

Koopman, Siem Jan; Lucas, Andre; Scharth, Marcel

**Working Paper**

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

Tinbergen Institute Discussion Paper, No. 11-057/4

**Provided in Cooperation with:**

Tinbergen Institute, Amsterdam and Rotterdam

*Suggested Citation:* Koopman, Siem Jan; Lucas, Andre; Scharth, Marcel (2011) : Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models, Tinbergen Institute Discussion Paper, No. 11-057/4, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/86828>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

TI 2011-057/4  
Tinbergen Institute Discussion Paper



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

*Siem Jan Koopman<sup>a,b</sup>*

*André Lucas<sup>a,b,c</sup>*

*Marcel Scharth<sup>a,b</sup>*

<sup>a</sup> *VU University Amsterdam;*

<sup>b</sup> *Tinbergen Institute;*

<sup>c</sup> *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<sup>(a,b)</sup> André Lucas<sup>(a,b,c)</sup> Marcel Scharth<sup>(a,b)</sup>

<sup>(a)</sup> VU University Amsterdam, The Netherlands

<sup>(b)</sup> Tinbergen Institute, The Netherlands

<sup>(c)</sup> Duisenberg school of finance, The Netherlands

January 2012

## Abstract

We introduce a new efficient importance sampler for nonlinear non-Gaussian state space models. We propose a general and efficient likelihood evaluation method for this class of models via the combination of numerical and Monte Carlo integration methods. Our methodology explores the idea that only a small part of the likelihood evaluation problem requires simulation. We refer to our new method as numerically accelerated importance sampling. The method is computationally and numerically efficient, facilitates parameter estimation for models with high-dimensional state vectors, and overcomes a bias-variance trade-off encountered by other sampling methods. An elaborate simulation study and an empirical application for U.S. stock returns reveal large efficiency gains for a range of models used in financial econometrics.

KEYWORDS : Kalman filter, Monte Carlo maximum likelihood, numerical integration, stochastic copula, stochastic conditional duration, stochastic volatility.

---

\*We would like to thank the seminar participants at the University of Mannheim, TI Amsterdam, Chicago Booth School of Business, Université Paris I, Sorbonne, and the 4<sup>th</sup> CSDA conference in London, for their insightful comments. Lucas acknowledges the financial support of the Dutch Science Foundation (NWO). Contact author: Siem Jan Koopman ([s.j.koopman@vu.nl](mailto:s.j.koopman@vu.nl)).

# 1 Introduction

The evaluation of analytically intractable likelihood functions is a challenging problem for a variety of statistical and econometric models. The difficulty is the numerical calculation of a high-dimensional integral, which we may typically carry out by the method of importance sampling. Advances in importance sampling over the past three decades have contributed to the interest in state space models that in many cases lack a tractable likelihood expression. Examples include stochastic volatility models as in Ghysels, Harvey, and Renault (1996), stochastic conditional intensity models as in Bauwens and Hautsch (2006), non-Gaussian unobserved components time series models as in Durbin and Koopman (2000), and flexible non-linear panel data models with unobserved heterogeneity as in Heiss (2008).

In this paper we propose a new importance sampling method with a high level of computational and numerical efficiency for a general class of nonlinear and non-Gaussian state space models. Our proposed methodology explores the idea that we can solve a substantial part of the likelihood evaluation problem by fast numerical integration rather than by Monte Carlo integration only. The contribution consists of two parts. First, we use numerical integration methods for the construction of an importance density that efficiently approximates the likelihood function. Second, we develop new control variables which we use as efficient variance reduction tools in evaluating the likelihood via importance sampling. Numerical integration is highly accurate when applicable, but its feasibility is typically limited to low dimensional problems. Monte Carlo integration, by contrast, is subject to simulation error but is more easily applicable in high-dimensional problems. By relying on both methods, we carry the virtues of numerical integration over to high-dimensional state space models. As a result, we depart from the numerical approaches of Kitagawa (1987) and Fridman and Harris (1998) as well as from the simulation based methods of Danielsson and Richard (1993) and Durbin and Koopman (1997). We refer to our new method as *numerically accelerated importance sampling* (NAIS).

We integrate two different importance sampling approaches into our method. The approach of Shephard and Pitt (1997) and Durbin and Koopman (1997) (referred to as SPDK) is based on an approximating linear Gaussian state space model that generates importance samples using Kalman filter and smoothing (KFS) methods.

The approximation model of SPDK is optimal in providing the mode estimate of the signal but it is also a local approximation of the entire likelihood integral. The efficient importance sampling (EIS) approach of Liesenfeld and Richard (2003) and Richard and Zhang (2007) establishes a global approximation of the likelihood. Koopman and Nguyen (2011) show how to implement the EIS method using the KFS methods of SPDK.

The NAIS method provides an accurate numerical solution for obtaining the optimal importance parameters leading to a global approximation of the likelihood function. The method of Liesenfeld and Richard (2003) and Richard and Zhang (2007) relies instead on Monte Carlo simulations for estimating the optimal sampling parameters. At the same time, the NAIS method generates additional computational efficiency via the combination of fast numerical integration techniques with the KFS methods of SPDK. Finally, by using NAIS we eliminate a bias-variance trade-off inherent to the EIS method. We show that the practice of re-using random numbers both to select the optimal importance parameters and to estimate the likelihood function induces a bias in the resulting estimate. We avoid this problem by replacing the first step by numerical integration.

We conduct an extensive simulation study to analyse the efficiency gains of the NAIS method. To validate the robustness of our results, we consider three different model specifications: the stochastic volatility model, for example, see Ghysels, Harvey, and Renault (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 we can efficiently implement our methods for each of these different models.

The Monte Carlo study reveals three major findings. First, we show that the linear state space model approximation always performs substantially faster than the standard implementation of the EIS method. This holds even without considering numerical acceleration. Second, when we increase the number of importance sampling trajectories, our NAIS method proves to be faster and more accurate than the standard EIS method: the NAIS method significantly improves the trade-off between computational and numerical efficiency in choosing the number of Monte Carlo samples. Third, for the different classes of models we consider in our simulation study we are able to

reduce the variance of the likelihood estimates by more than 40% with the use of our new control variables, relative to the use of antithetic variables as a variance reduction device.

To illustrate the NAIS method in an empirical setting, we consider a two-component stochastic volatility model for the time series of returns of a set of major U.S. stocks. The two-component structure of the volatility specification makes estimation by means of EIS a non-trivial and time-consuming operation which is frequently subject to numerical instability. However, we are able to implement the NAIS approach using standard hardware and software without further complications. The NAIS method reduces the estimation times in this application by as much as 90% and results in Monte Carlo standard errors for the estimated parameters which are small compared to the respective statistical standard errors. This application hence illustrates that we can use the NAIS method effectively for estimation and inference in many practical situations of interest.

The structure of the paper is as follows. Section 2 presents the nonlinear non-Gaussian state space model, introduces the necessary notation, and reviews the key importance sampling methods. Section 3 develops the new numerically accelerated importance sampling (NAIS) method, identifies the finite sample bias in EIS, and provides the details for the computation of the importance sampling parameters and our new control variables. Section 4 presents our simulation study and our empirical application. Section 5 concludes.

## **2 Importance sampling for state space models**

### **2.1 Nonlinear and non-Gaussian state space model**

The general ideas of importance sampling are well established and developed in the contributions of Kloek and van Dijk (1978), Ripley (1987), Geweke (1989) and others. Danielsson and Richard (1993), Shephard and Pitt (1997), Durbin and Koopman (1997) and others explore the implementation of importance sampling methods for the analysis of nonlinear non-Gaussian time series models. Richard and Zhang (2007) provide a short review of the literature with additional references. For the application of importance sampling in the context of time series, the main task is to evaluate the

likelihood function for the nonlinear non-Gaussian state space model as given by

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

where  $y_t$  is the  $p \times 1$  observation vector,  $\theta_t$  is the  $q \times 1$  signal vector,  $\alpha_t$  is the  $m \times 1$  state vector, and  $Z_t$  is the  $p \times m$  selection matrix; the  $m \times 1$  constant vector  $d_t$ , the  $m \times m$  transition matrix  $T_t$  and the  $m \times m$  variance matrix  $Q_t$  jointly determine the dynamic properties of the model. The system matrices  $Z_t$ ,  $T_t$  and  $Q_t$  are time-varying in a deterministic way. The unknown fixed parameter vector  $\psi$  contains the unknown coefficients in the observation density and in the system matrices.

The nonlinear non-Gaussian state space model as formulated in equation (1) allows the introduction of time-varying parameters in the density  $p(y_t|\theta_t; \psi)$ . The time-varying parameters depend on the signal  $\theta_t$  in a possibly nonlinear way. The signal vector  $\theta_t$  depends linearly on the state vector  $\alpha_t$ , for which we formulate a linear dynamic model. Our general framework can accommodate autoregressive moving average, long memory, random walk, cyclical and seasonal dynamic processes and combinations thereof. Harvey (1989) and Durbin and Koopman (2001) provide a detailed discussion of state space representations and unobserved components time series models.

We emphasise that the integration methodology we propose in this paper relies on the low dimensionality of the signal  $\theta_t$ , which contrasts with the typically higher dimensionality of the state vector  $\alpha_t$ . The shift of focus from  $\alpha_t$  to  $\theta_t$  enables large computational gains and eases many of the computational complications in the specification of the dynamic model for  $\theta_t$  when compared to existing methods.

## 2.2 Likelihood evaluation via importance sampling

Define  $\theta' = (\theta'_1, \dots, \theta'_n)$  and  $y' = (y'_1, \dots, y'_n)$ . If  $p(y_t|\theta_t; \psi)$  is a Gaussian density with mean  $\theta_t = Z_t\alpha_t$  and some variance  $H_t$ , for  $t = 1, \dots, n$ , Kalman filtering and smoothing methods analytically evaluate the likelihood and compute the minimum mean squared error estimates of the state vector  $\alpha_t$  together with its mean squared error matrix. In all other cases, the likelihood for (1) is given by the analytically



intractable integral

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

where  $p(\theta, y; \psi)$  is the joint density of  $y$  and  $\theta$  following from (1). Kitagawa (1987) has developed a numerical integration method for evaluating the likelihood integral above. This approach is in practice only practical when  $y_t$ ,  $\theta_t$  and  $\alpha_t$  are scalars.

To evaluate the likelihood function in a feasible manner by means of importance sampling, we consider the Gaussian importance density  $g(\theta, y; \psi) = g(y | \theta; \psi) g(\theta; \psi)$  where  $g(y | \theta; \psi)$  and  $g(\theta; \psi)$  are both Gaussian densities. It follows from (1) that  $p(\theta; \psi) = g(\theta; \psi)$ . We then express the likelihood function as

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

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

$$\omega(\theta, y; \psi) = p(y, \theta; \psi) / g(y, \theta; \psi) = p(y | \theta; \psi) / g(y | \theta; \psi). \quad (4)$$

The last equality is valid since  $p(\theta; \psi) = g(\theta; \psi)$ . We estimate the likelihood function (3) by generating  $S$  independent trajectories  $\theta^{(1)}, \dots, \theta^{(S)}$  from the importance density  $g(\theta | y; \psi)$  and by computing

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

where  $\omega_s$  is the realised importance weight function in (4) for  $\theta = \theta^{(s)}$ . Under standard regularity conditions, the weak law of large numbers ensures that

$$\widehat{L}(y; \psi) \xrightarrow{p} L(y; \psi), \quad (6)$$

when  $S \rightarrow \infty$ . A central limit theorem is applicable only when the variance of the importance weight function exists, see Geweke (1989). The failure of this condition leads to slow and unstable convergence of the estimate. Monahan (1993) and Koopman, Shephard, and Creal (2009) have developed diagnostic tests for validating the existence of the variance of the importance weights based on extreme value theory. Richard and Zhang (2007) discuss more informal methods for this purpose. We argue in Section 3.1 that the importance sampler proposed in this paper is robust to this problem.

### 2.3 The Gaussian importance density

We represent the Gaussian importance density as

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

where  $g(\alpha_t | \alpha_{t-1}; \psi)$  is the Gaussian density for  $\theta_t$  as implied by (1) and where

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

with  $a_t$ ,  $b_t$  and  $C_t$  defined as functions of the data vector  $y$  and the parameter vector  $\psi$ , for  $t = 1, \dots, n$ . The constants  $a_1, \dots, a_n$  ensure that  $g(\theta, y; \psi)$  integrates to one. The set of importance sampling parameters is

$$\chi = \{b_1, \dots, b_n, C_1, \dots, C_n\}. \quad (9)$$

Shephard and Pitt (1997) and Durbin and Koopman (1997), which we refer to as SPDK, treat the importance density (8) as equivalent to the density function associated with observation  $y_t^* = C_t^{-1} b_t$  and the linear Gaussian observation equation

$$y_t^* = \theta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, C_t^{-1}), \quad t = 1, \dots, n, \quad (10)$$

where we specify  $\theta_t$  as in (1). We can easily verify the equivalence of (8) with the

Gaussian logdensity  $\log g(y_t^*|\theta_t; \psi)$  for  $y_t^*$  in (10) since

$$\begin{aligned} \log g(y_t^*|\theta_t; \psi) &= -\frac{1}{2} \log 2\pi + \frac{1}{2} \log |C_t| - \frac{1}{2} \{(C_t^{-1} b_t - \theta_t)' C_t (C_t^{-1} b_t - \theta_t)\} \\ &= a_t + b_t' \theta_t - \frac{1}{2} \theta_t' C_t \theta_t, \end{aligned} \quad (11)$$

where the constant  $a_t$  collects all the terms that are not associated with  $\theta_t$ . It follows that  $g(y_t|\theta_t; \psi) \equiv g(y_t^*|\theta_t; \psi)$  for  $t = 1, \dots, n$ . Hence, we have  $g(\theta_t, y_t; \psi) \equiv g(\theta_t, y_t^*; \psi)$  for  $t = 1, \dots, n$ . Shephard and Pitt (1997) and Durbin and Koopman (1997) choose  $\chi$  such that the mean (or mode) estimate of  $\theta$  with respect to  $g(\theta|y^*; \psi)$  equals the mode estimate of  $\theta$  with respect to  $p(\theta|y; \psi)$ , where  $y^* = (y_1^{*'}, \dots, y_n^{*'})'$ .

The SPDK method is based on the linear state space model (10) with  $\theta_t$  specified by (1) to sample  $\theta_t$  from  $g(\theta|y^*; \psi)$ , for  $t = 1, \dots, n$ . de Jong and Shephard (1995) and Durbin and Koopman (2002) have developed simulation smoothing methods for the sampling of  $\theta$  from  $g(\theta|y^*; \psi)$  in a computationally efficient way. The vector  $\theta^{(s)}$  collects the simulations to compute the importance sampling weights  $\omega_s$  in (5) for  $s = 1, \dots, S$ . The evaluation of the Monte Carlo estimate of the likelihood is similar to (5) but with  $g(y; \psi) \equiv g(y^*; \psi)$ . The Kalman filter calculates  $g(y^*; \psi)$  via its evaluation of the likelihood function for the linear state space model (10).

Jungbacker and Koopman (2007) argue that the individual matrices  $C_t$  only need to be non-singular for the sampling density  $g(\theta|y^*; \psi)$  to be well defined. When any matrix  $C_t$  is not positive definite, we adopt their simulation smoothing scheme and modifications for computing the simulated likelihood function.

## 2.4 Selecting the importance sampling parameters

The choice of importance parameters in  $\chi$  determines the efficiency of the importance sampling procedure. Koopman and Nguyen (2011) apply the EIS method of Liesenfeld and Richard (2003) and Richard and Zhang (2007) to the SPDK importance model and consider the selection of the importance parameters in  $\chi$  of (9) based on a global approximation to  $p(y|\theta; \psi)$ . We obtain this approximation via the minimisation of the variance of the log-weights  $\log \omega(\theta, y; \psi)$  where we have defined  $\omega(\theta, y; \psi)$  in (4). The

variance minimisation problem is given by

$$\min_{\chi} \int \lambda^2(\theta, y; \psi) \omega(\theta, y; \psi) g(\theta|y; \psi) d\theta, \quad (12)$$

where

$$\lambda(\theta, y; \psi) = \log p(y|\theta; \psi) - \log g(y|\theta; \psi) - \lambda_0, \quad (13)$$

$g(y|\theta; \psi) = \prod_{t=1}^n g(y_t|\theta_t; \psi)$ , and  $g(y_t|\theta_t; \psi)$  is given by (8). The normalising constant  $\lambda_0$  sets the mean of  $\lambda(\theta, y; \psi)$  to zero.

The minimisation (12) is high-dimensional and numerically not feasible in most cases of interest. We follow Richard and Zhang (2007) and approximate the minimisation problem (12) by considering each time point  $t$  separately. The efficient importance parameters  $\chi_t = \{b_t, C_t\}$  are therefore the solutions of the minimisation problem

$$\min_{\chi_t} \int \lambda^2(\theta_t, y_t; \psi) \omega(\theta_t, y_t; \psi) g(\theta_t|y_t; \psi) d\theta_t \quad (14)$$

where

$$\omega(\theta_t, y_t; \psi) = \frac{p(y_t|\theta_t; \psi)}{g(y_t|\theta_t; \psi)}, \quad (15)$$

and

$$\lambda(\theta_t, y_t; \psi) = \log p(y_t|\theta_t; \psi) - \log g(y_t|\theta_t; \psi) - \lambda_{0t}, \quad (16)$$

for  $t = 1, \dots, n$ , where  $\lambda_{0t}$  is the normalising constant.

The EIS method of Richard and Zhang (2007) and its modification by Koopman and Nguyen (2011) rely on simulations and least squares computations for obtaining importance parameters; see Appendix A for further details.

### 3 Numerically Accelerated Importance Sampling

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

$$\int \varphi(x) dx, \quad (17)$$

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

The numerical evaluation of the integral in (17) via a Gauss-Hermite quadrature designates a set of  $M$  abscissae  $z_j$  and associated weights  $h(z_j)$  with  $j = 1, \dots, M$ . We compute the numerical approximation as

$$\int_{-\infty}^{\infty} \varphi(x) dx = \int_{-\infty}^{\infty} e^{-x^2} [e^{x^2} \varphi(x)] dx \approx \sum_{j=1}^M h(z_j) e^{z_j^2} \varphi(z_j), \quad (18)$$

where  $M$  is typically between 20 and 30. We can tabulate the weights  $h(z_j)$ . For a more detailed discussion on Gauss-Hermite quadrature, we refer to Monahan (2001). Even though we always refer to numerical integration in our discussion, we aim to work with Gaussian integrals for which analytical solutions may be available in specific applications. In such cases, the use of the analytical expression will bring further efficiency to the importance sampling procedure.

Our exposition below focuses on model (1) with a scalar signal  $\theta_t$ , that is  $q = 1$ . Although all results are valid for a high dimensional vector  $\theta_t$ , this setting brings additional computational challenges that are beyond the scope of this paper. We note, however, that our treatment below still allows for a high dimensional state vector  $\alpha_t$ . We provide an empirical illustration of this advantage in Section 4.

### 3.1 The construction of the importance sampler via numerical integration

Our importance sampler is the global approximation associated with the minimisation of the variance of the log-weights  $\log \omega(\theta_t, y_t; \psi)$  in (14). The key insight leading to our method is that the marginal density  $g(\theta_t | y^*; \psi)$  is available analytically for the linear state space approximation (10) from the output of the Kalman filter and smoother (KFS). This result allows us to directly minimise the low dimensional integral (14) for each time  $t$  by means of a Gauss-Hermite quadrature. In previous methods, only high dimensional importance densities  $g(\theta | y; \psi)$  or  $g(\alpha | y; \psi)$  were available for similar purposes; see Appendix A. In contrast with the resulting Monte Carlo approaches

based on simulated trajectories for the signals or the states, we therefore obtain our likelihood approximation entirely by numerical integration. In practical settings, our solution for (14) will be virtually exact.

For a given set of values in  $\chi = \chi^+ = \{b_1^+, \dots, b_n^+, C_1^+, \dots, C_n^+\}$  of (9), we have that the smoothed importance density  $g(\theta_t|y; \psi) = g(\theta_t|y^*; \psi)$  based on the linear Gaussian model (10) is given by

$$g(\theta_t|y^*; \psi) = N(\widehat{\theta}_t, V_t) = \exp\left\{-\frac{1}{2}V_t^{-1}(\theta_t - \widehat{\theta}_t)^2\right\} / \sqrt{2\pi V_t}, \quad (19)$$

where we compute  $\widehat{\theta}_t$  and  $V_t$  by KFS methods applied to the importance model (10) for  $y_t^* = (C_t^+)^{-1}b_t^+$  and  $\theta_t$  specified as in (1), for  $t = 1, \dots, n$ . For  $\chi = \chi^+$ , we evaluate the integral in (14) numerically as in (17) with  $x = \theta_t$  and

$$\varphi(\theta_t) = \lambda^2(\theta_t, y_t; \psi)\omega^*(\theta_t, y_t; \psi)g(\theta_t|y^*; \psi), \quad t = 1, \dots, n,$$

where  $\omega^*(\theta_t, y_t; \psi) = p(y_t|\theta_t; \psi) / g(y_t^*|\theta_t; \psi)$ . The minimisation is with respect to  $\chi_t$ .

We express the minimisation in (14) as

$$\min_{\chi_t} \sum_{j=1}^M \lambda^2(\widetilde{\theta}_{tj}, y_t; \psi)w_{tj}, \quad w_{tj} = g(\widetilde{\theta}_{tj}|y^*; \psi)\omega^*(\widetilde{\theta}_{tj}, y_t; \psi)h(z_j)e^{z_j^2}, \quad (20)$$

where  $\widetilde{\theta}_{tj} = \widehat{\theta}_t + V_t^{1/2}z_j$ , for  $j = 1, \dots, M$ . It follows from (19) that

$$g(\widetilde{\theta}_{tj}|y^*; \psi) = \exp\left\{-\frac{1}{2}z_j^2\right\} / \sqrt{2\pi}, \quad t = 1, \dots, n.$$

The minimisation (20) takes place via an iterative method. For a given  $\chi = \chi^+$ , we obtain  $\widehat{\theta}_t$  and  $V_t$  from the KFS applied to (10), for  $t = 1, \dots, n$ . Minimisation (20) for a scalar  $\widetilde{\theta}_{tj}$  reduces to weighted least squares computations, for each  $t$ , with dependent variable  $p(y_t|\widetilde{\theta}_{tj}; \psi)$ , explanatory variables  $\widetilde{\theta}_{tj}$ ,  $\widetilde{\theta}_{tj}^2$  (including a constant) and weights  $w_{tj}$ . We obtain the minimum in (20) by setting  $\chi_t = \{b_t, C_t\}$  equal to the least squares estimates associated with explanatory variables  $\widetilde{\theta}_{tj}$  and  $\widetilde{\theta}_{tj}^2$ , respectively. The new value for  $\chi_t$  becomes  $\chi_t^+$  in the next iteration. The iterative procedure terminates after convergence. We initialise the recursion by choosing an appropriate starting value for

$\chi$ . The mode estimate of SPDK provides efficient initial parameters for this problem. We can also set arbitrary starting values for the parameters in  $\chi$ ; convergence typically takes only a few iterations even if the initialisation is inaccurate.

Richard and Zhang (2007) argue that their EIS method becomes numerically more stable when they omit the term  $\omega^*(\tilde{\theta}_{tj}, y_t; \psi)$  from the weight  $w_{tj}$  at the initial iterations. We can also remove this term from  $w_{tj}$  as it does not lead to any important loss of numerical efficiency. This option has the advantage of increasing the computational speed of the algorithm. The weight  $w_{tj}$  becomes

$$w_{tj} = w_j = (2\pi)^{-\frac{1}{2}} h(z_j) \exp\left(\frac{1}{2} z_j^2\right), \quad j = 1, \dots, M.$$

The use of parallel computing techniques reduces the computing time required for the iterative optimisation procedure. For a given  $\chi = \chi^+$  and having  $\hat{\theta}_t$  and  $V_t$ , with  $t = 1, \dots, n$ , calculated by KFS, we can compute  $\tilde{\theta}_{tj}$  for  $j = 1, \dots, M$  in parallel over all  $t$ .

Our variance minimisation procedure based on numerical integration is guaranteed to evaluate the log importance weights at all the extremes of the sampling space. This implies that the NAIS algorithm is more robust to numerical instability than earlier approaches, since it is not directly confronted by unusual draws from the importance density at the sampling stage. In case the importance sampling procedure is altogether inappropriate, the auxiliary regressions are likely to crash. Hence we have implicitly introduced an automatic numerical diagnostic checking procedure for the infinite variance problem discussed in Section 2.2. However, we have not encountered this problem for any of the several importance samplers we construct in Section 4.

Given the optimal importance parameter values in  $\chi$ , we use model (10) for drawing samples  $\theta^{(s)} \sim g(\theta|y^*; \psi)$ , for  $s = 1, \dots, S$ , via a simulation smoothing method. We estimate the likelihood function of the nonlinear non-Gaussian state space model as in (5) on the basis of this set of draws. This procedure is the first ingredient of our numerically accelerated importance sampling (NAIS) method. We introduce the second ingredient of NAIS in Section 3.2.

The NAIS approach brings further advantages to the importance sampling estimation of likelihood (2). In Appendix A, we show that the linear state space approximation (10) leads to faster procedures in comparison to the EIS method of Richard and

Zhang (2007). In Appendix B, we show that the Monte Carlo approach to obtain the importance sampling parameters in the EIS method results in a bias-variance trade-off in the importance sampling procedure. The NAIS method does not entail the same issue.

### 3.2 Importance sampling and control variables

We introduce a new set of control variables to improve the numerical efficiency of our importance sampling procedure. The control variable is constructed from the simulated trajectories such that it is negatively correlated with the likelihood estimate. In contrast to earlier applications, which have relied on analytical results, we adopt control variables that evaluated by numerical integration. We then use the difference between the Monte Carlo estimate of the mean and the mean evaluated via numerical integration to reduce the variance of the importance sampling estimate. This method of variance reduction replaces the use of antithetic variables such as those developed by Ripley (1987) and Durbin and Koopman (2000).

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

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

for a sample of  $S$  draws of  $\theta$  we generate from the smooth importance density  $g(\theta|y^*; \psi)$ . We denote these draws by  $\theta^{(1)}, \dots, \theta^{(S)}$ , with  $\theta_t^{(s)}$  as the  $t$ th element of  $\theta^{(s)}$ . The Kalman filter computes the density  $g(y; \psi) = g(y^*; \psi)$ . The densities  $g(y^*; \psi)$  and  $g(\theta|y^*; \psi)$  refer to the importance model (10) with  $\theta_t$  specified as (1) and with the importance parameter set  $\chi$  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 control variates based on

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

The  $t$ th contribution of  $x(\theta, y; \psi)$  is 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  $\theta^{(1)}, \dots, \theta^{(S)}$ , we have  $x_s = \log(\omega_s) =$



$\sum_{t=1}^n x_{ts}$  where

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

for  $t = 1, \dots, n$  and  $s = 1, \dots, S$ . We can express the sample average of  $\omega_s$  in terms of  $x_s = \log \omega_s$  by means of a Taylor series around some value  $x$ , that is

$$\bar{\omega} = \exp(x) \frac{1}{S} \sum_{s=1}^S \left( 1 + [x_s - x] + \frac{1}{2}[x_s - x]^2 + \dots \right). \quad (21)$$

We adopt the terms involving  $x_{ts}$ ,  $t = 1, \dots, n$ , in this expansion as control variables. Our method consists of replacing the highest variance terms of the Taylor series by their probability limits, which we compute efficiently via the NAIS algorithm. This step further reduces the reliance of the method on simulation, improving the numerical efficiency of the importance sampling estimate at a low computational cost.

### 3.3 First new control variable

We base our first control variable on the first order term  $(x_s - x)$  of the Taylor expansion (21). Under the same regularity conditions required for importance sampling, we have

$$\bar{x} = \frac{1}{S} \sum_{s=1}^S x_s \xrightarrow{p} \hat{x}, \quad (22)$$

where  $\hat{x} = \mathbb{E}_g x(\theta, y; \psi)$  and where  $\mathbb{E}_g$  is expectation with respect to density  $g(\theta|y; \psi)$ . The Taylor expansion (21) around  $x = \hat{x}$  can now be used to construct a first order control variable.

Since

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

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

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

where  $\tilde{\theta}_{tj} = \hat{\theta}_t + V_t^{1/2} z_j$  and with the numerical evaluation as in (18). The Kalman filter and smoother computes  $\hat{\theta}_t$  and  $V_t$  for  $t = 1, \dots, n$ . Furthermore, we have  $\hat{x} = \sum_{t=1}^n \hat{x}_t$ .

The likelihood estimate (5) corrected for the first control variable is given by

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

It follows from (6) and (22) that

$$\hat{L}(y; \psi)_c \xrightarrow{p} L(y; \psi).$$

When the importance model (10) provides an accurate approximation to the likelihood,  $\omega_s$  is close to one and  $x_s$  is close to zero, such that  $\omega_s \approx 1 + x_s$ . Hence  $\omega_s$  and  $\exp(\hat{x})x_s$  are typically highly and positively correlated. When the importance model is a less accurate approximation, the positive correlation remains, but at a more moderate level. Therefore  $\hat{L}(y; \psi)_c$  is a more efficient estimate of the likelihood function compared to  $\hat{L}(y; \psi)$ .

### 3.4 Second new control variable

We base our second control variable on the second order term  $(x_s - x)^2$  of the Taylor expansion (21). We aim to correct for the sample variation of  $(x_{ts} - \hat{x}_t)^2$  within the sample of draws  $\theta_t^{(1)}, \dots, \theta_t^{(S)}$  for each  $t$  individually, where  $\hat{x}_t$  is the  $t$ th element of  $\hat{x}$ . Using the same arguments as in Section 3.3, we write

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

where

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

We compute the variance  $\widehat{\sigma}_t^2$  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 can 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 Taylor expansion (21) justifies the weights of 1 and  $\frac{1}{2}$  for the first and second control variables respectively. However, these values may 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  $\theta^{(1)}, \dots, \theta^{(S)}$ , we can estimate the variance minimising weights  $\beta_1$  and  $\beta_2$  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  $\beta_k$  are regression coefficients for  $k = 0, 1, 2$  and  $\epsilon_s$  is an error term. We denote the resulting estimator as  $\widehat{L}(y; \psi)_{cc}^*$ . The use of least squares estimates for assigning weights to control variables is due to Ripley (1987). A drawback of this modification is the introduction of a small sample bias, which arises because the least squares regression involves random independent variables.

## 4 Monte Carlo and empirical evidence

### 4.1 Likelihood estimation

We examine the performance of the importance sampling methods we list in Table 1 for likelihood estimation. The design of the simulation study is as follows. We consider fifty random time series of the three stochastic models we discuss below. We have taken fifty simulations to avoid the dependence of our conclusions on particular trajectories of the observed series. For each simulated time series, we estimate the loglikelihood function at the true parameters a hundred times using different common random numbers. Each cell in the subsequent tables therefore reflects 5,000 ( $= 50 \times 100$ ) simulations.

For each method, we report average bias, standard deviations, computation times, and root mean square error (Rmse) values over all 5,000 simulations. We use different sample sizes  $n = 1,000$  and  $n = 3,000$  and different numbers of importance samples  $S = 20$  and  $S = 200$ .

Table 1: IMPORTANCE SAMPLING METHODS.

The table presents the importance sampling methods with their acronyms that are adopted in the simulation and empirical studies.

SPDK	the method of Section 2.4 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 Koopman and Nguyen (2011) (see Appendix A).
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.

To measure the bias, standard deviation and Rmse for the estimated loglikelihood values, we require the true loglikelihood value which is unknown. We approximate it by the average of loglikelihood estimates from the NAIS and NAIScc methods for  $S = 200$ . This is similar to an NAIS likelihood estimate based on  $S = 200 \times 2 \times 100 = 40,000$  importance samples. The approximation error of the true likelihood is hence negligible.

We compute the reported statistics as given by

$$\begin{aligned}
 \text{Bias} &= 5000^{-1} \cdot \sum_{i=1}^{50} \sum_{j=1}^{100} \left( \log \widehat{L}^j(y^i; \psi) - \log L(y^i; \psi) \right), \\
 \text{Stand.dev} &= 50^{-1} \cdot \sum_{i=1}^{50} \left[ 100^{-1} \cdot \sum_{j=1}^{100} \left( \log \widehat{L}^j(y^i; \psi) - \log \bar{L}(y^i; \psi) \right)^2 \right]^{1/2}, \\
 \text{Rmse} &= 50^{-1} \cdot \sum_{i=1}^{50} \left[ 100^{-1} \cdot \sum_{j=1}^{100} \left( \log \widehat{L}^j(y^i; \psi) - \log L(y^i; \psi) \right)^2 \right]^{1/2},
 \end{aligned}$$

where  $y^i$  is the  $i$ th simulated time series,  $\log \widehat{L}^j(y^i; \psi)$  is the “true” loglikelihood value,  $\log \widehat{L}^j(y^i; \psi)$  is the  $j$ th estimate of the loglikelihood function for a particular method and  $\log \bar{L}(y^i; \psi) = 100^{-1} \sum_{j=1}^{100} \log L^j(y^i; \psi)$ . We denote Rmse\* as the ratio of Rmse in

relation to the Rmse of the MEIS method which we have selected as our benchmark.

Since computational efficiency is the main objective of importance sampling, we report the median computing times for each method and setting based on a machine equipped with an Intel Duo Core 2.5GHz. For our simulation study below we also present the Rmse ratio normalised by the associated computing times. This statistic summarises the key efficiency of each method. We calculate it as

$$TNR_i = \frac{\text{Rmse}_i \times \sqrt{T_i}}{\text{Rmse}_b \times \sqrt{T_b}},$$

where  $\text{Rmse}_i$  and  $T_i$  are Rmse and median computing time for method  $i$ , respectively, and with the benchmark method index  $i = b$ . All reported computing times include the fixed time costs required for obtaining the sampling parameters.

We implement all methods as described above. The number of nodes for numerical integration calculations is  $M = 20$ . We verify the sensitivity of our results to this choice at the end of Section 4.3 and in Table 7. Antithetic variables for location and scale, as proposed in Durbin and Koopman (2000), are the variance reduction tools in all likelihood evaluations, except for the NAIScc and NAIScc\* methods. We have found no evidence of importance sampling weights with an infinite variance for the models we discuss below; see the discussions in Koopman, Shephard, and Creal (2009). Our diagnostics include the verification of how sensitive the importance sampling weights are to artificial outliers as in Richard and Zhang (2007). We have implemented all methods using MATLAB and C.

## 4.2 Three models

Here we provide details of three stochastic models which are special cases of the non-linear and non-Gaussian state space model we discuss in Section 2.1.

### 4.2.1 Stochastic volatility model

The stochastic volatility (SV) model is an example of a nonlinear state space model. The key references to the development of the SV model are Tauchen and Pitts (1983), Taylor (1986) and Melino and Turnbull (1990). Ghysels, Harvey, and Renault (1996) and Shephard (2005) provide reviews of SV models. Liesenfeld and Richard (2003)

apply efficient importance sampling methods for the simulated maximum likelihood estimation of a wide range of stochastic volatility specifications. For a time series of log-returns  $y_t$ , we consider the model

$$\begin{aligned} y_t &\sim N(0, \sigma_t^2), & \sigma_t^2 &= \exp(\alpha_t), \\ \alpha_t &= d + T\alpha_{t-1} + \eta_t, & \eta_t &\sim N(0, Q), \\ \alpha_1 &\sim N(d/(1-T), Q/(1-T^2)), \end{aligned} \quad (23)$$

for  $t = 1, \dots, n$ , where  $d$  is a scalar constant,  $T$  is the autoregressive coefficient with  $|T| < 1$ , and  $Q$  is the variance of the disturbance  $\eta_t$  of the stochastic log-volatility process  $\alpha_t$ . 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). This allows us to investigate how importance sampling methods perform when the volatility process is less persistent.

#### 4.2.2 Stochastic conditional duration model

Bauwens and Veredas (2004) propose the stochastic conditional duration (SCD) model for modelling durations between high-frequency financial transactions. Bauwens and Galli (2009) study the efficient importance sampling estimation of SCD models. For a time series of durations  $y_t$ , we consider the model specification

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

for  $t = 1, \dots, n$ , where  $\lambda_t$  is the time varying scale parameter,  $\psi$  is the shape parameter of the Weibull distribution and with  $\alpha_t$  as the autoregressive process (23). 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, while the parameter set  $d = 0$ ,  $T = 0.9$ ,  $Q = 0.0225$ ,  $\psi = 1.2$  illustrates the performance of the methods for less persistent price durations.

### 4.2.3 Stochastic copula

Banachewicz (2009) and Hafner and Manner (2011) introduce the stochastic copula (SC) class of models for estimating and forecasting time-varying and possibly non-linear dependence between multiple time series. Schmidt (2006) provides a short introduction of the main concepts and results for (static) copulas, while Nelsen (1999) and Joe (1997) provide a more comprehensive discussion. Patton (2006) introduces dynamic copula models and extends the copula theory to specifications with conditionally time-varying parameters.

We consider a dynamic stochastic bivariate  $t$ -copula. Let  $u_{1t}$  and  $u_{2t}$  be two random variables with uniform  $(0, 1)$  marginal distributions. In our simulation study, we take  $u_t = (u_{1t}, u_{2t})'$  as probability integral transforms of two independent univariate series. The converse of Sklar's theorem implies that the combination of any set of univariate distributions together with a copula function characterises a well defined bivariate distribution. This result implies that the modelling of the dependence between the two random variables is completely disentangled from the modelling of the marginal distributions. We denote  $t_\nu$  as the standardised Student's  $t$  distribution and  $2 \times 2$  matrix  $P$  as the correlation matrix with unity values on the main diagonal and the correlation coefficient  $\rho$  on the two off-diagonal elements. The  $t$ -copula function  $C_{\nu,P}(u_t)$  describes the dependence structure for  $u_t$  and is given by

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

where  $T_{\nu,P}(a, b)$  is the cumulative density function associated with the standardised bivariate Student's  $t$  distribution with degrees of freedom  $\nu$  and correlation matrix  $P$  for any set of variables  $(a, b)$ . The copula is invariant under any standardisation of the marginal distributions. It follows that

$$C_{\nu,P}(u_t) = \int_{-\infty}^{t_\nu^{-1}(u_{1t})} \int_{-\infty}^{t_\nu^{-1}(u_{2t})} \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\nu/2)\sqrt{(\pi\nu)^2|P|}} \left(1 + \frac{x'P^{-1}x}{\nu}\right)^{-\frac{\nu+2}{2}} dx. \quad (24)$$

A possible state space model for the stochastic copula with a time-varying correlation

coefficient  $\rho_t$ , and hence a time-varying correlation matrix

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

is given by

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

for  $t = 1, \dots, n$ , where we model  $\alpha_t$  as the autoregressive process (23). We take our set of parameters from the empirical study of a bivariate financial time series of log-returns in Hafner and Manner (2011). The typical parameter estimates in their analysis are approximately equal to  $d = 0.017$ ,  $T = 0.98$ ,  $Q = 0.01$ . 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  $\nu = 5$  in order to introduce relevant tail dependence in our simulations.

### 4.3 Simulation results

Table 2 presents the results for the persistent stochastic volatility model specification. We summarise our findings as follows. For  $S = 20$ , the EIS and MEIS methods produce a substantial bias in their likelihood estimates; this bias is approximately 30% higher than the standard deviation of the estimates. Although the SPDK method is fast and the reported bias is moderate, it comes with a high variance. The relative low variance of the EIS and MEIS methods illustrates the numerical efficiency of the global approximation on which these methods are based. However, when we normalise the computing times by the Rmse statistic, the SPDK method turns out to be nearly as efficient as the EIS method for this problem. The MEIS method (EIS based on state space methods) is 50% faster in computing time than the EIS method, while its numerical efficiency is the same. For sample size  $n = 3000$  and  $S = 20$ , the MEIS likelihood evaluation procedure takes 0.11 seconds, while the EIS method takes 0.39 seconds, a computational saving of more than 70%.

The increase in the number of importance samples to  $S = 200$  mostly eliminates the bias in the EIS and MEIS methods, at the cost of a proportional tenfold increase in computing time. On the other hand, the reduction in standard deviation from the



Table 2: LOGLIKELIHOOD ERRORS FOR STOCHASTIC VOLATILITY I.

The table shows average bias, standard deviation and Rmse of loglikelihood estimation errors for different importance sampling methods. The reported Rmse\* statistic is the ratio of the Rmse over the Rmse of the MEIS method. We simulate 50 different realisations from the model. For each of these realisations, 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 NAIS and NAIScc methods (with  $S = 200$ ). We also report TNR as computing time normalised for an Rmse ratio of unity. The reported values are the average statistics across the 50 realisations. The methods (with their acronyms) are discussed in Table 1. 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  $t = 1, \dots, n$ .

$S = 20$	$n = 1000$					$n = 3000$				
	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

increase to  $S = 200$  in the EIS and MEIS methods falls short of the  $1/\sqrt{10}$  factor we expect in the absence of a bias-variance trade-off. For both  $S = 20$  and  $S = 200$  the NAIS method (without control variables) produces slightly higher variance and Rmse values when compared to 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 is able to compute the likelihood function four times faster, being therefore 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. This result is one of our main findings from the simulation study. We further highlight the minimal additional computational time required for the NAIS method to increase the number of samples from 20 to 200.

The 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 of the estimates 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 we can ignore the numerical integration error in constructing the control variables.

Table 3 for the stochastic volatility model with the autoregressive coefficient of 0.9 reveals new results. Our simulations suggest that the EIS bias becomes a larger problem for less persistent specifications of the state space model. The bias now almost completely dominates the Rmse of the EIS and MEIS methods when  $S = 20$ ; it is 10 times larger than the standard deviation for this low number of draws. Our new methods become comparatively more effective in this setting: by switching from the EIS method with  $S = 20$  to the NAIS method with  $S = 200$ , we obtain a 95% reduction in Rmse without any important increase in computational cost. The NAIScc method further reduces the variance of the likelihood estimate by 75% in relation to the NAIS

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  $t = 1, \dots, n$ .

$S = 20$	$n = 1000$					$n = 3000$				
	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

method.

Tables 4, 5 and 6 report the findings for the different specifications of the stochastic conditional duration and the stochastic copula models. Although the likelihood evaluation algorithms become more time consuming for these models, 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. 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 favour 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 favour the NAIS methods with a higher value for  $S$ .

Tables 7 and 8 present additional results. Table 7 reports 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. As  $S$  increases, the variance of the MEIS log importance sampling weights converges to the limiting

Table 4: LOGLIKELIHOOD ERRORS FOR STOCHASTIC CONDITIONAL DURATION I.

We report the same results as Table 2 for the model given by  $y_t \sim \text{Weibull}(\lambda_t, \psi = 1.2)$ ,  $\lambda_t = \exp(\alpha_t)$ ,  $\alpha_t = 0.98\alpha_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, Q = 0.15^2)$ .

$S = 20$	$n = 1000$					$n = 3000$				
	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

$S = 200$	$n = 1000$					$n = 3000$				
	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.

We report the same results as Table 2 for the model given by  $y_t \sim \text{Weibull}(\lambda_t, \psi)$ ,  $\lambda_t = \exp(\alpha_t)$ ,  $\alpha_t = T\alpha_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, Q)$ .

$S = 20$	$n = 1000$					$n = 1000$				
	$T = 0.96, Q = 0.1^2, \psi = 1.7$					$T = 0.9, Q = 0.15^2, \psi = 1.2$				
	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

$S = 200$	$n = 1000$					$n = 1000$				
	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.

We report the same results as Table 2 for the model given by  $u_{1t}, u_{2t} \sim C_{\nu=5, P_t}(u_t)$ ,  $\rho_t = (1 + \exp(-\alpha_t))^{-1}$ ,  $\alpha_t = 0.017 + 0.98\alpha_{t-1} + \eta_t$ ,  $\eta_t \sim N(0, Q = 0.1^2)$ .

$S = 20$	$n = 1000$					$n = 3000$				
	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

value obtained by our NAIS method. The table further illustrates the source of the EIS bias: for low values of  $S$ , the sample variance of the weights is artificially small in this method. 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, values of  $M$  between 20 and 30 guarantee a high degree of accuracy. A comparison between the results in Table 8 and in the previous tables confirms that the numerical integration error is negligible in relation to standard deviation of the likelihood estimates.

#### 4.4 Parameter estimation for a higher dimensional model

To further illustrate the performance of the NAIScc method, we consider the simulated maximum likelihood estimation of a multiple component stochastic volatility model in both a Monte Carlo exercise and an empirical application. We specify the model as in (23) but with

$$\sigma_t^2 = \exp(\theta_t), \quad \theta_t = d + \alpha_{1,t} + \dots + \alpha_{k,t},$$

Table 7: STANDARD DEVIATIONS FOR IMPORTANCE SAMPLING WEIGHTS.

For  $R = 100$  replications of the stochastic volatility specification as in Table 2 and the stochastic copula specification in Table 6, we obtain auxiliary coefficients for different numbers of Monte Carlo trajectories  $S$  (using the MEIS method) and different numbers of integration nodes  $M$  (using the numerical procedure of section of 3.1). We then simulate a thousand independent Monte Carlo 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-sample	Time	In-sample	Out-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  $M$  (see Section 3.1). We simulate 1,000 Monte Carlo paths from these coefficients and compute the variance of the resulting log importance sampling weights. We report average standard deviations over  $R$  replications. We denote  $\log(\widehat{L}(y; \psi)_{M=m})$  by  $\ell_m$  for  $m = 10, 20, 30,$ .

Model	$n = 1000$		$n = 3000$	
	Stoch Volatility	Stoch Copula	Stoch Volatility	Stoch Copula
Std. Dev. $\ell_{10}$	0.0075	0.0072	0.0236	0.0203
Std. Dev. $\ell_{20}$	0.0069	0.0067	0.0216	0.0189
Std. Dev. $\ell_{30}$	0.0069	0.0067	0.0216	0.0189
$ \ell_{30} - \ell_{10} $	0.0059	0.0032	0.0183	0.0102
$ \ell_{30} - \ell_{20} $	$6.24 \times 10^{-6}$	$5.55 \times 10^{-7}$	$1.91 \times 10^{-5}$	$1.94 \times 10^{-6}$

with  $k \times 1$  state vector  $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{k,t})'$ , with  $k \times k$  diagonal matrices  $T$  and  $Q$  given by

$$T = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_k \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma_{\eta,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\eta,k}^2 \end{bmatrix},$$

and with unknown coefficients  $|\phi_i| < 1$  and  $q_i > 0$ , for  $i = 1, \dots, k$ . We identify the model by imposing  $\phi_1 > \dots > \phi_k$ . The scalar signal  $\theta_t$  represents the log-volatility. Liesenfeld and Richard (2003) investigate their EIS method for a two component SV model. We can motivate the  $k$ -component SV model as a stochastic counterpart of the two component GARCH model of Engle and Lee (1999).

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. Maximise the loglikelihood function using an approximate but fast method. We suggest to take the NAIScc method with  $S = 0$  (no simulation, only numerical integration). We carry out the maximisation of the loglikelihood function by direct numerical optimisation.
4. Update the starting values for the importance parameters.
5. Re-start maximisation of loglikelihood function using the NAIScc method with  $S > 0$ .

The computational efficiency of this algorithm is due primarily to the accurate approximation of the loglikelihood function calculated by the NAIScc method using  $S = 0$ . As a result, the convergence of the maximisation in the last step is fast, requiring a small number of iterations. This two-step maximisation method gives the procedure the desirable property that we can set  $S$  at a high value with only a relatively small increase in computing time. Common random numbers ensure smoothness of the likelihood function, which is necessary for the application of numerical optimisation methods.

For our simulation exercise, we set the parameter values as  $d = 0.5$ ,  $\phi_1 = 0.99$ ,  $\phi_2 = 0.9$ ,  $\phi_3 = 0.4$ ,  $\sigma_{\eta,1}^2 = 0.005$ ,  $\sigma_{\eta,2}^2 = 0.016$ , and  $\sigma_{\eta,3}^2 = 0.05$ . The number of observations is set equal to  $n = 5000$ . The estimation of the multi-state specification requires a large time series dimension since the third volatility component has low persistence and may be hard to identify from a short sample. We set the number of simulated trajectories to  $S = 200$ . We draw 50 different time series realisations of the model. For each realised time series, we obtain 20 parameter estimates under different sets of common numbers and compute their Monte Carlo standard errors. We report the standard errors as the averages across the 50 realisations. Since we have set the true parameters ourselves, we also calculate the Rmse of the estimates, which allows us to directly compare the relative importance of the simulation and statistical errors in estimating the parameters.

We summarise the results in Table 9. The average estimation time for each realisation has been slightly under two minutes, despite the complexity of the model and the large sample size. Table 9 further presents the simulation errors, which are small for all parameters in absolute and in relative terms. In the estimation results, Monte Carlo standard errors represent only between 1% and 3% of the total Rmse.

## 4.5 Empirical application

Finally, we investigate whether the NAIScc method extends its good performance to empirical applications. Table 10 reports the estimation of a two-component stochastic volatility specification for the daily returns of six Dow Jones index stocks in the period between January 2001 and December 2010 (in a total of 2512 observations). As before, we set the number of simulated trajectories in the importance sampling estimation of the likelihood to  $S = 200$ . We repeat the estimation process a hundred times with different random numbers.

The results show that the Monte Carlo errors in the parameter estimates are virtually zero for persistent states, with autoregressive coefficients larger than about 0.9. For three of the stocks, parameter estimation has been challenging because the second component is weakly persistent and noisy. The Monte Carlo standard errors of the associated parameters have reached 10-20% of the statistical standard errors in these cases. However, the relatively low estimation times (between one and two minutes)



Table 9: THREE COMPONENT SV MODEL: MONTE CARLO ERRORS.

Based on 50 realisations of a three component stochastic volatility model, we obtain 20 simulated maximum likelihood parameter estimates based on difference random values and using the NAIScc method, for each realisation. We report the average Monte Carlo standard error (SE) across the 50 realisations. 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. We specify the model as  $y_t \sim N(0, \sigma_t^2)$ ,  $t = 1, \dots, n$ ,  $\sigma_t^2 = \exp(\theta_t)$ ,  $\theta_t = d + \alpha_{1,t} + \alpha_{2,t} + \alpha_{3,t}$ ,  $\alpha_t = T\alpha_{t-1} + \eta_t$ ,  $\alpha_1 \sim N(a_1, P_1)$ ,  $\eta_t \sim N(0, Q)$ , where  $T$  is a diagonal matrix with elements  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  and  $Q$  is a diagonal matrix with elements  $\sigma_{\eta,1}^2$ ,  $\sigma_{\eta,2}^2$  and  $\sigma_{\eta,3}^2$ .

Parameter	True Value	Monte Carlo SE	Rmse	Monte Carlo SE / Rmse
$d$	0.5	0.0016	0.103	0.016
$\phi_1$	0.99	0.0001	0.006	0.009
$\sigma_{\eta,1}^2$	0.005	0.0000	0.003	0.011
$\phi_2$	0.9	0.0006	0.049	0.011
$\sigma_{\eta,2}^2$	0.015	0.0002	0.010	0.018
$\phi_3$	0.4	0.0032	0.280	0.012
$\sigma_{\eta,3}^2$	0.05	0.0008	0.029	0.029

indicate that we may consider larger samples to better identify the second volatility component. Figure 1 illustrates the signal and state estimates we have obtained for the parameter estimates reported in Table 10.

To examine computational efficiency, we compare the NAIS and EIS methods for estimation of parameters in the stochastic volatility model. For the implementation of the EIS method we have followed Liesenfeld and Richard (2003). To avoid the bias problem and to make the computational burden comparable between the two methods, we consider  $S = 20$  simulated trajectories in the regression stage and  $S = 200$  samples for calculating the likelihood.

We present the estimation results for the EIS method in Table 10 as well. The averages of the parameter estimates are similar to the ones we have obtained with the NAIS method. However, the use of the EIS method for this problem has three important disadvantages. First, the EIS method produces large Monte Carlo standard errors for parameters associated with non-persistent states. Second, the EIS method

is computationally less efficient. We obtain up to 90% computing time reductions by using the NAIS method. This finding is partly due to the fact that the EIS method simulates the two-dimensional state, rather than the one-dimensional signal; the auxiliary EIS regressions are based on six regressors, against three in the NAIS method (see Appendix A). Third, the low number of samples in the EIS regressions make the algorithm numerically unstable, leading to many numerical errors during the estimation. Numerical crashes affected between 15% and 80% of the EIS replications, while the NAIS method has not crashed in any of the replications. This result illustrates both the practical difficulties with importance sampling and the stability of the NAIS method. Finally, we note that increasing the number of samples in the first step of the EIS algorithm is very costly, as it requires the computation of a large number of auxiliary regressions.

## 5 Conclusion

We have developed a new efficient importance sampling method for the evaluation of the likelihood function of nonlinear non-Gaussian state space models. The numerically accelerated importance sampling (NAIS) approach is a non-trivial mix of numerical and Monte Carlo integration methods. We use Gauss-Hermite quadratures for constructing the importance sampler. The Monte Carlo evaluation of the likelihood function is primarily based on Kalman filtering and smoothing methods. We introduce new control variables to further reduce the sampling variance of the Monte Carlo estimate of the likelihood function. We have carried out a comprehensive simulation study to verify the performance of our approach relative to other importance sampling methods for a variety of financial time series models. Our empirical application to U.S. stock returns has shown that the NAIS produces reliable results in a numerically and computationally efficient way.

Table 10: TWO COMPONENT SV MODEL: ESTIMATION RESULTS.

We estimate a two-component stochastic volatility specification for the daily returns of six Dow Jones index stocks in the period between January 2001 and December 2010 (2512 observations). We repeat the NAIS and EIS estimation methods a hundred times with different random numbers. The Monte Carlo and statistical standard errors are in brackets [ ] and parentheses ( ) respectively. We specify the model as  $y_t \sim N(0, \sigma_t^2)$ ,  $t = 1, \dots, n$ ,  $\sigma_t^2 = \exp(\theta_t)$ ,  $\theta_t = d + \alpha_{1,t} + \alpha_{2,t}$ ,  $\alpha_t = T\alpha_{t-1} + \eta_t$ ,  $\alpha_1 \sim N(0, P_1)$ ,  $\eta_t \sim N(0, Q)$ , where  $T$  is a diagonal matrix with elements  $\phi_1$ , and  $\phi_2$  and  $Q$  is a diagonal matrix with elements  $\sigma_{\eta,1}^2$  and  $\sigma_{\eta,2}^2$ .

	GE		JP Morgan		Coca-Cola		AT&T		Wal-Mart		Exxon	
	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS
$d$	0.847 [0.004] (0.690)	0.848 [0.008]	1.273 [0.002] (0.458)	1.270 [0.006]	0.005 [0.004] (0.393)	0.009 [0.008]	0.654 [0.002] (0.652)	0.653 [0.004]	0.374 [0.002] (0.413)	0.377 [0.005]	0.555 [0.000] (0.193)	0.555 [0.001]
$\phi_1$	0.997 [0.000] (0.002)	0.997 [0.000]	0.997 [0.000] (0.002)	0.997 [0.000]	0.995 [0.000] (0.003)	0.995 [0.001]	0.997 [0.000] (0.002)	0.997 [0.000]	0.996 [0.000] (0.002)	0.996 [0.000]	0.991 [0.000] (0.006)	0.991 [0.001]
$\sigma_{1,\eta}^2$	0.007 [0.000] (0.003)	0.007 [0.000]	0.008 [0.000] (0.003)	0.008 [0.000]	0.006 [0.000] (0.003)	0.006 [0.001]	0.005 [0.000] (0.003)	0.005 [0.000]	0.005 [0.000] (0.002)	0.005 [0.000]	0.008 [0.000] (0.008)	0.008 [0.001]
$\phi_2$	0.457 [0.020] (0.271)	0.463 [0.061]	0.839 [0.003] (0.065)	0.839 [0.011]	0.489 [0.032] (0.164)	0.494 [0.090]	0.897 [0.001] (0.081)	0.897 [0.006]	0.273 [0.013] (0.118)	0.263 [0.064]	0.936 [0.000] (0.048)	0.937 [0.004]
$\sigma_{1,\eta}^2$	0.262 [0.010] (0.091)	0.250 [0.019]	0.085 [0.002] (0.032)	0.085 [0.006]	0.307 [0.016] (0.077)	0.291 [0.035]	0.027 [0.000] (0.018)	0.027 [0.001]	0.263 [0.006] (0.054)	0.259 [0.015]	0.026 [0.000] (0.013)	0.026 [0.001]
Time	2m12s	14m03s	1m30s	14m12s	1m51s	11m52	1m10s	12m05s	1m59s	8m26s	1m00s	7m58s
Improvement	-84%		-89%		-84%		-90%		-76%		-87%	
Crashes	0%	36%	0%	78%	0%	21%	0%	13%	0%	40%	0%	16%

## References

- Banachewicz, K. (2009). *A collection of problems in credit risk modeling*. Ph. D. thesis, VU University Amsterdam.
- Bauwens, L. and F. Galli (2009). Efficient importance sampling for ml estimation of scd models. *Computational Statistics and Data Analysis* 53(6), 1974–1992.
- Bauwens, L. and N. Hautsch (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics* 4(3), 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(2), 381–412.
- 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* (84), 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. *Journal of the Royal Statistical Society, Series B* (62), 3–56.
- Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* (89), 603–616.
- Engle, R. and G. Lee (1999). A long-run and short-run component model of stock return volatility. In R. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, pp. 475–497. Oxford University Press.

- Fridman, M. and L. Harris (1998). A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business and Economic Statistics* 16(3), 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 G. Maddala and C. Rao (Eds.), *Handbook of Statistics, Vol 14*. Elsevier, Amsterdam.
- Hafner, C. and H. Manner (2011). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics*. forthcoming.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Heiss, F. (2008). Sequential numerical integration in nonlinear state space models for microeconomic panel data. *Journal of Applied Econometrics* 23(3), 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* (94), 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). Fast efficient importance sampling by state space methods. Working paper, VU University Amsterdam.
- Koopman, S. J., N. Shephard, and D. Creal (2009). Testing the assumptions behind importance sampling. *Journal of Econometrics* 149(1), 2–11.
- Liesenfeld, R. and J. Richard (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance* 10(4), 505–531.
- Melino, A. and S. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 239–265.

- Monahan, J. F. (1993). Testing the behaviour of importance sampling weights. *Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface*, 112–117.
- Monahan, J. F. (2001). *Numerical methods of statistics*. Cambridge: Cambridge University Press.
- 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.
- 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. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Shephard, N. and M. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika* 84, 653–667.
- 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) applied to (1) considers the Gaussian importance model density

$$g(\alpha, y; \psi) = \prod_{t=1}^n g(y_t | \alpha_t; \psi) g(\alpha_t | \alpha_{t-1}; \psi), \quad (25)$$

where

$$g(y_t | \alpha_t; \psi) = \exp \left\{ a_t + b'_t \alpha_t - \frac{1}{2} \alpha'_t C_t \alpha_t \right\}. \quad (26)$$

We can merge the expressions for  $g(y_t | \alpha_t; \psi)$  and  $g(\alpha_t | \alpha_{t-1}; \psi)$  and interpret them as a Gaussian density  $g(\alpha_t | y_t, \alpha_{t-1}; \psi)$  by completing the squares. The mean vector and variance matrix are given by

$$\hat{\alpha}_t = V_t (b'_t + (d_t + T_t \alpha_{t-1})' Q_t^{-1})', \quad V_t = (Q_t^{-1} + C_t)^{-1}, \quad (27)$$

respectively, provided that we set the constant  $a_t = a_t(\chi_t, \alpha_{t-1})$  equal to

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

We may then sample the state vectors sequentially from  $g(\alpha_t | y_t, \alpha_{t-1}; \psi)$  for a given set of importance parameters  $\chi$ . In contrast to the importance density (11) we adopt in Section 2.4, the constant  $a_t$  depends on  $\alpha_{t-1}$  in the EIS approach. This property marks the essential difference between the EIS algorithm and the modified EIS method of Koopman and Nguyen (2011), leading to different simulation frameworks for selecting  $\chi$ . Liesenfeld and Richard (2003) have originally suggested the recursive structure we apply in Section 3.1. Let  $k$  index the iterations. The EIS method minimises

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

backwards from  $t = n$  to  $t = 1$ , where

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

with constant  $\lambda_{0t}$  and where we define the weight  $\omega(\alpha_t, y_t; \psi)$  in (15). In this algorithm it is fundamental that the updated integration constant  $a_{t+1}(\chi_{t+1}^{[k+1]}, \alpha_t)$  appears in the period  $t$  regression as above. Richard and Zhang (2007) argue that the integration constants  $a_t$ ,  $t = 1, \dots, n$ , capture the dynamic structure of model (1). For this reason, it is necessary to base the EIS regressions on the state  $\alpha_t$  rather than the signal  $\theta_t$ .

The EIS and MEIS methods rely on similar importance density approximations of  $p(y_t | \theta_t; \psi)$  as given by (25) and (8) respectively. This implies that the two procedures are numerically close to each other, although not identical because of the role of the integration constants in the EIS method. The simulation results presented in Section 4.3 confirm this observation. From a computational perspective, however, we point out that the SPDK importance model brings four advantages. First, it avoids the large number of computations required by the EIS method to track the integration constants  $a_t$ . Second, it relies on fast linear state space methods such as the Kalman filter. Third, it is based on directly simulating the signal  $\theta_t$  rather than the possibly high dimensional state  $\alpha_t$ . It also leads to the estimation of a smaller number of importance parameters when the state has higher dimension than the signal. Fourth, it enables the use of parallel computing for running the auxiliary regressions

## B The bias in the EIS method

The EIS method is subject to a finite sample bias when the same set of random numbers is used for obtaining the importance parameters  $\chi$  via the sampling variance minimisation (29) and for computing the likelihood estimate (5) via importance sampling. We denote the common random numbers by a vector  $u$ . The choice of  $\chi$  depends on  $u$  in the EIS method, that is  $\chi = \chi(u)$ . The simulated signal  $\theta^{(s)}$  also depends on  $u$ , that is  $\theta^{(s)} = \theta^{(s)}(u)$  for  $s = 1, \dots, S$ . Hence, we argue that  $g(\theta^{(s)}(u) | \chi(y, u); \psi)$  is not well defined as an importance density.

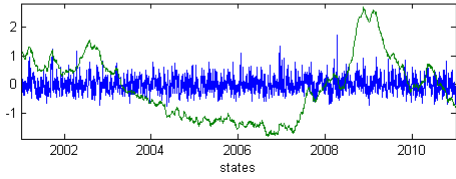
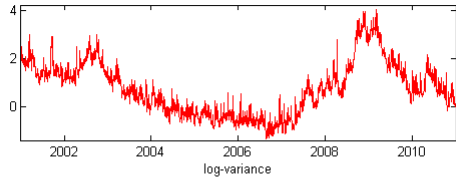
Consider the Taylor series expansion of the likelihood estimate round some value  $x$



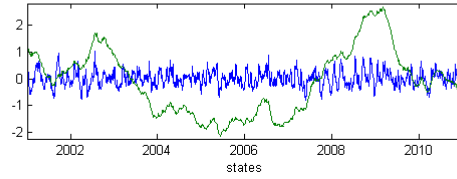
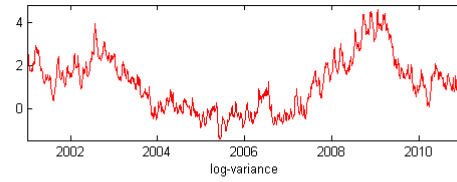
given by (21)

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

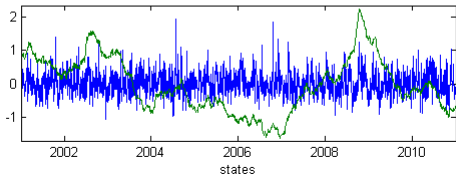
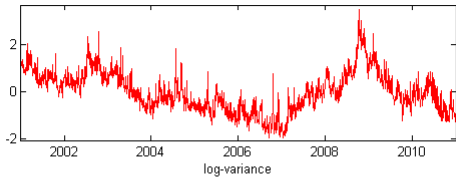
where  $x_s = \log \omega(\theta^{(s)}, y; \psi)$  and  $\theta^{(s)}$  does not necessarily depend on  $u$ . For the EIS method,  $\chi(u)$  is explicitly selected to minimise the sample variance of the log importance sampling weights  $\log \omega[\theta^{(s)}(u), y; \psi]$ . The minimised variance of  $\log \omega[\theta^{(s)}(u), y; \psi]$  is therefore artificially low when we compare it with the variance evaluated over the full support of  $g(\theta|y; \psi)$ . As a result, the third term in the Taylor expansion (30) is biased downwards for  $\theta^{(s)} = \theta^{(s)}(u)$ . This problem also contaminates the other terms, with ambiguous net effects for the likelihood estimate.



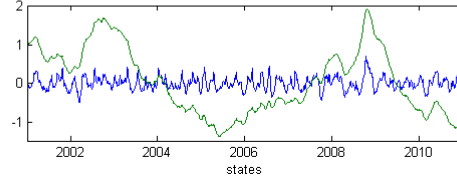
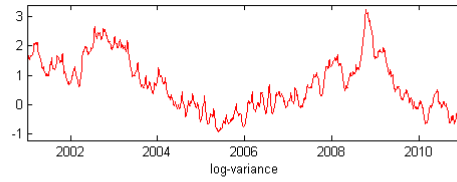
(a) General Electric



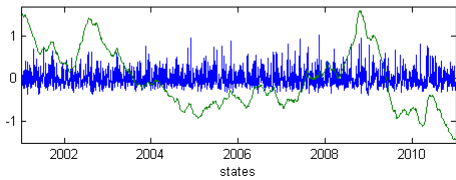
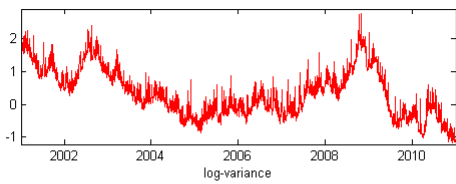
(b) JP Morgan



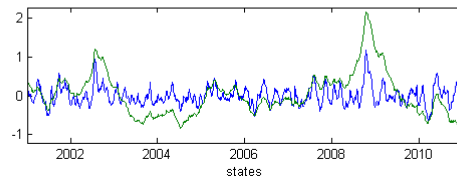
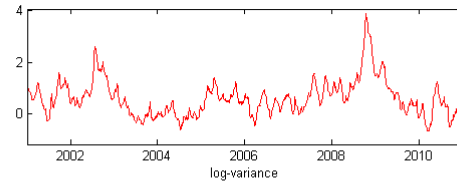
(c) Coca-Cola



(d) AT&T



(e) Wal-Mart



(f) Exxon

Figure 1: Estimated log-variance (top) and states (bottom) for six Dow Jones stocks.