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

*G. Mesters*[a]

*S.J. Koopman*[b,c]

*M. Ooms*[b]

[a] *Netherlands Institute for the Study of Crime and Law Enforcement;*
[b] *Dept. of Econometrics, Faculty of Economics and Business Administration, VU University of Amsterdam;*
[c] *Tinbergen Institute, The Netherlands.*

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

G. Mesters$^{(a,b)}$,   S.J. Koopman$^{(b,c)*}$ and  M. Ooms$^{(b)}$

$^{(a)}$ Netherlands Institute for the Study of Crime and Law Enforcement,
$^{(b)}$ Department of Econometrics, VU University Amsterdam,
$^{(c)}$ Tinbergen Institute, Amsterdam

June 24, 2011

---

$^{*}$Corresponding author: prof.dr. S.J. Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email : `s.j.koopman@vu.nl`

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

Geert Mesters, Siem Jan Koopman and Marius Ooms

**Abstract**

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 log-return series from the S&P 500 stock index by univariate and multivariate long-memory stochastic volatility models.

*Some Keywords* : Fractional Integration; Importance Sampling; Kalman Filter; Latent Factors; Stochastic Volatility.

# 1   Introduction

In this paper 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, Crato & De Lima (1998) and Wright (1999), and the long-memory censored Gaussian model as in Brockwell & Chan (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 presented in this paper 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 paper 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 the prediction error decomposition and the Durbin-Levinson algorithm, see Durbin (1960). Computational refinements of this method are proposed in Doornik & 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 en 6 below. Another example is the long-memory censored Gaussian model of Brockwell & Chan (2006). 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 & Pitt (1997) and Durbin & 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.

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 become 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 paper 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 long-memory stochastic volatility 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. This illustration is extended in section 6 where we examine a multivariate long-memory stochastic volatility model. In the final section 7 we summarize and present some directions for further research.

## 2 Modeling framework

Consider a vector time series $Y_t$, for $t = 1, \ldots, n$, 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 $Y = (Y_1', \ldots, Y_n')'$ and where $p(Y|Z)$ can be any density function for $Y$ given the latent vector $Z = (Z_1', \ldots, Z_n')'$, which is often referred to as the signal. In this paper, 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 autoregressive fractional integrated moving average (ARFIMA) process as given by

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

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 $\varepsilon_{it}$ is a serially uncorrelated and normally distributed sequence with zero mean and variance $\sigma_i^2$, which will be restricted in multivariate settings such that $\mathrm{Var}(X_{it}) = 1$. The disturbances $\varepsilon_{it}$ are mutually and serially uncorrelated at all time periods $t$ and for all $i$. The backshift polynomials are given by

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

for known non-negative integer values $p$ and $q$, unknown autoregressive coefficients $\phi_{i,j}$ and unknown moving average coefficients $\theta_{i,k}$ with $i = 1, \ldots, p$ and $k = 1, \ldots, q$, 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

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

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 $\eta_{jt}$ is similarly defined as $\varepsilon_{it}$ 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

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

where

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

with $V_t = \mathrm{E}[(Z_t - \hat{Z}_t)(Z_t - \hat{Z}_t)' | Z_{t-1}, \ldots, Z_1]$ 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

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

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

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

For a realization $y$ of $Y$, the likelihood function is defined as $\ell(\psi) = p(y)$ where $\psi$ 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 $\{z^{(1)}, \ldots, z^{(M)}\}$, 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 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

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

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 $\{z^{(1)}, \ldots, z^{(M)}\}$ 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, Kolmogorovs strong law of large numbers implies that $\hat{\ell}(\psi) \to \ell(\psi)$ as $M \to \infty$. 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, Shephard & Creal (2009). We will present a selection of these diagnostic statistics when we study our choice of importance densities for long-memory stochastic volatility 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 & Pitt (1997) and Durbin & Koopman (1997, 2000). So (2003) and Jungbacker & 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

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

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

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

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 $t = 1, \ldots, n$. 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

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

for $t = 1, \ldots, n$, 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 $\hat{Z} = \mathrm{E}_g(Z|Y)$ 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 & MacKinnon (2004) although Algorithm A is based on a second-order Taylor expansion. The computations can be carried out for any realisation $Y = y$.

## 4.2 Long-memory approximation

When $\mathcal{A} = 0$ in (2), the approximation method can be implemented as described in detail by Jungbacker & 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 & Moore (1979) and Durbin & Koopman (2001, Chapter 4). However, in our general model with $\mathcal{A} \neq 0$, the long-memory process cannot be formulated in state space form with a finite state vector, see the discussion in Chan & Palma (1998). The approximating model $g(Y, Z)$ in (12) is linear and Gaussian nonetheless and a standard lemma insists that

$$\mathrm{E}_g(Z|Y) = \mathrm{E}_g(Z) + \mathrm{Cov}_g(Z, Y)\mathrm{Var}_g(Y)^{-1}[Y - \mathrm{E}_g(Y)].$$

When variance matrix $\mathrm{Var}_g(Y)$ has a convenient structure, the computations can exploit the structure and calculating $\mathrm{E}_g(Z|Y)$ 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 & Ooms (2003). The model (12) implies however a variance matrix $\mathrm{Var}_g(Y)$ that is equal to the sum of a Toeplitz and a block-diagonal matrix. A computationally efficient algorithm for computing $\mathrm{E}_g(Z|Y)$ is unfortunately not available when $\mathrm{Var}_g(Y)$ 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 $\delta(B) = 1 - \delta_1 B - \ldots - \delta_m B^m$. 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 $\delta_1, \ldots, \delta_m$ 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 treshold 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 & 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 & 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

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

where $g_s()$ 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 realisation $Y = y$, the Monte Carlo estimate of the likelihood function is given by

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

8

where the samples $\{z^{(1)}, \ldots, z^{(M)}\}$ are drawn from the importance density $g_s(Z|y)$ that is obtained from Algorithm B. Since $g_s()$ is the approximating model (12) where $Z_t$ has short-memory, we can represent equations (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 & Kohn (1994), de Jong & Shephard (1995) and Durbin & Koopman (2002). The convergence of the estimator $\tilde{\ell}(\psi) \to \ell(\psi)$ as $M \to \infty$ depends on the existence of the variance of the importance weights $w(Y, Z)$ as given by

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

## 4.4   Parameter estimation

Given a particular model specification for (1) and (2), together with a realisation of the time series $Y$ and a particular value for the parameter vector $\psi$, 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 $\psi$. A change of the parameter vector $\psi$ leads to a different value of likelihood function when the parameters are properly identified. The value of $\psi$ 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 $\psi$. In the simulation exercises and in the empirical studies below, we make use of the BFGS algorithm, see Nocedal & 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. During the estimation process of $\psi$, the same set of random values is therefore used for each likelihood evaluation.

Elements of $\psi$ 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 $\psi$ around $\tilde{\psi}$. The standard errors of elements of $\psi$ are given by the square root of the diagonal elements of $\tilde{\Sigma}$.

## 4.5   Signal Extraction

Given parameter vector $\psi$ 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

$$\mathrm{E}_p(h(Z)|Y) = \int h(Z) \cdot p(Y, Z)\mathrm{d}Z = \frac{\int h(Z)w(Y, Z)g_s(Z|Y)\mathrm{d}Z}{\int w(Y, Z)g_s(Z|Y)\mathrm{d}Z}, \tag{17}$$

where $w(Y, Z)$ is given by (16). The estimation of $\mathrm{E}_p(h(Z)|Y)$ by importance sampling, for a realisation $Y = y$, 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 long-memory stochastic volatility

We illustrate the methods developed in the previous sections to 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.

## 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

$$y_t = \exp(X_t/2)\xi_t, \qquad \xi_t \sim \mathrm{N}(0, \sigma_\xi^2), \qquad t = 1, \ldots, n, \tag{19}$$

where $X_t$ is modeled as the ARFIMA process (3). The assumption of a normal density for the disturbances $\xi_t$ can be replaced by the assumption of a Student's $t$-density. The resulting models are referred to as the LMSV model, with Gaussian disturbances $\xi_t$, and the LMSV-$t$ model, with Student's $t$ disturbances $\xi_t$. The parameters of the models are collected in vectors $\psi$ and $\psi^t$ respectively. The LMSV model is introduced by Breidt et al. (1998) and Harvey (1998). Their estimation methods for $\psi$ are based on quasi-maximum likelihood methods. Wright (1999) also considers the LMSV model and proposes to estimate $\psi$ via the general method of moments based on the estimated log-periodogram of Geweke & Porter-Hudak (1983). A comparison of the different estimation methods for the LMSV model is provided by Deo, Hurvich & Lu (2006) where also an enhanced quasi-maximum likelihood method is proposed. Bayesian estimation methods for the LMSV model are considered by So (2002), Hsu & Breidt (2003) and Jensen (1983).

Table 1: Parameter vectors used for simulating observations in the Monte Carlo study

| $\psi$ | $\psi^{(1)}$ | $\psi^{(2)}$ | $\psi^{(3)}$ | $\psi^{(4)}$ | $\psi^{(5)}$ | $\psi^{(6)}$ | $\psi^{(7)}$ | $\psi^{(8)}$ |
|---|---|---|---|---|---|---|---|---|
| $d$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.1 | 0.2 | 0.3 | 0.4 |
| $\phi$ | – | – | – | – | 0.9 | 0.9 | 0.9 | 0.9 |
| $\sigma$ | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |

| $\psi^t$ | $\psi^{(1t)}$ | $\psi^{(2t)}$ | $\psi^{(3t)}$ | $\psi^{(4t)}$ | $\psi^{(5t)}$ | $\psi^{(6t)}$ | $\psi^{(7t)}$ | $\psi^{(8t)}$ |
|---|---|---|---|---|---|---|---|---|
| $d$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.1 | 0.2 | 0.3 | 0.4 |
| $\phi$ | – | – | – | – | 0.9 | 0.9 | 0.9 | 0.9 |
| $\sigma$ | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| $\nu$ | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |

## 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 Gaussian case, we obtain parameter vector $\psi = (d, \phi, \sigma)$ where $d = d_1$, $\phi = \phi_{1,1}$ and $\sigma^2 = \sigma_1^2$. In the Student's $t$ case, we have $\psi^t = (d, \phi, \sigma, \nu)$, where $\nu$ is the degrees of freedom for the Student's $t$-density.

Our Monte Carlo study is based on generating $n = 2000$ 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 $\sigma$ 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 & Jubinski (1999) and Ray & Tsay (2000). When simulating observations from the Student's $t$-density, we consider only $\nu = 10$ since other values for $\nu$ 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 $n = 2000$ and with a particular value of $\psi$ or $\psi^t$ 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 $x_1, \ldots, x_r$ 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 $i = 1, \ldots, r$, where unknown parameters $a$ and $b$ determine the shape and scale of the density, respectively. For an appropriately chosen threshold and 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 $n \to \infty$ 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 Figure 1, we report the test statistics based on the importance weights from the Yule-Walker and ARMA approximations, for the parameter values $\psi^{(i)}$ and $\psi^{(it)}$ from Table 1, for $i = 1, \ldots, 8$. 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 model (19) with different parameter settings. We set $\nu = 0$ to obtain the Gaussian LMSV model, while with $\nu > 2$ we obtain the model with Student's $t$ disturbances. For each parameter vector value from Table 1, we simulate 100 time series of length $n = 2000$ 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 $\psi$. For each element in $\psi$, we report the average estimation bias and standard

Figure 1: Importance sampling diagnostics for parameters $\psi^{(i)}$ and $\psi^{(it)}$, for $i = 1, \ldots, 8$, 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 $x_1, \ldots, x_r$ included. We have taken $0.01 = r/100000$, $0.025 = r/100000$, $0.05 = r/100000$, until $0.5 = r/100000$.
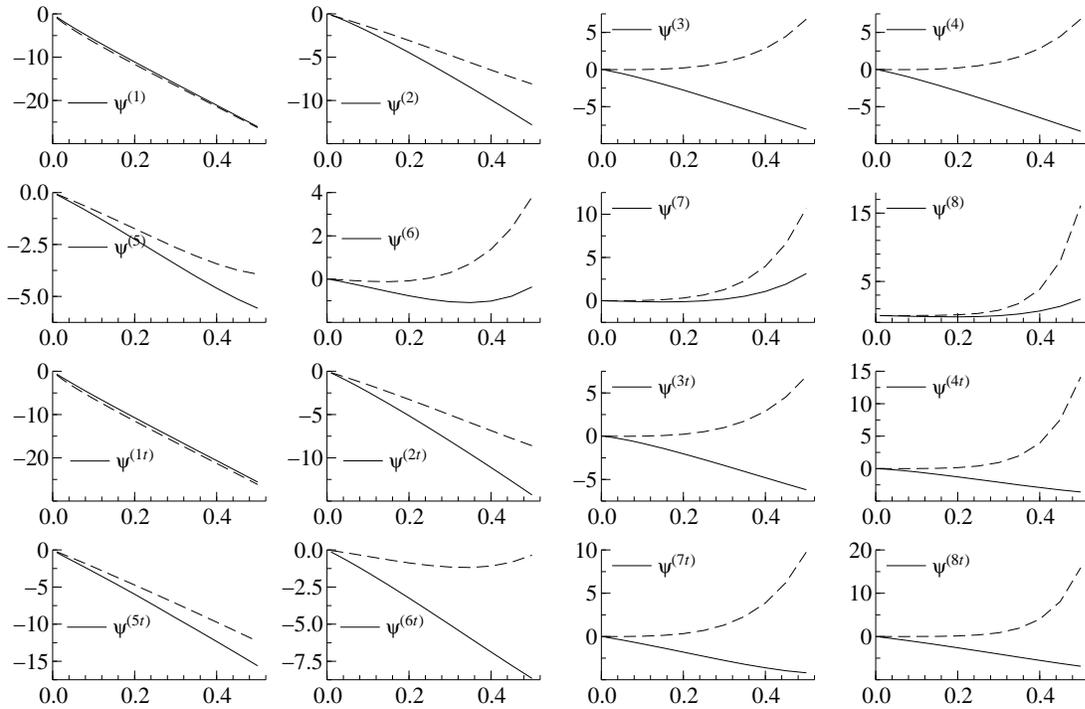
Table 2: Descriptive statistics of estimated parameters from the simulation study. For each parameter vector the average estimation bias and standard deviation (as subscript) is computed as follows. We sample 100 time-series of length $n = 2000$ from the univariate long-memory stochastic volatility model 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.

| $\psi$ | $\psi^{(1)}$ | $\psi^{(2)}$ | $\psi^{(3)}$ | $\psi^{(4)}$ |
|---|---|---|---|---|
| $d$ | -0.039 $_{0.153}$ | -0.018 $_{0.111}$ | -0.019 $_{0.098}$ | -0.033 $_{0.107}$ |
| $\phi$ | — | — | — | — |
| $\sigma$ | 0.001 $_{0.080}$ | -0.002 $_{0.048}$ | 0.008 $_{0.032}$ | -0.011 $_{0.064}$ |

| $\psi$ | $\psi^{(5)}$ | $\psi^{(6)}$ | $\psi^{(7)}$ | $\psi^{(8)}$ |
|---|---|---|---|---|
| $d$ | -0.018 $_{0.125}$ | -0.022 $_{0.145}$ | -0.036 $_{0.109}$ | -0.032 $_{0.106}$ |
| $\phi$ | -0.001 $_{0.035}$ | -0.009 $_{0.065}$ | 0.000 $_{0.045}$ | 0.002 $_{0.047}$ |
| $\sigma$ | 0.031 $_{0.052}$ | 0.024 $_{0.051}$ | 0.011 $_{0.043}$ | 0.017 $_{0.042}$ |

| $\psi$ | $\psi^{(1t)}$ | $\psi^{(2t)}$ | $\psi^{(3t)}$ | $\psi^{(4t)}$ |
|---|---|---|---|---|
| $d$ | -0.031 $_{0.123}$ | -0.004 $_{0.135}$ | -0.016 $_{0.111}$ | -0.012 $_{0.101}$ |
| $\phi$ | — | — | — | — |
| $\sigma$ | 0.021 $_{0.061}$ | 0.008 $_{0.072}$ | 0.031 $_{0.034}$ | 0.001 $_{0.059}$ |
| $\nu$ | 0.461 $_{3.169}$ | 0.215 $_{4.126}$ | 0.345 $_{3.059}$ | 0.826 $_{2.180}$ |

| $\psi$ | $\psi^{(5t)}$ | $\psi^{(6t)}$ | $\psi^{(7t)}$ | $\psi^{(8t)}$ |
|---|---|---|---|---|
| $d$ | -0.009 $_{0.102}$ | -0.012 $_{0.122}$ | -0.029 $_{0.102}$ | -0.024 $_{0.098}$ |
| $\phi$ | 0.003 $_{0.047}$ | 0.002 $_{0.050}$ | 0.004 $_{0.040}$ | 0.002 $_{0.038}$ |
| $\sigma$ | 0.013 $_{0.062}$ | -0.004 $_{0.041}$ | 0.005 $_{0.042}$ | 0.003 $_{0.033}$ |
| $\nu$ | 1.226 $_{4.543}$ | 1.920 $_{4.586}$ | 0.850 $_{2.981}$ | -0.150 $_{2.485}$ |

deviation in Table 2. To make a comparison with the standard SV model, we also consider the settings $d = 0$ and $\nu = 0$ to obtain the Gaussian SV model and $d = 0$ and $\nu > 2$ to obtain the Student's $t$ SV model. For the parameter vectors $\psi^{(8)}$ and $\psi^{(8t)}$, we present the resulting sample histograms in Figure 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.

Figure 2: Sample densities of estimated parameters in the Monte Carlo study. We present a $4 \times 4$ matrix of density plots from a sample of 100 estimates of parameters in the SV model (19). The four columns are associated with parameters $d$, $\phi$, $\sigma^2$ and $\nu$, respectively. The four rows are associated with the LMSV model with $\nu = 0$ (LMSV Gaussian), $\nu > 2$ (LMSV Student's $t$), $d = 0, \nu = 0$ (SV Gaussian) and $d = 0, \nu > 2$ (SV Student's $t$), respectively. The simulations are based on parameter vectors $\psi^{(8)}$: $d = 0.4$, $\phi = 0.9$, $\sigma = 0.2$ and $\psi^{(8t)}$: $d = 0.4$, $\phi = 0.9$, $\sigma = 0.2$, $\nu = 10$.
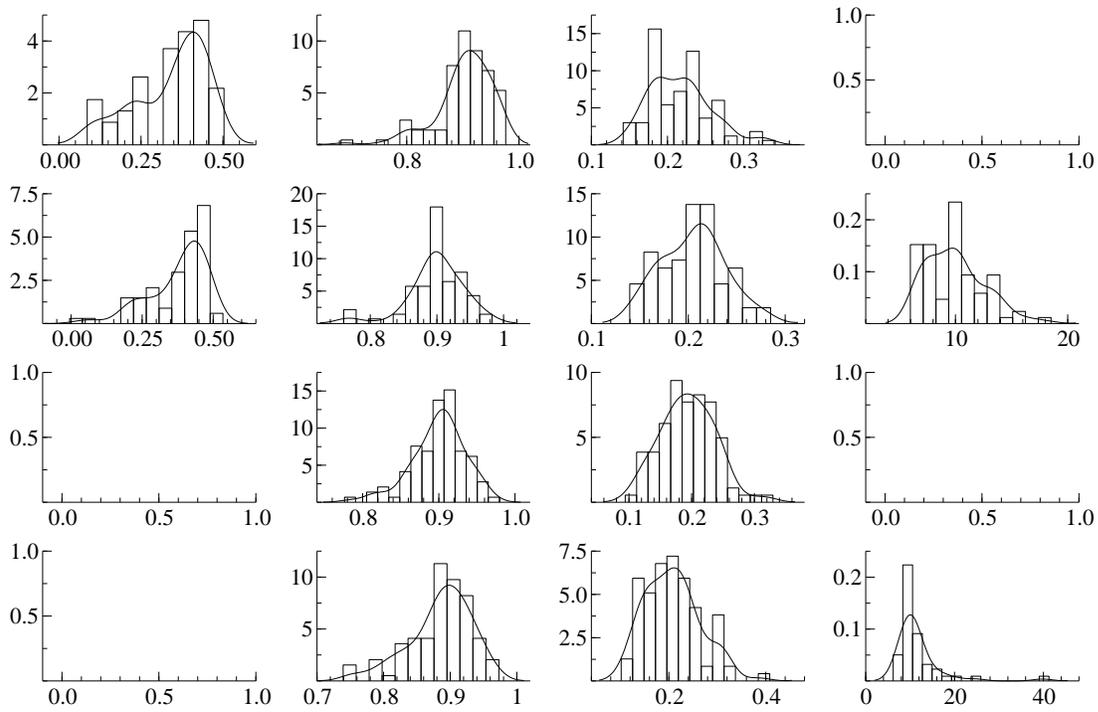
## 5.5 Empirical evidence for ten S&P 500 stocks

The presence of long-memory in the volatility of the S&P 500 stock index is documented in Ding, Granger & Engle (1993) and further analyzed for different components of the index by Ray & 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 Figures 3.a and 3.b we present, for $t = 1, \ldots, n = 1745$, 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 & Kwok (2006). Most of the sample autocorrelation functions for $\log y_t^2$ show hyperbolic 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 4.a and 4.b. We consider the unrestricted LMSV model, the LMSV model with restriction $d = 0$ and the LMSV model with restriction $\phi = 0$. The three LMSV models are also considered with the Gausian density for $\xi_t$ replaced by the Student's $t$ density (denoted by LMSV-$t$).

For all time series the log-likelihood value is highest for the LMSV-$t$ model where no restrictions are placed on $d$ or $\phi$. 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 $\phi$ 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 $\nu$ are often low indicating substantially larger tails compared to the Gaussian distribution. In Figure 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 4.a and 4.b. The computations are given by (18) and based on $M = 400$ simulations from the importance density. Many of the estimated log-volatilities in Figure 4 have similar patterns. We may therefore consider the reduction of the number of volatility components underlying the 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 the next section.

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 |

Figure 3.a: Data descriptives on constituents 1 to 5 by market cap of the S&P 500 stock index. Each series contains $n = 1745$ 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, $\log y_t^2$.
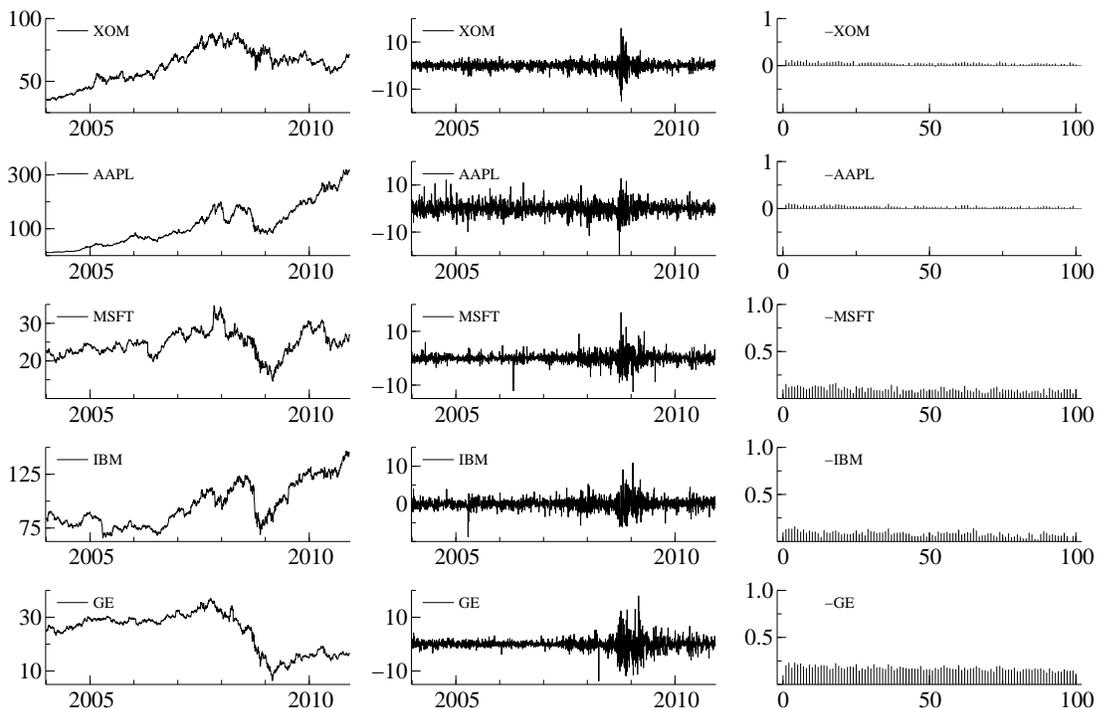
Figure 3.b: Data descriptives on constituents 6 to 10 by market cap of the S&P 500 stock index. Each series contains $n = 1745$ trading days between 01-01-2004 and 06-12-2010. The columns show adjusted daily closing prices, $P_t$; the daily differences of the log adjusted closing prices, $y_t$; and the sample autocorrelation function of the squared daily differences of the log adjusted closing prices, $\log y_t^2$.
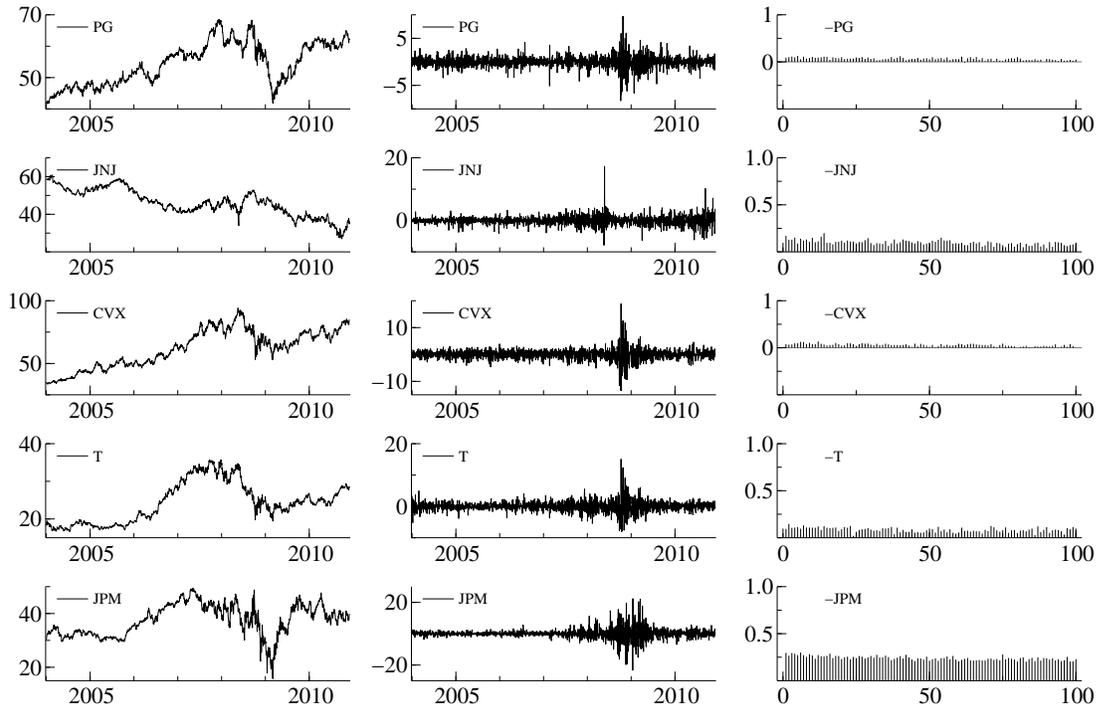
Table 4.a: Estimated parameters for univariate LMSV and LMSV-$t$ models using log-returns from constituents 1 to 5 by market cap of the S&P 500 index, for days from 01-01-2004 to 6-12-2010 ($n = 1745$ trading days). The standard errors of the estimates are given as subscripts. The method of Monte Carlo 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 $\tilde{\psi}$.

| | $d$ | $\phi$ | $\sigma$ | $\nu$ | $\log \tilde{\ell}(\tilde{\psi})$ | AIC |
|---|---|---|---|---|---|---|
| | | | XOM | | | |
| 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 |
| 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 |
| | | | AAPL | | | |
| 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 |
| 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 |
| | | | MSFT | | | |
| 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 |
| 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 |
| | | | IBM | | | |
| 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 |
| 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** |
| | | | GE | | | |
| 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 |
| 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** |

Table 4.b: 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 ($n = 1745$ 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 $\tilde{\psi}$.

| | $d$ | $\phi$ | $\sigma$ | $\nu$ | $\log\tilde{\ell}(\tilde{\psi})$ | AIC |
|---|---|---|---|---|---|---|
| | | | PG | | | |
| 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 |
| 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 |
| | | | JNJ | | | |
| 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 |
| 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** |
| | | | CVX | | | |
| 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 |
| 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 |
| | | | T | | | |
| 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 |
| 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** |
| | | | JPM | | | |
| 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 |
| 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 |

Figure 4: Estimated volatility paths $\exp(\tilde{X}_t)$ of constituents 1 to 10 by market cap of the S&P 500 stock index. Log-variance $X_t$ follows the ARFIMA process of equation (3) with $p = 1$ and $q = 0$. Log-returns $y_t$ are modelled by equation (19) where $\xi_t$ 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 4.a and 4.b and $M = 400$ simulations from the importance density.
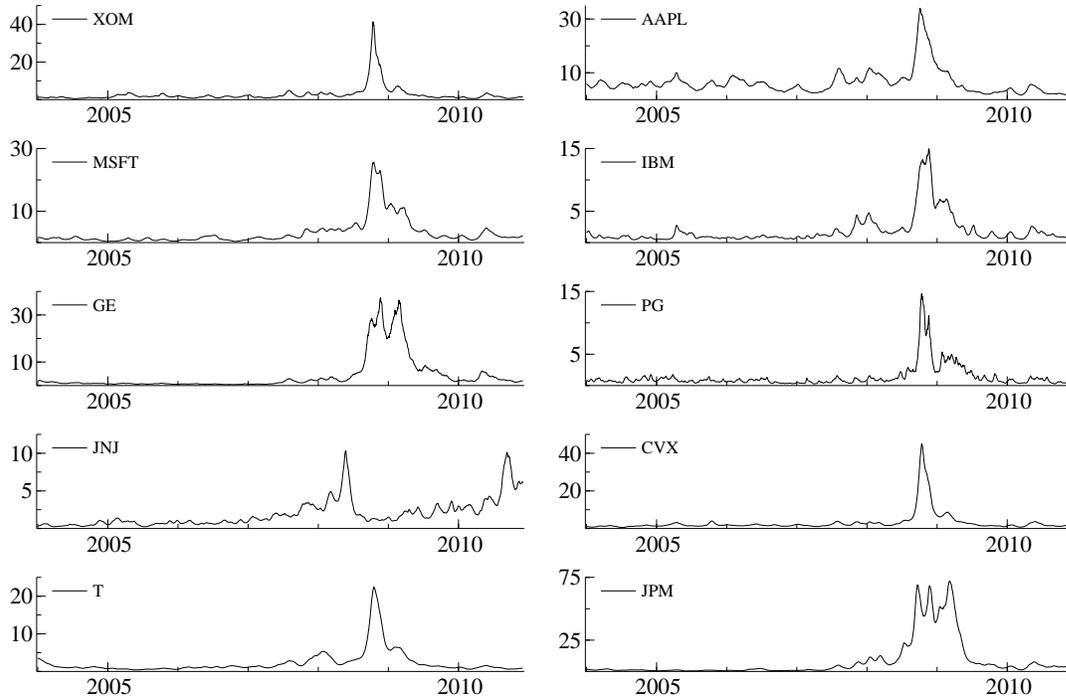
# 6  Multivariate long-memory stochastic volatility

Different specifications for a multivariate extension of the stochastic volatility model can be considered, see, for example, Asai, McAleer & Yu (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 $y_t = (y_{1t}, \ldots, y_{kt})'$ for $t = 1, \ldots, n$. 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

$$y_t = \mathcal{Z}_t \xi_t, \qquad \xi_t \sim \mathrm{N}(0, \Sigma_\xi), \qquad t = 1, \ldots, n, \tag{20}$$

where $\mathcal{Z}_t$ is a $k \times k$ diagonal matrix with elements $\exp(Z_{it}/2)$, for $i = 1, \ldots, k$, on the diagonal. Latent $k \times 1$ vector $Z_t$, as given in equation (2), allows for the variance to depend on $l \times 1$ vector $X_t$ of independent long-memory processes, where holds $l << k$. We restrict $\mathcal{B} = 0$ since our main interest is in long-memory components. The $k \times 1$ disturbance vector $\xi_t$ 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 rapildly 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 $\mathrm{Var}(X_{it}) = 1$ for $i = 1, \ldots, l$. 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 & Tsay (2000). The estimation of the parameters is based on quasi-maximum likelihood and spectral regression methods, see also So & Kwok (2006).

We continue our study with the ten volatility series from the S&P500 index. We study

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 ($n = 1745$). Log-variances $X_{it}$ are specified by equation (3), with $p = 1$ and $q = 0$. The log-returns vector $y_t$ is modelled by equation (20) using Student's $t$ distributions for $\xi_t$. Estimates are computed as described in section 4.5 using optimized parameters from Table 5 and $M = 400$ simulations from the importance density.
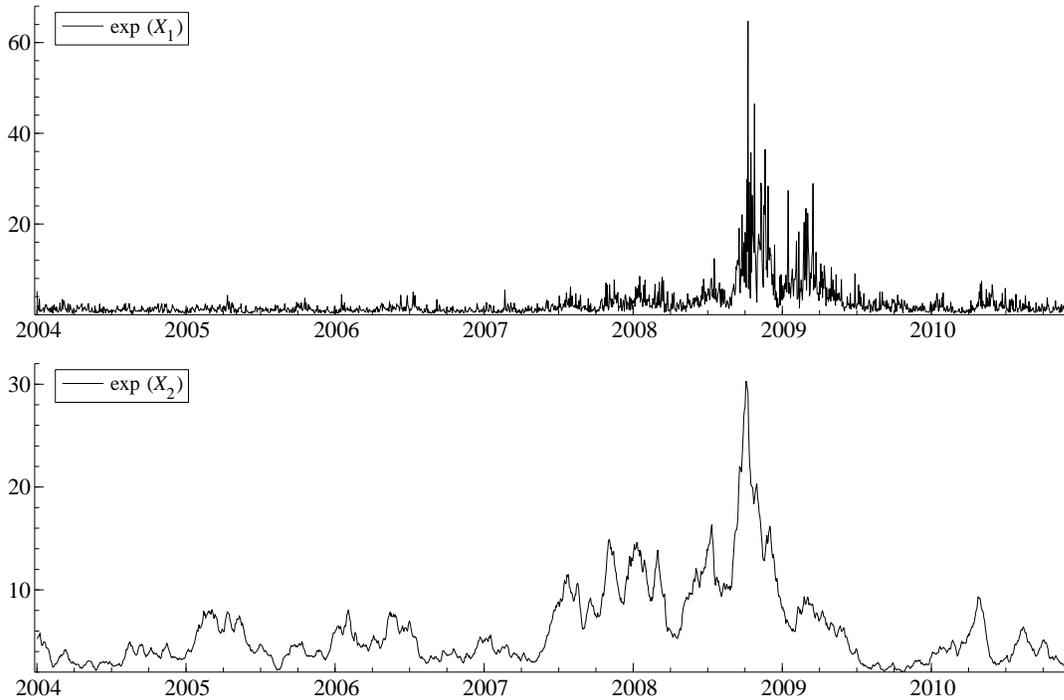


the $10 \times 1$ vector $y_t = (y_{1t}, \ldots, y_{10,t})'$ simultaneously with $y_{it} = 100 \log(P_{i,t} / \log P_{i,t-1})$ and $P_{it}$ is the daily adjusted closing price of stock $i$. The stocks are ordered as in Table 3. We implement one and two factor versions of the model and for each $X_{it}$ 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 $\xi_t$ normally and Student's $t$ distributed, respectively, are presented in Table 5. 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 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 Figure 5. The first factor is clearly more noisy while the other factor may represent more long-term changes in volatility.

Table 5: 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 ($n = 1745$). 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 $\tilde{\psi}$.

| | LMSV | | LMSV-$t$ | |
| --- | --- | --- | --- | --- |
| | $l = 1$ | $l = 2$ | $l = 1$ | $l = 2$ |
| $d_1$ | $0.462_{\ 0.031}$ | $0.452_{\ 0.034}$ | $0.463_{\ 0.032}$ | $0.461_{\ 0.036}$ |
| $d_2$ | - | $0.376_{\ 0.045}$ | - | $0.454_{\ 0.105}$ |
| $\phi_1$ | $-0.266_{\ 0.060}$ | $-0.206_{\ 0.076}$ | $-0.251_{\ 0.074}$ | $-0.209_{\ 0.083}$ |
| $\phi_2$ | - | $0.823_{\ 0.122}$ | - | $0.943_{\ 0.192}$ |
| $\nu$ | - | - | $5.943_{\ 0.268}$ | $10.511_{\ 1.161}$ |
| $a_{1,1}$ | 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** |

# 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.

# References

Anderson, B. D. O. & Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall, Englewood Cliffs.

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.

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

Bollerslev, T. & Jubinski, D. (1999), 'Equity trading volume and volatility: latent information 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.

Davidson, R. & MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press, New York.

de Jong, P. & Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**, 339–350.

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. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

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.

Harvey, A. C. (1998), Long memory in stochastic volatility, *in* T. Knight & S. Satchell, eds, 'Forecasting Volatility in Financial Markets', Butterworth-Heineman, Oxford, 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., 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 University Press, Cambridge.

Nocedal, J. & Wright, S. J. (1999), *Numerical Optimization*, Springer Verlag, New York.

Palma, W. (2007), *Long Memory Time Series*, John Wiley & Sons, New York.

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*, John Wiley & Sons, New York.

Robinson, P. (1994), Time series with strong dependence, *in* C. A. Sims, ed., 'Advances in Econometrics, Sixth World Congress', Cambridge University Press, Cambridge, pp. 47–95.

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.