TI 2012-008/4
Tinbergen Institute Discussion Paper

Modified Efficient Importance Sampling using State Space Methods

Siem Jan Koopman[1]
Rutger Lit[1]
Thuy Minh Nguyen[2]

[1] *Faculty of Economics and Business, VU University Amsterdam, and Tinbergen Institute, the Netherlands;*
[2] *Deutsche Bank, London, United Kingdom.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Modified Efficient Importance Sampling using State Space Methods*

Siem Jan Koopman[(a,b)], *Rutger Lit*[(a)] *and* *Thuy Minh Nguyen*[(c)]

[(a)]Department of Econometrics, VU University Amsterdam & Tinbergen Institute [(b)]CREATES, Aarhus University, Denmark

[(c)]Deutsche Bank, London

Abstract

A successful construction of an importance density for nonlinear non-Gaussian state space models is crucial when Monte Carlo simulation methods are used for likelihood evaluation, signal extraction of dynamic latent factors and forecasting. The method of efficient importance sampling is successful in this respect but we show that it can be implemented more conveniently using standard Kalman filter and smoothing methods. We further obtain computational gains by simulating directly from the signal equation rather than simulating from the usually higher dimensional state equation. Our results provide some new insights but they primarily lead to a more simple and fast method for efficient importance sampling. In a simulation study we provide some evidence of the computational gains. Our new methods are illustrated for a stochastic volatility model with a Student's t distribution.

Key words: Kalman filter; Monte Carlo maximum likelihood; Nonlinear Non-Gaussian state space models; Simulation smoothing; Stochastic Volatility.

*We would like to thank André Lucas, Marcel Scharth and István Barra for giving comments on an earlier version of this paper. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78) at Aarhus University, Denmark, and funded by the Danish National Research Foundation. The views expressed in this paper are those of the authors and do not necessarily reflect those of Deutsche Bank. The usual disclaimer applies. Corresponding email : `s.j.koopman@vu.nl`

1 Introduction

For the modelling of an observed time series $y_1, \ldots, y_n$, we consider a parametric model that we formulate conditionally on a dynamic latent factor or a time-varying parameter vector α_t, for time index $t = 1, \ldots, n$. The conditional model for the observations is given by

$$y_t|\alpha_t \overset{ind}{\sim} p(y_t|\alpha_t; \psi), \qquad t = 1, \ldots, n, \tag{1}$$

where $\overset{ind}{\sim}$ is notation for serially independently distributed, ψ is an unknown and fixed parameter vector, and $p(y_t|\alpha_t; \psi)$ is an observation density that is possibly non-Gaussian and may represent a nonlinear relation between y_t and α_t. The density $p(y_t|\alpha_t; \psi)$ is for a random variable y_t that is conditional on the latent random variable α_t and is a function of parameter vector ψ. Conditional on the sequence $\alpha_1, \ldots, \alpha_n$, the observations $y_1, \ldots, y_n$ are serially independently distributed. The time-varying parameters in α_t can represent different features of the model including mean, variance and regression effects that may not be constant over time. Different dynamic specifications for the parameters in α_t can be adopted. In our analysis, the conditional observation density and the dynamic model for α_t must be specified and both may depend on the fixed parameter vector ψ.

When (i) the observation density for y_t conditional on α_t is Gaussian, (ii) the relation between y_t and α_t is linear and (iii) the dynamic model for α_t is linear and Gaussian, our time series modelling framework reduces to the linear Gaussian state space model as discussed and reviewed, for example, in Durbin and Koopman (2012, Part I). In this framework, we can rely on the celebrated Kalman filter and its related smoothing method for the signal extraction of α_t, the evaluation of the likelihood function for a specific value of ψ and the forecasting of y_t. These methods provide minimum mean squared error (MMSE) estimates for α_t and MMSE forecasts for y_t under correct model specification. Such optimal properties for estimates produced by the Kalman filter methods are not obtained when we depart from one of the three given assumptions above. The resulting nonlinear non-Gaussian state space models require other methods to obtain the optimal estimates. In almost all cases of practical interest, we require numerical methods which often pose various computational challenges. For example, to evaluate the likelihood function for a nonlinear non-Gaussian state space

model, we cannot use the Kalman filter or a related analytical filter. We need to evaluate the high-dimensional integrals of the likelihood function directly using numerical methods. In this paper we focus on Monte Carlo methods, in particular on importance sampling methods.

The general ideas of importance sampling are established in statistics and econometrics, see Kloek and van Dijk (1978), Ripley (1987), and Geweke (1989). Importance sampling techniques for state space models have been explored by Danielsson and Richard (1993), Shephard and Pitt (1997), Durbin and Koopman (1997), So (2003) and Jungbacker and Koopman (2007). A textbook treatment is given by Durbin and Koopman (2012, Part II). The performance of the Monte Carlo estimation method relies on the successful construction of an importance density. Several methods for designing an importance density for time series modelling have been proposed. For example, Shephard and Pitt (1997) and Durbin and Koopman (1997) adopt an importance density based on the mode of the conditional density.

In this paper we consider the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003), Richard and Zhang (2007) and Jung, Liesenfeld, and Richard (2011) where the sampling is based on a global approximation of the original model. We show that EIS can be implemented using standard Kalman filter methods. It leads to a simple and fast procedure for efficient importance sampling. We discuss how our modified efficient importance sampling (MEIS) procedure is related to the procedure of Shephard and Pitt (1997) and Durbin and Koopman (1997), hereafter referred to as SPDK. It is shown by Koopman, Lucas, and Scharth (2014) that similar modifications can also be used for the introduction of numerical integration methods in importance sampling.

The remainder of the paper is organized as follows. In Section 2 we briefly introduce our generalized time-varying parameter model. In Section 3 we introduce our new MEIS method by reviewing the EIS method for constructing the importance density and show how the EIS method can be implemented using state space methods. In Section 4 we discuss parameter estimation, signal extraction and forecasting. The simulation study presented in Section 5 provides the evidence of the computational efficiency gains. In Section 6 we apply our new methodology to four time series of financial returns which are analysed on the basis of the stochastic volatility model with a Student's t density. The empirical results are of interest generally when analyzing volatilities in stock markets. Section 7 concludes.

2 A generalized time-varying parameter model

The dynamic model specification under consideration is given by the observation density $p(y_t|\alpha_t;\psi)$ as introduced in (1) and with the stochastically time-varying parameter vector α_t. The linear Gaussian dynamic process for α_t is given by

$$\alpha_{t+1} = d_t + T_t\alpha_t + R_t\eta_t, \qquad \eta_t \overset{ind}{\sim} \mathrm{N}(0, Q_t), \qquad \alpha_1 \sim \mathrm{N}(a_1, Q_0), \tag{2}$$

where the elements of the transition matrix T_t, the selection matrix R_t, and the variance matrices Q_t and Q_0 are known except that some elements have a possible dependence on parameter vector ψ, for $t = 1, \ldots, n$. The disturbances η_t are normally and independently distributed and do not depend on the normally distributed initial state vector α_1. All stochastic and non-stochastic variables have appropriate dimensions and the dimensions will only be given when it is necessary. The observation y_t is typically a scalar but the methods presented in Section 3 are also applicable for a vector of observations y_t. Illustrations of special cases of our general modelling framework are given below.

2.1 Signal plus heavy-tailed noise model

When the time series observations y_t are randomly contaminated by noise with large shocks, we may wish to remove the noise from the signal and to model the noise explicitly by a heavy-tailed density. We then may consider the model

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \qquad \varepsilon_t \overset{ind}{\sim} \tau(0, \sigma^2, \nu), \qquad t = 1, \ldots, n, \tag{3}$$

where signal function $Z_t(\cdot)$ is fixed and known, and may also depend on the parameter vector ψ, the stochastically time-varying state vector α_t is specified in (2) and $\tau(\mu, \sigma^2, \nu)$ is the Student's t density with mean μ, variance σ^2 and degrees of freedom ν. The model clearly fits in our general framework with observation equation (1) given by

$$p(y_t|\alpha_t;\psi) = \tau(\theta_t, \sigma^2, \nu), \qquad \theta_t = Z_t(\alpha_t), \qquad t = 1, \ldots, n,$$

where θ_t is the signal. Other heavy tailed densities for ε_t can also be considered.

A well-known example of a signal plus noise model is the basic structural time series model of Harvey (1989) in which the univariate time series y_t can be decomposed into trend, seasonal and noise components,

$$y_t = \mu_t + \gamma_t + \varepsilon_t,$$

where μ_t is the trend component, γ_t is the seasonal component, and both are elements of the state vector α_t. The dynamic specifications of the two components can be formulated jointly in the form of (2). The signal for this model is linear and is given $\theta_t = Z_t(\alpha_t) = \mu_t + \gamma_t$ for $t = 1, \ldots, n$. When the noise component ε_t is normally distributed, the time series analysis, including the maximum likelihood estimation of ψ, the signal extraction of θ_t (or, additionally, μ_t and γ_t) and the forecasting of y_t can be based on the Kalman filter and related methods. However, when the noise ε_t is non-Gaussian as in (3), the resulting estimates of the Kalman filter have no longer optimal properties. For such cases, we adopt importance sampling methods.

2.2 Stochastic volatility model

A time series of financial returns is often subject to clusters of volatility changes which can effectively be modelled by a dynamic process for the variance. A basic version of the stochastic volatility model for a time series of returns y_t is given by

$$y_t = \mu + \exp(\frac{1}{2}\theta_t)\varepsilon_t, \qquad \theta_t = Z_t(\alpha_t), \qquad \varepsilon_t \overset{ind}{\sim} \mathrm{N}(0, \sigma^2), \qquad t = 1, \ldots, n, \tag{4}$$

where μ is a constant, the signal θ_t represents the time-varying log-variance of y_t, and ε_t is the normally distributed noise term. The specification for θ_t can be formulated as in the previous illustration. However, a more appropriate formulation for the signal is the stationary process

$$Z_t(\alpha_t) = \sum_{j=1}^{p} \alpha_{jt}, \qquad \alpha_{j,t+1} = \phi_j \alpha_{jt} + \eta_{jt}, \tag{5}$$

where α_{jt} is the jth element of α_t with autoregressive coefficient $0 < \phi_j < 1$, and η_{jt} is the jth element of η_t, for $j = 1, \ldots, p$, and where α_t and η_t are specified as in (2), for $t = 1, \ldots, n$. It follows that the transition matrix T_t and the variance matrix Q_t in (2) are diagonal matrices, for $t = 1, \ldots, n$, with their jth diagonal elements equal to ϕ_j and $\sigma^2_{\eta,j}$, for $j = 1, \ldots, p$. The

conditional observation density (1), in logs, for this stochastic volatility model is given by

$$\log p(y_t|\theta_t;\psi) = -\frac{1}{2}\log(2\pi\,\sigma^2) - \frac{1}{2}\theta_t - \frac{1}{2\sigma^2}\exp(-\theta_t)(y_t-\mu)^2, \qquad t = 1,\ldots,n.$$

The stochastic volatility model can be extended in many ways. For example, leverage effects can be introduced by having dependence between ε_t and η_{jt}, for $j = 1,\ldots,p$. Also, heavy-tailed density functions can be considered for the noise term ε_t. We refer to Shephard (2005) for extensive discussions on stochastic volatility models.

2.3 Time-varying model for counts

Time series of counts can be modelled by the Poisson density with the intensity parameter as a function of the time-varying signal θ_t that we can specify as (1) and (2). The observation log-density function is then given by

$$\log p(y_t|\alpha_t;\psi) = y_t\log\theta_t - \theta_t - \log(y_t!), \qquad \theta_t = Z_t(\alpha_t), \qquad t = 1,\ldots,n, \tag{6}$$

where the signal θ_t is defined in the same way as in the earlier illustrations. Other densities from the exponential family can also be considered such as the Binomial distribution, the negative Binomial distribution and the Skellam distribution (difference of two Poisson variables).

3 Modified efficient importance sampling

We discuss the EIS and MEIS methods by considering likelihood evaluation. For a discussion about other applications in which MEIS plays an important role including signal extraction, maximum likelihood estimation of ψ and forecasting of future observations y_t we refer to Section 4.

The likelihood function of the model (1) and (2) for the observed vector $y = (y_1',\ldots,y_n')'$ and as a function of parameter vector ψ is given by

$$L(y;\psi) = \int p(y,\alpha;\psi)\,\mathrm{d}\alpha = \int p(y|\alpha;\psi)p(\alpha;\psi)\,\mathrm{d}\alpha, \tag{7}$$

where $\alpha = (\alpha_1', \ldots, \alpha_n')'$. Analytical expressions for the typically high-dimensional integral are only available in specific cases. An example is the linear Gaussian state space model for which the Kalman filter can be used to evaluate the likelihood value for a given value of ψ. Numerical evaluation is usually dismissed because of the high dimensional vector α. A Monte Carlo evaluation of the likelihood function is often explored as a feasible alternative. A basic version of a Monte Carlo estimate of (7) is

$$\widehat{L}(y;\psi) = \sum_{i=1}^{M} p(y|\alpha^{(i)};\psi), \qquad \alpha^{(i)} \sim p(\alpha;\psi), \tag{8}$$

where $\alpha^{(i)}$ refers to the ith simulated sample of α that is generated from the unconditional density $p(\alpha;\psi)$ with $i = 1, \ldots, M$. The standard law of large numbers (LLN) insists that $\widehat{L}(y;\psi)$ converges to $L(y;\psi)$ as $M \to \infty$. Since the simulation of α has no reference to data vector y, the efficiency of the estimate is very low and we need M to be extremely large.

An efficient Monte Carlo method for the evaluation of integrals such as (7) is based on importance sampling techniques. Simulation-based methods are explored in statistics for different models and purposes. We will review importance sampling methods in the context of time series models. It is useful to express the likelihood function (7) in terms of individual time series observations. Given the serial independence properties for the observations y_t conditional on α_t and for the disturbances η_t, we have

$$L(y;\psi) = \int \left[\prod_{t=1}^{n} p(y_t|\alpha_t;\psi) p(\alpha_t|\alpha_{t-1};\psi) \right] \mathrm{d}\alpha_t, \tag{9}$$

with $p(\alpha_1|\alpha_0;\psi) = p(\alpha_1;\psi)$. We also have $p(\alpha_t|\alpha_{t-1};\psi) = p(\eta_{t-1};\psi) = \mathrm{N}(0, Q_{t-1})$ for $t = 1, \ldots, n$.

3.1 Importance density

For an introduction to Monte Carlo simulation methods and in particular the technique of importance sampling, we refer to Ripley (1987). To evaluate (9) via importance sampling, we introduce the importance density based on the linear Gaussian joint density $g(y, \alpha;\psi)$ with properly defined mean vector and variance matrix. The dependence of ψ originates

from our model (1) and (2). We adopt the decomposition $g(y, \alpha; \psi) = g(y|\alpha; \psi)\, g(\alpha; \psi)$ and have

$$g(y|\alpha;\psi) = \prod_{t=1}^{n} g(y_t|\alpha_t;\psi), \qquad g(\alpha;\psi) = p(\alpha;\psi) = \prod_{t=1}^{n} p(\alpha_t|\alpha_{t-1};\psi) = \prod_{t=1}^{n} p(\eta_{t-1};\psi),$$

where $\eta_{t-1} \overset{ind}{\sim} \mathrm{N}(0, Q_{t-1})$, for $t = 1, \ldots, n$. Since the dynamic specification for the state vector in (2) is linear and Gaussian, the equality $g(\alpha; \psi) = p(\alpha; \psi)$ is valid. The Gaussian observation density can be expressed by

$$g(y_t|\alpha_t;\psi) = \exp\left(a_t + b_t\theta_t - \frac{1}{2}c_t\theta_t^2\right), \qquad t = 1, \ldots, n. \tag{10}$$

where $\theta_t = Z_t(\alpha_t)$ represents the signal and with $Z_t(\alpha_t)$ being the link function that connects the observation y_t with the state vector α_t. Examples of link functions for the signal θ_t are presented in the illustrations of Section 2. It follows that the variables b_t and c_t are functions of the observations $y_1, \ldots, y_n$ and parameter vector ψ for $t = 1, \ldots, n$. The constant a_t ensures that $g(y_t|\alpha_t; \psi)$ integrates to unity and hence it is a deterministic function of b_t and c_t. An effective importance sampler is obtained by selecting appropriate values for b_t and c_t for $t = 1, \ldots, n$. The design of the importance sampler is elegantly reduced to a choice for b_t and c_t that determine the mean and variance implied by $g(y_t|\alpha_t; \psi)$.

The importance density $g(y_t|\alpha_t; \psi)$ can alternatively be expressed in terms of constructed variable $x_t = b_t/c_t$ and the linear and Gaussian model

$$x_t = \theta_t + u_t, \qquad u_t \overset{ind}{\sim} \mathrm{N}\left(0, \frac{1}{c_t}\right), \qquad t = 1, \ldots, n, \tag{11}$$

which has the conditional observation log-density function

$$\log g(x_t|\alpha_t;\psi) = -\frac{1}{2}\log 2\pi + \frac{1}{2}\log c_t - \frac{1}{2}\left(x_t - \theta_t\right)^2 c_t$$

for $t = 1, \ldots, n$. The artificial variable x_t is defined as $x_t = b_t/c_t$ and we can substitute it

into the observation log-density to obtain

$$\begin{aligned}\log g(x_t|\alpha_t;\psi) &= -\frac{1}{2}\log 2\pi + \frac{1}{2}\log c_t - \frac{1}{2}\left(b_t/c_t - \theta_t\right)^2 c_t \\ &= a_t + b_t\theta_t - \frac{1}{2}c_t\theta_t^2,\end{aligned}$$

where the constant a_t collects all terms that are not associated with θ_t and is given by

$$a_t = -\frac{1}{2}\left(\log 2\pi - \log c_t + b_t^2/c_t\right).$$

It follows immediately that

$$\log g(y|\alpha;\psi) = \sum_{t=1}^{n}\log g(y_t|\alpha_t;\psi) = \sum_{t=1}^{n}\log g(x_t|\alpha_t;\psi) = \sum_{t=1}^{n}\left(a_t + b_t\theta_t - \frac{1}{2}c_t\theta_t^2\right),$$

where it is assumed that x_t is modelled by (11) for $t = 1, \ldots, n$. This is a key result for the developments below. It implies that the analysis concerning $g(y, \alpha; \psi) = g(y|\alpha;\psi)g(\alpha;\psi)$ can be based on the model (11) for which we can use the Kalman filter and related methods.

3.2 Likelihood evaluation via importance sampling

The actual importance density for the evaluation of (7) is chosen as

$$g(\alpha|y;\psi) = g(y|\alpha;\psi)\, g(\alpha;\psi) \,/\, g(y;\psi),$$

where $g(\alpha;\psi) = p(\alpha;\psi)$. The likelihood function (7) with the importance sampling density incorporated is given by

$$L(y;\psi) = \int \frac{p(y|\alpha;\psi)p(\alpha;\psi)}{g(\alpha|y;\psi)} g(\alpha|y;\psi)\,\mathrm{d}\alpha.$$

After some minor manipulations, we can express the likelihood function as

$$L(y;\psi) = g(y;\psi)\int\left[\prod_{t=1}^{n} w(y_t,\alpha_t;\psi)\right] g(\alpha|y;\psi)\,\mathrm{d}\alpha, \qquad w(y_t,\alpha_t;\psi) = \frac{p(y_t|\alpha_t;\psi)}{g(y_t|\alpha_t;\psi)}, \tag{12}$$

where $w(y_t, \alpha_t; \psi)$ is referred to as the importance weight, for $t = 1, \ldots, n$.

The evaluation of the likelihood function by means of importance sampling takes place by simulating state vectors from the importance density $g(\alpha|y;\psi)$ which we denote by

$$\alpha^{(i)} = \left(\alpha_1^{(i)\prime}, \ldots, \alpha_n^{(i)\prime}\right)' \sim g(\alpha|y;\psi), \qquad i = 1, \ldots, M,$$

where vector $\alpha^{(i)}$ is independently drawn for $i = 1, \ldots, M$. Since we can represent $g(y, \alpha; \psi)$ by the linear Gaussian state space model (11) and (2), we can simulate α from the conditional state density $g(\alpha|y;\psi)$ via the simulation smoothing method; see, for example, Fruhwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). Given the simulated realisations $\alpha^{(i)}$, for $i = 1, \ldots, M$, the likelihood function is computed by

$$\widehat{L}(y;\psi) = g(y;\psi)M^{-1}\sum_{i=1}^{M}\prod_{t=1}^{n} w_{it}, \qquad w_{it} = w(y_t, \alpha_t^{(i)};\psi). \tag{13}$$

Some practical issues on computing $\widehat{L}(y;\psi)$ for the purpose of estimating parameters are discussed in Section 4.1. It can be expected that the Monte Carlo estimate (13) is more efficient than the estimate (8) since we simulate $\alpha_t^{(i)}$ with a reference to the data vector y.

3.3 Implementation of modified efficient importance sampling

Here we introduce our modified efficient importance sampling method. It is based on the efficient importance sampling of Richard and Zhang (2007) but we show that the method can be implemented using the simulation smoothing method. Hence we obtain a computationally fast and more convenient implementation of efficient importance sampling for time series models.

The values for b_t and c_t, with $t = 1, \ldots, n$, need to be determined before the calculation of (13) can start. Here we follow Richard and Zhang (2007) and adopt their efficient importance sampling method. They propose to choose b_t and c_t such that the criterion

$$I_t = \int \lambda_t^2(y_t, \alpha_t; \psi) p(y_t, \alpha_t; \psi)\, \mathrm{d}\alpha_t, \qquad \lambda_t(y_t, \alpha_t; \psi) = \log w(y_t, \alpha_t; \psi) - \bar{\lambda}_t, \tag{14}$$

is minimized for each t separately and where $\bar{\lambda}_t$ is the normalizing constant such that the

expectation of λ_t with respect to the true model $p(y_t, \alpha_t; \psi)$ is zero. We therefore interpret I_t as the variance of the logged importance weight function with respect to $p(y_t, \alpha_t; \psi)$. We notice that the variables b_t and c_t determine $g(y_t|\alpha_t; \psi)$ that is part of $w(y_t, \alpha_t; \psi)$ and hence of $\lambda_t(y_t, \alpha_t; \psi)$. The function I_t cannot be evaluated analytically for the same reason as the likelihood function (7) cannot be evaluated analytically. Hence we follow the same approach of introducing the importance density $g(\alpha_t|y; \psi)$. The criterion to be minimized can then be expressed as

$$\begin{aligned} I_t &= \int \lambda_t^2(y_t, \alpha_t; \psi) \frac{p(y_t, \alpha_t; \psi)}{g(\alpha_t|y; \psi)} g(\alpha_t|y; \psi)\, \mathrm{d}\alpha_t \\ &\propto g(y; \psi) \int \lambda_t^2(y_t, \alpha_t; \psi) \frac{p(y_t|\alpha_t; \psi)}{g(y_t|\alpha_t; \psi)} g(\alpha_t|y; \psi)\, \mathrm{d}\alpha_t \\ &\propto I_t^*, \end{aligned}$$

where

$$I_t^* = \int \lambda_t^2(y_t, \alpha_t; \psi) w(y_t, \alpha_t; \psi) g(\alpha_t|y; \psi)\, \mathrm{d}\alpha_t. \tag{15}$$

The statements above are valid since b_t and c_t only have an impact on y_t and $\theta_t = Z_t(\alpha_t)$, they have no impact on y_j and $\theta_j = Z_j(\alpha_j)$ with $j \neq t$. Also, we have

$$g(\alpha_t|y; \psi) \propto g(y_t|\alpha_t; \psi) g(\alpha_t; \psi),$$

with $g(\alpha_t; \psi) = p(\alpha_t; \psi)$. Hence the minimization of I_t with respect to (b_t, c_t) is equivalent to the minimization of I_t^*. The evaluation and minimization of I_t^* takes place via importance sampling. We minimize

$$\widehat{I_t^*} = M^{-1} \sum_{i=1}^{M} \lambda_t^2(y_t, \alpha_t^{(i)}; \psi) w(y_t, \alpha_t^{(i)}; \psi),$$

with respect to (b_t, c_t), where $\alpha_t^{(i)}$ is obtained by sampling from $g(\alpha|y; \psi)$. This minimization of $\widehat{I_t^*}$ leads to the weighted least squares solution. In case $\theta_t = Z_t(\alpha_t)$ is a scalar, we can define the regression coefficient vector $\beta_t = (a_t^*, b_t, c_t)'$ where a_t^* is the intercept and can be

regarded as a deterministic function of b_t and c_t. The minimum is then obtained at

$$\widehat{\beta}_t = \left(\sum_{i=1}^{M} w_{it} v_{it} v_{it}' \right)^{-1} \sum_{i=1}^{M} w_{it} v_{it} p_{it}, \tag{16}$$

where w_{it} is defined in (13) and where

$$v_{it} = (1, \theta_t^{(i)}, -\tfrac{1}{2}\theta_t^{(i)\,2})', \qquad p_{it} = \log p(y_t|\theta_t^{(i)}; \psi), \qquad \theta_t^{(i)} = Z_t(\alpha_t^{(i)}),$$

for $t = 1, \ldots, n$. The sampling of $\alpha_t^{(i)}$ from $g(\alpha_t|y; \psi)$ requires applying the simulation smoother on the model $g(y, \alpha; \psi) = g(y|\alpha; \psi)p(\alpha; \psi)$ that we typically represent by the model (11) and (2). However, observation equation (11) requires values for b_t and c_t, for $t = 1, \ldots, n$, which we want to establish via the least squares solution (16). Since the Gaussian kernel of the log-density $\log g(y|\alpha; \psi)$ acts effectively as a second order Taylor approximation to $\log p(y|\alpha; \psi)$, around some value of θ_t, we can carry out the minimization iteratively as follows. We set values for $b_1, \ldots, b_n$ and $c_1, \ldots, c_n$ initially. A search for good starting values can be conducted but in many cases of practical interest, any set of initial values work sufficiently well. Next we simulate $\theta_t^{(i)}$ by means of simulation smoothing applied to the linear Gaussian model (11) and (2), based on the current set of values for (b_t, c_t) with $t = 1, \ldots, n$. A new set of values can be obtained from (16). This iterative scheme continues until some level of convergence is obtained. It is assumed that at each iteration when samples are generated from $g(\alpha|y; \psi)$ using a new set of values for (b_t, c_t) with $t = 1, \ldots, n$, the same random numbers are used for computing $\alpha^{(i)}$ so that a smooth convergence process takes place. The random numbers can be saved on the computer or they can be generated by using the the same random seed.

The MEIS implementation can be summarized in algorithmic form as follows:

(i) Set $k = 1$ and set values for $\beta_t^k = (a_t^*, b_t, c_t)'$ for $t = 1, \ldots, n$.

(ii) Construct the linear Gaussian state space model (11) for $x_t = b_t/c_t$ based on β_t^k; simulate $\theta_1^{(i)}, \ldots, \theta_n^{(i)}$, for $i = 1, \ldots, M$, with $\theta_t^{(i)} = Z_t(\alpha_t^{(i)})$, by means of a simulation smoothing algorithm where the seed of the random number generator is the same for every iteration k.

(iii) Set $\beta_t^{k+1} = \widehat{\beta}_t$, for $t = 1, \ldots, n$, as obtained from the regression (16) using the simulated values $\theta_1^{(i)}, \ldots, \theta_n^{(i)}$, for $i = 1, \ldots, M$, from the previous step.

(iv) If $||(\beta_t^{k+1} - \beta_t^k) \oslash \beta_t^k||_\infty < \epsilon$, for $t = 1, \ldots, n$ and some threshold value ϵ, the algorithm has converged and the algorithm can be terminated; else set $k = k + 1$ and return to step (ii) and (iii). $\oslash$ denotes the Hadamard division (point wise division).

3.4 A comparison with EIS

Our proposed modification of the efficient importance sampling method, the MEIS method, is clearly different than the original EIS method although the objective function is the same. The key insight that we explore is the representation of $g(y, \alpha; \psi)$ by the linear Gaussian state space model (11) and (2) for the constructed variable x_t. This allows us to treat the EIS method on the basis of the computationally efficient Kalman filter and its related smoothing methods including the simulation smoother.

Richard and Zhang (2007) have proposed the minimization of (14) and have provided the solution (16). The key difference is how the draws $\alpha_t^{(i)}$ are generated. In their implementation of EIS, they adopt an approximate backwards scheme, starting from $t = n$ towards $t = 1$, and need to track an integration constant so that each density at time t integrates to unity. We circumvent this time-consuming process since we interpret the density as a well-defined model for x_t and apply the simulation smoothing method of Durbin and Koopman (2002) for computing the draws $\alpha_t^{(i)}$, for $t = 1, \ldots, n$, directly.

Another key development of our MEIS method is that the simulations are based on the signal vector θ_t, this in contrast to EIS where the simulations are with respect to the state vector α_t. In many empirical models of interest, the state vector is typically of a higher dimension than the signal vector which has the same dimension of y_t. We therefore expect that in many studies, our implementation will gain computational efficiency. We emphasize here that in situations where the dimension of the state vector is lower than that of the signal vector we simply base our simulations on the state vector so that we can always obtain draws with the least computational effort.

4 Nonlinear non-Gaussian state space analysis

In Section 3.2 we have shown how the likelihood function can be evaluated by the method of importance sampling. In this section we briefly illustrate other applications of (modified) efficient importance sampling and provide the details for an effective implementation.

4.1 Maximum likelihood estimation of ψ

The maximum likelihood estimate (MLE) of parameter vector ψ can be simply obtained via a numerical optimization method. Quasi-Newton methods are often used for this task. It may be clear that analytical expressions for the MLE are not available in almost all cases.

A number of numerical issues need to be addressed before the actual maximization of the likelihood function can take place. We evaluate the likelihood function as a Monte Carlo estimate. The use of different sets of random values for generating the importance draws of α^i, with $i = 1, \ldots, M$, leads clearly to different estimates of the likelihood function $L(y; \psi)$. Since numerical optimization methods require smooth functions, we evaluate the likelihood functions using the same set of random values. In other words, the same "seed" of the random number generator is taken for each likelihood evaluation. The likelihood is then a smooth function of ψ only.

In practice, the log-likelihood function is maximized. However, the log of the estimate (13) is not equal to the estimate of the log-likelihood function. The bias in the log of the estimate can be approximately corrected on the basis of a second-order Taylor expansion. We therefore maximize the bias-corrected log-likelihood estimate

$$\widehat{\ell(y;\psi)} = \log \widehat{L}(y;\psi) + \frac{1}{2M}\bar{w}^{-2}s_w^2, \qquad s_w^2 = (M-1)^{-1}\sum_{i=1}^{M}(w_i - \bar{w})^2,$$

where $\ell(y;\psi) = \log L(y;\psi)$, $w_i = \prod_{t=1}^{n} w_{it}$ and $\bar{w} = M^{-1}\sum_{i=1}^{M} w_i$; see Durbin and Koopman (1997) for more details.

The bias-corrected log-likelihood estimate can be expressed as

$$\widehat{\ell(y;\psi)} = \log g(y;\psi) + \log \bar{w} + \frac{1}{2M}\bar{w}^{-2}s_w^2, \tag{17}$$

The computation of w_i, $\log \bar{w}$ and $\bar{w}^{-2} s_w^2$ requires modifications for a numerically feasible and stable implementation. Define

$$a_i = \log w_i = \sum_{t=1}^{n} \log p(y_t|\alpha_t^{(i)};\psi) - \log g(y_t|\alpha_t^{(i)};\psi), \qquad \bar{a} = M^{-1} \sum_{j=1}^{M} a_j,$$

for $i = 1, \ldots, M$. The computation of a_i and $\bar{a}$ is numerically stable. However, the computation of $w_i = \exp(a_i)$ can lead to numerical overflow problems whereas the computation of $u_i = \exp(a_i - \bar{a})$ is numerical stable. It follows that $w_i = \exp(\bar{a}) u_i$. After some further minor manipulations, it can be shown that

$$\log \bar{w} = \bar{a} + \log \bar{u}, \qquad \text{and} \qquad \bar{w}^{-2} s_w^2 = \bar{u}^{-2} s_u^2,$$

where

$$u_i = \exp(a_i - \bar{a}), \qquad \bar{u} = M^{-1} \sum_{i=1}^{M} u_i, \qquad s_u^2 = (M-1)^{-1} \sum_{i=1}^{M} (u_i - \bar{u})^2. \tag{18}$$

The bias-corrected log-likelihood estimate (17) is computed in a numerically feasible manner using these results.

4.2 Signal extraction: estimation of state and signal vectors

The estimation of α_t is based on the evaluation of the integral

$$\tilde{\alpha} = \int \alpha p(\alpha|y;\psi)\, \mathrm{d}\alpha.$$

We have argued that also the evaluation of such integral in a computational efficient way can be carried out by efficient importance sampling. The construction of a Monte Carlo estimate for $\tilde{\alpha}$ is based on

$$\begin{aligned} \tilde{\alpha} &= \int \alpha \left[p(\alpha|y;\psi) \,/\, g(\alpha|y;\psi) \right] g(\alpha|y;\psi)\, \mathrm{d}\alpha \\ &= \left[g(y;\psi) \,/\, p(y;\psi) \right] \int \alpha w(y,\alpha;\psi) g(\alpha|y;\psi)\, \mathrm{d}\alpha, \end{aligned} \tag{19}$$

since $g(\alpha;\psi) = p(\alpha;\psi)$, where

$$w(y,\alpha;\psi) = \frac{p(y|\alpha;\psi)}{g(y|\alpha;\psi)} = \prod_{t=1}^{n} w(y_t,\alpha_t;\psi), \tag{20}$$

with $w(y_t,\alpha_t;\psi)$ as defined in (12). The density $p(y;\psi)$ reflects the likelihood function (12) and its substitution in (19) leads to the equation

$$\tilde{\alpha} = \frac{\int \alpha w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha}{\int w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha}.$$

The two integrals can be evaluated by Monte Carlo simulation. The estimate of $\tilde{\alpha}$ is then given by

$$\widehat{\tilde{\alpha}} = \frac{\sum_{i=1}^{M} \alpha^{(i)} w_i}{\sum_{i=1}^{M} w_i},$$

where $w_i = \prod_{t=1}^{n} w_{it}$ with w_{it} defined as in (13) and where both $\alpha^{(i)}$ and w_i are based on the draws from the importance density, that is

$$\alpha^{(i)} \sim g(\alpha|y;\psi), \qquad i = 1,\ldots,M.$$

The draws are obtained by using the method of efficient importance sampling described in Section 3.3. The nominator and denominator are typically computed by using the same random numbers and therefore we can base the estimate on normalized weights, that is

$$\widehat{\tilde{\alpha}} = \sum_{i=1}^{M} \alpha^{(i)} w_i^*, \qquad w_i^* = \frac{w_i}{\sum_{k=1}^{M} w_k}.$$

The signal is a function of the state vector, we have $\theta_t = Z_t(\alpha_t)$ and $\theta = Z(\alpha)$ where $Z(\alpha) = [Z_1(\alpha_1)', \ldots, Z_n(\alpha_n)']'$. Using the same arguments as above, the estimate of θ is given by

$$\tilde{\theta} = \frac{\int \theta w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha}{\int w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha},$$

and we evaluate it via the efficient importance sampling method to obtain

$$\widehat{\tilde{\theta}} = \sum_{i=1}^{M} Z(\alpha^{(i)}) w_i^*,$$

where the normalized weight w_i^* is defined as above. These arguments are also valid for any other known function of α, see also Durbin and Koopman (2000).

4.3 Forecasting

Forecasting requires the estimation of the state vector α_t at a time period t after n, that is $t > n$. The same principles of signal extraction can be applied. We denote the forecast of α_{n+j} by $\tilde{\alpha}_{n+j}$, for $j = 1, 2, \ldots$, and we compute it by importance sampling methods. It follows that

$$\tilde{\alpha}_{n+j} = \int \alpha_{n+j} p(\alpha^+|y;\psi)\,\mathrm{d}\alpha^+, \qquad j = 1, 2, \ldots,$$

where $\alpha^+ = (\alpha'\,,\, \alpha_{n+1}\,, \ldots,\, \alpha_{n+j})'$. We notice that $p(y|\alpha^+;\psi) = p(y|\alpha;\psi)$ and $g(y|\alpha^+;\psi) = g(y|\alpha;\psi)$ since the future states $\alpha_{n+1}, \ldots, \alpha_{n+j}$ do not affect y. Similarly as in (19), we then have

$$\tilde{\alpha}_{n+j} = [g(y;\psi)\,/\,p(y;\psi)] \int \alpha_{n+j} w(y,\alpha;\psi) g(\alpha^+|y;\psi)\,\mathrm{d}\alpha^+, \qquad j = 1, 2, \ldots,$$

with $w(y,\alpha;\psi)$ defined in (20). The forecast estimate is then obtained as for signal extraction via importance sampling and is given by

$$\widehat{\tilde{\alpha}}_{n+j} = \sum_{i=1}^{M} \alpha_{n+j}^{(i)} w_i^*, \qquad w_i^* = \frac{w_i}{\sum_{k=1}^{M} w_k}, \qquad j = 1, 2, \ldots.$$

The computation of the draws $\alpha_{n+j}^{(i)} \sim g(\alpha_{n+j}|y;\psi)$, for $j = 1, 2, \ldots$, is facilitated by the simulation smoothing algorithm. It requires the extension of the data vector y with missing values for the time periods $n+1, \ldots, n+j$ to obtain the draws from $g(\alpha^+|y;\psi)$ as required; see Durbin and Koopman (2012, Chapter 4). The forecasting of signal and observations vectors is carried out in a similar way. Observations forecasts can then be based on the signal forecasts.

5 Simulation study

Before we can discuss the computational gains that can be obtained by simulating draws from the signal θ_t rather than the state α_t we first have to verify that the EIS and MEIS methods have roughly the same simulation efficiency. This can partly be verified by looking at the variance of the importance sampling weights $\omega(y_t, \alpha_t; \psi)$ as given in (12) or by the variance of the de-meaned importance weights s_u^2 in (18). We notice that $\omega(y_t, \alpha_t; \psi)$ and s_u^2 of EIS and MEIS should theoretically be reasonably close to each other because of the same objective function in (14). We also expect parameter estimates and standard errors of both methods to be reasonably close to each other. To provide some simulation evidence of the claims above we present the results of a simulation study. We consider the stochastic volatility model with multiple volatility factors of Section 2.2 where we replace the Gaussian density $p(\varepsilon_t)$ of equation (4) by the t-distribution to obtain a stochastic volatility model with heavier tails. Since we have $\mathrm{E}(y_t|\theta_t) = \mu$ and $\mathrm{Var}(y_t|\theta_t) = \exp(\theta_t)$, it follows that the log-density is given by

$$\log p(y_t|\theta_t) = \text{constant} - \frac{1}{2}\left[\theta_t + (\nu+1)\,\log(1+\kappa_t)\right], \qquad \kappa_t = \exp(-\theta_t)\frac{(y_t-\mu)^2}{\nu-2}, \tag{21}$$

where, constant $= \log\Gamma(\frac{\nu}{2}+\frac{1}{2}) - \log\Gamma(\frac{\nu}{2}) - \frac{1}{2}\log\left[(\nu-2)\pi\right]$. We let the signal consist of the sum of the individual state elements as specified in (5). In the case of $p = 2$, we can associate the first state element with long-run dependence and the second state element with short-run dependence, see for example Durham and Gallant (2002). The $p \times 1$ state vector $\alpha_t = (\alpha_{1,t}, \ldots, \alpha_{p,t})'$ as defined in (2) has $p \times p$ constant over time system matrices given by

$$T = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_p \end{bmatrix}, \qquad Q = \begin{bmatrix} \sigma^2_{\eta,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\eta,p} \end{bmatrix}, \tag{22}$$

with $|\phi_i| < 1$ and $\sigma^2_{\eta,i} > 0$ for $i = 1, \ldots, p$. The model is identified by imposing $\phi_1 > \ldots > \phi_p$. For more information about the sum of autoregressive and moving average processes, we refer to Granger and Morris (1976). This dynamic specification in the context of the stochastic volatility model (21) is also adopted by Koopman and Scharth (2013).

We have simulated $S = 500$ return series of length $n = 5000$ with true values of ψ as presented in Table 1 where the value of μ in (21) is fixed at 0. After simulation, estimates of ψ are obtained from the simulated series for the SPDK method and the EIS and MEIS methods as described in Section 3. All calculations are carried out in the Ox computing language of Doornik (2007) and the additional SsfPack which is a set of C routines for Kalman filter and related methods, see Koopman, Shephard, and Doornik (2008). We notice that the implemented routines in Ox make callbacks to C for its matrix computations but it does not do this if "for loops" are used. Therefore, we programmed time consuming for loops of the EIS method in C to provide an honest comparison in speed between all three methods. For SPDK and MEIS we used the simulation smoother of Durbin and Koopman (2002) to obtain draws from the signal θ_t while for EIS the draws from the state α_t were obtained by implementing the methods described in Jung et al. (2011). The maximum likelihood estimates of the parameter vector ψ are obtained by the BFGS algorithm where the starting values were set to 90% of the true parameter values as given in Table 1.

Table 1: We report values of ψ that are used for simulating time series of returns with t-distributed error terms. For each method (SPDK, EIS, MEIS) and for each state dimension p we simulate $S = 500$ return series of length $n = 5000$. The estimation results are presented in Table 2. The sum of the variance of the state components is kept constant for all p (column 8) which means that the individual variances of the states do vary with p.

p	ϕ_1	ϕ_2	ϕ_3	$\sigma^2_{\eta,1}$	$\sigma^2_{\eta,2}$	$\sigma^2_{\eta,3}$	$\sum_{i=1}^{p}\sigma^2_{\eta,i}$	$\sigma_{\eta,1}$	$\sigma_{\eta,2}$	$\sigma_{\eta,1}$	ν
1	0.98	-	-	0.0225	-	-	0.0225	0.15	-	-	10
2	0.60	0.98	-	0.0169	0.0056	-	0.0225	0.13	0.0748	-	10
3	0.50	0.90	0.99	0.0121	0.0064	0.0040	0.0225	0.11	0.0800	0.0632	10

We present the results of our first simulation study in Table 2. We can conclude from the table that the simulation efficiency of EIS and MEIS are comparable as can be seen by comparing the variance of the de-meaned importance weights s^2_u and the sample standard errors of the parameter estimates. This is an important result because if EIS would have produced a more efficient importance density then our claim of being more computationally efficient would have been less strong.

Although SPDK is much faster and the mean of the parameter estimates are often comparable to EIS and MEIS, the method is not a good candidate in this setting because of the (much too high) variance of the importance sampling weights. We show in a second

Table 2: We report results of the simulation study in which $S = 500$ time series of length $n = 5000$ are simulated from the Student's t as given by equation (21). After simulation, estimates of ψ are obtained for three methods (SPDK, EIS, MEIS). Details of the simulation study are as follows; the computations are carried out on a i7-2600, 3.40 GHz desktop PC using one core. The EIS and MEIS method are initialized by 3 SPDK iterations. The convergence criterion of all three methods is set to 10^{-3} and the algorithms converge most of the time in $4-6$ iterations. One likelihood evaluation is based on $M = 100$ draws. Starting values are set to 90% of the true parameter values as given in Table 1. $\bar{\nu}$, $\bar{\phi}_p$ and $\bar{\sigma}_{\eta,p}$, $p = 1, 2, 3$, are the means of the estimated parameters. $\bar{s}_u^2$ denotes the mean of the variance of the importance sampling weights, i.e. $\frac{1}{S}\sum_{i=1}^{500} s_u^2$. The numbers in parenthesis are the sample standard errors of the estimates. Time in the last column is the average computing time (in seconds) it took to maximize the log-likelihood.

$n = 5000$	p	$\bar{\phi}_1$	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\sigma}_{\eta,1}$	$\bar{\sigma}_{\eta,2}$	$\bar{\sigma}_{\eta,3}$	$\bar{\nu}$	$\bar{s}_u^2$	time
SPDK	1	0.98	-	-	0.15	-	-	10.29	213.84	14.34
		(0.01)	-	-	*(0.02)*	-	-	*(1.71)*	*(551.20)*	*(1.25)*
	2	0.54	0.98	-	0.11	0.08	-	10.56	1.63E+05	63.00
		(0.13)	*(0.01)*	-	*(0.09)*	*(0.02)*	-	*(2.15)*	*(3.55E+06)*	*(19.68)*
	3	0.44	0.80	0.99	0.11	0.06	0.06	10.89	2.04E+07	178.95
		(0.13)	*(0.14)*	*(0.01)*	*(0.10)*	*(0.06)*	*(0.02)*	*(2.99)*	*(3.67E+08)*	*(94.40)*
EIS	1	0.98	-	-	0.15	-	-	10.34	0.52	148.17
		(0.01)	-	-	*(0.02)*	-	-	*(1.76)*	*(0.24)*	*(13.76)*
	2	0.56	0.98	-	0.130	0.08	-	10.84	0.20	417.73
		(0.11)	*(0.01)*	-	*(0.10)*	*(0.02)*	-	*(2.68)*	*(0.63)*	*(152.50)*
	3	0.47	0.81	0.99	0.12	0.06	0.06	11.13	0.80	1440.52
		(0.07)	*(0.09)*	*(0.01)*	*(0.10)*	*(0.05)*	*(0.05)*	*(2.90)*	*(2.26)*	*(748.87)*
MEIS	1	0.98	-	-	0.15	-	-	10.35	0.64	145.5
		(0.01)	-	-	*(0.016)*	-	-	*(1.77)*	*(0.38)*	*(19.85)*
	2	0.55	0.98	-	0.113	0.075	-	10.71	0.23	415.58
		(0.13)	*(0.01)*	-	*(0.096)*	*(0.018)*	-	*(2.43)*	*(0.62)*	*(143.92)*
	3	0.43	0.80	0.99	0.14	0.06	0.06	12.31	0.76	1351.78
		(0.13)	*(0.13)*	*(0.01)*	*(0.13)*	*(0.06)*	*(0.02)*	*(5.06)*	*(3.07)*	*(750.52)*

simulation experiment the effect of a poorly fitting importance density on parameter and likelihood estimates. The effect of a high variance in the importance weights is presented in Table 3. Here, we simulated one data set of length $n = 5000$ from the Student's t as given by equation (21) and estimated ψ a 100 times where each time a different random seed is used. This procedure allows us to investigate the sensitivity of the methods to random numbers. Needless to say we expect better performances for methods with a low variance in the importance weights. This is confirmed by the sample standard errors of the parameter estimates and likelihood estimate as presented in Table 3. The sample standard error of the average likelihood value is 18 times larger for SPDK compared to both EIS and MEIS.

Table 3: We report the sensitivity of the three methods to random numbers. We learn from the sample standard errors that the EIS and MEIS perform much better in terms of precision.

$n = 5000$	p	$\bar{\phi}_1$	$\bar{\phi}_2$	$\bar{\sigma}_{\eta,1}$	$\bar{\sigma}_{\eta,2}$	$\bar{\nu}$	$\bar{s}_u^2$	$\overline{L}(y;\psi)$
SPDK	2	0.569	0.978	0.172	0.085	8.76	1.62E+07	-7245.480
		(0.090)	*(0.001)*	*(0.056)*	*(0.002)*	*(0.67)*	*(1.12E+08)*	*(0.734)*
EIS	2	0.598	0.978	0.134	0.084	8.35	0.113	-7246.021
		(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.005)*	*(0.022)*	*(0.040)*
MEIS	2	0.598	0.978	0.133	0.084	8.35	0.110	-7246.024
		(0.000)	*(0.000)*	*(0.000)*	*(0.000)*	*(0.005)*	*(0.020)*	*(0.039)*

MEIS is faster than EIS for the state dimensions considered in Table 2, however, not much faster. EIS becomes computationally less efficient for $p > 1$ since more computations are necessary to draw state vectors for increasing state vector dimensions. We expect difference in computation time between EIS and MEIS to become larger if state dimensions increase further than the dimensions used in our first experiment. Therefore we compared the time to calculate the likelihood for increasingly larger state vectors and increasing time series length in Table 4. The table clearly shows that the computational gains are small for small dimensions of the state vector but they become significant as the dimension of the state or the length of the time series increases. The irregularities in some patterns of the fractions as n or p increases are due to our efficient computer implementations using multiple core platforms. We therefore take the reported fractions as indicative.

Given the computational savings and the convenient state space representation of MEIS we regard MEIS as a viable candidate for non-Gaussian state space models, especially when the dimension of the state vector and/or the time series length increase.

Table 4: We report the fraction $t(EIS)/t(MEIS)$ where $t(x)$ is the time in seconds for method x to calculate the likelihood, for different state dimensions p and different time series lengths n.

$p \backslash n$	100	200	1000	5000	10000	40000
1	1.00	1.00	1.43	1.06	1.00	1.46
2	1.00	1.00	1.50	1.05	1.01	1.43
4	1.00	1.00	1.16	1.08	1.12	1.42
8	1.00	1.26	2.25	1.53	1.62	2.19
16	2.10	3.21	1.71	2.09	2.10	2.66
32	2.37	2.87	4.46	2.89	3.46	3.67
64	4.04	3.11	5.24	4.36	4.19	7.13

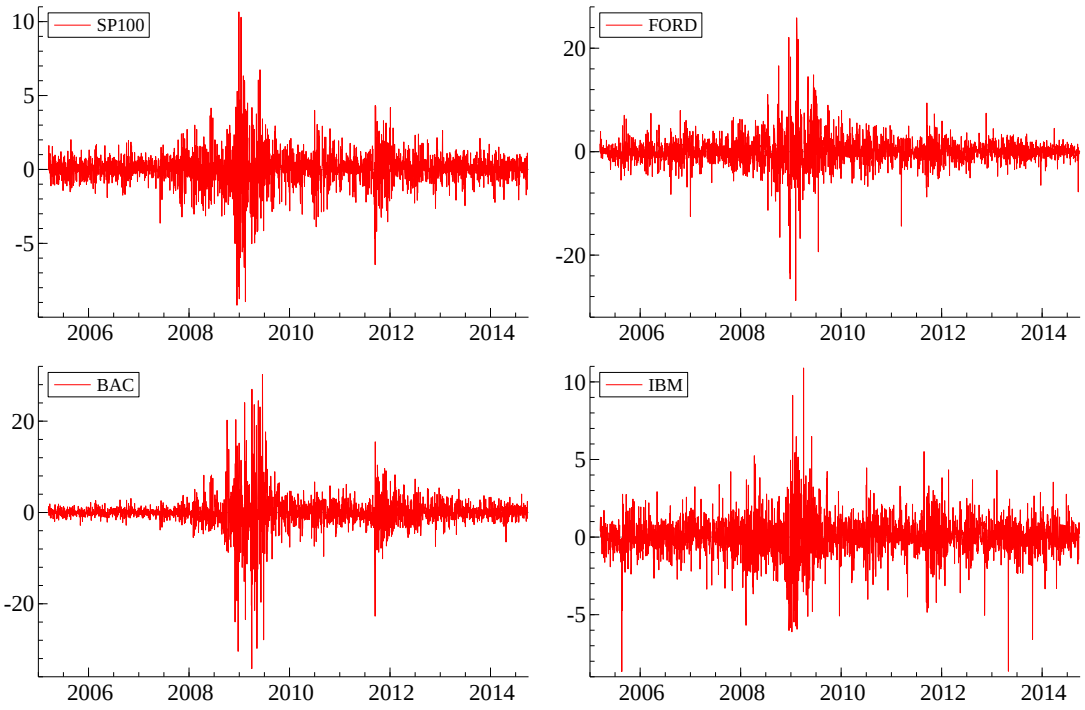

Figure 1: Daily financial returns from 7 October 2004 up to 6 October 2014 (ten years) for the S&P100 stock index and the stocks Ford, Bank of America (BAC) and IBM, all traded at the New York Stock Exchange and based on close prices.

6 Volatility measurement from daily stock returns

To investigate whether our modified efficient importance sampling method has relevance in an empirical study with the purpose of measuring time-varying volatility in daily stock returns, we consider the stochastic volatility model (4) with Student's t density (21) and log-volatility signal (5), in short the SV-t model. It is the same model specification that is used in our Monte Carlo study of the previous section. We first estimate the parameters ϕ_j and $\sigma_{\eta,j}$, for $j = 1, \ldots, p$ from (5) or (22), and ν from (21), simultaneously by maximum likelihood. Then we extract log-volatility θ_t as specified in (5) using the SV-t model with estimated parameters.

In our empirical study we consider four daily stock returns, the S&P100 composite index and three key U.S. firms (Ford, Bank of America and IBM) which are all traded at the New York Stock Exchange (NYSE). The time series are obtained from Yahoo Finance and our ten-year sample is from 7 October 2004 up to 6 October 2014 ($T = 2500$). The four stock return series are graphically displayed in Figure 1. Each return series are modelled

by the SV-t model with $p = 1, 2, 3$ in order to investigate whether log-volatility can be decomposed into long-term and short-term dynamics. The estimated parameters for the two SV-t specifications and for the four return series are reported in Table 5 together with the maximized log-likelihood function values and the corresponding Akaike (AIC) information criterion.

The parameters of the SV-t model specifications with $p = 1, 2, 3$ have all been estimated without any numerical problem for all considered series in this study. We have also shortened and lengthened the time series sample but numerical errors have not been encountered. The results provide some interesting insights. Although we have found some evidence for the decomposition of the log-volatility into two autoregressive components of different persistency levels, we have not found much evidence for more than two log-volatility components. In terms of smallest AIC values, the stock returns of Ford and Bank of America appear to favor two log-volatility components in the SV-t model while the daily volatilities in S&P100 and IBM returns appear to be best modelled by a single autoregressive component. In the case of Ford and Bank of America, the first component is highly persistent as ϕ_1 is virtually equal to unity while the second component is less but still highly persistent. We therefore can interpret the first component as long-term volatility while the second component is short-term volatility. The implied unconditional variances of the two log-volatility components indicate that most of the variation is explained by the first component. We have also computed the Schwarz' Bayesian (BIC) information criterion together with the AIC. The BIC points to single component log-volatilities for all four stock return series. Hence our empirical evidence of a decompostion of log-volatility into multiple components is not very strong. Finally we have found that the SV-t model with $p = 3$ is difficult to justify empirically, also because it appears to be harder to estimate the model parameters.

The estimated log-volatilities from the SV-t model with $p = 2$ and for the four return series are presented in Figure 2. The log-volatility signal estimates are computed using the MEIS procedure for which the details are presented in Section 4.2. The signal estimates are very similar, or in some cases virtually indistinguishable, amongst SV-t models with $p = 1$, $p = 2$ or $p = 3$. We display the three volatility signal estimates for the firms Ford, Bank of America and IBM in three separate plots and, in each of them, together with the S&P100 volatility signal estimate. From these plots we can easily compare between the

Table 5: We report the estimation results for the parameters of the SV-t model and for the return series of the S&P100 stock index and the stocks Ford, Bank of America (BAC) and IBM as displayed in Figure 1. The SV-t model is given by (4) with Student's t density (21) and log-volatility signal (5) with $p = 1, 2, 3$. The estimate of parameter vector ψ is obtained using our MEIS method after the initialization of 3 SPDK iterations. The computations are carried out on a i7-2600, 3.40 GHz desktop PC using one core. Each likelihood evaluation is based on $M = 100$ draws. The maximized log-likelihood values are reported together with the AIC information criterion.

	p	$\bar{\phi}_1$	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\sigma}_{\eta,1}$	$\bar{\sigma}_{\eta,2}$	$\bar{\sigma}_{\eta,3}$	ν	Log-L	AIC
S&P100	1	0.990	-	-	0.149	-	-	8.918	-3289.66	**6585.3**
	2	0.990	0.878	-	0.148	0.019	-	8.959	-3289.65	6589.3
	3	0.990	0.990	0.227	0.149	0.000	0.00049	8.942	-3289.65	6593.3
FORD	1	0.998	-	-	0.103	-	-	6.858	-5492.96	10991.9
	2	0.999	0.883	-	0.063	0.243	-	9.261	-5485.49	**10981.0**
	3	0.999	0.884	0.534	0.063	0.241	0.0420	9.258	-5485.48	10985.0
BAC	1	0.998	-	-	0.125	-	-	8.511	-5124.38	10254.8
	2	0.998	0.868	-	0.103	0.190	-	10.508	-5122.39	**10254.8**
	3	0.998	0.878	0.528	0.102	0.184	0.0427	10.484	-5122.36	10258.7
IBM	1	0.985	-	-	0.139	-	-	6.702	-3890.47	**7786.9**
	2	0.994	0.939	-	0.072	0.167	-	7.120	-3888.72	7787.4
	3	0.994	0.933	0.554	0.075	0.170	0.0309	7.176	-3888.71	7791.4

individual volatility pattern of the firm and the one of the market as represented by the S&P100 index. We learn from Figure 2 that the stock volatility patterns of both Ford and IBM are generally in common with the market. It is interesting that the same applies to Bank of America (BAC) until the financial crisis has started. At the heighth of the financial crisis in 2009 the BAC volatility has been higher than the S&P100 volatility. During the aftermath of the financial crisis up to recently, the BAC volatility have remained relatively high.

7 Conclusions

We have presented a new modification of the efficient importance sampling (EIS) method for the analysis of nonlinear non-Gaussian state space models which include a wide range of time series models of interest. For the original EIS method of Richard and Zhang (2007) the construction of the importance density relies on an iterative method for which at each step simulation samples of the state vector needs to computed. The simulation of the state

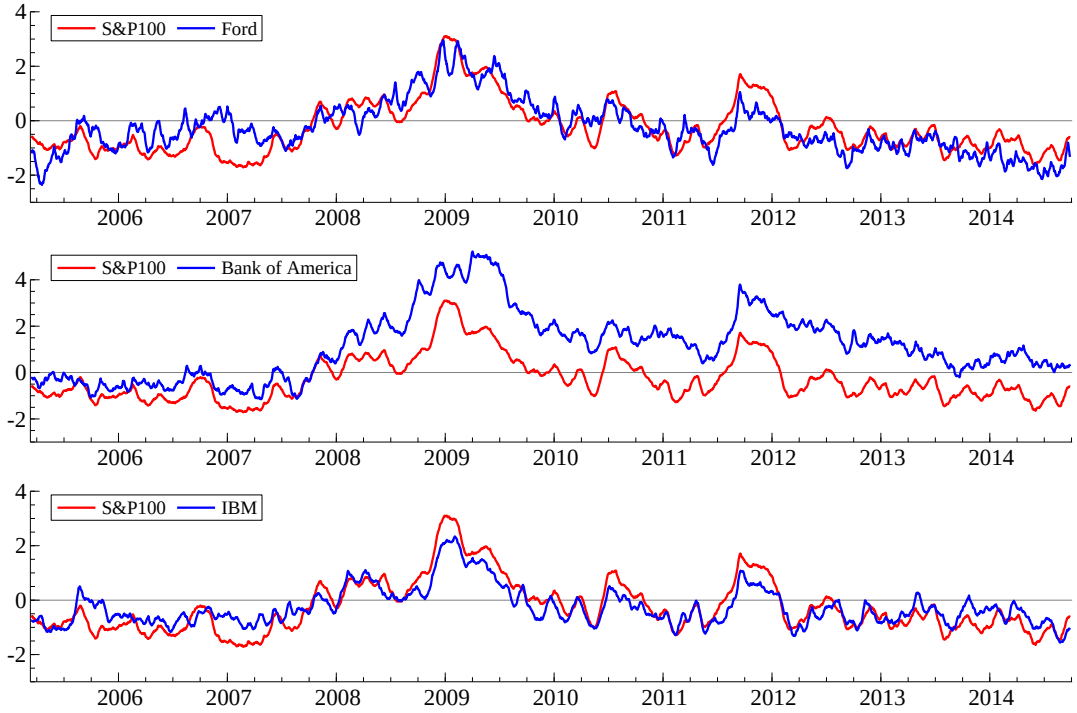

Figure 2: Estimated log-volatilities from 7 October 2004 up to 6 October 2014 (ten years) for the S&P100 stock index and the stocks Ford, Bank of America and IBM, based on the SV-t model given by (4) with Student's t density (21) and log-volatility signal (5) with $p = 2$.

vectors in the EIS method can become computationally involved and not straightforward when models require a large state vector with time-varying parameters. In the modified efficient importance sampling (MEIS) we construct the same importance density based on a similar simulation method. However, we show that the EIS method can also fully rely on computationally efficient Kalman filter and smoothing methods. The modification therefore leads to a fast efficient importance sampling method, especially for large state dimensions. The details of this development have been documented in this paper. We further show in a Monte Carlo study that the computational gains can be substantial for larger time series length and state dimensions. To show that our method also has relevance in an empirical study, we analyze the volatility patterns in four U.S. stock return series and we comment on our findings.

References

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81(3)*, 541–553.

Danielsson, J. and J. F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *J. Applied Econometrics 8*, S153–S174.

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

Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84(3)*, 669–684.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *J. Royal Statistical Society B 62(1)*, 3–56.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89(3)*, 603–615.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2nd ed.). Oxford: Oxford University Press.

Durham, A. G. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Business and Economic Statist. 20*, 297–316.

Fruhwirth-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*, 1317–39.

Granger, C. W. J. and M. J. Morris (1976). Time series modelling and interpretation. *J. Royal Statistical Society A 139(2)*, 246–257.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Jung, R. C., R. Liesenfeld, and J. F. Richard (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *J. Business and Economic Statist. 29*, 73–85.

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika 94(4)*, 827–839.

Kloek, T. and H. K. van Dijk (1978). Bayesian estimation of equation system parameters: an application of integration by monte carlo. *Econometrica 47*, 1–20.

Koopman, S., A. Lucas, and M. Scharth (2014). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. *J. Business and Economic Statist. 32*, forthcoming.

Koopman, S. and M. Scharth (2013). The analysis of stochastic volatility in the presence of daily realised measures. *J. Financial Econometrics 11*, 76–115.

Koopman, S. J., N. Shephard, and J. A. Doornik (2008). *SsfPack 3.0: Statistical algorithms for models in state space form*. London: Timberlake Consultants Press.

Liesenfeld, R. and J. F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance 10*, 505–531.

Richard, J. F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*, 1385–1411.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.

Shephard, N. (2005). *Stochastic volatility: Selected Readings*. New York: Oxford University Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84(3)*, 653–667.

So, M. K. P. (2003). Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica 13*, 255–274.