Numerically accelerated importance sampling for nonlinear non-Gaussian state space models

Koopman, S.J.; Lucas, A.; Scharth Figueiredo Pinto, M.

2011

TI 2011-057/4
Tinbergen Institute Discussion Paper

Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models

Siem Jan Koopman[a,b]
André Lucas[a,b,c]
Marcel Scharth[a,b]

[a] *VU University Amsterdam;*
[b] *Tinbergen Institute;*
[c] *Duisenberg school of finance.*

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

NUMERICALLY ACCELERATED IMPORTANCE SAMPLING FOR NONLINEAR NON-GAUSSIAN STATE SPACE MODELS

Siem Jan Koopman,[a,b] **André Lucas**[a,b,c] **and Marcel Scharth**[a,b,*]

[a] *VU University Amsterdam*
[b] *Tinbergen Institute, The Netherlands*
[c] *Duisenberg school of finance, The Netherlands*

This version: March 2011

Abstract. We introduce a new efficient importance sampler for nonlinear non-Gaussian state space models. By combining existing numerical and Monte Carlo integration methods, we obtain a general and efficient likelihood evaluation method for this class of models. Our approach is based on the idea that only a small part of the likelihood evaluation problem requires simulation, even in high dimensional settings. We refer to this method as Numerically Accelerated Importance Sampling. Computational gains of our efficient importance sampler are obtained by relying on Kalman filter and smoothing methods associated with an approximated linear Gaussian state space model. Our approach also leads to the removal of the bias-variance tradeoff in the efficient importance sampling estimator of the likelihood function. We illustrate our new methods by an elaborate simulation study which reveals high computational and numerical efficiency gains for a range of well-known models.

Keywords: State space models, importance sampling, simulated maximum likelihood, stochastic volatility, stochastic copula, stochastic conditional duration.

1. Introduction

The evaluation of an analytically intractable likelihood function is a challenging problem for a variety of econometric models. The core difficulty is the numerical evaluation of a high-dimensional integral. Importance sampling methods are typically used to approximate such integrals. The advances in importance sampling over the past three decades have contributed to the interest in nonlinear non-Gaussian state space models including stochastic volatility models by Ghysels et al. (1996), stochastic conditional intensity models by Bauwens and Hautsch (2006), non-Gaussian unobserved components time series models by Durbin and Koopman (2000), and flexible non-linear panel data models with unobserved heterogeneity by Heiss (2008).

We propose a new importance sampler with a high level of numerical and computational efficiency for a general class of nonlinear non-Gaussian state space models. We first show that the likelihood

*Contact author: Marcel Scharth (mscharth@feweb.vu.nl).
Acknowledgements: the authors are thankful for the helpful comments of seminar participants at Mannheim University, Tinbergen Institute Amsterdam, Chicago Booth School of Business, Université Paris I, Sorbonne, and the 4^{th} CSDA conference in London.

evaluation problem for this class of models can be in large part approached by numerical integration. It is known that numerical integration is fast and accurate but its applicability is typically limited to lower dimensional problems. Monte Carlo integration is subject to simulation error but widely applicable to high-dimensional problems. By combining the two approaches, we are able to carry the virtues of numerical integration over to high-dimensional state space models for long time series. We therefore depart from the numerical approaches of Kitagawa (1987) and Fridman and Harris (1998) but also from the simulation methods of Danielsson and Richard (1993) and Durbin and Koopman (1997). We will refer to our method as *numerically accelerated importance sampling* (NAIS).

We next show that the two different importance sampling approaches can be integrated into the implementation of NAIS in a computationally efficient way. The approach of Shephard and Pitt (1997) and Durbin and Koopman (1997), which we refer to as SPDK, is based on an approximating linear Gaussian state space model from which importance samples can be generated by fast methods based on the Kalman filter and smoother (KFS). The approximation of SPDK is optimal as a mode estimator but is a local approximation of the likelihood integral. We adopt the device in Koopman and Nguyen (2011) to show that the efficient importance sampling method of Liesenfeld and Richard (2003) and Richard and Zhang (2007), which we refer to as EIS, can be implemented using the KFS methods used in SPDK. The approximation of EIS is a global approximation of the likelihood integral. The implementation of our NAIS attains a nearly exact numerical solution for obtaining the global approximate likelihood function which we will use as an importance density. The computations rely fully on the computationally efficient KFS methods.

Our method also eliminates the bias-efficiency tradeoff in the EIS method. It is shown that the standard implementation of EIS method is subject to finite sample bias as the algorithm uses the same random numbers for estimating the optimal importance sampling coefficients and for computing the likelihood function estimate. This procedure leads to a violation of the assumption that the parameters of the importance samplers are fixed for sampling purposes. On the other hand, the immediate alternative of using distinct random draws across the two steps may lead to an increase in the variance of the likelihood estimator if the number of Monte Carlo samples used for estimating the optimal sampling coefficients is small. The iterative solution of hundreds of auxiliary regressions within the EIS algorithm therefore imposes a large cost on a numerically efficient implementation of the method. It will be shown that the bias problem can be completely eliminated by the use of NAIS for selecting the importance sampling coefficients. The bias removal comes with a substantial increase in computational efficiency.

We conduct a simulation study to analyze the efficiency gains of the new methods proposed in this paper. To check for the robustness of our results, we consider three different model specifications: the stochastic volatility model, see for example Ghysels et al. (1996), the stochastic duration model of Bauwens and Veredas (2004), and the stochastic copula model of Hafner and Manner (2011). Each of these models requires likelihood evaluation by numerical techniques such as importance sampling. We show that all our methods can be efficiently implemented in each of these different contexts. We obtain three main results. First, we show that the linear state space model approximation

always performs substantially faster than a standard implementation of the EIS method. This holds even without considering numerical acceleration. Second, as we increase the number of importance sampling trajectories our NAIS method proves significantly faster and more accurate than standard EIS. The NAIS method unambiguously improves the tradeoff between computational and numerical efficiency in the number of samples. Finally, for all models considered in our simulation study we are able to reduce the variance of our likelihood estimates by more than 40% with the use of efficiently computed NAIS based control variates, relative to the use of antithetic variables as a variance reduction device.

The paper is structured as follows. Section 2 introduces the notation and the importance sampling concepts used throughout the paper. Section 3 introduces the new importance sampler. Section 4 compare the different IS methods in an extensive simulation study. Section 5 concludes.

2. Importance sampling for nonlinear non-Gaussian state space models

The main ideas of importance sampling are presented in Kloek and van Dijk (1978), Ripley (1987) and Geweke (1989), among others. Importance sampling techniques for state space models have been examined in Danielsson and Richard (1993), Shephard and Pitt (1997), Durbin and Koopman (1997) and other papers. A textbook treatment can be found in Durbin and Koopman (2001). A short review of the literature with additional references is also provided by Richard and Zhang (2007).

Our aim is to evaluate the likelihood function for the nonlinear non-Gaussian state space model

$$\begin{aligned} &y_t|\theta_t \sim p(y_t|\theta_t;\psi), \qquad \theta_t = Z_t\alpha_t, \qquad t = 1,\ldots,n, \\ &\alpha_t = d_t + T_t\alpha_{t-1} + \eta_t, \qquad \alpha_1 \sim \mathrm{N}(a_1, P_1), \qquad \eta_t \sim \mathrm{N}(0, Q_t), \end{aligned} \tag{1}$$

where y_t is a $p \times 1$ observation vector, θ_t is the signal vector, and α_t is an $m \times 1$ state vector. The $m \times 1$ constant vector d_t, the $m \times m$ transition matrix T_t and the variance matrix Q_t determine the dynamic properties of the model. The observation density $p(y_t|\theta_t;\psi)$ may rely on a non-linear transformation of the signal θ_t. The unknown fixed parameter vector ψ contains the unknown coefficients in the observation density as well as unknown coefficients on which the system matrices depend.

When we assume that $p(y_t|\theta_t;\psi)$ is a Gaussian density with mean $\theta_t = Z_t\alpha_t$ and some variance V_t, for $t = 1,\ldots,n$, we can adopt Kalman filter and smoothing methods for evaluating the likelihood function and the minimum mean squared error estimates of the state vector α_t together with its mean square error. However, this paper focuses on non-Gaussian densities for $p(y_t|\theta_t;\psi)$ with parameters that possibly depend on θ_t in a possibly nonlinear fashion.

In our treatment we focus on the signal θ_t. Define $\theta' = (\theta_1', \ldots, \theta_n')$ and $y' = (y_1', \ldots, y_n')$ and let ψ be a fixed and unknown parameter vector. The likelihood for (1) is given by the analytically intractable integral

$$L(y;\psi) = \int p(\theta, y;\psi)\mathrm{d}\theta = \int \prod_{t=1}^{n} p(y_t|\theta_t;\psi)p(\theta_t|\alpha_{t-1};\psi)\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n, \tag{2}$$

where $p(\theta, y; \psi)$ is the joint model density for y and θ that are generated by (1). Kitagawa (1987) has developed a numerical integration method for the evaluation of the likelihood integral. This approach is only feasible when the dimensions of θ and y are small.

For the evaluation of the likelihood function by means of importance sampling, we consider the importance model density $g(\theta, y; \psi) = g(y|\theta; \psi)g(\theta; \psi)$ that we assume to be Gaussian, including $g(y|\theta; \psi)$ and $g(\theta; \psi)$, where $g(\theta; \psi) = p(\theta; \psi)$ is the density for θ from model (1). The likelihood function can then be expressed by

$$\begin{aligned} L(y; \psi) &= \int \frac{p(\theta, y; \psi)}{g(\theta|y; \psi)} g(\theta|y; \psi) \mathrm{d}\theta \\ &= g(y; \psi) \int \frac{p(\theta, y; \psi)}{g(\theta, y; \psi)} g(\theta|y; \psi) \mathrm{d}\theta \\ &= g(y; \psi) \int \omega(\theta, y; \psi) g(\theta|y; \psi) \mathrm{d}\theta, \end{aligned} \tag{3}$$

where $g(y; \psi)$ is the likelihood function of the Gaussian importance model and

$$\omega(\theta, y; \psi) \equiv \frac{p(y|\theta; \psi)}{g(y|\theta; \psi)}, \tag{4}$$

is referred to as the importance weight function. The last equality in (3) is valid since $p(\theta; \psi) = g(\theta; \psi)$. By generating S independent trajectories $\theta^{(1)}, \ldots, \theta^{(S)}$ from the importance density $g(\theta|y; \psi)$, the likelihood function (3) can be estimated by the sample mean

$$\widehat{L}(y; \psi) = g(y; \psi)\bar{\omega}, \qquad \bar{\omega} = \frac{1}{S} \sum_{s=1}^{S} \omega_s, \qquad \omega_s = \omega(\theta^{(s)}, y; \psi), \tag{5}$$

where ω_s is the realized importance weight for $\theta^{(s)}$. Under suitable regularity conditions, the law of large numbers ensures that

$$\widehat{L}(y; \psi) \xrightarrow{p} L(y; \psi). \tag{6}$$

Geweke (1989) has argued that importance sampling should only be used in settings where the variance of the importance weights $\omega_1, \ldots, \omega_S$ is known to exist. Failure of this condition can lead to slow and unstable convergence of the estimator as the central limit theorem governing convergence fails to hold. Koopman et al. (2009) have discussed methods to check for this condition.

The Gaussian importance model density function can be represented as

$$g(\theta, y; \psi) = \prod_{t=1}^{n} g(y_t|\theta_t; \psi) g(\theta_t|\alpha_{t-1}; \psi), \tag{7}$$

where $g(\theta_t|\alpha_{t-1}; \psi)$ is the Gaussian density for θ_t as given by (1) and where Gaussian density $g(y_t|\theta_t; \psi)$ can be expressed as

$$g(y_t|\theta_t; \psi) = \exp\left\{ a_t + b_t' \theta_t - \frac{1}{2} \theta_t' C_t \theta_t \right\}, \tag{8}$$

with a_t, b_t and C_t defined as functions of data vector y and parameter vector ψ, for $t = 1, \ldots, n$. The constants $a_1, \ldots, a_n$ are chosen such that the model density $g(\theta, y; \psi)$ integrates to unity while

the mean and variance of the density function (8) are determined by b_t and C_t only, for $t = 1, \ldots, n$. The variables b_t and C_t therefore determine the properties of importance sampling density. The set of importance sampling parameters is given by

$$\chi = \{b_1, \ldots, b_n, C_1, \ldots, C_n\}. \tag{9}$$

2.1. **SPDK method.** The importance sampling method of Shephard and Pitt (1997) and Durbin and Koopman (1997), which we refer to as SPDK, chooses the variables in χ implicitly via the second-order Taylor expansion of $\log p(y_t|\theta_t;\psi)$ in (1), around a fixed pre-determined value for θ_t, say $\widehat{\theta}_t$, that is

$$\log p(y_t|\theta_t;\psi) \approx \widehat{x}_{0t} + \widehat{x}'_{1t}\,(\theta_t - \widehat{\theta}_t) + \frac{1}{2}(\theta_t - \widehat{\theta}_t)'\,\widehat{x}_{2t}\,(\theta_t - \widehat{\theta}_t), \tag{10}$$

where $\widehat{x}_{it} = x_{it}(\widehat{\theta}_t)$, $i = 0, 1, 2$, with

$$x_{0t}(s_t) = \log p(y_t|\theta_t;\psi)|_{\theta_t = s_t}, \quad x_{1t}(s_t) = \frac{\partial \log p(y_t|\theta_t;\psi)}{\partial \theta_t}|_{\theta_t = s_t}, \quad x_{2t}(s_t) = \frac{\partial^2 \log p(y_t|\theta_t;\psi)}{\partial \theta_t \partial \theta'_t}|_{\theta_t = s_t},$$

for any $p \times 1$ vector s_t and $t = 1, \ldots, n$. The three right-handside terms of the expansion (10) can be represented by the Gaussian importance density function (8) where

$$b_t = \widehat{x}_{1t} - \widehat{x}_{2t}\,\widehat{\theta}_t, \qquad C_t = -\widehat{x}_{2t}, \qquad t = 1, \ldots, n. \tag{11}$$

The integrating constants do not play a role in the SPDK method. The importance density (8) is equivalent to the density function associated with the linear Gaussian observation equation for $y_t^* = C_t^{-1} b_t$ as given by

$$y_t^* = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, C_t^{-1}), \qquad t = 1, \ldots, n, \tag{12}$$

with θ_t specified as in (1). The Gaussian logdensity $\log g(y_t^*|\theta_t;\psi)$ for (12) is given by

$$\begin{aligned} \log g(y_t^*|\theta_t;\psi) &= -\frac{1}{2}\log 2\pi + \frac{1}{2}\log|C_t| - \frac{1}{2}\{(C_t^{-1}\,b_t - \theta_t)'\,C_t\,(C_t^{-1}\,b_t - \theta_t)\} \\ &= a_t + b_t'\,\theta_t - \frac{1}{2}\theta_t'\,C_t\,\theta_t, \end{aligned} \tag{13}$$

which is (8) in logs and therefore $g(y_t|\theta_t;\psi) \equiv g(y_t^*|\theta_t;\psi)$ for $t = 1, \ldots, n$. By considering the linear Gaussian "approximating" model (12), with $y_t^* = C_t^{-1} b_t$ where b_t and C_t are given by (11), for $t = 1, \ldots, n$, we effectively take the second-order Taylor expansion of $p(y|\theta;\psi)$ as the observation equation of the importance density $g(\theta, y;\psi)$. The expansion is around $\widehat{\theta}_t$, for $t = 1, \ldots, n$, and is therefore regarded as a local approximation to $p(y|\theta;\psi)$.

The importance variables b_t and C_t are functions of $\widehat{\theta}_t$, for $t = 1, \ldots, n$. The choice of $\widehat{\theta}_1, \ldots, \widehat{\theta}_n$ therefore determines the accuracy of the Taylor approximation of $\log p(y|\theta;\psi)$. The optimal choice for $\widehat{\theta} = (\widehat{\theta}'_1, \ldots, \widehat{\theta}'_n)'$ is the mode of the smooth density $\log p(\theta|y;\psi)$ which can be obtained by a Newton-Raphson procedure; see So (2003) and Jungbacker and Koopman (2007). It consists of repeatedly estimating θ given $y^* = (y_1^{*\prime}, \ldots, y_n^{*\prime})'$ for the linear Gaussian model (12) using the

Kalman filter and smoothing methods. The estimate $\widehat{\theta}_t$ allows the computation of new values for b_t, C_t and y_t^*, for $t = 1, \ldots, n$. A new estimate $\widehat{\theta}_t$ can then be obtained from (12) for the new y_t^*'s.

After convergence of this recursive process, the linear Gaussian model (12), with θ_t specified by (1), is used to sample θ_t from $g(\theta_t|y;\psi)$, $t = 1, \ldots, n$, by a simulation smoothing method; see, for example, de Jong and Shephard (1995) and Durbin and Koopman (2002). This step follows by noting that $g(\theta|y^*) \propto g(y^*, \theta)$. Jungbacker and Koopman (2007) show that the individual matrices C_t only need to be non-singular for $g(\theta|y^*)$ to be well defined. However, when any of the matrices C_t fails to be negative definite the alternative sampling scheme developed therein is required. The simulations are collected in $\theta^{(s)}$. This sampling step is repeated S times. The vector $\theta^{(s)}$ is used to compute the importance sampling weight ω_s in (5) for $s = 1, \ldots, S$. The Monte Carlo estimate of the likelihood function is then evaluated as in (5).

2.2. **Modified EIS method.** In the EIS method of Liesenfeld and Richard (2003) and Richard and Zhang (2007), the importance sampling density is obtained by focusing on the importance weights directly. Here we discuss a modification of the EIS method which we refer to as the modified EIS (MEIS) and is proposed by Koopman and Nguyen (2011). The modification consists of an alternative method for the simulation of θ_t based on a linear state space approximation. The objective is to (i) increase computational efficiency (ii) implement the EIS algorithm for a sampler in which the marginal distribution of θ_t is available, as required by the NAIS method of Section 3. The reader is referred to Appendix A for a review of the approach proposed by Richard and Zhang (2007).

The aim of EIS is to choose the importance parameters in χ of (9) such that the variance of the log-weights $\log \omega(\theta, y, \psi)$ is minimized where the weight $\omega(\theta, y, \psi)$ is defined in (4). It is expected that the resulting importance sampling density accurately approximates the smooth density of interest. We have

$$\min_{\chi} \int \lambda^2(\theta, y; \psi) p(\theta|y;\psi) \mathrm{d}\theta, \tag{14}$$

where

$$\lambda(\theta, y; \psi) = \log p(y|\theta;\psi) - \log g(y|\theta;\psi) - \lambda_0, \tag{15}$$

where $g(y|\theta;\psi) = \prod_{t=1}^n g(y_t|\theta_t;\psi)$ and $g(y_t|\theta_t;\psi)$ is given by (8). The normalizing constant λ_0 ensures that $\lambda(\theta, y;\psi)$ has mean zero. The resulting importance density that we obtain from this minimization can be regarded as a global approximation of $p(y|\theta;\psi)$.

The minimization (14) encounters three problems: (i) the minimization is high-dimensional and numerically not feasible in most cases of interest; (ii) the evaluation of the integral in (14) is only feasible via importance sampling based on a given value of χ (iii) Monte Carlo methods for approximating the minimization problem entail a bias-variance trade-off for estimating the likelihood function (3). To address problem (i), we follow Richard and Zhang (2007) and carry out the minimization separately for each time point t. For a given set of importance parameters $\chi = \chi^*$ and for

each t, we minimize

$$\min_{\chi_t} \int \lambda^2(\theta_t, y_t; \psi) p(\theta_t|y;\psi) \quad \Leftrightarrow \quad \min_{\chi_t} \int \lambda^2(\theta_t, y_t; \psi)\omega(\theta_t, y_t; \psi) g(\theta_t|y;\psi), \tag{16}$$

where

$$\lambda(\theta_t, y_t; \psi) = \log p(y_t|\theta_t;\psi) - \log g(y_t|\theta_t;\psi) - \lambda_{0t}, \qquad \omega(\theta_t, y_t;\psi) = \frac{p(y_t|\theta_t;\psi)}{g(y_t|\theta_t;\psi)},$$

to obtain new values for $\chi_t = \{b_t\,,\, C_t\}$ where λ_{0t} is the normalizing constant. The equivalence of both minimizations in (16) follows from the usual identities such as $p(\theta_t|y_t;\psi) = p(\theta_t;\psi)p(y_t|\theta_t;\psi)/p(y_t;\psi)$ and from $p(\theta_t;\psi) = g(\theta_t;\psi)$ while $g(y_t;\psi)/p(y_t;\psi)$ can be treated as a fixed scaling constant for the minimization (16).

To address problem (ii), we follow an approach similar to Richard and Zhang (2007). The integral in (16) is evaluated by importance sampling and therefore the minimization (16) is replaced by

$$\min_{\chi_t} \frac{1}{S}\sum_{s=1}^{S} \lambda_t^2(\theta_t^{(s)}, y_t;\psi)\omega_{ts},$$

where $\omega_{ts} = \omega(\theta_t^{(s)}, y_t;\psi)$ and the draw $\theta_t^{(s)}$ is from $g(\theta_t|y;\psi)$, that is the Gaussian importance density model (12) for $y_t^* = C_t^{-1}b_t$ with b_t and C_t taken from χ_t^*. Since $\log g(y_t|\theta_t;\psi)$ is of the form (8), $\lambda_t(\theta_t, y_t;\psi)$ is a function of b_t and C_t due to logdensity $\log g(y_t|\theta_t;\psi)$ only. The minimization is with respect to both b_t and C_t and can be carried out by standard least squares calculations once we have computed sufficient S draws $\theta_t^{(1)}, \ldots, \theta_t^{(S)}$. The least squares estimates can be taken as new values for b_t and C_t. The minimization can then be repeated by taking the new values of b_t and C_t as χ_t^*. It leads to a recursive algorithm that typically converges quickly. In each step of the recursion, we require to simulate S draws from $g(\theta|y;\psi)$ and we use a simulation smoothing method that uses the same underlying random values for each step in the recursion and for $t = 1, \ldots, n$. The simulation smoothing method of SPDK can be adopted for this purpose.

The bias-variance trade off is referred to as problem (iii) for EIS and is due to the possible use of the same set of random numbers for obtaining χ and for computing the loglikelihood function. When the same random numbers are used for determining χ and for computing $\theta^{(s)}$, which is the typical case in applied work, a finite sample bias is introduced in the estimation of the loglikelihood function; the details are discussed in the Appendix B. This implementation minimizes the sample variance of importance weights and therefore reduces the variance of the likelihood estimate. The simulation results reported in Section 4 indicate that the bias is substantive in practice, although the biased estimator is mean-square efficient in most cases of interest.

3. Numerically Accelerated Importance Sampling

In our exposition below we take θ_t univariate. Even though extensions to higher dimensional signals follow immediately from the notation, cases where θ_t is multivariate bring additional computational challenges that are out of the scope of this paper. We note that a moderate dimension for signal vector θ_t can still rely on a high-dimensional state vector α_t in (1).

When some continuous function $\varphi(x)$ is known analytically for any x, we can efficiently evaluate integrals of the form

$$\int \varphi(x)\mathrm{d}x, \tag{17}$$

by numerical integration methods which are fast, reliable, and accurate. Numerical integration is not prone to simulation uncertainty and can be computed to any degree of precision.

The numerical evaluation of integral (17) via Gauss-Hermite quadrature designates a set of M abscissae z_j and associated weights $h(z_j)$ with $j = 1, \ldots, M$. The numerical approximation is then based on

$$\int_{-\infty}^{\infty} \varphi(x)\mathrm{d}x = \int_{-\infty}^{\infty} e^{-x^2}[e^{x^2}\varphi(x)]\mathrm{d}x \approx \sum_{j=1}^{M} h(z_j)e^{z_j^2}\varphi(z_j),$$

where M is typically between 20 and 30. The weights $h(z_j)$ can be tabulized. For a more detailed discussion on Gauss-Hermite quadrature, we refer to Press et al. (1992).

3.1. **Construction of importance sampler via numerical integration.** The SPDK and MEIS methods are designed to construct an importance sampler by computing an appropriate value for χ in (9). The values in χ are used in (8) for drawing θ's from $g(\theta|y;\psi)$ using a simulation smoothing method. The local SPDK determines χ from a Taylor expansion of $p(y|\theta;\psi)$ around the mode $\widehat{\theta} = (\widehat{\theta}_1', \ldots, \widehat{\theta}_n')'$ of the smooth density $p(\theta|y;\psi)$ while the global MEIS is based on the variance minimization (16).

Here we propose to obtain χ via the minimization (16) as in MEIS but we evaluate the integral in via Gauss-Hermite quadrature which is possible since $g(\theta_t|y;\psi)$ is known analytically. In effect, the integral in (16) is equivalent to (17) with $x = \theta_t$ and

$$\varphi(\theta_t) = \lambda^2(\theta_t, y_t;\psi)\omega(\theta_t, y_t;\psi)g(\theta_t|y;\psi).$$

Since $p(y_t|\theta_t;\psi)$, $g(y_t|\theta_t;\psi)$ and $g(\theta_t|y;\psi)$ are known analytically, we can adopt the Gauss-Hermite method. We notice that

$$g(\theta_t|y;\psi) = \mathrm{N}(\widehat{\theta}_t\,,\, V_t) = \exp\left\{-\frac{1}{2}V_t^{-1}(\theta_t - \widehat{\theta}_t)^2\right\} \,/\, \sqrt{2\pi\, V_t},$$

where $\widehat{\theta}_t$ and V_t are computed by the Kalman filter and smoother applied to the importance model (12) for $y_t^* = C_t^{-1}b_t$ and for θ_t specified as (1). The minimization (16) reduces to

$$\min_{\chi_t} \sum_{j=1}^{M} \lambda_t^2(\widetilde{\theta}_{tj}, y_t;\psi)\, g(\widetilde{\theta}_{tj}|y;\psi)\, \omega_{tj}\, h(z_j)e^{z_j^2}, \tag{18}$$

where $\widetilde{\theta}_{tj} = \widehat{\theta}_t + V_t^{1/2}z_j$ and $\omega_{tj} = \omega(\widetilde{\theta}_{tj}, y_t;\psi)$. It follows that $g(\widetilde{\theta}_{tj}|y;\psi) = \exp\left\{-\frac{1}{2}z_j^2\right\} \,/\, \sqrt{2\pi}$.

The minimization (18) takes place in the same recursive manner as described for the MEIS method in Section 2.2. For a given $\chi = \chi^*$, we obtain $\widehat{\theta}_t$ and V_t from the Kalman filter and smoother for $t = 1, \ldots, n$. The minimization (18) is solved via least squares computations and the estimates for b_t and C_t replace their values in χ^* in the next recursive step. These two steps are repeated until convergence. The computing time can be reduced when parallel computing techniques are used.

Our method opens this possibility since no dependency exist between the computations for different time periods once $\widehat{\theta}_t$ and V_t, $t = 1, \ldots, n$, are computed by the Kalman filter and smoother.

When the importance parameters in χ are obtained, importance sampling based on $g(\theta|y;\psi)$ takes place as usual. For example, the likelihood function can be estimated by importance sampling as in (5). Draws for θ are obtained by simulation smoothing methods based on the importance density model (12) with parameter set χ. Because we solve (16) by numerical integration, the NAIS algorithm is not subject to the bias-efficiency problem of the EIS method as described in Appendix B.

3.2. **Introducing control variables.** Antithetic variables can be used to improve the efficiency of the likelihood estimate computed by importance sampling; see, for example, Ripley (1987) and Durbin and Koopman (2000). Here we introduce new control variates that reduce variances of estimates computed by NAIS. A control variate is a variable whose mean is known such that the difference between its sample estimate and the true mean can be used to adjust the importance sampling estimate. In contrast to earlier applications, we adopt control variates that can be quickly computed numerically.

The likelihood estimate (5) is the sample average $\bar{\omega} = S^{-1}\sum_{s=1}^{S}\omega_s$ multiplied by $g(y;\psi)$ where

$$\omega_s = \omega(\theta^{(s)}, y;\psi) = \prod_{t=1}^{n}\omega_{ts}, \qquad \omega_{ts} = \omega(\theta_t^{(s)}, y_t;\psi), \quad t = 1,\ldots,n, \quad s = 1,\ldots,S,$$

for a sample of S draws of θ generated from the smooth density $g(\theta|y;\psi)$ and denoted by $\theta^{(1)},\ldots,\theta^{(S)}$, where the tth element of $\theta^{(s)}$ is denoted by $\theta_t^{(s)}$. The density $g(y;\psi)$ is computed by the Kalman filter. The densities $g(y;\psi)$ and $g(\theta|y;\psi)$ refer to the importance model (12) with θ_t specified as (1) and with the importance parameter set χ obtained as in Section 3.1. The variance of the sample average $\bar{\omega}$ determines the efficiency of the importance sampling likelihood estimate (5).

To reduce the variance of $\bar{\omega}$ we construct a control variate based on

$$x(\theta, y;\psi) = \log\omega(\theta, y;\psi) = \log p(y|\theta;\psi) - \log g(y|\theta;\psi).$$

The tth element of these functions are given by $x(\theta_t, y_t;\psi) = \log\omega(\theta_t, y_t;\psi)$ such that $x(\theta, y;\psi) = \sum_{t=1}^{n} x(\theta_t, y_t;\psi)$ Given the draws $\theta^{(1)},\ldots,\theta^{(S)}$, we have $x_s = \log(\omega_s) = \sum_{t=1}^{n} x_{ts}$ where

$$x_{ts} = \log(\omega_{ts}), \qquad \omega_{ts} = \exp(x_{ts}),$$

for $t = 1,\ldots,n$ and $s = 1,\ldots,S$. The sample average of ω_s can be written in terms of $x_s = \log\omega_s$ by means of a Taylor series around some value x^*, that is

$$\bar{\omega} = \exp(x^*)\,\frac{1}{S}\sum_{s=1}^{S}\left(1 + [x_s - x^*] + \frac{1}{2}[x_s - x^*]^2 + \ldots\right). \tag{19}$$

The second and higher order terms of this expansion can be used as control variables. The use of these control variables is equivalent to replacing the highest variance terms of the Taylor series by their probability limits which are computed to the desired degree of precision.

3.3. **A first control variable.** Under suitable regularity conditions, we have

$$\bar{x} = \frac{1}{S}\sum_{s=1}^{S} x_s \xrightarrow{p} \widehat{x}, \tag{20}$$

where $\widehat{x} = \mathbb{E}_g x(\theta, y; \psi)$ and where $\mathbb{E}_g$ is expectation with respect to density $g(\theta|y;\psi)$. It follows that an accurate Taylor expansion (19) is around $x^* = \widehat{x}$. Since

$$\widehat{x} = \sum_{t=1}^{n} \mathbb{E}_g \left[x(\theta_t, y_t; \psi) \right],$$

we can evaluate $\widehat{x}$ by means of the Gauss-Hermite quadrature method for each t separately as discussed in Section 3.1, that is

$$\widehat{x}_t = \mathbb{E}_g x(\theta_t, y_t; \psi) = \int x(\theta_t, y_t; \psi)\, g(\theta_t|y;\psi)\, \mathrm{d}\theta_t \approx \sum_{j=1}^{M} x(\widetilde{\theta}_{tj}, y_t; \psi)\, g(\widetilde{\theta}_{tj}|y;\psi) h(z_j) e^{z_j^2},$$

where $\widetilde{\theta}_{tj} = \widehat{\theta}_t + V_t^{1/2} z_j$ with $\widehat{\theta}_t$ and V_t computed by the Kalman filter smoother for $t = 1, \ldots, n$. It follows that $\widehat{x} = \sum_{t=1}^{n} \widehat{x}_t$.

The likelihood estimate (5) corrected for the control variable that is based on the second term of the Taylor expansion (19) around $x^* = \widehat{x}$ is given by

$$\begin{aligned}\widehat{L}(y;\psi)_c &= g(y;\psi) \left(\exp(\widehat{x})\widehat{x} + \frac{1}{S}\sum_s \left[\omega_s - \exp(\widehat{x}) x_s\right] \right) \\ &= \widehat{L}(y;\psi) + g(y;\psi)\, \exp(\widehat{x})\, (\widehat{x} - \bar{x}).\end{aligned}$$

It follows that

$$\widehat{L}(y;\psi)_c \xrightarrow{p} L(y;\psi),$$

given (6) and (20). When the importance model is accurately designed, we expect ω_s to be close to one and x_s to be close to zero. The two-term Taylor series of an exponent function insists that $\omega_s \approx 1 + x_s$. Hence, we expect ω_s and $\exp(\widehat{x})x_s$ to be highly and positively correlated. When the importance model is less accurate, the positive correlation remains. Therefore we expect in practice that $\widehat{L}(y;\psi)_c$ is a more efficient estimate of the likelihood function compared to $\widehat{L}(y;\psi)$.

3.4. **A second control variable.** Our second control variable is based on the third term of the Taylor expansion (19). We aim to correct for the sample variation of $(x_{ts} - \widehat{x}_t)^2$ within the sample of draws $\theta_t^{(1)}, \ldots, \theta_t^{(S)}$ for each t individually, where $\widehat{x}_t$ is the tth element of $\widehat{x}$. Based on the same arguments of Section 3.3, we have

$$\bar{\sigma}_t^2 \xrightarrow{p} \widehat{\sigma}_t^2,$$

where

$$\bar{\sigma}_t^2 = \frac{1}{S}(x_{ts} - \widehat{x}_t)^2, \qquad \widehat{\sigma}_t^2 = \mathbb{E}_g (x_{ts} - \widehat{x}_t)^2 = \int (x_{ts} - \widehat{x}_t)^2 g(\theta_t|y;\psi) \mathrm{d}\theta_t.$$

The mean variance $\widehat{\sigma}_t^2$ can be computed using the Gauss-Hermite quadrature. Define

$$\widehat{L}(y;\psi)_{cc} = \widehat{L}(y;\psi)_c + \frac{1}{2} g(y;\psi)\, \exp(\widehat{x}) \sum_{t=1}^{n} (\widehat{\sigma}_t^2 - \bar{\sigma}_t^2),$$

from which it follows that $\widehat{L}(y;\psi)_{cc} \xrightarrow{p} L(y;\psi)$. Since we replace the sample variation of $(x_{ts} - \widehat{x}_t)^2$ by its probability limit, we can expect estimate $\widehat{L}(y;\psi)_{cc}$ to be more efficient than $\widehat{L}(y;\psi)$ and $\widehat{L}(y;\psi)_c$.

The estimate $\widehat{L}(y;\psi)_{cc}$ can be written as

$$\widehat{L}(y;\psi)_{cc} = \widehat{L}(y;\psi) + g(y;\psi) \exp(\widehat{x})\, \frac{1}{S} \sum_{s=1}^{S} \sum_{t=1}^{n} \tau_{ts}, \qquad \tau_{ts} = (\widehat{x}_t - x_{ts}) + \frac{1}{2}(\widehat{\sigma}_t^2 - [\widehat{x}_t - x_{ts}]^2),$$

where τ_{ts} can be regarded as a Taylor approximation of $\exp(-\widehat{x})\omega_{ts} - 1$.

The weights of 1 and $\frac{1}{2}$ for the two terms in τ_{ts} are justified by the Taylor expansion (19). However, these values might not be optimal as they do not fully take into account the covariances between $\widehat{L}(y;\psi)$, $\bar{x}$ and $\sum_{t=1}^{n} \bar{\sigma}_t$. For a finite sample $\theta^{(1)}, \ldots, \theta^{(S)}$, we can estimate the variance minimizing weights by ordinary least squares applied to the regression equation

$$\exp(-\widehat{x})\omega_s = \beta_0 + \beta_1(\widehat{x} - x_s) + \beta_2 \sum_{t=1}^{n} (\widehat{\sigma}_t^2 - [\widehat{x}_t - x_{ts}]^2) + \epsilon_s,$$

where β_k are regression coefficients for $k = 0, 1, 2$ and ϵ_s is the disturbance. We denote the resulting estimate as $\widehat{L}(y;\psi)^*_{cc}$. The use of least squares estimates of this type for assigning weights to control variables dates back to Ripley (1987). As pointed out by the author, a drawback of the current modification is the introduction of a small sample bias, which arises because the OLS regression involves independent variables which are random.

4. Simulation Study

4.1. **Likelihood estimation.** We examine the performances of the importance sampling methods listed in Table 1 for likelihood estimation. The design of the simulation study is as follows. We consider fifty randomly generated time series from the models discussed below. For each of simulated time series, we estimate the loglikelihood function at the true parameters hundred times and each time with a different choice of a random seed. For each method, we report results based on $S = 20$ and $S = 200$ Monte Carlo trajectories. We are interested in the bias, the variance and the root mean squared error of these estimates. Although the true loglikelihood value is unknown, we approximate it as the average of loglikelihood estimates from the NAIS and NAIScc methods. This corresponds to a NAIS likelihood estimate with $S = 40000$ and therefore the error in this approximation is imperceptible in the arithmetic precision of the reported results.

The reported values are the average statistics across the fifty simulated time series. We have taken fifty simulations to avoid the dependence of our conclusion on particular trajectories. We have take hundred estimates to obtain reliable statistics for bias, standard deviation and root mean

square error. The reported statistics are computed as

$$\text{Bias} = 0.0002 \sum_{i=1}^{50} \sum_{j=1}^{100} \left(\log \widehat{L}^j(y^i;\psi) - \log L(y^i;\psi) \right), \tag{21}$$

$$\text{Stand.dev} = 0.002 \sum_{i=1}^{50} \left[\sum_{j=1}^{100} \left(\log \widehat{L}^j(y^i;\psi) - \log \overline{L}(y^i;\psi) \right)^2 \right]^{1/2}, \tag{22}$$

$$\text{Rmse} = 0.002 \sum_{i=1}^{50} \left[\sum_{j=1}^{100} \left(\log \widehat{L}^j(y^i;\psi) - \log L(y^i;\psi) \right)^2 \right]^{1/2}, \tag{23}$$

where $\log \overline{L}(y^i;\psi) = 0.01 \sum_{j=1}^{100} \log L^j(y^i;\psi)$. The Rmse is reported as a ratio with the MEIS method as the benchmark in each setting.

Since computational efficiency is the main objective of importance sampling, for each method and setting, we report the median computing times based on a machine equipped with an Intel Duo Core 2.5GHz. For our simulation study below we also report a statistic that takes into account both the numerical and computational efficiencies for each method. We therefore also report the Rmse ratio normalized by the associated computing times, that is

$$TNR_i = \frac{\widehat{Rmse_i} \times \sqrt{T_i}}{\widehat{Rmse_b} \times \sqrt{T_b}},$$

where T_i is the median computing time for method i and b indexes the benchmark. All reported computing times include the fixed time costs required for obtaining the sampling parameters.

We consider time series lengths of 1000 and 3000 observations. The antithetic variables for location and scale, as in Durbin and Koopman (2001), are used for variance reduction in all cases, except for the NAIScc and NAIScc* methods. The number of nodes for the methods that involve numerical integration is kept at 20 in all cases. We verify the sensitivity of this choice of 20 nodes at the end of §4.3 and in Table 7. All methods have been implemented using MATLAB and C.

4.2. **Models.**

4.2.1. *Stochastic volatility model.* The stochastic volatility (SV) model is a special case of a non-Gaussian state space model; key references are Tauchen and Pitts (1983), Taylor (1986) and Melino and Turnbull (1990). Ghysels et al. (1996) have provided a review of SV models. Liesenfeld and Richard (2003) analyzes the simulated maximum likelihood estimation via efficient importance sampling for a wide range of stochastic volatility specifications. For a time series of log-returns y_t, we consider the model specification

$$\begin{aligned} y_t \sim N(0,\sigma_t^2), \qquad & \sigma_t^2 = \exp(\alpha_t), \\ & \alpha_t = d + T\alpha_{t-1} + \eta_t, \qquad \eta_t \sim N(0,Q), \\ & \alpha_1 \sim N(d/(1-T), Q/(1-T^2)), \end{aligned} \tag{24}$$

for $t = 1, \ldots, n$, where d is a scalar constant, T is the autoregressive coefficient with $|T| < 1$, and Q is the variance of the disturbance η_t of the stochastic log-volatility process α_t. We have two sets of parameter values for the unknown coefficients of the SV model. The first set consists of $d = 0.01$, $T = 0.98$ and $Q_t = 0.01$ which reflects a typical set of parameters found for daily stock returns. The second set is the same but with a lower value for the autoregressive coefficient, $T = 0.9$ (the constant is set to $d = 0.05$ to imply the same unconditional mean). It will reveal how importance sampling methods perform when the volatility process is less persistent.

4.2.2. *Stochastic conditional duration model.* The stochastic conditional duration (SCD) model is proposed by Bauwens and Veredas (2004) for modelling high-frequency durations between financial transactions. Efficient importance sampling estimation for the SCD model has been studied by Bauwens and Galli (2009). For a time series of durations y_t, we consider the model specification

$$y_t \sim \text{Weibull}(\lambda_t, \psi), \qquad \lambda_t = \exp(\alpha_t),$$

for $t = 1, \ldots, n$, where λ_t is the time varying scale parameter, ψ is the shape parameter of the Weibull distribution and with α_t modelled as the autoregressive process (24). The set of parameters is chosen to reflect the estimation results of Bauwens and Galli (2009), that is $d = 0$, $T = 0.98$, $Q = 0.0225$, and $\psi = 1.2$. The choice of $d = 0$, $T = 0.96$, $Q = 0.01$, $\psi = 1.7$ approximates the parameters for volume durations in the same paper. The choice of $d = 0$, $T = 0.9$, $Q = 0.0225$, $\psi = 1.2$ is taken to illustrate the performance of the methods for less persistent price durations.

4.2.3. *Stochastic copula.* Stochastic copula (SC) models are designed to capture time-varying and possibly non-linear dependence between multiple time series and have been proposed by Banachewicz and Lucas (2008) and Hafner and Manner (2011). A short introduction of the main concepts and results for (static) copulas can be found in Schmidt (2006), while Nelsen (1999) and Joe (1997) provide a more comprehensive discussion. Dynamic copula models have been introduced by Patton (2006) who extends the copula theory for models with conditionally time-varying parameters.

We consider a dynamic stochastic bivariate t-copula. Let u_{1t}, u_{2t} be two random variables with uniform $(0, 1)$ marginal distributions. The dependence structure for the random variables is described by the t-copula $C_{\nu,P}(u_t)$, with $u_t = (u_{1t}, u_{2t})'$ and 2×2 correlation matrix P, and is given by

$$C_{\nu,P}(u_t) = t_{\nu,P}(t_{\nu,1}^{-1}(u_{1t}), t_{\nu,1}^{-1}(u_{2t})),$$

where t_{ν,P^*} is the cumulative density function of the standardized t distribution with ν degrees of freedom for variables with its dimension implied by the correlation matrix P^* for $P^* = P$ and $P = 1$. The copula is invariant under any standardization of the marginal distributions. It follows that

$$C_{\nu,P}(u_t) = \int_{-\infty}^{t_\nu^{-1}(u_{1t})} \int_{-\infty}^{t_\nu^{-1}(u_{2t})} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\nu/2)\sqrt{(\pi\nu)^2|P|}} \left(1 + \frac{x'P^{-1}x}{\nu}\right)^{-\frac{\nu+2}{2}} \mathrm{d}x. \tag{25}$$

In our simulation study, we take u_t as probability integral transforms of two independent univariate series. The converse of Sklar's theorem implies that by combining any set of univariate distributions

with a copula, a bivariate distribution is defined. The modeling of the marginal variables and the dependence between them can be completely disentangled.

The state space representation of the stochastic copula with the time-varying correlation coefficient ρ_t and with the time-varying correlation matrix

$$P_t = \begin{bmatrix} 1 & \rho_t \\ \rho_t & 1 \end{bmatrix},$$

is given by

$$u_t \sim C_{\nu,P_t}(u_t), \qquad \rho_t = (1+\exp(-\alpha_t))^{-1},$$

for $t = 1, \ldots, n$, with α_t modeled as the autoregressive process (24). The set of parameters are taken from the empirical study of a bivariate financial time series of log-returns in Hafner and Manner (2011) and is approximately equal to the set $d = 0.017$, $T = 0.98$, $Q = 0.01$, and $\nu = 5$. The constant d implies an unconditional correlation coefficient of approximately 0.7. Since Hafner and Manner (2011) do not consider a t-copula, we take the degrees of freedom $\nu = 5$ to have sufficient tail dependence.

4.3. **Simulation results.** The results for the persistent stochastic volatility model specification are presented on Table 2. The first important finding is that the MEIS method (the EIS based on an approximating linear state space model) is computationally much more efficient than the EIS method itself. The increase in computing speed comes without loss of numerical efficiency. While the EIS and MEIS methods are numerically equivalent, the time needed for evaluating the likelihood is reduced by at least 50% when the approximating linear state space model is used for MEIS. For sample size $n = 3000$ and $S = 20$, the MEIS likelihood evaluation procedure takes 0.11 seconds while the EIS takes 0.39 seconds, a computational saving of more than 70%. The fast computations for $S = 20$ come however at the cost of substantial bias; for the EIS and MEIS methods, the bias is approximately 30% higher than the standard deviation. The low variance of the EIS method compared to the SPDK method shows the numerical efficiency of the global approximation used in the EIS method. However, when we normalize the computing times by the Rmse statistic, the SPDK method turns out to be nearly as efficient as the EIS method.

By increasing the number of importance samples to $S = 200$ the bias in the EIS and MEIS methods is mostly eliminated, at the cost of a tenfold increase in computing time. On the other hand, by increasing S from 20 to 200, we expect a $1/\sqrt{10}$ reduction in standard deviation but we obtain a much smaller reduction for the EIS and MEIS methods. For both $S = 20$ and $S = 200$ the NAIS method (without control variables) produces slightly higher variance and Rmse values when compared the MEIS method. This result suggests that the biased EIS algorithm is mean square efficient in the SV case. However, when $S = 200$ the NAIS method compute the likelihood function four times faster and hence is much more efficient in real time. By obtaining the optimal sampling coefficients at a small and fixed cost, the NAIS method significantly improves the trade off between numerical and computational efficiency in the number of samples relative to the SPDK, EIS and MEIS methods. We regard this result as one of our main findings from the simulation study. We

also highlight the minimal additional computational time required for the NAIS method to increase S from 20 to 200.

Another interesting finding is that NAIScc and NAIScc* methods are substantially more efficient in Rmse compared to the NAIS algorithm, while likelihood evaluation is just as fast. For $S = 20$, the results show that the control variates reduce the Rmse by 30%, relative to the NAIS method with antithetic variables. The relative efficiency of the NAIScc method also improves in the number of samples. For $S = 200$, the NAIScc is 36% more efficient in Rmse for $n = 1000$ and 42% more efficient for $n = 3000$. The results show that the NAIScc* method further improves the Rmse by around 10% relative to the NAIScc procedure. Although this difference is small, the NAIScc* method is still an useful extension when S is large since the additional computational cost is low. Finally, we note that the bias statistics suggest no systematic differences between the average estimates for the NAIS and NAIScc methods, supporting the claim that the numerical integration error in constructing the control variables can be ignored.

Table 3 for the stochastic volatility model with the autoregressive coefficient of 0.9 presents some revealing results. For less persistent specifications of the state space model, the EIS bias becomes a larger problem. The bias almost completely dominates the Rmse of the EIS and MEIS methods when $S = 20$; it is 10 times larger then the standard deviation. However the results for our methods become even more convincing. The reduction in Rmse is 95% when we compare the EIS method for $S = 20$ with the NAIS method for $S = 200$. The NAIScc method is again more efficient than the NAIS method; the variance of the likelihood estimate is reduced by 75%.

Tables 4, 5 and 6 present the findings for the different specifications of the stochastic conditional duration and the stochastic copula models. Although these models are more complex and the likelihood evaluation algorithms become more time consuming, the results confirm our previous findings. For $S = 200$, the NAIScc and NAIScc* methods consistently bring Rmse reductions of 25-50% or more when compared to the simpler NAIS alternative. Similar results hold for $S = 20$. For the SCD model and $S = 20$, we find that the EIS and MEIS methods produce estimates with the lowest Rmse but with substantial biases. Hence, certain parameter combinations may favor the MEIS method if S is low. However, the cost of increasing the number of simulations from $S = 20$ to $S = 200$ is small for all NAIS methods. We therefore conclude that the results strongly favor the NAIS methods with a higher value for S, since the additional computational cost is low and the properties of the estimates are much improved.

Tables 7 and 8 present additional results. Table 7 presents the standard deviation of the log importance sampling weights for different choices of S under the MEIS method, both in sample and out of sample, for the SV and SC models. The standard deviation of the log importance sampling weights, also when the importance parameters are selected through the numerical method, are also reported. When S increases, the variance of the MEIS log importance sampling weights converges to a limiting value that is obtained by our NAIS method. The table also further illustrates the source of the EIS bias. For low values of S, the sample variance of the weights is artificially small. Table 8 focuses on the robustness of the NAIS method with respect to the choice of the number of numerical

integration nodes. As typical in other applications of Gauss Hermite integration, any value of M between 20 and 30 guarantees a high degree of accuracy. A comparison between Table 8 and the previous results confirms that the numerical integration error is negligible in relation to standard deviation of the likelihood estimates.

4.4. **Parameter estimation.** To further illustrate the performance of our preferred NAIScc method, we consider the simulated maximum likelihood estimation for a three component stochastic volatility model which we specify as (24) but with

$$\sigma_t^2 = \exp(\theta_t), \qquad \theta_t = d + \alpha_{1,t} + \alpha_{2,t} + \alpha_{3,t},$$

with 3×1 state vector $\alpha_t = (\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t})'$, with 3×3 diagonal matrices T and Q given by

$$T = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ 0 & 0 & \phi_3 \end{bmatrix}, \qquad Q = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix},$$

and with unknown coefficients $|\phi_i| < 1$ and $q_i > 0$, for $i = 1, 2, 3$. The model is identified by imposing $\phi_1 > \phi_2 > \phi_3$. The scalar signal θ_t represents log-volatility. The initial condition for α_1 is adjusted straightforwardly.

The three component SV model is motivated from the two component GARCH model of Engle and Lee (1999). We have extended the volatility specification by an additional component to capture the weakly serially correlated noise that is typically found in realized volatility studies; see, for example, Bollerslev et al. (2009). Liesenfeld and Richard (2003) investigate their EIS method for a two component SV model. The parameter values are set as $d = 0.5$, $\phi_1 = 0.99$, $\phi_2 = 0.9$ $\phi_3 = 0.4$, $q_1 = 0.005$, $q_2 = 0.016$ $q_3 = 0.05$.

We adopt the following steps for parameter estimation:

(1) Set starting values for the parameter vector.
(2) Set starting values for the sampling coefficients.
(3) Maximize the loglikelihood function using an approximate but fast method. We suggest to take the NAIScc method with $S = 0$ (no simulation, only numerical integration). The maximization of the loglikelihood function is carried out by direct numerical optimization.
(4) Re-start maximization of loglikelihood function using NAIScc method with $S > 0$.

The efficiency of this algorithm is primarily due to the accurate approximation of the loglikelihood function with the NAIScc method using $S = 0$. As a result convergence of the maximization in the last step will be very fast, only requiring a small number of iterations. It gives the procedure the desirable property that S can be set at a high value with only a relatively small increase of computing time. We have a set of common random numbers that we use for each likelihood evaluation to ensure smoothness of the likelihood function. The smoothness with respect to the parameter vector is necessary for numerical optimization methods. It is also important to work with transformed parameters so that they stay within their admissible ranges.

In our illustration we set the number of observations equal to $n = 5000$. The large time series dimension is required for being able to estimate the parameters in the model because the third volatility component has low persistence and may be hard to identify from a short sample. We set the number of samples to $S = 200$. We take 20 different time series realizations of the model. For each realized time series, we obtain parameter estimates and compute their Monte Carlo standard errors. The reported standard errors are averages across those from the 20 realizations. Since we have set the true parameters ourselves, we can also calculate the Rmse of the estimates. It allows us to compare the relative importance of the simulation and statistical errors in estimating the parameters.

The results are summarized in Table 9. The average estimation time for each realization has been slightly under two minutes. Table 9 further presents the simulation errors which are small for all parameters, in absolute and in relative terms. In the estimation results the Monte Carlo standard errors represent only between 1% and 3% of the total Rmse. We conclude that the NAIS method can be successfully applied for the estimation of fairly demanding model specifications.

5. Conclusion

We have developed a new efficient importance sampling method for a likelihood-based analysis of nonlinear non-Gaussian state space models. The numerically accelerating importance sampling method is a mix of numerical integration and Monte Carlo integration. We adopt the Gauss-Hermite quadrature method for computing the importance sampling coefficients and the importance sampling method for the actual evaluation of the likelihood function. New control variables are introduced to reduce the sampling variance of the Monte Carlo estimate of the likelihood function. We have carried out a comprehensive Monte Carlo study for different model classes to illustrate the performance of our approach, relative to earlier importance sampling methods. Overall we can conclude that our methods lead to an outstanding performance in terms of numerical and computational efficiency.

References

Banachewicz, K. and A. Lucas (2008): "Parameter Driven Copula Models," Working paper.

Bauwens, L. and F. Galli (2009): "Efficient Importance Sampling for ML Estimation of SCD Models," *Computational Statistics and Data Analysis*, 53, 1974–1992.

Bauwens, L. and N. Hautsch (2006): "Stochastic Conditional Intensity Processes," *Journal of Financial Econometrics*, 4, 450–493.

Bauwens, L. and D. Veredas (2004): "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations," *Journal of Econometrics*, 119, 381–412.

Bollerslev, T., U. Kretschmer, C. Pigorsch, and G. Tauchen (2009): "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," *Journal of Econometrics*, 150, 151–166.

Danielsson, J. and J. Richard (1993): "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, 153–174.

de Jong, P. and N. Shephard (1995): "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350.

Durbin, J. and S. J. Koopman (1997): "Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models," *Biometrika*, 669–684.

——— (2000): "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives," *Journal of the Royal Statistical Society, Series B*, 3–56.

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

——— (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 603–616.

Engle, R. and G. Lee (1999): "A Long-Run and Short-Run Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, ed. by R. Engle and H. White, Oxford University Press, 475–497.

Fridman, M. and L. Harris (1998): "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 16, 284–291.

Geweke, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1739.

Ghysels, E., A. Harvey, and E. Renault (1996): "Stochastic Volatility," in *Handbook of Statistics, Volume 14*, ed. by G. Maddala and C. Rao, Elsevier, Amsterdam.

Hafner, C. and H. Manner (2011): "Dynamic Stochastic Copula Models: Estimation, Inference and Applications," *Journal of Applied Econometrics*, forthcoming.

Heiss, F. (2008): "Sequential numerical integration in nonlinear state space models for microeconometric panel data," *Journal of Applied Econometrics*, 23, 373–389.

Joe, H. (1997): *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.

Jungbacker, B. and S. J. Koopman (2007): "Monte Carlo Estimation for Nonlinear non-Gaussian State Space Models," *Biometrika*, 827–839.

Kitagawa, G. (1987): "Non-Gaussian State Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032–63.

Kloek, T. and H. van Dijk (1978): "Bayesian Estimation of Equation System Parameters: an Application by Monte Carlo," *Econometrica*, 47, 1–20.

Koopman, S. J. and T. M. Nguyen (2011): "A Modified Efficient Importance Sampler for Nonlinear non-Gaussian State Space Models," Working paper.

Koopman, S. J., N. Shephard, and D. Creal (2009): "Testing the assumptions behind importance sampling," *Journal of Econometrics*, 149, 2–11.

Liesenfeld, R. and J. Richard (2003): "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance*, 10, 505–531.

Melino, A. and S. Turnbull (1990): "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–265.

Nelsen, R. (1999): *An Introduction to Copulas*, New York: Springer Verlag.

Patton, A. (2006): "Modeling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527–556.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling (1992): *Numerical Recipes in C (2nd ed.): The Art of Scientific Computing*, Cambridge University Press.

Richard, J. and W. Zhang (2007): "Efficient High-Dimensional Importance Sampling," *Journal of Econometrics*, 141, 1385–1411.

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

Schmidt, T. (2006): "Coping with Copulas," Manuscript.

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

So, M. K. P. (2003): "Posterior Mode Estimation For Nonlinear And Non-Gaussian State Space Models," *Statistica Sinica*, 255–274.

Tauchen, G. and M. Pitts (1983): "The Price Variability-Volume Relationship in Speculative Markets," *Econometrica*, 51, 485–505.

Taylor, S. J. (1986): *Modelling Financial Time Series*, Chichester, UK: John Wiley.

Appendix A. The high-dimensional EIS method

The high-dimensional EIS method of Liesenfeld and Richard (2003) and Richard and Zhang (2007) considers the Gaussian importance model density characterized by equations (7) and (8) of Section 2. The period t expressions $g(y_t|\theta_t;\psi)$ and $g(\theta_t|\alpha_{t-1};\psi)$ can be merged and interpreted as a Gaussian density $g(\alpha_t|y_t,\alpha_{t-1})$ with covariance matrix

$$V_t = (Q_t^{-1} + Z_t' C_t Z_t)^{-1} \tag{26}$$

and mean

$$\widehat{\alpha}_t = V_t^{-1}(b_t' Z_t + (d_t + T_t\alpha_{t-1})' Q_t^{-1})', \tag{27}$$

provided that the constant a_t is set to

$$a_t(\chi_t,\alpha_{t-1}) = \frac{1}{2}\log(|Q|/|V_t|) + \frac{1}{2}(d_t + T_t\alpha_{t-1})' Q_t^{-1}(d_t + T_t\alpha_{t-1}) - \frac{1}{2}\widehat{\alpha}_t' V_t^{-1}\widehat{\alpha}_t \tag{28}$$

See for example Bauwens and Galli (2009) for a derivation of this result.

The state vectors are therefore sampled sequentially from $g(\alpha_t|y_t,\alpha_{t-1})$ for a given set of importance parameters χ_t. In contrast with the approximating linear state space model approach of sections 2.1 and 2.2, the constant a_t now depends on α_{t-1}. This property marks the fundamental difference between the EIS and MEIS and leads to a different setup for the recursive procedure for selecting χ. Let k index the EIS iterations. The method minimizes backwards from $t = n$ to $t = 1$

$$\min_{\chi_t^{[k+1]}} \int \lambda^2(\theta_t, y_t;\psi)\omega(\theta_t, y_t;\psi) g(\theta_t|y;\psi), \tag{29}$$

where

$$\lambda(\theta_t, y_t;\psi) = \log p(y_t|\theta_t;\psi) + a_{t+1}(\chi_{t+1}[k],\alpha_t) - \log g(y_t|\theta_t;\psi) - \lambda_{0t}, \qquad \omega(\theta_t, y_t;\psi) = \frac{p(y_t|\theta_t;\psi)}{g(y_t|\theta_t;\psi)},$$

Richard and Zhang (2007) argue that the dynamic structure of the model is captured by the integration constant a_t, which is shifted through time. The fact that the EIS and MEIS methods rely on the same basic approximation to $p(y_t|\theta_t;\psi)$ given by (8) implies that the procedures are largely equivalent numerically (though not identical). This observation is confirmed the the simulation results of Section 4.1. From a computational perspective, however, the MEIS method has two advantages in addition to the ones we have discussed in Section 2. First, it avoids the large number of computations required by the EIS method to track the constants a_t. Second, it enables the direct simulation of the signal vector θ_t rather then the possibly higher dimensional state vector α_t.

Appendix B. The Bias in the standard implementation of the EIS method

In section 2.2 we have argued that a finite sample bias arises in the EIS algorithm when the same set of common random numbers (CRN) is used for obtaining the importance parameters χ via the MC variance minimization problem (29) and computing the likelihood estimate (5). As the CRNs u determine both $\chi(y) = \chi(y, u)$ and $\theta^{(s)} = \theta^{(s)}(u)$, we obtain that $g(\theta^{(s)}(u)|\chi(y, u); \psi)$ is not well defined as an importance density. Consider the Taylor series expansion of the likelihood estimate round some value x^* given by (19)

$$\bar{\omega} = \exp(x^*) \frac{1}{S} \sum_{s=1}^{S} \left(1 + [x_s - x^*] + \frac{1}{2}[x_s - x^*]^2 + \dots\right), \tag{30}$$

where $x_s = \log \omega(\theta^{(s)}, y, \psi)$ and $\theta^{(s)}$ does not necessarily depend on u.

In the MEIS algorithm described in Section 2.2 and the EIS method detailed in the previous appendix, $\chi(y, u)$ is explicitly selected to minimize the sample variance of the log importance sampling weights $\log \omega(\theta^{(s)}(u), y, \psi)$. Consequently, the minimized variance of $\log \omega(\theta^{(s)}(u), y, \psi)$ is artificially low in relation to what is the case in the full support of $g(\theta|\chi(y, u); \psi)$. It follows that the third term in the Taylor expansion (30) is biased downwards when $\theta^{(s)} = \theta^{(s)}(u)$. This problem also contaminates the other terms, with ambiguous net effects for the likelihood estimate.

Appendix C. Tables

Table 1. Inportance sampling methods.

The table reviews the importance sampling methods and provide their acronyms.

SPDK	the method of Section 2.1 by Shephard and Pitt (1997) and Durbin and Koopman (1997).
EIS	the high-dimensional efficient importance sampling method by Richard and Zhang (2007) and described in Appendix A.
MEIS	the method of Section 2.2 by Koopman and Nguyen (2011).
NAIS	the method of Section 3.1.
NAIScc	the method of Section 3.1 with the two control variables of Sections 3.3 and 3.4.
NAIScc*	the estimate $\widehat{L}(y;\psi)^*_{cc}$ of Section 3.4.

Table 2. Loglikelihood Errors for Stochastic Volatility I.

The table shows average bias, standard deviation and Rmse of loglikelihood estimation errors for different IS methods. The Rmse statistic is reported as its ratio with respect to the MEIS method. We simulate 20 different realizations from the model. For each of these realizations, we obtain loglikelihood estimates for 100 different sets of random numbers and then calculate the bias, variance and Rmse with the unknown loglikelihood being approximated by the average of estimates of the best unbiased methods. We also report TNR as computing time normalized for an Rmse ratio of unity. The reported values are the average statistics across the 20 realizations. The methods (with their acronyms) are discussed in Table. The stochastic volatility model is specified as: $y_t \sim N(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$ where $\eta_t \sim N(0, Q = 0.1^2)$ for $t = 1, \ldots, n$.

	$n = 1000$					$n = 3000$				
$S = 20$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.014	0.163	5.52	1.99	0.01	-0.107	0.423	4.93	1.66	0.01
EIS	0.025	0.016	0.99	1.74	0.13	0.079	0.042	1.00	1.88	0.39
MEIS	0.026	0.015	1.00	1.00	0.04	0.079	0.043	1.00	1.00	0.11
NAIS	-0.001	0.035	1.19	1.25	0.05	-0.002	0.094	1.06	1.08	0.12
NAIScc	-0.001	0.026	0.91	0.89	0.04	-0.006	0.068	0.77	0.80	0.12
NAIScc*	-0.007	0.024	0.84	0.82	0.04	-0.024	0.066	0.79	0.82	0.12
$S = 200$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.001	0.069	5.53	1.93	0.05	-0.023	0.203	6.00	2.05	0.13
EIS	0.003	0.012	1.00	1.61	0.98	0.009	0.034	1.02	1.69	3.02
MEIS	0.003	0.012	1.00	1.00	0.38	0.009	0.033	1.00	1.00	1.10
NAIS	0.000	0.014	1.15	0.54	0.08	-0.001	0.039	1.15	0.53	0.23
NAIScc	0.000	0.009	0.71	0.33	0.08	0.000	0.023	0.67	0.32	0.25
NAIScc*	-0.001	0.008	0.61	0.28	0.08	-0.001	0.021	0.62	0.30	0.25

Table 3. Loglikelihood Errors for Stochastic Volatility II.

We report the same results as Table 2 for a less persistent stochastic volatility model given by $y_t \sim N(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.05 + 0.9\alpha_{t-1} + \eta_t$ where $\eta_t \sim N(0, Q = 0.1^2)$ for $t = 1, \ldots, n$.

	$n = 1000$					$n = 3000$				
$S = 20$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.001	0.025	1.01	0.41	0.00	-0.004	0.069	0.94	0.37	0.01
EIS	0.024	0.003	1.00	1.72	0.09	0.074	0.007	1.00	1.81	0.26
MEIS	0.024	0.003	1.00	1.00	0.03	0.074	0.007	1.00	1.00	0.08
NAIS	0.000	0.006	0.23	0.26	0.04	0.000	0.014	0.19	0.20	0.09
NAIScc	0.000	0.004	0.15	0.16	0.03	0.000	0.010	0.14	0.15	0.09
NAIScc*	-0.001	0.004	0.16	0.16	0.03	-0.003	0.011	0.15	0.16	0.09
$S = 200$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	0.000	0.009	1.99	0.78	0.05	0.000	0.028	2.07	0.82	0.13
EIS	0.004	0.002	1.00	1.57	0.74	0.013	0.005	1.01	1.62	2.12
MEIS	0.004	0.002	1.00	1.00	0.30	0.012	0.005	1.00	1.00	0.83
NAIS	0.000	0.002	0.51	0.25	0.07	0.000	0.006	0.43	0.22	0.20
NAIScc	0.000	0.001	0.25	0.13	0.07	0.000	0.003	0.24	0.12	0.22
NAIScc*	0.000	0.001	0.25	0.12	0.07	0.000	0.003	0.23	0.12	0.22

Table 4. Loglikelihood Errors for Stochastic Conditional Duration I.

The model is specified as: $y_t \sim \text{Weibull}(\lambda_t, \psi = 1.2)$, $\lambda_t = \exp(\alpha_t)$, $\alpha_t = 0.98\alpha_{t-1} + \eta_t$, $\eta_t \sim N(0, Q = 0.15^2)$.

	$n = 1000$					$n = 3000$				
$S = 20$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.144	0.493	6.27	2.68	0.02	-0.492	0.616	6.54	2.45	0.04
EIS	0.061	0.054	0.99	1.44	0.22	0.093	0.079	1.02	1.48	0.67
MEIS	0.062	0.054	1.00	1.00	0.10	0.093	0.077	1.00	1.00	0.32
NAIS	-0.008	0.115	1.41	1.07	0.06	-0.032	0.160	1.37	1.05	0.19
NAIScc	-0.005	0.086	1.06	0.84	0.06	-0.029	0.138	1.18	0.95	0.21
NAIScc*	-0.026	0.086	1.10	0.87	0.06	-0.064	0.134	1.24	1.00	0.20
$S = 200$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.039	0.246	6.08	1.93	0.09	-0.197	0.386	6.45	1.86	0.25
EIS	0.002	0.043	1.04	1.52	1.84	-0.002	0.065	0.96	1.33	5.73
MEIS	0.001	0.041	1.00	1.00	0.87	0.002	0.067	1.00	1.00	3.03
NAIS	-0.002	0.049	1.18	0.45	0.12	-0.006	0.076	1.13	0.40	0.38
NAIScc	0.000	0.032	0.77	0.30	0.13	-0.003	0.056	0.83	0.32	0.43
NAIScc*	-0.002	0.030	0.74	0.29	0.13	-0.007	0.051	0.77	0.29	0.43

Table 5. Loglikelihood Errors for Stochastic Conditional Duration II.

The model is specified as: $y_t \sim \text{Weibull}(\lambda_t, \psi)$, $\lambda_t = \exp(\alpha_t)$, $\alpha_t = T\alpha_{t-1} + \eta_t$, $\eta_t \sim N(0, Q)$.

	$n = 1000$					$n = 1000$				
	$T = 0.96, Q = 0.1^2, \psi = 1.7$					$T = 0.9, Q = 0.15^2, \psi = 1.2$				
$S = 20$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.110	0.422	5.76	2.50	0.02	-0.060	0.331	3.56	1.46	0.01
EIS	0.061	0.043	0.98	1.44	0.19	0.089	0.036	1.01	1.47	0.16
MEIS	0.062	0.044	1.00	1.00	0.09	0.088	0.035	1.00	1.00	0.08
NAIS	-0.008	0.094	1.25	0.98	0.06	-0.003	0.075	0.80	0.78	0.07
NAIScc	-0.001	0.074	0.98	0.80	0.06	-0.001	0.056	0.60	0.58	0.07
NAIScc*	-0.020	0.071	0.97	0.80	0.06	-0.014	0.057	0.63	0.62	0.07
$S = 200$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.034	0.197	5.81	1.82	0.09	-0.009	0.150	5.32	1.76	0.08
EIS	0.004	0.035	1.02	1.44	1.74	0.009	0.026	0.99	1.41	1.45
MEIS	0.004	0.035	1.00	1.00	0.88	0.009	0.027	1.00	1.00	0.72
NAIS	-0.001	0.041	1.17	0.43	0.12	-0.001	0.031	1.12	0.47	0.13
NAIScc	0.000	0.025	0.72	0.28	0.13	0.000	0.018	0.64	0.28	0.14
NAIScc*	-0.002	0.023	0.67	0.26	0.13	-0.001	0.017	0.60	0.26	0.14

Table 6. Loglikelihood Errors for Stochastic Copula.

The model is specified as: $u_{1t}, u_{2t} \sim C_{\nu=5,P_t}(u_t)$, $\rho_t = (1 + \exp(-\alpha_t)^{-1}$, $\alpha_t = 0.017 + 0.98\alpha_{t-1} + \eta_t$, $\eta_t \sim N(0, Q = 0.1^2)$.

	$n = 1000$					$n = 3000$				
$S = 20$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.014	0.148	4.58	6.64	0.16	-0.092	0.342	3.90	5.12	0.35
EIS	-0.032	0.015	1.00	1.45	0.16	-0.093	0.037	1.01	1.55	0.48
MEIS	-0.032	0.015	1.00	1.00	0.07	-0.093	0.037	1.00	1.00	0.20
NAIS	-0.003	0.035	1.04	0.82	0.05	-0.001	0.095	0.96	0.77	0.13
NAIScc	-0.001	0.026	0.75	0.63	0.05	0.000	0.078	0.82	0.72	0.15
NAIScc*	-0.008	0.019	0.59	0.50	0.05	-0.021	0.062	0.67	0.58	0.15
$S = 200$	Bias	SD	Rmse*	TNR	Time	Bias	SD	Rmse*	TNR	Time
SPDK	-0.002	0.063	4.56	2.32	0.22	-0.015	0.161	4.56	2.35	0.55
EIS	-0.007	0.013	1.02	1.34	1.46	-0.023	0.032	1.01	1.42	4.05
MEIS	-0.007	0.013	1.00	1.00	0.85	-0.022	0.032	1.00	1.00	2.05
NAIS	0.000	0.016	1.12	0.41	0.12	-0.001	0.041	1.09	0.44	0.34
NAIScc	0.000	0.012	0.81	0.31	0.13	-0.001	0.026	0.68	0.29	0.37
NAIScc*	-0.002	0.007	0.51	0.21	0.13	-0.003	0.021	0.55	0.23	0.38

Table 7. Standard Deviations for the logged Importance Sampling Weights.

For $R = 100$ replications of the stochastic volatility specification of Table 2 and the stochastic copula specification of Table 6, we obtain auxiliary coefficients under different numbers of MC trajectories S (for the MEIS method) and different numbers of integration nodes M (for the numerical procedure of section of 3.1). We then simulate a thousand independent MC paths from these coefficients and compute the variance of the resulting log importance sampling weights. The displayed results are the average standard deviations across the R replications.

		SV, $n = 3000$			SC, $n = 3000$		
		In Sample	Out of Sample	Time	In Sample	Out of Sample	Time
MEIS	$S = 20$	0.4179	0.7815	0.85s	0.3675	0.7070	1.68s
	$S = 40$	0.4728	0.7294	0.97s	0.4143	0.6528	1.90s
	$S = 80$	0.5254	0.6871	1.18s	0.4719	0.6185	2.25s
	$S = 200$	0.5761	0.6495	2.07s	0.5275	0.5870	3.48s
	$S = 1000$	0.6025	0.6262	5.78s	0.5556	0.5678	10.34s
Numerical	$M = 10$	-	0.6190	0.86s	-	0.5609	1.88s
	$M = 20$	-	0.6190	0.88s	-	0.5609	1.90s
	$M = 30$	-	0.6190	0.95s	-	0.5609	2.00s

Table 8. Robustness to Choice of Number Integration Nodes.

For $R = 1000$ replications of the stochastic volatility specification of Table 2 and the stochastic copula specification of Table 6, we obtain auxiliary coefficients under different numbers of integration nodes M (for the numerical procedure of section of 3.1). We then simulate a thousand independent MC paths from these coefficients and compute the variance of the resulting log importance sampling weights. The displayed results are the average standard deviations across the R replications.

	$n = 1000$		$n = 3000$	
	SV	SC	SV	SC
Std. Dev. $\log(\widehat{L}(y;\psi)_{M=10})$	0.0075	0.0072	0.0236	0.0203
Std. Dev. $\log(\widehat{L}(y;\psi)_{M=20})$	0.0069	0.0067	0.0216	0.0189
Std. Dev. $\log(\widehat{L}(y;\psi)_{M=30})$	0.0069	0.0067	0.0216	0.0189
$\lvert\log(\widehat{L}(y;\psi)_{M=30}) - \log(\widehat{L}(y;\psi)_{M=10})\rvert$	0.0059	0.0032	0.0183	0.0102
$\lvert\log(\widehat{L}(y;\psi)_{M=30}) - \log(\widehat{L}(y;\psi)_{M=20})\rvert$	6.24×1^{-6}	5.55×1^{-7}	1.91×1^{-5}	1.94×1^{-6}

Table 9. Three Component SV Model Estimation.

We simulate 20 different realizations of a three component stochastic volatility model. For each of these realizations, we obtain simulated maximum likelihood parameter estimates using the NAIScc method for for 20 different sets of common random numbers. We report the average MC standard error across the 20 realizations. The Rmse column reports the total root mean squared error by comparing the estimates to the true parameters. The number of observations is n=5000. The number of MC trajectories is $S = 200$. Average estimation time: 116 seconds. The model is specified as $y_t \sim N(0, \sigma_t^2)$, $t = 1, \ldots, n$, $\sigma_t^2 = \exp(\theta_t)$, $\theta_t = d + \alpha_{1,t} + \alpha_{2,t} + \alpha_{3,t}$, $\alpha_t = T\alpha_{t-1} + \eta_t$, $\alpha_1 \sim \mathrm{N}(a_1, P_1)$, $\eta_t \sim \mathrm{N}(0, Q)$, where T is a diagonal matrix with elements ϕ_1, ϕ_2 and ϕ_3 and Q is a diagonal matrix with elements ν_1, ν_2 and ν_3.

Parameter	True	MC Error	Rmse	MC Error/Rmse
d	0.5	0.0016	0.103	0.016
ϕ_1	0.99	0.0001	0.006	0.009
ν_1	0.005	0.0000	0.003	0.011
ϕ_2	0.9	0.0006	0.049	0.011
ν_2	0.015	0.0002	0.010	0.018
ϕ_3	0.4	0.0032	0.280	0.012
ν_3	0.05	0.0008	0.029	0.029