

# Modified Efficient Importance Sampling for partially non-Gaussian State Space Models\*

*Siem Jan Koopman*<sup>(a,b)</sup>, *Rutger Lit*<sup>(a)</sup> and *Thuy Minh Nguyen*<sup>(c)</sup>

<sup>(a)</sup>Vrije Universiteit Amsterdam & Tinbergen Institute

<sup>(b)</sup>CREATES, Aarhus University, Denmark

<sup>(c)</sup>Deutsche Bank, London

## Abstract

The construction of an importance density for partially non-Gaussian state space models is crucial when simulation methods are used for likelihood evaluation, signal extraction and forecasting. The method of efficient importance sampling is successful in this respect but we show that it can be implemented in a computationally more efficient manner using standard Kalman filter and smoothing methods. Efficient importance sampling is generally applicable for a wide range of models but it is typically a custom-built procedure. For the class of partially non-Gaussian state space models, we present a general method for efficient importance sampling. Our novel method makes the efficient importance sampling methodology more accessible because it does not require the computation of a (possibly) complicated density kernel that needs to be tracked for each time period. The new method is illustrated for a stochastic volatility model with a Student's t distribution.

*Key words:* Kalman filter; Monte Carlo maximum likelihood; Simulation smoothing; Efficient Importance Sampling; Non-Gaussian dynamics models.

---

\*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 usual disclaimer applies. Corresponding email : [s.j.koopman@vu.nl](mailto:s.j.koopman@vu.nl)

# 1 Introduction

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

$$y_t | \alpha_t \stackrel{ind}{\sim} p(y_t | \alpha_t; \psi), \quad t = 1, \dots, n, \quad (1)$$

where  $\stackrel{ind}{\sim}$  is notation for serially independently distributed,  $\psi$  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  $\alpha_t$ . The density  $p(y_t | \alpha_t; \psi)$  is for a random variable  $y_t$  that is conditional on the latent random variable  $\alpha_t$  and is a function of the parameter vector  $\psi$ . Conditional on the sequence  $\alpha_1, \dots, \alpha_n$ , the observations  $y_1, \dots, y_n$  are serially independently distributed. The time-varying parameters in  $\alpha_t$  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  $\alpha_t$  can be adopted. In our analysis, the conditional observation density and the dynamic model for  $\alpha_t$  must be specified and both may depend on the fixed parameter vector  $\psi$ .

When (i) the observation density for  $y_t$  conditional on  $\alpha_t$  is Gaussian, (ii) the relation between  $y_t$  and  $\alpha_t$  is linear and (iii) the dynamic model for  $\alpha_t$  is linear and Gaussian, our time series modelling framework reduces to the linear Gaussian state space model as discussed and reviewed in, for example, Harvey (1989) and 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  $\alpha_t$ , the evaluation of the likelihood function for a specific value of  $\psi$  and the forecasting of  $y_t$ . These methods provide minimum mean squared error (MMSE) estimates for  $\alpha_t$  and MMSE forecasts for  $y_t$ . 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. For the class of partially non-Gaussian state space models, we replace the conditions (i) and (ii) by the condition that the observation density for  $y_t$  conditional on  $\alpha_t$  can be non-Gaussian and that the relation between  $y_t$  and  $\alpha_t$  can be nonlinear. Under these more general conditions, we cannot rely on linear methods to obtain

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 partially non-Gaussian state space model, in most cases we cannot use the Kalman filter or a related analytical filter. We need to evaluate the high-dimensional integral 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.

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). The EIS method constructs the parameters of the importance density in such a way that the variance of the importance sampling weights (in logs) is minimized. Although this leads to a globally efficient importance sampling process, the method can be complicated due to the necessary computation of a density kernel for the importance density at each time period. Moreover, EIS is often custom-build which makes it less accessible. Although Richard and Zhang (2007) have shown that EIS can be implemented for state space models, we show that EIS in the case of partially non-Gaussian state space models can be implemented using standard Kalman filter methods in a similar way as in Shephard and Pitt (1997) and Durbin and Koopman (1997). There is no need for new methods or major adjustments of existing methods. The implementation of our proposed modification of the EIS method for this class of models is therefore straightforward. Furthermore we show that our modified EIS (MEIS) method can lead to some considerable reductions in computing time. We emphasize however that both implementations of the EIS method should lead to the same estimation results, up to a level of numerical accuracy. It is shown by Koopman, Lucas, and Scharth (2014) that similar modifications can also be explored for the use of numerical integration methods in importance sampling.

The remainder of the paper is organized as follows. In Section 2 we briefly introduce partially non-Gaussian state space models. 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. In Section 5 we verify that EIS and MEIS produce similar estimation results in a Monte Carlo study and we indicate the reductions in computing time by MEIS. 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 reviews the developments and explores future research.

## 2 Partially non-Gaussian state space models

The specification of the observation density in a partially non-Gaussian state space model is given by  $p(y_t|\alpha_t; \psi)$  and its specification can be given as in (1) with the stochastically time-varying parameter vector  $\alpha_t$ . The linear Gaussian dynamic process for  $\alpha_t$  is given by

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

where the elements of the vector of intercepts  $d_t$ , 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  $\psi$ , for  $t = 1, \dots, n$ . The disturbances  $\eta_t$  are normally and independently distributed and do not depend on the normally distributed initial state vector  $\alpha_1$ . 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  $y_t$ . This class of models have been originally studied by Shephard (1994). Illustrations of special cases of our partially non-Gaussian dynamic 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

$$y_t = Z_t(\alpha_t) + \varepsilon_t, \quad \varepsilon_t \stackrel{ind}{\sim} \tau(0, \sigma^2, \nu), \quad t = 1, \dots, n, \quad (3)$$

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

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

where  $\theta_t$  is the signal. Other heavy tailed densities for  $\varepsilon_t$  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  $\mu_t$  is the trend component,  $\gamma_t$  is the seasonal component, and both are elements of the state vector  $\alpha_t$ . 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 by  $\theta_t = Z_t(\alpha_t) = \mu_t + \gamma_t$  for  $t = 1, \dots, n$ . When the noise component  $\varepsilon_t$  is normally distributed, the time series analysis, including the maximum likelihood estimation of  $\psi$ , the signal extraction of  $\theta_t$  (or, additionally,  $\mu_t$  and  $\gamma_t$ ) and the forecasting of  $y_t$  can be based on the Kalman filter and related methods. However, when the noise  $\varepsilon_t$  is non-Gaussian as in (3), the resulting estimates of the Kalman filter are no longer optimal.

## 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

$$y_t = \mu + \exp\left(\frac{1}{2}\theta_t\right) \varepsilon_t, \quad \theta_t = Z_t(\alpha_t), \quad \varepsilon_t \stackrel{ind}{\sim} N(0, \sigma^2), \quad t = 1, \dots, n, \quad (4)$$

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

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

where  $\alpha_{jt}$  is the  $j$ th element of  $\alpha_t$  with autoregressive coefficient  $0 < \phi_j < 1$ , and  $\eta_{jt}$  is the  $j$ th element of  $\eta_t$ , for  $j = 1, \dots, p$ , and where  $\alpha_t$  and  $\eta_t$  are specified as in (2), for  $t = 1, \dots, n$ . It follows that the transition matrix  $T_t$  and the variance matrix  $Q_t$  in (2) are diagonal matrices, for  $t = 1, \dots, n$ , with their  $j$ th diagonal elements equal to  $\phi_j$  and  $\sigma_{\eta,j}^2$ , for  $j = 1, \dots, p$ . The conditional observation density (1), in logs, for this stochastic volatility model is given by

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

The stochastic volatility model can be extended in many ways. For example, leverage effects can be introduced by having dependence between  $\varepsilon_t$  and  $\eta_{jt}$ , for  $j = 1, \dots, p$ . Also, heavy-tailed density functions can be considered for the noise term  $\varepsilon_t$ . 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  $\theta_t$  that we can specify as (1) and (2). The observation

log-density function is then given by

$$\log p(y_t|\alpha_t; \psi) = y_t \log \theta_t - \theta_t - \log(y_t!), \quad \theta_t = Z_t(\alpha_t), \quad t = 1, \dots, n, \quad (6)$$

where the signal  $\theta_t$  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  $\psi$  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  $y = (y'_1, \dots, y'_n)'$  and as a function of parameter vector  $\psi$  is given by

$$L(y; \psi) = \int p(y, \alpha; \psi) d\alpha = \int p(y|\alpha; \psi)p(\alpha; \psi) d\alpha, \quad (7)$$

where  $\alpha = (\alpha'_1, \dots, \alpha'_n)'$ . 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  $\psi$ . Numerical evaluation is usually dismissed because of the high dimensional vector  $\alpha$ . 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

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

where  $\alpha^{(i)}$  refers to the  $i$ th simulated sample of  $\alpha$  that is generated from the unconditional density  $p(\alpha; \psi)$  with  $i = 1, \dots, M$ . The standard law of large numbers (LLN) insists that

$\widehat{L}(y; \psi)$  converges to  $L(y; \psi)$  as  $M \rightarrow \infty$ . Since the simulation of  $\alpha$  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  $\alpha_t$  and for the disturbances  $\eta_t$ , we have

$$L(y; \psi) = \int \left[ \prod_{t=1}^n p(y_t | \alpha_t; \psi) p(\alpha_t | \alpha_{t-1}; \psi) \right] d(\alpha_1, \dots, \alpha_n), \quad (9)$$

with  $p(\alpha_1 | \alpha_0; \psi) = p(\alpha_1; \psi)$ . We also have  $p(\alpha_t | \alpha_{t-1}; \psi) = N(d_{t-1} + T_{t-1}\alpha_{t-1}, R_{t-1}Q_{t-1}R'_{t-1})$  given the state equation (2), for  $t = 1, \dots, n$ .

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

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

where  $\eta_{t-1} \stackrel{ind}{\sim} N(0, Q_{t-1})$ , for  $t = 1, \dots, n$ . 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

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

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  $\alpha_t$ . Examples of link functions for the signal  $\theta_t$  are



presented in the illustrations of Section 2. It follows that the variables  $b_t$  and  $c_t$  are functions of the observations  $y_1, \dots, y_n$  and parameter vector  $\psi$  for  $t = 1, \dots, n$ . 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  $t = 1, \dots, n$ . 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

$$x_t = \theta_t + u_t, \quad u_t \stackrel{ind}{\sim} \text{N} \left( 0, \frac{1}{c_t} \right), \quad t = 1, \dots, n, \quad (11)$$

which has the conditional observation log-density function

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

for  $t = 1, \dots, n$ . 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

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

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

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

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  $t = 1, \dots, n$ . 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 the likelihood function (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 can be written as

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

where  $g(\alpha|y; \psi)$  is the Gaussian importance density and density  $p(y|\alpha; \psi)$  can be non-Gaussian. After some minor manipulations, we can express the likelihood function as

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

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

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

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

where vector  $\alpha^{(i)}$  is independently drawn for  $i = 1, \dots, M$ . Since we can represent  $g(y, \alpha; \psi)$  by the linear Gaussian state space model (11) and (2), we can simulate  $\alpha$  from the conditional state density  $g(\alpha|y; \psi)$  via simulation smoothing; 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  $i = 1, \dots, M$ , the likelihood estimate is

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

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  $t = 1, \dots, n$ , 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

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

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

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

where

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

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

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

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

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

for  $t = 1, \dots, n$ . The sampling of  $\alpha_t^{(i)}$  from  $g(\alpha_t|y; \psi)$  leads to a main difference with the EIS method. In our MEIS method, the computations for (16) are carried out by 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  $t = 1, \dots, n$ , 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  $\theta_t$ , we can carry out the minimization iteratively as follows. We set values for  $b_1, \dots, b_n$  and  $c_1, \dots, c_n$  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  $t = 1, \dots, n$ . 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  $t = 1, \dots, n$ , 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  $t = 1, \dots, n$ .
- (ii) Construct the linear Gaussian state space model (11) for  $x_t = b_t/c_t$  based on  $\beta_t^k$ ; simulate  $\theta_1^{(i)}, \dots, \theta_n^{(i)}$ , for  $i = 1, \dots, M$ , 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  $t = 1, \dots, n$ , as obtained from the regression (16) using the simulated values  $\theta_1^{(i)}, \dots, \theta_n^{(i)}$ , for  $i = 1, \dots, M$ , from the previous step.
- (iv) If  $\|(\beta_t^{k+1} - \beta_t^k) \oslash \beta_t^k\|_\infty < \epsilon$ , for  $t = 1, \dots, n$  and some threshold value  $\epsilon$ , the algorithm has converged and the algorithm can be terminated; else set  $k = k + 1$  and return to step (ii) and (iii). We notice that  $\oslash$  denotes 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 when applied to model (1) and (2) although the objective function is the same. Hence both methods will produce similar results, up to a level of numerical accuracy and simulation error. However we argue that the implementation of MEIS is much simpler and more straightforward in comparison with EIS. Only standard state space methods are required for MEIS. In higher dimensions, MEIS will be computationally more efficient since it requires the simulation of the signal  $\theta_t$  rather than the state vector  $\alpha_t$ ; in most models of practical interest, the dimension of the signal is much smaller.

The key insight is that we explore 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 the way the importance density is constructed. In their implementation of EIS, the researcher needs to do algebra with density kernels which can differ per application and hence EIS is custom-build. They adopt a 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  $t = 1, \dots, n$ , directly.

Another key development of our MEIS method is that the simulations are based on the signal vector  $\theta_t$ , this in contrast to EIS where the simulations are with respect to the state vector  $\alpha_t$ . 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.

Richard and Zhang (2007) show that an analytical solution to the variance minimization equation in (14) can be obtained by OLS if one restricts the choice of the importance density to the exponential family. Although this restriction is a computationally convenient one, it is not a necessary restriction as in other cases, equation (14) can be minimized by nonlinear methods. It should be noted that MEIS is restricted to Gaussian importance samplers, however, this is only a small limitation since Gaussian importance samplers are by far the most popular candidates due to their attractive properties.

Finally, both EIS and MEIS methods can be generalized to state space models with a possibly nonlinear and non-Gaussian state vector updating equation. It requires an extension of the regression computations in step (iii) of MEIS and simulations from another linear Gaussian state space model which is discussed in Durbin and Koopman (2012, Part II).

## 4 Applications of MEIS

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 $\psi$

The maximum likelihood estimate (MLE) of the parameter vector  $\psi$  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 except some rare 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  $\alpha^i$ , with  $i = 1, \dots, M$ , 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  $\psi$  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

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

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

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

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

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

for  $i = 1, \dots, M$ . 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 numerically stable. It follows that  $w_i = \exp(\bar{a}) u_i$ . After some further minor manipulations, it can be shown that

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

where

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

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  $\alpha_t$  is based on the evaluation of the integral

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

### 4.3 Forecasting

Forecasting requires the estimation of the state vector  $\alpha_t$  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  $\alpha_{n+j}$  by  $\tilde{\alpha}_{n+j}$ , for  $j = 1, 2, \dots$ , and we compute it by importance sampling methods. It follows that

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

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

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

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

$$\hat{\tilde{\alpha}}_{n+j} = \sum_{i=1}^M \alpha_{n+j}^{(i)} w_i^*, \quad w_i^* = \frac{w_i}{\sum_{k=1}^M w_k}, \quad j = 1, 2, \dots$$

The computation of the draws  $\alpha_{n+j}^{(i)} \sim g(\alpha_{n+j} | y; \psi)$ , for  $j = 1, 2, \dots$ , is facilitated by the simulation smoothing algorithm. It requires the extension of the data vector  $y$  with missing values for the time periods  $n+1, \dots, n+j$  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

We study the performances of the EIS and MEIS methods in a simulation study. We first verify whether both methods produce roughly the same estimation results with the same simulation efficiency. We expect the parameter estimates and the associated standard errors of both methods to be reasonably close to each other because both methods rely on the same objective function (14). We further provide an indication of the computational gains of MEIS. We expect these gains because MEIS requires simulation draws from the signal  $\theta_t$  rather than draws from the state  $\alpha_t$  in EIS.

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, in short the SVt model. Since we have  $E(y_t|\theta_t) = \mu$  and  $\text{Var}(y_t|\theta_t) = \exp(\theta_t)$ , it follows that the log density is given by

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

where,  $\text{constant} = \log \Gamma(\frac{\nu}{2} + \frac{1}{2}) - \log \Gamma(\frac{\nu}{2}) - \frac{1}{2} \log [(\nu - 2)\pi]$ . 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  $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{p,t})'$  as defined in (2) has  $p \times p$  constant over time system matrices given by

$$T = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_p \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma_{\eta,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\eta,p}^2 \end{bmatrix}, \quad (22)$$

with  $|\phi_j| < 1$  and  $\sigma_{\eta,j}^2 > 0$  for  $j = 1, \dots, p$ . The model is identified by imposing  $\phi_1 > \dots > \phi_p$ . 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  $\psi$  as presented in Table 1 where the value of  $\mu$  in (21) is fixed at 0. For all simulated time series,

Table 1: We report values of  $\psi$  that are used for simulating time series of returns with  $t$ -distributed error terms. For the EIS and MEIS methods 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  $p$ .

$p$	$\phi_1$	$\phi_2$	$\phi_3$	$\sigma_{\eta,1}^2$	$\sigma_{\eta,2}^2$	$\sigma_{\eta,3}^2$	$\sum_{i=1}^p \sigma_{\eta,i}^2$	$\sigma_{\eta,1}$	$\sigma_{\eta,2}$	$\sigma_{\eta,3}$	$\nu$
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

estimates of  $\psi$  are obtained from the EIS and MEIS methods as described in Section 3. The calculations are carried out in the Ox computing language of Doornik (2007) and the SsfPack library which consists of 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 have programmed the time consuming “for loops” of the EIS method in C to provide a fair comparison in speed between the two methods. For MEIS we have used the simulation smoother of Durbin and Koopman (2002) to obtain draws from the signal  $\theta_t$  while for EIS the draws from the state  $\alpha_t$  are obtained by implementing the methods described in Jung et al. (2011). The maximum likelihood estimation of the parameter vector  $\psi$  is based on the BFGS algorithm where the starting values are set equal to the true parameter vector as given in Table 1. We present the results of this simulation study in Table 2. We can conclude from the reported estimation results obtained by the EIS and MEIS methods that they are very similar and in some cases identical, up to numerical precision, and when also the standard errors of the estimates are considered in the comparisons. After comparing the sample standard errors of the parameter estimates and the variances of the de-meaned importance weights,  $s_u^2$ , we can also conclude that the simulation efficiencies of both methods are very similar overall. Hence both methods produce results that are precise and numerically accurate.

Further we can learn from Table 2, that the MEIS method is somewhat faster than the EIS method, but not much faster. The EIS method becomes computationally less efficient for  $p > 1$  since more computations are necessary to draw the  $p \times 1$  state vectors when  $p$  increases. Hence we expect difference in computation time between EIS and MEIS to become

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  $\psi$  are obtained for the EIS and MEIS methods. 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 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, that is  $\frac{1}{S} \sum_{i=1}^{500} s_u^2$ . The numbers in parenthesis are the sample standard errors of the estimates. Time in the last column is the average computing time (in seconds) for parameter estimation.

$n = 5000$	$p$	$\bar{\phi}_1$	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\sigma}_{\eta,1}$	$\bar{\sigma}_{\eta,2}$	$\bar{\sigma}_{\eta,3}$	$\bar{\nu}$	$\bar{s}_u^2$	time
<b>EIS</b>	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)
<b>MEIS</b>	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)

larger when the state dimension increases. Therefore in our next exercise we compare the computing time for calculating the likelihood for larger  $p$  dimensions and for longer time series. In Table 3 it is clearly shown that the computational gains are small for small dimensions of the state vector but the gains become much larger as the dimension of the state  $p$  or the length of the time series  $n$  increases. The irregular patterns in the reported gain fractions as  $n$  increases are due to the numerical iterative procedures which rely on convergence criteria and are due to our efficient computer implementations using multiple core platforms. We therefore take the reported fractions in Table 3 as indicative.

Table 3: We report the fraction  $t(EIS)/t(MEIS)$  where  $t(x)$  is the time in seconds required for method  $x$  in computing the log likelihood function, for different state dimensions  $p$  and different time series lengths  $n$ .

	$n = 100$	200	1000	5000	10000	40000
$p = 2$	1.00	1.00	1.50	1.05	1.01	1.43
8	1.00	1.26	2.25	1.53	1.62	2.19
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

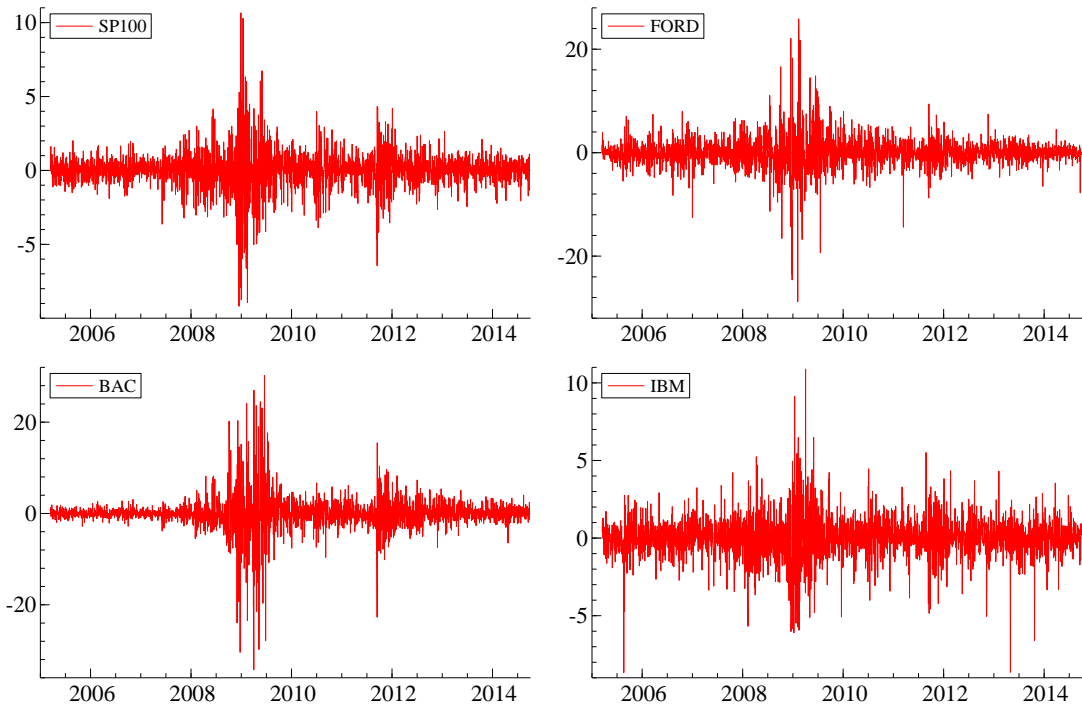


Figure 1: Daily financial returns from 7 October 2004 up to 6 October 2014 (ten years) for the S&P100 stock index and the stocks Ford, Bank of America (BAC) and IBM, all traded at the New York Stock Exchange and based on close prices.

## 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 with multiple volatility factors as adopted in Section 5. We first estimate the parameters  $\phi_j$  and  $\sigma_{\eta,j}$ , for  $j = 1, \dots, p$  from (5) or (22), and  $\nu$  from (21), simultaneously by maximum likelihood. Then we extract log-volatility  $\theta_t$  as specified in (5) and based on the model with estimated parameters.

In our empirical study we consider four daily stock returns, the S&P100 composite index and from 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 is modelled by the SVt 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

SVt specifications and for the four return series are reported in Table 4 together with the maximized log likelihood function values and the corresponding Akaike (AIC) information criterion.

The parameters of the SVt model specifications with  $p = 1, 2, 3$  have all been estimated without any numerical problems 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. There is some evidence found for the decomposition of the log-volatility into two autoregressive components of which the persistency levels are of different orders. However, we have not found much evidence for more than two log-volatility components. The stock returns of Ford and Bank of America appear to favor the two-components log-volatility SVt model in terms of AIC while the daily volatilities in S&P100 and IBM returns can possibly be captured by a single autoregressive component. In the case of Ford and Bank of America, the first component is highly persistent as  $\phi_1$  is virtually equal to unity while the second component is less but still strongly 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 in all four stock returns. Hence our empirical evidence of multiple log-volatility components is not very strong. Finally we have found that the SVt model with  $p = 3$  is difficult to justify empirically, also because it is harder to empirically identify or estimate the model parameters.

The estimated log-volatilities for the four return series analyzed by the SVt model with  $p = 2$  are presented in Figure 2. The 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, for SVt models with  $p = 1$  or  $p = 3$ . We display the three volatility signal estimates for the firms Ford, Bank of America and IBM in three separate plots but together with the S&P100 volatility signal in each of them. Hence we can easily compare between the 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

Table 4: We report the estimation results for the parameters of the SVt 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 SVt 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  $\psi$  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.

	$p$	$\bar{\phi}_1$	$\bar{\phi}_2$	$\bar{\phi}_3$	$\bar{\sigma}_{\eta,1}$	$\bar{\sigma}_{\eta,2}$	$\bar{\sigma}_{\eta,3}$	$\nu$	LogL	AIC
<b>S&amp;P100</b>	1	0.990	-	-	0.149	-	-	8.918	-3289.66	<b>6585.3</b>
	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
<b>FORD</b>	1	0.998	-	-	0.103	-	-	6.858	-5492.96	10991.9
	2	0.999	0.883	-	0.063	0.243	-	9.261	-5485.49	<b>10981.0</b>
	3	0.999	0.884	0.534	0.063	0.241	0.0420	9.258	-5485.48	10985.0
<b>BAC</b>	1	0.998	-	-	0.125	-	-	8.511	-5124.38	10254.8
	2	0.998	0.868	-	0.103	0.190	-	10.508	-5122.39	<b>10254.8</b>
	3	0.998	0.878	0.528	0.102	0.184	0.0427	10.484	-5122.36	10258.7
<b>IBM</b>	1	0.985	-	-	0.139	-	-	6.702	-3890.47	<b>7786.9</b>
	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

that the same applies to Bank of America (BAC) until the financial crisis has started. At the height 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.

The empirical illustration can be extended in various ways, we mention two possibilities. First we can carry out a multivariate stochastic volatility analysis in which the stock index and the three individual stocks in the  $4 \times 1$  observation vector  $y_t$  are then jointly modeled with a common stochastic log-volatility signal. In this modeling approach, the individual stock volatility deviations from the common log-volatility signal can be measured directly. When the signal remains univariate, MEIS can still efficiently carry out the computations, whatever the dimension of  $y_t$ . Second we can consider a heavy-tailed distribution for the state disturbances  $\eta_t$  in order to capture breaks in volatility (such as for the stock Ford) more effectively. In addition we can consider the incorporation of the so-called leverage effect. Such challenging extensions require further amendments for both the EIS and MEIS methods. We regard these considerations as interesting directions for future research projects.



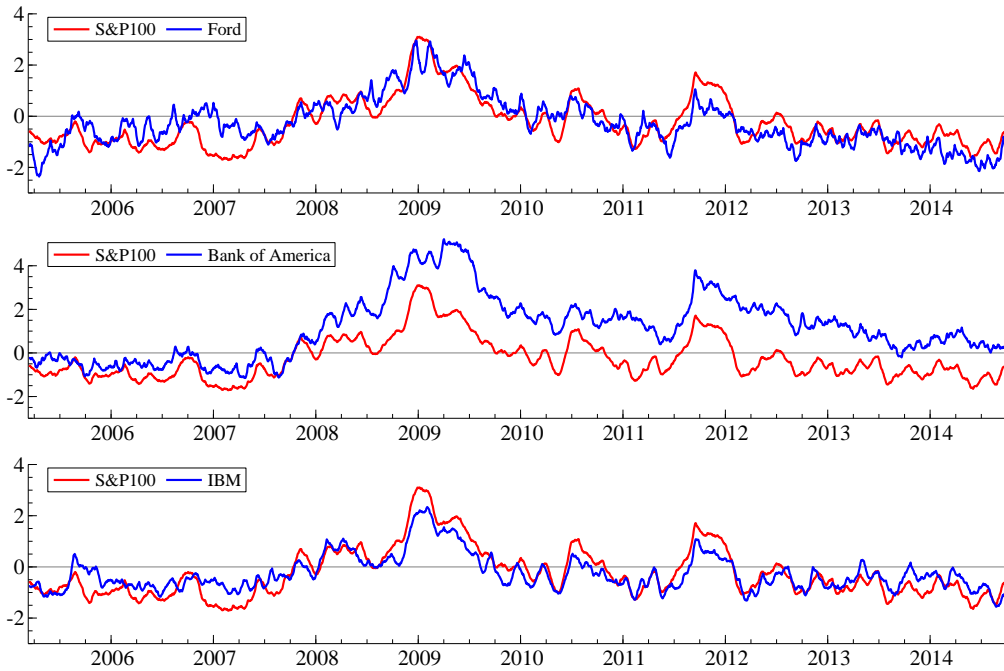


Figure 2: Estimated log-volatilities from 7 October 2004 up to 6 October 2014 (ten years) for the S&P100 stock index and the stocks Ford, Bank of America and IBM, based on the SVt model given by (4) with Student's  $t$  density (21) and log-volatility signal (5) with  $p = 2$ .

## 7 Conclusions

We have presented a new modification of the efficient importance sampling (EIS) method for the analysis of partially 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 an importance density kernel needs to be computed. This process can be algebraically complicated and the simulation of the state vectors in EIS can become computationally involved and not straightforward when models require a large state vector with time-varying parameters. In the MEIS method 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 simpler importance sampling method, especially for large state dimensions. The details of this development have been presented here. To show the relevance of these methods in an empirical study, we have analyzed the volatility patterns in four U.S. stock return series and we have commented 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.
- Fruhworth-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. J., A. Lucas, and M. Scharth (2014). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. *Journal of Business and Economic Statistics* *33(1)*, 114–127.
- Koopman, S. J. 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. (1994). Partial non-Gaussian state space. *Biometrika* *81(1)*, 115–131.
- 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.