and 95% MSC, respectively

	MSE	p -value	MSE	p -value
		XOM		PG
LMSV- t	46.784	0.061*	11.903	0.740**
LMSV- t ($\phi = 0$)	60.958	0.000	12.028	0.740**
LMSV- t ($d = 0$)	52.024	0.007	12.350	0.740**
LMSV	45.154	1.000**	12.011	0.740**
LMSV ($\phi = 0$)	60.940	0.000	12.313	0.740**
LMSV ($d = 0$)	48.921	0.012	11.025	1.000**
		AAPL		JNJ
LMSV- t	114.22	0.194**	5.242	0.201**
LMSV- t ($\phi = 0$)	150.02	0.012	5.277	0.164**
LMSV- t ($d = 0$)	117.47	0.194**	5.875	0.164**
LMSV	122.72	0.194**	6.206	0.091*
LMSV ($\phi = 0$)	148.14	0.183**	6.262	0.074*
LMSV ($d = 0$)	81.728	1.000**	4.713	1.000**
		MSFT		CVX
LMSV- t	31.765	1.000**	60.377	1.000**
LMSV- t ($\phi = 0$)	45.254	0.001	80.416	0.000
LMSV- t ($d = 0$)	33.037	0.001	65.079	0.061*
LMSV	47.446	0.001	60.402	0.061*
LMSV ($\phi = 0$)	47.047	0.001	80.447	0.000
LMSV ($d = 0$)	38.879	0.001	65.038	0.061*
		IBM		T
LMSV- t	28.430	0.851**	43.951	1.000**
LMSV- t ($\phi = 0$)	30.783	0.045	56.089	0.000
LMSV- t ($d = 0$)	28.416	1.000**	44.509	0.536**
LMSV	30.129	0.048	45.199	0.015
LMSV ($\phi = 0$)	29.483	0.048	55.878	0.000
LMSV ($d = 0$)	30.871	0.048	44.511	0.046
		GE		JPM
LMSV- t	128.27	0.324**	222.75	1.000**
LMSV- t ($\phi = 0$)	158.55	0.000	340.55	0.026
LMSV- t ($d = 0$)	121.68	1.000**	222.75	0.995**
LMSV	150.89	0.000	271.76	0.026
LMSV ($\phi = 0$)	158.89	0.000	334.71	0.026
LMSV ($d = 0$)	123.91	0.423**	227.86	0.236**

smaller than the p -value of the previously eliminated models. Intuitively this implies that we cannot reject a particular model with a higher level of confidence when compare to models that were previously eliminated. In this way, we can determine for any level of significance which models should be included in the MSC.

The results for all 10 stocks are shown in Table 5. We report for each stock and each model the mean squared error loss and the corresponding MSC p -value. We denote the models that are included in the 90% and 95% MSC by * and **, respectively. We find that the LMSV- t model is always in the 90% model confidence set and, except for one stock (Exxon Mobil Corp), always in the 95% model confidence set. When compared to other model specifications, we find that the LMSV- t model is on average the best performing model specification. In particular, no other model specification is also always in the 90% confidence set. The SV- t model, without the long-memory component also performs relatively well. For eight stocks it is included in the 90% model confidence set, and for seven stocks in the 95% model confidence set. For some stocks, all models were included in the 90% model confidence set.

6. MULTIVARIATE LONG-MEMORY STOCHASTIC VOLATILITY

Different specifications for a multivariate extension of the stochastic volatility model can be considered, see, for example Asai et al. (2006). We illustrate the multivariate capabilities of our estimation methodology by means of a long-memory stochastic volatility model where the log-volatility depends on a small set of multiplicative factors that are modeled independently.

In general, we consider time series of k asset daily log-returns, denoted by the $k \times 1$ vector $y_t = (y_{1t}, \dots, y_{kt})'$ for $t = 1, \dots, n$. We assume that y_t has mean zero and a time-varying variance matrix depending on a small number of ARFIMA processes with Gaussian innovations. The model for y_t is given by

$$y_t = \mathcal{L}_t \xi_t, \quad \xi_t \sim N(0, \Sigma_\xi), \quad t = 1, \dots, n, \tag{24}$$

where \mathcal{L}_t is a $k \times k$ diagonal matrix with elements $\exp(Z_{it}/2)$, for $i = 1, \dots, k$, on the diagonal. Latent $k \times 1$ vector Z_t , as given in Eq. (2), allows for the variance to depend on $l \times 1$ vector X_t of independent long-memory processes, where holds $l \ll k$. We restrict $\mathcal{B} = 0$ since our main interest is in long-memory components. The $k \times 1$ disturbance vector ξ_t has its variance matrix equal to the unity matrix, that is $\Sigma_\xi = I_k$. The l components follow either normal or Student's t distributions. We consider models with $l = 1$ and $l = 2$ long-memory volatility factors. Models with more than two factors become numerically more challenging as for each long-memory factor a short-memory process need to be found for its approximation. The state vector increases rapidly when the Yule-Walker approximation requires, say, ten lags. The dimension of the approximating linear Gaussian state space model increases rapidly and likelihood evaluation via importance sampling becomes computationally demanding. The number of observed time series k is much less relevant for computational efficiency.

The matrix \mathcal{A} in (2) becomes a factor loading matrix and is constrained for identification purposes, see Geweke, Zhou (1996). For example, in our illustration below, we have $k = 10$ and $l = 2$, and specify the loading matrix as

$$\mathcal{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ a_{3,1} & a_{3,2} \\ \vdots & \vdots \\ a_{10,1} & a_{10,2} \end{bmatrix}$$

The unrestricted elements $a_{i,j}$ of matrix \mathcal{A} are estimated together with the other parameters. We further restrict $\text{Var}(X_{it}) = 1$ for $i = 1, \dots, l$. This can be done using the exact auto covariance formulas as presented in Sowell (1992). A one factor version of this model, with $\phi_1(B) = 1$ and $\theta_1(B) = 1$, is proposed by Ray and Tsay (2000). The estimation of the parameters is based on quasi-maximum likelihood and spectral regression methods, see also So and Kwok (2006).

We continue our study with the ten volatility series from the S&P500 index. We study the 10×1 vector $y_t = (y_{1t}, \dots, y_{10,t})'$ simultaneously with $y_{it} = 100 \log(P_{i,t} / \log P_{i,t-1})$ and

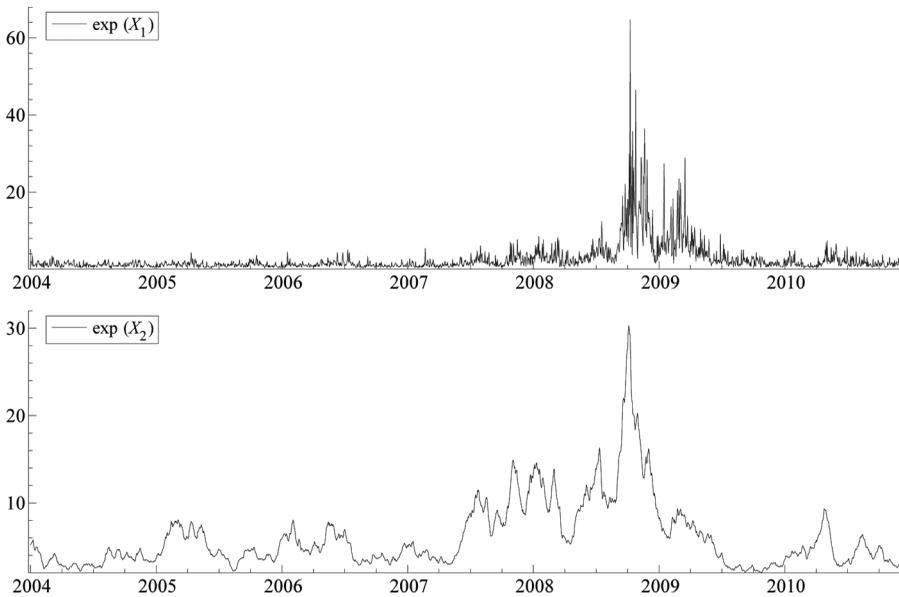


FIGURE 5 Estimated long-memory factors $\exp(\tilde{X}_{1t})$ and $\exp(\tilde{X}_{2t})$ underlying the volatility of the log-returns of ten constituents by market cap of the S&P 500 stock index from 01-01-2004 until 6-12-2010 ($n = 1,745$). Log-variances X_{it} are specified by Eq. (3), with $p = 1$ and $q = 0$. The log-returns vector y_t is modelled by Eq. (24) using Student's t distributions for ζ_t . Estimates are computed as described in Section 4.5 using optimized parameters from Table 6 and $M = 400$ simulations from the importance density.

P_{it} is the daily adjusted closing price of stock i . The stocks are ordered as in Table 3. We implement one and two factor versions of the model, and for each X_{it} , we set $p_i = 1$ and $q_i = 0$, with $i = 1, 2$.

The parameter estimates for the multivariate LMSV and LMSV- t models, for ξ_t normally and Student's t distributed, respectively, are presented in Table 6. The estimated factors appear to have strong long-memory features together with either slight negative or high positive stationary persistence. The log-likelihood value of the LMSV- t model with

TABLE 6

Multivariate Long-memory Estimation Results for the Log-returns of $k = 10$ Stocks of the S&P 500 Index From 01-01-2004 Until 6-12-2010 ($n = 1,745$). Estimation Results are Presented for the Multivariate LMSV and LMSV- t Models with $l = 1, 2$ Factors. The Standard Errors of the Estimates are Given as Subscripts. The Method of Monte Carlo Maximum Likelihood is Based on Importance Sampling Using the Yule-Walker Approximation With Ten Lags, for Each Long-memory Component. The Number of Importance Simulations for Likelihood Evaluation is $M = 400$. The Akaike Information Criterion is Compute as $AIC = 2\mathcal{P} - 2\log \tilde{\ell}(\tilde{\psi})$, where \mathcal{P} Denotes the Number of Elements in $\tilde{\psi}$. The last two rows report the maximized loglikelihood and AIC values; the bold values are the respective maximum and minimum values over the four different models

	LMSV		LMSV- t	
	$l = 1$	$l = 2$	$l = 1$	$l = 2$
d_1	0.462 _{0.031}	0.452 _{0.034}	0.463 _{0.032}	0.461 _{0.036}
d_2	—	0.376 _{0.045}	—	0.454 _{0.105}
ϕ_1	-0.266 _{0.060}	-0.206 _{0.076}	-0.251 _{0.074}	-0.209 _{0.083}
ϕ_2	—	0.823 _{0.122}	—	0.943 _{0.192}
v	—	—	5.943 _{0.268}	10.511 _{1.161}
$a_{1,1}$	1	1	1	1
$a_{2,1}$	1.785 _{0.057}	0	1.564 _{0.047}	0
$a_{3,1}$	1.103 _{0.042}	1.186 _{0.056}	1.037 _{0.039}	1.236 _{0.115}
$a_{4,1}$	0.834 _{0.037}	1.067 _{0.052}	0.824 _{0.037}	1.122 _{0.063}
$a_{5,1}$	1.318 _{0.045}	1.450 _{0.056}	1.235 _{0.041}	1.563 _{0.042}
$a_{6,1}$	0.607 _{0.034}	1.098 _{0.053}	0.575 _{0.032}	1.108 _{0.083}
$a_{7,1}$	0.825 _{0.043}	-0.158 _{0.054}	0.512 _{0.032}	1.090 _{0.055}
$a_{8,1}$	1.052 _{0.039}	0.860 _{0.043}	1.084 _{0.041}	0.897 _{0.038}
$a_{9,1}$	0.886 _{0.038}	0.945 _{0.051}	0.882 _{0.038}	0.993 _{0.054}
$a_{10,1}$	1.710 _{0.052}	1.505 _{0.063}	1.627 _{0.049}	1.671 _{0.036}
$a_{1,2}$	—	0	—	0
$a_{2,2}$	—	1	—	1
$a_{3,2}$	—	-0.026 _{0.042}	—	-0.110 _{0.052}
$a_{4,2}$	—	-0.147 _{0.038}	—	-0.220 _{0.045}
$a_{5,2}$	—	-0.097 _{0.046}	—	-0.225 _{0.064}
$a_{6,2}$	—	-0.383 _{0.040}	—	-0.420 _{0.077}
$a_{7,2}$	—	0.464 _{0.242}	—	0.297 _{0.032}
$a_{8,2}$	—	0.167 _{0.031}	—	0.152 _{0.048}
$a_{9,2}$	—	-0.023 _{0.038}	—	-0.072 _{0.043}
$a_{10,2}$	—	0.193 _{0.048}	—	0.043 _{0.033}
$\log \tilde{\ell}(\tilde{\psi})$	-47120	-46527	-46887	-46430
AIC	94262	93093	93797	92901

two factors is given by $-46,430$ and is significantly higher than the sum of all univariate log-likelihood estimates which is given by $-46,605$. The estimated factors of our two-factor LMSV- t model are presented in Fig. 5. The first factor is clearly more noisy while the other factor may represent more long-term changes in volatility.

7. CONCLUSIONS

We have shown that a general class of nonlinear non-Gaussian time series models with latent long-memory components can be treated successfully by Monte Carlo maximum likelihood methods based on importance sampling techniques. The estimation method is based on exact maximum likelihood, but it is subject to Monte Carlo error. The importance sampling method is based on a linear Gaussian approximation model that also approximated the long-memory process by a stationary autoregressive process with a large number of lags. This is a new development, and it illustrates the flexibility of the general methodology. We have implemented a computationally efficient method for evaluating the Monte Carlo estimate of the loglikelihood value. The methodology is studied in detail via a set of Monte Carlo simulation studies in which we show that, for a range of models, the underlying true parameter values can be estimated accurately.

We further show that the methodology can be used in empirical analyses. We illustrate the methods by fitting stochastic volatility models to ten components of the S&P 500 stock index. Although it is empirically challenging to empirically identify long-memory and a Student's t density simultaneously in a stochastic volatility model, we have shown that it is possible within our framework. The extension to multivariate analysis is shown to be analytically relatively easy. However, the computational implications of including many latent long-memory components in the model are high. The introduction of time-varying correlation between the latent long-memory components is an interesting research project for future consideration.

FUNDING

S. J. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

REFERENCES

- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96:42–55.
- Anderson, B. D. O., Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs N.J: Prentice-Hall.
- Asai, M., McAleer, M., Medeiros, M. (2012). Modelling and forecasting noisy realized volatility. *Computational Statistics and Data Analysis* 56:217–230.
- Asai, M., McAleer, M., Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews* 25:145–175.

- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73:5–59.
- Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial econometrics. *Journal of the Royal Statistical Society, Series B* 63:167–241.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. London: Chapman and Hall.
- Bollerslev, T., Jubinski, D. (1999). Equity trading volume and volatility: latent in-formation arrivals and common long-run dependencies. *Journal of Business and Economic Statistics* 17:9–21.
- Breidt, F. J., Crato, N., De Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83:325–348.
- Brockwell, A. E. (2007). Likelihood-based analysis of a class of generalized long-memory time series models. *Journal of Time Series Analysis* 28:386–407.
- Brockwell, A. E., Chan, N. H. (2006). Long-memory dynamic tobit models. *Journal of Forecasting* 25: 351–367.
- Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81:541–553.
- Chan, N. H., Palma, W. (1998). State space modeling of long-memory processes. *Annals of Statistics* 26: 719–740.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 26:174–196.
- Davidson, R., MacKinnon, J. G. (2004). *Econometric Theory and Methods*. New York: Oxford University Press.
- de Jong, P., Shephard, N. (1995). The simulation smoother for time series models. *Biometrika* 82:339–350.
- Deo, R., Hsieh, M., Hurvich, C., Soulier, P. (2006). Long memory in nonlinear processes. In: Bertail, P., Soulier, P., Doukhan, P. eds. *Dependence in Probability and Statistics*. Vol. 187, Lecture Notes in Statistics. New York: Springer, pp. 221–244.
- Deo, R. S., Hurvich, C. M., Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics* 131:29–58.
- Ding, Z., Granger, C. W. J., Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1:83–106.
- Doornik, J. A., Ooms, M. (2003). Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics & Data Analysis* 42:333–348.
- Durbin, J. (1960). The fitting of time series models. *International Statistical Review* 28:233–244.
- Durbin, J., Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika* 84:669–684.
- Durbin, J., Koopman, S. J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Series B* 62:3–56.
- Durbin, J., Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. 2nd ed. Oxford: Oxford University Press.
- Durbin, J., Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89:603–616.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15:183–202.
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57:1317–1339.
- Geweke, J. F., Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4:221–238.
- Geweke, J. F., Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* 9:557–587.
- Granger, C. W. J., Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1:15–29.
- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics* 23:365–380.
- Hansen, P. R., Lunde, A., Nason, J. M. (2011). The model confidence set. *Econometrica* 79:453–497.

- Harvey, A. C. (1998). Long memory in stochastic volatility. In: Knight, T., Satchell, S. eds. *Forecasting Volatility in Financial Markets*. Oxford: Butterworth-Heinemann, pp. 307–320.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* 68:165–176.
- Hsu, N. J., Breidt, F. J. (2003). Bayesian analysis of fractionally integrated ARMA with additive noise. *Journal of Forecasting* 22:491–514.
- Jensen, M. J. (1983). Semiparametric Bayesian inference of long-memory stochastic volatility models. *Journal of Time Series Analysis* 25:895–922.
- Jungbacker, B., Koopman, S. J. (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika* 94:827–839.
- Koopman, S. J., Jungbacker, B., Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12:445–475.
- Koopman, S. J., Scharth, M. (2013). The analysis of stochastic volatility in the presence of daily realized measures. *Journal of Financial Econometrics* 11:76–115.
- Koopman, S. J., Shephard, N., Creal, D. D. (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics* 149:2–11.
- Mandelbrot, B. B. (1969). Long-run linearity, locally Gaussian processes, H-spectra and infinite variances. *International Economic Review* 10:82–111.
- Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge: Cambridge University Press.
- Nocedal, J., Wright, S. J. (1999). *Numerical Optimization*. New York: Springer Verlag.
- Palma, W. (2007). *Long Memory Time Series*. New York: John Wiley & Sons.
- Ray, B. K., Tsay, R. S. (2000). Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics* 18:254–262.
- Ripley, B. D. (1987). *Stochastic Simulation*. New York: John Wiley & Sons.
- Robinson, P. (1994). Time series with strong dependence. In: Sims, C. A., ed. *Advances in Econometrics, Sixth World Congress*. Cambridge: Cambridge University Press. pp. 47–95.
- Sela, R. J., Hurvich, C. M. (2009). Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis* 30:631–651.
- Shephard, N., Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84:653–667.
- So, M. K. P. (2002). Bayesian analysis of long memory stochastic volatility models. *The Indian Journal of Statistics* 64:1–10.
- So, M. K. P. (2003). Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica* 13:255–274.
- So, M. K. P., Kwok, S. W. Y. (2006). A multivariate long memory stochastic volatility model. *Physica A: Statistical Mechanics and Its Applications* 362:450–464.
- Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics* 53:165–188.
- Tiao, G. C., Tsay, R. S. (1994). Some advances in non-linear and adaptive modeling in time series. *Journal of Forecasting* 13:109–131.
- Wright, J. H. (1999). A new estimator of the fractionally integrated stochastic volatility model. *Economic Letters* 63:295–303.