Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

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

Siem Jan Koopman[ab], André Lucas[a] & Marcel Scharth[c]
[a] VU University Amsterdam and Tinbergen Institute, The Netherlands
[b] CREATES, Aarhus University, Denmark
[c] Australian School of Business, University of New South Wales, Australia
Accepted author version posted online: 12 Jun 2014.

To cite this article: Siem Jan Koopman, André Lucas & Marcel Scharth (2014): Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2014.925807

To link to this article: http://dx.doi.org/10.1080/07350015.2014.925807

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

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

[(a)] VU University Amsterdam and Tinbergen Institute, The Netherlands

[(b)] CREATES, Aarhus University, Denmark

[(c)] Australian School of Business, University of New South Wales, Australia

Abstract

We propose a general likelihood evaluation method for nonlinear non-Gaussian state space models using the simulation based method of efficient importance sampling. We minimise the simulation effort by replacing some key steps of the likelihood estimation procedure by numerical integration. We refer to this method as numerically accelerated importance sampling. We show that the likelihood function for models with a high-dimensional state vector and a low-dimensional signal can be evaluated more efficiently using the new method. We report many efficiency gains in an extensive Monte Carlo study as well as in an empirical application using a stochastic volatility model for U.S. stock returns with multiple volatility factors.

1 Introduction

The evaluation of an analytically intractable likelihood function is a challenging task for many statistical and econometric time series models. The key challenge is the computation of a high-dimensional integral which is typically carried out by importance sampling methods. Advances in

*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 4th CSDA conference in London, for their insightful comments. S.J. Koopman acknowledges the support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. A. Lucas acknowledges the financial support of the Dutch National Science Foundation (NWO, grant VICI453-09-005). M. Scharth acknowledges partial support from the Australian Research Council grant DP0667069. Contact author: Siem Jan Koopman (s.j.koopman@vu.nl).

importance sampling over the past three decades have contributed to the interest in nonlinear non-Gaussian state space models that in most 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 nonlinear panel data models with unobserved heterogeneity as in Heiss (2008).

We propose a new numerically and computationally efficient importance sampling method for nonlinear non-Gaussian state space models. We show that a major part of the likelihood evaluation procedure can be done by fast numerical integration rather than Monte Carlo integration only. Our main contribution consists of two parts. First, a numerical integration scheme is developed to construct an efficient importance density that minimises the log-variance of the simulation error. Our approach is based on the efficient importance sampling (EIS) method of Richard and Zhang (2007), which relies on Monte Carlo simulations to obtain the importance density. Second, we propose new control variables that eliminate the first order simulation error in evaluating the likelihood function via importance sampling.

Numerical integration is generally highly accurate but its feasibility is limited to low dimensional problems. Although the Monte Carlo integration method is applicable to high dimensional problems, it is subject to simulation error. These properties are typical to applications in time series modelling. Here we adopt both methods and we show how to carry the virtues of numerical integration over to high dimensional state space models. 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).

Following Shephard and Pitt (1997) and Durbin and Koopman (1997) (referred to as SPDK), we base our importance sampling method on an approximating linear Gaussian state space model and use computationally efficient methods for this class of models. We explore two different, but

numerically equivalent, sets of algorithms. The first approach follows SPDK and considers Kalman filtering and smoothing (KFS) methods. The second reinterprets and extends the EIS sampler in Jung, Liesenfeld, and Richard (2011) as a backward-forward simulation smoothing method for a linear state space model. We clarify the relations between these two methods. We further conduct a Monte Carlo and empirical study to analyse the efficiency gain of the NAIS method when applied to the stochastic volatility model as, for example, in Ghysels, Harvey, and Renault (1996). We present results for other model specifications in an online appendix: the stochastic duration model of Bauwens and Veredas (2004); the stochastic copula model of Hafner and Manner (2011); and the dynamic factor model for multivariate count data of Jung, Liesenfeld, and Richard (2011).

The Monte Carlo study reveals three major findings. First, we show that our NAIS method can provide 40-50% reductions in simulation variance for likelihood evaluation, when compared to a standard implementation of the EIS. Second, the use of the new control variables further decreases the variance of the likelihood estimates by 20-35% relative to the use of antithetic variables as a variance reduction device. The use of antithetic variance reduction techniques may even lead to a loss of computational efficiency, especially for models with multiple state variables. Third, by taking the higher computational efficiency of the NAIS method into account, we find 70-95% gains in variance reduction of the likelihood estimates, when compared to the standard EIS method and after we have normalised the gains in computing time. Similar improvements are obtained when we compare the EIS algorithm of Richard and Zhang (2007) with the local approximation method of SPDK.

To illustrate the NAIS method in an empirical setting, we consider a two-factor stochastic volatility model applied to time series of returns for a set of major U.S. stocks. The two-factor structure of the volatility specification makes estimation by means of importance sampling a non-trivial task. However, we are able to implement the NAIS approach using standard hardware and software facilities without complications. The NAIS method reduces the computing times in this application by as much as 66% and leads to Monte Carlo standard errors for the estimated

parameters which are small compared to their statistical standard errors. This application illustrates that we are able to 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. Our main methodological contributions are in Section 2.4 and Section 3, which present our numerically accelerated importance sampling (NAIS) method and the corresponding new control variables, respectively. Section 4 discusses the results of the Monte Carlo and empirical studies. 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) and Geweke (1989), among others. Danielsson and Richard (1993), Shephard and Pitt (1997) and Durbin and Koopman (1997) have explored 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. Our main task is to evaluate the likelihood function for the nonlinear non-Gaussian state space model

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

where y_t is the $p \times 1$ observation vector, θ_t is the $q \times 1$ signal vector, α_t is the $m \times 1$ state vector, and Z_t is the $q \times m$ selection matrix; the dynamic properties of the stochastic vectors y_t, θ_t and α_t

are determined by the $m \times 1$ constant vector d_t, the $m \times m$ transition matrix T_t, and $m \times m$ variance matrix Q_t. The Gaussian innovation series η_t is serially uncorrelated. The initial mean vector a_1 and variance matrix P_1 are determined from the unconditional properties of the state vector α_t. The system variables Z_t, d_t, T_t and Q_t are only time-varying in a deterministic way. The unknown fixed parameter vector ψ contains, or is a function of, the unknown coefficients associated with the observation density $p(y_t|\theta_t;\psi)$ and the system variables.

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 θ_t in a possibly nonlinear way. The signal vector θ_t depends linearly on the state vector α_t, for which we formulate a linear dynamic model. Our general framework accommodates combinations of autoregressive moving average, long memory, random walk, cyclical and seasonal dynamic processes. Harvey (1989) and Durbin and Koopman (2012) provide a detailed discussion of state space model representations and unobserved components time series models.

2.2 Likelihood evaluation via importance sampling

Define and $y' = (y_1', \ldots, y_n')$, $\theta' = (\theta_1', \ldots, \theta_n')$ and $\alpha' = (\alpha_1', \ldots, \alpha_n')$. If $p(y_t|\theta_t;\psi)$ is a Gaussian density with mean $\theta_t = Z_t\alpha_t$ and covariance matrix H_t, for $t = 1, \ldots, n$, Kalman filtering and smoothing methods evaluate the likelihood and compute the minimum mean squared error estimates of the state vector α_t together with its mean squared error matrix. In all other cases, the likelihood for (1) is given by the analytically intractable integral

$$L(y;\psi) = \int p(\alpha, y;\psi)\,\mathrm{d}\alpha = \int \prod_{t=1}^{n} p(y_t|\theta_t;\psi)p(\alpha_t|\alpha_{t-1};\psi)\,\mathrm{d}\alpha_1 \ldots\, \mathrm{d}\alpha_n, \tag{2}$$

where $p(\alpha, y;\psi)$ is the joint density of y and α following from (1). Kitagawa (1987) has developed a numerical integration method for evaluating the likelihood integral in (2). This approach is only feasible when n is small and y_t, θ_t and α_t are scalars.

We aim to evaluate the likelihood function by means of importance sampling. For this purpose, we consider the Gaussian importance density

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

where $g(y|\theta;\psi)$, $g(\alpha;\psi)$ and $g(y;\psi)$ are all Gaussian densities and where $g(y;\psi)$ can be interpreted as a normalising constant. It is implied from (1) that $g(\alpha;\psi) \equiv p(\alpha;\psi)$. We can express the likelihood function as

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

where the importance weight function is given by

$$\omega(\theta,y;\psi) = p(y|\theta;\psi) \,/\, g(y|\theta;\psi). \tag{3}$$

We evaluate the likelihood function using S independent trajectories $\theta^{(1)},\ldots,\theta^{(S)}$ that we sample from the signal importance density $g(\theta|y;\psi)$. The likelihood estimate is given by

$$\widehat{L}(y;\psi) = g(y;\psi)\times\bar{\omega}, \qquad \bar{\omega} = \frac{1}{S}\sum_{s=1}^{S}\omega_s, \qquad \omega_s = \omega(\theta^{(s)},y;\psi), \tag{4}$$

where ω_s is the importance weight (3) evaluated at $\theta = \theta^{(s)}$. The estimate (4) relies on the typically low-dimensional signal vector θ_t rather than the typically high-dimensional state vector α_t. Hence the computations can be implemented efficiently.

Under standard regularity conditions, the weak law of large numbers ensures that

$$\widehat{L}(y;\psi) \xrightarrow{p} L(y;\psi), \tag{5}$$

when $S \to \infty$. 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.

2.3 Importance density as a linear state space model

The Gaussian importance density for the state vector can be represented as $g(\alpha|y;\psi) = g(\alpha, y;\psi)/g(y;\psi)$ with

$$g(\alpha, y;\psi) = \prod_{t=1}^{n} g(y_t|\theta_t;\psi)g(\alpha_t|\alpha_{t-1};\psi), \tag{6}$$

where $g(\alpha_t|\alpha_{t-1};\psi)$ is the Gaussian density for the state transition equation in (1), and with

$$g(y_t|\theta_t;\psi) = \exp\left\{a_t + b_t'\,\theta_t - \frac{1}{2}\theta_t'\,C_t\,\theta_t\right\}, \tag{7}$$

where scalar a_t, vector b_t and matrix C_t are defined as functions of the data vector y and the parameter vector ψ, for $t = 1, \ldots, n$. The constants $a_1, \ldots, a_n$ are chosen such that (6) integrates to one. The set of unique importance sampling parameters is therefore given by

$$\{b, C\} = \{b_1, \ldots, b_n, C_1, \ldots, C_n\}. \tag{8}$$

The state transition density $g(\alpha_t|\alpha_{t-1};\psi)$ in (6) represents the dynamic properties of α_t and is the same as in the original model (1) because $g(\alpha;\psi) \equiv p(\alpha;\psi)$. Hence the importance density only varies with $\{b, C\}$. We discuss proposals for computing the importance parameter set $\{b, C\}$ in Section 2.4.

It is shown by Koopman, Lit, and Nguyen (2014) that the observation density $g(y_t|\theta_t;\psi)$ in (7) can be represented by a linear state space model for the artificial observations $y_t^* = C_t^{-1}b_t$ which are computed for a given importance parameter set $\{b, C\}$ in (8). The observation equation for y_t^* is

given by

$$y_t^* = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, C_t^{-1}), \qquad t = 1, \ldots, n, \tag{9}$$

where θ_t is specified as in (1) and the innovation series ε_t is assumed to be serially and mutually uncorrelated with the innovation series η_t in (1). The Gaussian logdensity $\log g(y_t^*|\theta_t; \psi)$ is equivalent to the log of (7) since

$$\begin{aligned} \log g(y_t^*|\theta_t; \psi) &= -\frac{1}{2}\log 2\pi + \frac{1}{2}\log|C_t| - \frac{1}{2}(C_t^{-1} b_t - \theta_t)' C_t (C_t^{-1} b_t - \theta_t) \\ &= a_t + b_t' \theta_t - \frac{1}{2}\theta_t' C_t \theta_t, \end{aligned} \tag{10}$$

where the constant a_t collects all the terms that are not associated with θ_t. It follows that $a_t = (\log|C_t| - \log 2\pi - b_t' y_t^*)/2$. We conclude that $g(\theta, y; \psi) \equiv g(\theta, y^*; \psi)$, and hence $g(\alpha, y; \psi) \equiv g(\alpha, y^*; \psi)$, with $y^{*\prime} = (y_1^{*\prime}, \ldots, y_n^{*\prime})$.

The linear model representation allows the use of computationally efficient methods for signal extraction and simulation smoothing. In particular, for the Monte Carlo evaluation of the likelihood as in (4), we require simulations from $g(\theta|y; \psi) \equiv g(\theta|y^*; \psi)$ for a given set $\{b, C\}$ where b and C are functions of y and ψ. We have two options that lead to numerically equivalent results. The first option is to apply the state space methods to model (9) as described in Durbin and Koopman (2012). Signal extraction relies on the Kalman filter smoother and simulation smoothing may rely on the methods of de Jong and Shephard (1995) or Durbin and Koopman (2002). The second option is to use a modification of the importance sampling method of Jung, Liesenfeld, and Richard (2011) which we refer to as the backward-forward (BF) method. We show in Appendix A that the approach of Jung, Liesenfeld, and Richard (2011) can be reshaped and extended to obtain a new algorithm for signal extraction and simulation smoothing applied to model (9). This completes the discussion of likelihood evaluation via importance sampling. Several devices for simulation variance reduction, including antithetic variables, can be incorporated in the computations of (4).

Next we discuss a new method for choosing the importance parameter set $\{b, C\}$ in (8). Existing methods are reviewed in Section 2.5.

2.4 Numerically accelerated importance sampling

The numerically accelerated importance sampling (NAIS) method constructs optimal values for the importance parameter set $\{b, C\}$ using the same criterion for the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003) and Richard and Zhang (2007). The values for b and C in the EIS are chosen such that the variance of the log importance weights $\log \omega(\theta, y; \psi)$ is minimised, that is

$$\min_{b,C} \int \lambda^2(\theta, y; \psi)\omega(\theta, y; \psi)g(\theta|y; \psi)\,\mathrm{d}\theta, \tag{11}$$

where

$$\lambda(\theta, y; \psi) = \log \omega(y, \theta; \psi) = \log p(y|\theta; \psi) - \log g(y|\theta; \psi), \tag{12}$$

with $g(y|\theta; \psi) = \prod_{t=1}^{n} g(y_t|\theta_t; \psi)$ and with $g(y_t|\theta_t; \psi)$ given by (7). The minimisation in (11) is high-dimensional and numerically not feasible in most cases of interest. The minimisation in (11) is therefore reduced to a series of minimisations for b_t and C_t, for $t = 1, \ldots, n$. Hence we obtain the importance parameters $\{b_t, C_t\}$ by

$$\min_{b_t,C_t} \int \lambda^2(\theta_t, y_t; \psi)\omega(\theta_t, y_t; \psi)g(\theta_t|y; \psi)\,\mathrm{d}\theta_t, \qquad t = 1, \ldots, n, \tag{13}$$

where $\omega(\theta_t, y_t; \psi) = p(y_t|\theta_t; \psi) \,/\, g(y_t|\theta_t; \psi)$ and $\lambda(\theta_t, y_t; \psi) = \log \omega(y_t, \theta_t; \psi)$.

The NAIS method is based on the insight that the smoothing density $g(\theta_t|y; \psi) \equiv g(\theta_t|y^*; \psi)$ is available analytically for the linear Gaussian state space model (9). For a given set $\{b, C\}$ and for a scalar signal θ_t, we have

$$g(\theta_t|y^*; \psi) = \mathrm{N}(\widetilde{\theta}_t\,,\, V_t) = \exp\left\{-\frac{1}{2}V_t^{-1}(\theta_t - \widetilde{\theta}_t)^2\right\} \,/\, \sqrt{2\pi\, V_t}, \qquad t = 1, \ldots, n, \tag{14}$$

where the conditional mean $\widetilde{\theta}_t$ and variance V_t are obtained from the Kalman filter and smoother (KFS) or the backward-forward (BF) smoothing method of Appendix A, in both cases applied to model (9).

This result allows us to directly minimise the low dimensional integral (13) for each time period t via the method of numerical integration. For most cases of practical interest, the method produces a virtually exact solution. Numerical integration is reviewed in Monahan (2001). We adopt the Gauss-Hermite quadrature method that is based on a set of M abscissae z_j with associated weights $h(z_j)$, for $j = 1, \ldots, M$. The value of M is typically between 20 and 30. The values for the weights $h(z_j)$ are predetermined for any value of M. The Gauss-Hermite quadrature approximation of the minimisation (13) is then given by

$$\min_{b_t, C_t} \sum_{j=1}^{M} h(z_j) \exp(z_j^2) \varphi(\widetilde{\theta}_{tj}), \tag{15}$$

where $\varphi(\widetilde{\theta}_{tj}) = \lambda^2(\widetilde{\theta}_{tj}, y_t; \psi)\, \omega(\widetilde{\theta}_{tj}, y_t; \psi)\, g(\widetilde{\theta}_{tj}|y^*; \psi)$, with $\widetilde{\theta}_{tj} = \widetilde{\theta}_t + V_t^{1/2} z_j$, for $j = 1, \ldots, M$. The conditional mean $\widetilde{\theta}_t$ and variance V_t are defined below equation (14) from which it follows that $g(\widetilde{\theta}_{tj}|y^*; \psi) = \exp(-0.5 z_j^2)/\sqrt{2\pi V_t}$. Further details of our numerical implementation are provided in the Online Appendix.

The minimisation (15) takes place iteratively as in the original EIS method. For a given importance parameter set $\{b, C\}$, we obtain mean $\widetilde{\theta}_t$ and variance V_t from (14), for $t = 1, \ldots, n$. New values for b_t and C_t are then obtained from the minimisation (15) that is reduced to a weighted least squares computation for each t. We take $\log p(y_t|\widetilde{\theta}_{tj}; \psi)$ as the "dependent" variable, vector $(1, \widetilde{\theta}_{tj}, -0.5\,\widetilde{\theta}_{tj}^2)'$ as the "explanatory" variable and $\exp(z_j^2/2) h(z_j) \omega(\widetilde{\theta}_{tj}, y_t; \psi)$ as the "weight" for which all terms that do not depend on z_j are dropped. The regression computations are based on sums over $j = 1, \ldots, M$. The second and third regression estimates, those associated with $\widetilde{\theta}_{tj}$ and $\widetilde{\theta}_{tj}^2$, represent the new values for b_t and C_t, respectively. The computations are repeated for each t or are done in parallel. Once a new set for $\{b, C\}$ is constructed, we can repeat this procedure. Convergence towards a final set $\{b, C\}$ typically takes a few iterations.

We start the procedure by having initial values for $\{b, C\}$. A simple choice is $b_t = 0$ and $C_t = 1$, for $t = 1, \ldots, n$. We can also initialise $\{b, C\}$ by the local approximation of Durbin and Koopman (1997), see Section 2.5. Finally, Richard and Zhang (2007) argue that we can omit the term $\omega(\widetilde{\theta}_{tj}, y_t; \psi)$ for computing the regression "weight" without much loss of numerical efficiency. We prefer to adopt this modification because it is computationally convenient. The NAIS procedure for selecting the importance parameter set is reviewed in Figure 1.

2.5 Relation to previous methods

We show in Section 4 that the NAIS method is an efficient method from both computational and numerical perspectives. We compare the NAIS with two related methods. The first is the local approximation method of Shephard and Pitt (1997) and Durbin and Koopman (1997) which we refer to as SPDK. It is based on a second-order Taylor expansion of $\log p(y_t|\theta_t; \psi)$ around the conditional mode of $p(\theta_t|y; \psi)$ which can be computed iteratively using KFS or BF applied to model (9) for evaluating $\widetilde{\theta}_t$ and V_t as in the algorithm in Figure 1. The iterations require analytic expressions for the first two derivatives of $\log p(y_t|\theta_t; \psi)$, with respect to θ_t, for $t = 1, \ldots, n$. The convergence towards the mode of $p(\theta_t|y; \psi)$ typically takes a few iterations, see Durbin and Koopman (2012).

The second method is the efficient importance sampling (EIS) algorithm as developed by Richard and Zhang (2007). This method approximates the minimisation problem in (13) via the sampling of the signal $\theta_t = Z_t\alpha_t$ via simulation smoothing from $g(\alpha|y; \psi)$, for $t = 1, \ldots, n$, rather than by numerical integration as in NAIS. In the modelling framework (1), the sampling is carried out by the backward-forward sampler of Jung, Liesenfeld, and Richard (2011). The BF sampler in Appendix A is a modification and a more efficient implementation of this sampler. Furthermore, the introduction of antithetic variables halves the simulation effort. Koopman, Lit, and Nguyen (2014) develop an implementation of EIS that is based on the KFS. We emphasise that the EIS uses simulation methods both for selecting $\{b, C\}$ and for evaluating the likelihood function. In

Section 4 we consider the methods of SPDK and EIS using BF and KFS. These methods for selecting the importance parameter set are reviewed in detail in the Online Appendix.

3 New control variables for NAIS

We introduce a new set of control variables to improve the numerical efficiency of likelihood evaluation using importance sampling based on NAIS. We develop control variables that are based on specific Taylor series expansions. The control variables can be evaluated by numerical integration using Gauss-Hermite quadrature. We exploit the differences between the estimates that are obtained from simulation and numerical integration methods with the sole purpose of reducing the variance of importance sampling estimates. This approach of variance reduction in the context of NAIS can replace the use of the antithetic variables proposed by Ripley (1987) and Durbin and Koopman (2000).

The likelihood estimate (4) is the sample average $\bar{\omega} = S^{-1}\sum_{s=1}^{S}\omega_s$ multiplied by $g(y;\psi)$, where

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

with the importance sample weights $\omega(\theta_t, y_t; \psi)$ defined below (13), $\theta^{(s)}$ generated from the importance density $g(\theta|y;\psi)$, and $\theta_t^{(s)}$ denoting the tth element of $\theta^{(s)}$, for $t = 1,\ldots,n$. The density $g(y;\psi)$ can be evaluated by the Kalman filter applied to the linear Gaussian model (9) for some value of χ. The variance of the sample average $\bar{\omega}$ determines the efficiency of the importance sampling likelihood estimate (4).

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 tth 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 $\theta^{(1)},\ldots,\theta^{(S)}$, we define

$$x_s = \log(\omega_s) = \sum_{t=1}^{n} x_{ts}, \qquad s = 1,\ldots,S,$$

where $x_{ts} = \log(\omega_{ts})$, and hence $\omega_{ts} = \exp(x_{ts})$ for $t = 1, \ldots, n$. We can express the sample average of ω_s in terms of $x_s = \log \omega_s$ by means of a Taylor series around some value x, that is

$$\bar{\omega} = \exp(x) \frac{1}{S} \sum_{s=1}^{S} \left(1 + [x_s - x] + \frac{1}{2}[x_s - x]^2 + \ldots\right). \tag{16}$$

We adopt the terms involving x_{ts}, $t = 1, \ldots, n$, 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 clearly leads to a further reduction of the importance sampling variance and to an improvement of the numerical efficiency at a low computational cost.

3.1 Construction of the first new control variable

We base our first control variable on the first order term $(x_s - x)$ of the Taylor series expansion (16). Under the same regularity conditions required for importance sampling, we have

$$\bar{x} = \frac{1}{S} \sum_{s=1}^{S} x_s \xrightarrow{p} \widehat{x}, \tag{17}$$

where $\widehat{x} = \mathbb{E}_g x(\theta, y; \psi)$ and where $\mathbb{E}_g$ is expectation with respect to density $g(\theta|y; \psi)$. The Taylor series expansion (16) around $x = \widehat{x}$ can now be used to construct a first order control variable.

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 time index t separately as discussed in Section 2.4, that is

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

where $\widetilde{\theta}_{tj} = \widehat{\theta}_t + V_t^{1/2} z_j$ and with the numerical evaluation as in (15). The Kalman filter and smoother computes $\widehat{\theta}_t$ and V_t for $t = 1, \ldots, n$. Furthermore, we have $\widehat{x} = \sum_{t=1}^{n} \widehat{x}_t$.

The likelihood estimate (4) corrected for the first control variable is given by

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

It follows from (5) and (17) that $\widehat{L}(y;\psi)_c \xrightarrow{p} L(y;\psi)$. When the importance model (9) provides an accurate approximation to the likelihood, ω_s is close to one and x_s is close to zero, such that $\omega_s \approx 1 + x_s$. Hence ω_s and $\exp(\widehat{x})x_s$ 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 $\widehat{L}(y;\psi)_c$ is a more efficient estimate of the likelihood compared to $\widehat{L}(y;\psi)$.

3.2 Construction of the second new control variable

We base our second control variable on the second order term $(x_s - x)^2$ of the Taylor series expansion (16). 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 $\widehat{x}$. Using the same arguments as in Section 3.1, we have $\bar{\sigma}_t^2 \xrightarrow{p} \widehat{\sigma}_t^2$ where

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

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 $\widehat{L}(y;\psi)_c$.

Monte Carlo and empirical evidence of the importance of the two control variables in reducing the simulation variance of the likelihood estimate obtained from the NAIS method is provided in

Section 4 for the stochastic volatility model. More evidence for other models is presented in the Online Appendix.

4 Monte Carlo and empirical evidence for SV model

4.1 Design of Monte Carlo study

We consider the stochastic volatility (SV) model to illustrate the performance of NAIS in comparison to the alternative importance sampling methods discussed in Section 2, with or without the control variables introduced in Section 3. The SV model may easily be one of the most widely studied nonlinear state space models. The references to some key developments in the SV model literature are Tauchen and Pitts (1983), Taylor (1986) and Melino and Turnbull (1990). Ghysels, Harvey, and Renault (1996) and Shephard (2005) provide detailed reviews of SV models.

For a time series of log-returns y_t, we consider the k-factor univariate stochastic volatility model of Liesenfeld and Richard (2003) and Durham (2006), that is given by

$$y_t \sim \mathrm{N}(0, \sigma_t^2), \qquad \sigma_t^2 = \exp(c + \theta_t), \qquad \theta_t = \alpha_{1,t} + \ldots + \alpha_{k,t}, \qquad t = 1, \ldots, n,$$

with scalar constant c, scalar signal θ_t and $k \times 1$ state vector $\alpha_t = (\alpha_{1,t}, \ldots, \alpha_{k,t})'$. The state vector is modelled by (1) with $k \times k$ diagonal time-invariant matrices T_t and Q_t given by

$$T_t = \begin{bmatrix} \phi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \phi_k \end{bmatrix}, \qquad Q_t = \begin{bmatrix} \sigma_{\eta,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\eta,k}^2 \end{bmatrix},$$

and with unknown coefficients $|\phi_i| < 1$ and $\sigma_{\eta,i}^2 > 0$, for $i = 1, \ldots, k$. We identify the parameters by imposing $\phi_1 > \ldots > \phi_k$. The signal θ_t represents the log-volatility.

For the purpose of likelihood evaluation by importance sampling, we examine the performance of the SPDK method, two implementations of the standard EIS method and four implementations

of our NAIS method. The SPDK method is based on the mode approximation that is computed iteratively using the Kalman filter and smoother (KFS), see Section 2.5. The likelihood function is evaluated using the simulation smoother of de Jong and Shephard (1995) or Durbin and Koopman (2002) which we refer to as JSDK. The EIS method is based on a Monte Carlo approximation to obtain the importance parameter set via the minimisation (13), see Section 2.5. Our proposed NAIS method is introduced in Section 2.4. The likelihood evaluation using EIS and NAIS can both be based on the JSDK simulation smoother or the BF sampler of Appendix A. Finally, both NAIS implementations can be extended with the use of the new control variables introduced in Section 3. Table 1 reviews the methods and their different implementations.

The design of the Monte Carlo study is as follows. We consider 500 random time series for each of the two SV models in our study. The first SV model has a single log-volatility component, $k = 1$, with coefficients

$$c = 1, \quad \phi_1 = 0.98, \quad \sigma^2_{\eta,1} = 0.0225,$$

which are typical values in empirical studies for daily stock returns. The second SV model has $k = 2$ with

$$c = 1, \quad \phi_1 = 0.99, \quad \sigma^2_{\eta,1} = 0.005, \quad \phi_2 = 0.9, \quad \sigma^2_{\eta,2} = 0.03.$$

We take 500 simulated time series to avoid the dependence of our conclusions on particular trajectories of the simulated states and series. For each simulated time series, we estimate the log-likelihood function at the true parameters a hundred times using different common random numbers. Hence each cell in the tables presented below reflects $50,000$ likelihood evaluations. We report the results for two different sample sizes, $n = 1,000$ and $n = 5,000$, and we use $S = 200$ importance samples for each likelihood evaluation and for each method. The number of nodes for the numerical integration calculations is set to $M = 20$.

We start by estimating the variance and the bias associated with each importance sampling

method. We compute the reported statistics as

$$\begin{aligned} \text{Bias} &= \tfrac{1}{50{,}000} \cdot \textstyle\sum_{i=1}^{500} \sum_{j=1}^{100} \left(\log \widehat{L}^j(y^i;\psi) - \log L(y^i;\psi) \right), \\ \text{Variance} &= \tfrac{1}{500} \cdot \textstyle\sum_{i=1}^{500} \tfrac{1}{100} \cdot \sum_{j=1}^{100} \left(\log \widehat{L}^j(y^i;\psi) - \log \overline{L}(y^i;\psi) \right)^2, \end{aligned} \tag{18}$$

where y^i is the ith simulated time series, $\log L(y^i;\psi)$ is the "true" log-likelihood value, $\log \widehat{L}^j(y^i;\psi)$ is the estimate of the log-likelihood function for a particular method and for the jth set of common random numbers, $j = 1, \ldots, 100$, and $\log \overline{L}(y^i;\psi) = 100^{-1} \sum_{j=1}^{100} \log L^j(y^i;\psi)$. The true log-likelihood value is unknown but we take its "true" value as the log of the average of likelihood estimates from the NAIS method with $S = 200 \times 100 = 20,000$ importance samples. We expect the approximation error with respect to the true likelihood to be small. We compute the mean squared error (MSE) as the sum of the variance and the squared bias. The variance and the MSE are reported as a ratios with respect to the EIS-BF method, see Table 1.

The numerical efficiency of estimation via simulation can be increased by generating additional samples. A systematic comparison between methods must therefore take into account computational efficiency. We report the average computing times for each method on the basis of an Intel Duo Core 2.5GHz processor. We record the times required for constructing the importance sampling parameter set (task I) and for jointly generating importance samples and computing the likelihood estimate (task II). In case of NAIS with control variables, we include the additional time for task III in the total time of task II. Our key summary statistic is the time normalised variance ratio of method a against the benchmark method b and it is given by

$$\text{Variance}_{a/b} \times \left(1 + \frac{\text{Time}^a_{\text{I+II}} - \text{Time}^b_{\text{I+II}}}{\text{Time}^a_{\text{II}}} \right)^{-1} \tag{19}$$

where $\text{Variance}_{a/b}$ is the ratio of the variance defined in (18) for methods a and b and Time^m_j is the time length of task j, for $j = \text{I}, \text{II}, \text{I+II}$, by method m, for $m = a, b$. We have excluded Time^a_{I} from the denominator because it is a fixed cost and not relevant for drawing additional samples.

To have the computing times of EIS and NAIS implementations comparable to each other, we initialise the minimisation of (11) by the local approximation for $\{b, C\}$ of the SPDK method. To reduce the simulation variance for all methods, we use antithetic variables for location as in Durbin and Koopman (2000) except for the NAIS methods that include the control variables of Section 3. We have found no evidence of importance sampling weights that constitute an infinite variance in our study, see the discussions in Koopman, Shephard, and Creal (2009). Our diagnostic procedure includes the verification of how sensitive the importance sampling weights are to artificial outliers as in Richard and Zhang (2007). We have efficiently implemented all methods using MATLAB and C.

4.2 Monte Carlo results: log-likelihood estimation

Table 2 presents the results for the stochastic volatility model with $k = 1$. The results show that our numerical integration method for constructing the importance density leads to 40-50% reductions in the variance of the log-likelihood estimates compared to the EIS method. In all cases the numerical gains become larger when we increase the time series dimension from $n = 1,000$ to $n = 5,000$. The use of the control variable further increases the efficiency of the NAIS method as it generates a further 20-35% gain in producing a smaller variance. We find no deterioration in relative performance for the MSE measure, confirming the accuracy of the numerical integration method for obtaining the control variables.

The results further show that the NAIS method can also achieve substantial gains in computational efficiency. It is able to construct the importance density 70% faster than the EIS method. This result is partly due to the ability of working with the marginal densities and the scope for optimising the computer code in this setting. In all cases the EIS method takes longer to construct the importance density than it does to generate the importance samples and to compute the log-likelihood estimate based on $S = 200$ simulation samples. When we normalise the variances by

the length of computing time, we also obtain gains for the NAIS method. Table 2 presents a total improvement of 70-80% by the NAIS method with control variables. The gain in performance is comparable when the default EIS method is compared with the SPDK method. The results further indicate that the computing times for the BF and JSDK simulation methods are equivalent for the SV model with $k = 1$.

Table 3 reports the findings for the SV model with two log-volatility components, $k = 2$. These results are overall similar to the ones for the SV model with $k = 1$ in Table 2. However, we find two important differences. Firstly, for a model with multiple states, as in the SV model with $k = 2$, we achieve a small gain in computational efficiency by switching from the BF sampler to the JSDK sampler. The main reason is that JSDK is able to simulate the univariate signal θ_t directly whereas the BF sampler needs to simulate the signal via the (multiple) state vector α_t. Hence we obtain around 83% reductions in the variance normalised by time when using the NAIS-JSDK method instead of the EIS-BF method.

The second difference between the results for the SV model with $k = 1$ and $k = 2$ concerns the variance reduction that we obtain by using control variables in the likelihood estimation. For the case of $k = 2$, the variance reductions due to the control variables are similar to those for $k = 1$, but the relative increase in the computational cost for step 2 is higher. This is due to the fact that antithetic variables reduce the computational effort for simulation.

4.3 Parameter estimation for the SV model using NAIS

To illustrate the performance of our proposed NAIS method in more detail, we consider the simulated maximum likelihood estimation of the unknown parameters in a multi-factor stochastic volatility model. We report our findings from both a Monte Carlo study and an empirical study below. We carry out parameter estimation using the following consecutive steps:

1. Set a starting value for the estimate of parameter vector ψ.

2. Set starting values for the importance sampling parameter χ in (8).

3. Maximise the log-likelihood function with respect to ψ using the NAIS-JSDK method with $S = 0$ and only the first control variate; use the estimates as starting values for the next step.

4. Maximise the log-likelihood function with respect to ψ using the NAIS-JSDK-Ctrl method with $S > 0$.

The estimation in Step 3 concerns an approximate log-likelihood function. It is fast and requires no simulation, only numerical integration. The computational efficiency of this procedure is primarily due to the accurate approximation of the log-likelihood function calculated by the NAIS-JSDK method with $S = 0$ and the first control variate from Section 3. As a result, the convergence of the maximisation in the last step is fast as it only requires a small number of iterations. The consequence is that we can set S at a high value in Step 4 as it only marginally increases the required computing time. The maximisations in Steps 3 and 4 are based on a quasi-Newton method. The use of common random numbers in Step 4 for each likelihood estimation leads to a smooth likelihood function in ψ, which is necessary for the application of the quasi-Newton optimisation methods. In both studies below we set $S = 200$ in Step 4.

4.3.1 Monte Carlo evidence

We consider the k-factor stochastic volatility model with $k = 3$ for time series of lengths $n = 5,000$ and $n = 10,000$. The true parameter values are set to $d = 0.5$, $\phi_1 = 0.99$, $\phi_2 = 0.9$ $\phi_3 = 0.4$, $\sigma^2_{\eta,1} = 0.005$, $\sigma^2_{\eta,2} = 0.016$, and $\sigma^2_{\eta,3} = 0.05$. We take these parameter values also as the starting values for parameter estimation. We draw 100 different time series. For each realised time series, we obtain 20 parameter estimates under different sets of common numbers and compute the Monte Carlo variance. We report the Monte Carlo standard error as the square root of the average sample variance across the 100 realisations. Since we set the true parameters ourselves, we also

calculate the mean square error (MSE) of the estimates. It allows us to directly compare the relative importance of the simulation and statistical errors in estimating the parameters.

Table 4 summarises the results. The average computing time for estimation is 65 and 123 seconds for $n = 5,000$ and $n = 10,000$, respectively. We learn from Table 4 that despite the complexity of the model and the large sample sizes, the simulation errors in parameter estimates are small both in absolute and in relative terms. For $n = 5,000$ ($n = 10,000$), the Monte Carlo variances represent between 0.1% (0.1%) and 3% (9%) of the total MSEs. Except for the autoregressive coefficient ϕ_3, we observe no relevant deterioration in the Monte Carlo variances for the case in which $n = 10,000$. It emphasises the effectiveness of the NAIS-JSDK method for high-dimensional problems.

4.3.2 Empirical evidence

We investigate whether the NAIS method extends its Monte Carlo performance to real data. Table 5 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. The sample covers 2512 observations. We repeat the estimation process for each series a hundred times with different random numbers.

The empirical results confirm the Monte Carlo results for parameter estimation. For example, the simulation errors in parameter estimates are negligible for persistent state processes, that is, for autoregressive coefficients larger than 0.9. For three of the stocks, parameter estimation has been more challenging because the second volatility factor is weakly persistent and noisy. The Monte Carlo variances of the parameters corresponding to the second factor are approximately 5% of the statistical variance. However, the relatively low estimation times (between 46 and 86 seconds) indicate that we can consider larger importance samples, which enable us to better identify the second volatility component. Figure 2 presents the signal and factor estimates based on parameter estimates from Table 5.

Table 5 also presents the estimated parameters obtained from the EIS-BF method. The estimates obtained from the two methods are identical for almost all parameters. The computation times, however, are not the same. Table 6 reports ratios between the Monte Carlo variances of the estimated parameters obtained from the NAIS-JSDK and EIS-BF methods. We again find that the NAIS-JSDK method leads to substantial relative gains in numerical efficiency. However the differences have only practical relevance for persistent state variables. Substantial reductions in computing time are also achieved by the NAIS-JSDK algorithm.

4.4 Further evidence for other models

We have extended the Monte Carlo study for the SV model as reported above for a range of other dynamic models, including the stochastic conditional duration model of Bauwens and Veredas (2004), the stochastic copula model of Hafner and Manner (2011) and the dynamic one-factor model for multivariate Poisson counts of Jung, Liesenfeld, and Richard (2011). All of these models can be readily handled by the importance methods proposed in this paper. As the main findings do not deviate substantially from the ones reported for the SV model, we provide the Monte Carlo results for these additional models in the Online Appendix.

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 shows that the NAIS method produces reliable results in a numerically and computationally efficient way.

References

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. F. 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 (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* (89), 603–616.

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

Durham, G. B. (2006, July). Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. *Journal of Econometrics 133*(1), 273–305.

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. C. 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 microeconometric panel data. *Journal of Applied Econometrics 23*(3), 373–389.

Jung, R., R. Liesenfeld, and J. F. Richard (2011). Dynamic Factor Models for Multivariate Count Data: An Application to Stock-Market Trading Activity. *Journal of Business and Economic Statistics 29*(1), 73–85.

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. K. van Dijk (1978). Bayesian estimation of equation system parameters: an application by Monte Carlo. *Econometrica 47*, 1–20.

Koopman, S. J., R. Lit, and T. M. Nguyen (2014). Modified Efficient Importance Sampling for State Space Models with an Application. 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. F. 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.

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

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

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. 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

Backward-forward (BF) smoothing and simulation

We present new smoothing and simulation smoothing algorithms that can be regarded as alternatives to the Kalman filter smoother (KFS) and to the JSDK simulation smoother, respectively. The new algorithms directly stem from the Gaussian EIS sampler developed by Jung, Liesenfeld, and Richard (2011). We propose a more efficient formulation, extend it to the more general model (1) and we show how it can be used for computing the smoothed estimates of the state vector. The algorithms are numerically equivalent to the KFS and JSDK, but their computational efficiencies can be different.

We derive the backward-forward (BF) method using the importance model given by (6) and (7). We suppress the dependence of all densities on the parameter vector ψ. The sampler is based on the following decomposition

$$g(\alpha|y) = \prod_{t=1}^{n} g(\alpha_t|\alpha_{t-1}, y) = \prod_{t=1}^{n} g_0(y_t|\theta_t) g(\alpha_t|\alpha_{t-1}) \chi_t(\alpha_{t-1}, b_t, C_t) \zeta_t(\alpha_t, b_{t+1}, C_{t+1}),$$

where $g_0(y_t|\theta_t)$ is defined as in (7) with scalar $a_t = 0$, $g(\alpha_t|\alpha_{t-1})$ is the state vector transition density as implied by (1), and the functions $\chi_t(\cdot)$ and $\zeta_t(\cdot)$ are chosen such that $g(\alpha_t|\alpha_{t-1}, y)$, the individual importance density at time t, integrates to one. For the function $\chi_t(\alpha_{t-1}, b_t, C_t)$, this is accomplished by arguments b_t and C_t. The function

$$\zeta_t(\alpha_t, b_{t+1}, C_{t+1}) \propto 1 \,/\, \chi_{t+1}(\alpha_t, b_{t+1}, C_{t+1}), \tag{20}$$

counters the integration constant for the next period $t+1$. Hence function $\zeta_t(\cdot)$ is of key importance in the backward-forward sampler as it implies that

$$\prod_{t=1}^{n} g(\alpha_t|\alpha_{t-1}, y) \propto \prod_{t=1}^{n} g_0(y_t|\theta_t) g(\alpha_t|\alpha_{t-1}), \tag{21}$$

represents the smoothing density $g(\alpha|y)$ in the same way the KFS applied to the approximating linear state space model (9) represents $g(\alpha|y)$.

Constructing the sampler: We follow Jung, Liesenfeld, and Richard (2011) by adopting an induction argument and defining the integration constant as

$$\chi_t(\alpha_{t-1}, b_t, C_t) = \exp\left\{r_t - q_t' \alpha_{t-1} + \frac{1}{2}\alpha_{t-1}' P_t \alpha_{t-1}\right\}, \tag{22}$$

where scalar r_t, vector q_t and matrix P_t have closed form expressions as functions of parameter vector ψ and the importance parameters b_t and C_t, but do not depend on the state vectors, for example, $P_t = P_t(b_t, C_t; \psi)$. Suppose we have selected $\{b_{t+1}, C_{t+1}\}$ such that (22) holds at time $t+1$. We then use $\zeta_t(\cdot)$ to derive $g(\alpha_t|\alpha_{t-1}, y)$ and verify that it is in the form (20). For the next period $t+1$ we have

$$\zeta_t(\alpha_t, b_{t+1}, C_{t+1}) = \exp\left(q_{t+1}' \alpha_t - \frac{1}{2}\alpha_t' P_{t+1} \alpha_t\right). \tag{23}$$

Hence r_t in (22) ensures that $g(\alpha_t|\alpha_{t-1}, y)$ integrates to one. For a given set $\{b_{t+1}, C_{t+1}\}$, it follows that $g(\alpha_t|\alpha_{t-1}, y)$ is proportional to the product of Gaussian kernels

$$\begin{aligned} g(\alpha_t|\alpha_{t-1}, y) \quad &\propto \; g(y_t|\alpha_t)\zeta_t(\alpha_t, b_{t+1}, C_{t+1})g(\alpha_t|\alpha_{t-1}) \\ &\propto \; \exp\left\{b_t' Z_t\alpha_t - \frac{1}{2}\alpha_t' Z_t' C_t Z_t\alpha_t,\right\} \exp\left\{q_{t+1}' \alpha_t - \frac{1}{2}\alpha_t' P_{t+1} \alpha_t\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\alpha_t - d_{t-1} - T_{t-1}\alpha_{t-1})' Q_{t-1}^{-1}(\alpha_t - d_{t-1} - T_{t-1}\alpha_{t-1})\right\}, \end{aligned} \tag{24}$$

which is a Gaussian kernel as well. The normal distribution with the canonical form $\exp(\delta_t\alpha_t - \alpha_t'\Omega_t\alpha_t/2)$ has mean vector $\mu_t = \Omega_t^{-1}\delta_t$ and variance matrix $\Sigma_t = \Omega_t^{-1}$ for $m \times 1$ vector δ_t and $m \times m$ symmetric positive definite (precision) matrix Ω_t. After some minor manipulation, we can show that the mean vector and variance matrix of $g(\alpha_t|\alpha_{t-1}, y)$ are given by

$$\mu_t = \Sigma_t(Z_t' b_t + Q_{t-1}^{-1}(d_{t-1} + T_{t-1}\alpha_{t-1}) + q_{t+1}), \quad \Sigma_t = (Q_{t-1}^{-1} + Z_t' C_t Z_t + P_{t+1})^{-1}, \tag{25}$$

respectively. Given the expressions for the mean μ_t and the variance Σ_t, it follows immediately that $g(\alpha_t|\alpha_{t-1}, y)$ can be represented by the linear model equation

$$\alpha_t = d^*_{t-1} + T^*_{t-1}\alpha_{t-1} + \eta^*_{t-1}, \qquad \eta^*_t \sim \mathrm{N}(0, \Sigma_t), \qquad t = 2, \ldots, n, \tag{26}$$

where $d^*_{t-1} = \Sigma_t(Z'_t b_t + Q^{-1}_{t-1} d_{t-1} + q_{t+1})$ and $T^*_{t-1} = \Sigma_t Q^{-1}_{t-1} T_{t-1}$. The initial condition is $\alpha_1 \sim \mathrm{N}(d^*_0, \Sigma_1)$. The updating or transition equation (26) for α_t suggests a forward recursion to simulate from $g(\alpha_t|y)$. The implementation is simple and computationally efficient using this construction.

Backward pass: Next, we need to show that (22) holds and how the integration constant $\chi_t(\cdot)$ is computed. In particular, we require an expression for r_t in (22) for the actual calculation of $g(\alpha_t|\alpha_{t-1}, y)$ which is needed to compute the importance weights. The function $\chi_t(\cdot)$ is given by the integration constant for a Gaussian kernel with mean μ_t and variance Σ_t divided by the terms in $g(\alpha_t|\alpha_{t-1})$ that do not depend on α_t:

$$\begin{aligned}\chi_t(\alpha_{t-1}, b_t, C_t) \;=\; & \sqrt{\frac{|Q_{t-1}|}{|\Sigma_t|}} \exp\left(\frac{1}{2}(d_{t-1} + T_{t-1}\alpha_{t-1})' Q^{-1}_{t-1}(d_{t-1} + T_{t-1}\alpha_{t-1})\right) \\ & \times \exp\left(-\frac{1}{2}\mu'_t \Sigma^{-1}_t \mu_t\right).\end{aligned} \tag{27}$$

By substituting μ_t of (25) into (27) and after some manipulation, we obtain $\chi_t(\cdot)$ in the form of (22) with

$$\begin{aligned} P_t \;&= T'_{t-1} Q^{-1}_{t-1} T_{t-1} - T'_{t-1} Q^{-1}_{t-1} \Sigma_t Q^{-1}_{t-1} T_{t-1}, \\ q_t \;&= T'_{t-1} Q^{-1}_{t-1} \Sigma_t (Z'_t b_t + Q^{-1}_{t-1} d_{t-1} + q_{t+1}) - T'_{t-1} Q^{-1}_{t-1} d_{t-1}, \\ r_t \;&= \frac{1}{2}\log(|Q_{t-1}|/|\Sigma_t|) + \frac{1}{2} d'_{t-1} Q^{-1}_{t-1} d_{t-1} \\ &\quad - \frac{1}{2}(Z'_t b_t + Q^{-1}_{t-1} d_{t-1} + q_{t+1})' \Sigma_t (Z'_t b_t + Q^{-1}_{t-1} d_{t-1} + q_{t+1}), \end{aligned} \tag{28}$$

with Σ_t defined in (25), thus completing the induction argument. We evaluate r_t, q_t and P_t by a backward recursion that we initialise with $\zeta_n(\cdot) = 1$. It implies that we can initialise the recursion (28) by $q_{n+1} = 0$ and $P_{n+1} = 0$.

Simulation smoothing (forward pass): For a given set of importance parameters, we determine $g(\alpha_t|\alpha_{t-1}, y)$ by the backward recursion (28); the order of computations are Σ_t, P_t, q_t and r_t. We use d^*_{t-1}, T^*_{t-1} and Σ_t computed during the backward pass for simulating from $g(\alpha|y)$ by means of the forward pass (26).

Evaluation of likelihood function for the importance density model: The simulated likelihood estimate is given by (4) where it follows from the results above that

$$g(y) = \prod_{t=1}^{n} \exp(-r_t). \tag{29}$$

State smoothing (forward pass): For the implementation of the NAIS method using the backward-forward method, we require the smoothed estimate of $\theta_t = Z_t\alpha_t$ from the Gaussian importance density $g(\alpha_t|y)$. We obtain the smoothed mean $\widehat{\alpha}_t = \mathrm{E}(\alpha_t|y)$ and variance $V_t = \mathrm{Var}(\alpha_t|y)$, for $t = 1, \ldots, n$, by the forward pass

$$\widehat{\alpha}_t = d^*_{t-1} + T^*_{t-1}\widehat{\alpha}_{t-1}, \qquad V_t = T^*_{t-1}V_{t-1}T^{*\prime}_{t-1} + \Sigma_t, \qquad t = 2, \ldots, n, \tag{30}$$

with initialisation $\widehat{\alpha}_1 = d^*_0$ and $V_1 = \Sigma_1$. The recursion for V_t follows immediately from (26).

Table 1: IMPORTANCE SAMPLING METHODS FOR STATE SPACE MODELS.
Likelihood evaluation via importance sampling: I, obtain importance sampling (IS) parameter set $\{b, C\}$ using methods described in Sections 2.4 and 2.5; II, sampling from the importance density to construct likelihood function; III, in case of NAIS, with use of control variables in Section 3 or not.

	Task	SPDK	EIS-BF	EIS-JSDK	NAIS-BF (-Ctrl)	NAIS-JSDK (-Ctrl)
I	IS parameter set $\{b, C\}$ via	local mode	(13)	(13)	(13)	(13)
	constructed by simulation	—	✓	✓	—	—
	or by numerical integration	—	—	—	✓	✓
	using method	KFS	BF	JSDK	BF	KFS
II	Simulation method	JSDK	BF	JSDK	BF	JSDK
III	Control variables (optional)	—	—	—	(✓)	(✓)

Table 2: Log-likelihood Errors for Stochastic Volatility model, $k = 1$.
The table presents the numerical and computational performance of different IS methods for log-likelihood estimation. We simulate 500 different realisations from the model. For each of these realisations, we obtain log-likelihood estimates for 100 different sets of random numbers. We estimate the variance associated with each method as the average sample variance across the 500 realisations. We define the mean-square error (MSE) as the sum of the variance and the square of the average bias across the 500 realisations. We show these statistics as ratios with the standard implementation of the EIS-BF method as the benchmark. The time for step 1 column gives the fixed time cost for obtaining the parameters of the importance density, while the time for step 2 refers to the computational cost of sampling from the importance density and calculating the likelihood estimate. The TNVAR column reports the time normalised variance ratio according to (19). Table 1 lists the methods used and their acronyms. NAIS-BF-Ctrl and NAIS-JSDK-Ctrl refer to the NAIS methods with use of the control variables of Section 3. We specify the stochastic volatility model as: $y_t \sim N(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_{t+1} = 0.98\alpha_t + \eta_t$ where $\eta_t \sim N(0, \sigma_\eta^2 = 0.0225)$ for $t = 1, \ldots, n$.

$n = 1000, S = 200$	Variance	MSE	Time Step 1 (×10)	Time Step 2 (×10)	TNVAR
SPDK	12.779	12.837	0.022	0.173	5.576
EIS-BF	1.000	1.000	0.232	0.187	1.000
EIS-JSDK	1.009	1.009	0.247	0.172	1.005
NAIS-BF	0.595	0.595	0.073	0.174	0.299
NAIS-JSDK	0.594	0.594	0.070	0.175	0.297
NAIS-BF-Ctrl	0.405	0.406	0.073	0.192	0.224
NAIS-JSDK-Ctrl	0.415	0.416	0.073	0.180	0.216

$n = 5000, S = 200$	Variance	MSE	Time Step 1 (×10)	Time Step 2 (×10)	TNVAR
SPDK	15.025	18.594	0.052	0.908	6.152
EIS-BF	1.000	1.000	1.278	0.990	1.000
EIS-JSDK	0.997	0.998	1.246	0.908	0.885
NAIS-BF	0.503	0.501	0.364	0.928	0.245
NAIS-JSDK	0.501	0.499	0.340	0.908	0.236
NAIS-BF-Ctrl	0.380	0.381	0.418	1.002	0.206
NAIS-JSDK-Ctrl	0.375	0.376	0.398	0.939	0.189

Table 3: Log-likelihood Errors for Stochastic Volatility model, $k = 2$.

We refer to the description of Table 2. We specify the two-factor stochastic volatility model as: $y_t \sim N(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_{t,1} + \alpha_{t,2})$, $\alpha_{t+1,1} = 0.99\alpha_{t,1} + \eta_{t,1}$, $\alpha_{t+1,1} = 0.9\alpha_{t,2} + \eta_{t,2}$, where $\eta_{t,1} \sim N(0, \sigma_{\eta,1}^2 = 0.005)$, $\eta_{t,2} \sim N(0, \sigma_{\eta,2}^2 = 0.03)$, for $t = 1, \ldots, n$.

$n = 1000$, $S = 200$	Variance	MSE	Time Step 1 (×10)	Time Step 2 (×10)	TNVAR
SPDK	17.523	17.539	0.223	0.238	4.572
EIS-BF	1.000	1.000	0.763	0.374	1.000
EIS-JSDK	1.011	1.001	0.578	0.314	0.569
NAIS-BF	0.562	0.564	0.569	0.304	0.301
NAIS-JSDK	0.562	0.554	0.364	0.237	0.172
NAIS-BF-Ctrl	0.374	0.378	0.592	0.408	0.280
NAIS-JSDK-Ctrl	0.374	0.373	0.385	0.307	0.153

$n = 5000$, $S = 200$	Variance	MSE	Time Step 1 (×10)	Time Step 2 (×10)	TNVAR
SPDK	18.915	26.283	1.341	1.293	5.305
EIS-BF	1.000	1.000	3.988	1.962	1.000
EIS-JSDK	0.972	0.963	3.067	1.673	0.564
NAIS-BF	0.483	0.484	3.010	1.619	0.266
NAIS-JSDK	0.474	0.465	1.976	1.298	0.155
NAIS-BF-Ctrl	0.368	0.371	3.089	2.152	0.277
NAIS-JSDK-Ctrl	0.365	0.363	2.050	1.655	0.155

Table 4: Parameter Estimation for Stochastic Volatility model, $k = 3$.
We simulate 100 trajectories of a three-factor stochastic volatility model. For each of these realisations, we obtain 20 simulated maximum likelihood parameter estimates based on different sets of common random numbers and using the NAIS-JSDK method. We first show the average parameter estimates across the 500 replications. The Monte Carlo (MC) standard error column reports the square root of the average of the sample variance of the parameter estimates across the 100 realisations. The total standard error column shows the square-root of the sum of the MC variance and the variance of the average estimates across the 100 trajectories. Finally, we obtain the MC variance ratio by dividing the MC variance by the total variance. The starting values are the true parameters. The average computing time was 65s and 123s for $n = 5,000$ and $n = 10,000$ respectively. $y_t \sim N(0, \sigma_t^2)$, $t = 1, \ldots, n$, $\sigma_t^2 = \exp(\theta_t)$, $\theta_t = d + \alpha_{1,t} + \alpha_{2,t} + \alpha_{3,t}$, $\alpha_t = T\alpha_{t-1} + \eta_t$, $\alpha_1 \sim \mathrm{N}(a_1, P_1)$, $\eta_t \sim \mathrm{N}(0, Q)$, where T is a diagonal matrix with elements ϕ_1, ϕ_2 and ϕ_3 and Q is a diagonal matrix with elements $\sigma^2_{\eta,1}$, $\sigma^2_{\eta,2}$ and $\sigma^2_{\eta,3}$.

			$n = 5,000$		
	True	Mean	MC Std. Error	Total Std. Error	MC Variance Ratio
c	0.500	0.489	0.003	0.098	0.001
ϕ_1	0.990	0.989	0.000	0.005	0.010
$\sigma^2_{1,\eta}$	0.005	0.005	0.000	0.002	0.007
ϕ_2	0.900	0.883	0.009	0.055	0.029
$\sigma^2_{2,\eta}$	0.015	0.014	0.001	0.007	0.009
ϕ_3	0.400	0.314	0.025	0.224	0.012
$\sigma^2_{3,\eta}$	0.050	0.054	0.003	0.032	0.006

			$n = 10,000$		
	True	Mean	MC Std. Error	Total Std. Error	MC Variance Ratio
c	0.500	0.500	0.003	0.078	0.001
ϕ_1	0.990	0.990	0.000	0.003	0.008
$\sigma^2_{1,\eta}$	0.005	0.005	0.000	0.002	0.010
ϕ_2	0.900	0.893	0.005	0.034	0.021
$\sigma^2_{2,\eta}$	0.015	0.015	0.001	0.005	0.046
ϕ_3	0.400	0.334	0.061	0.205	0.090
$\sigma^2_{3,\eta}$	0.050	0.054	0.004	0.023	0.035

Table 5: Two-Factor Stochastic Volatility Model: Empirical study.

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-JSDK and EIS-BF 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 $y_t \sim N(0, \sigma_t^2)$, $t = 1, \ldots, n$, $\sigma_t^2 = \exp(\theta_t)$, $\theta_t = c + \alpha_{1,t} + \alpha_{2,t}$, $\alpha_t = T\alpha_{t-1} + \eta_t$, $\alpha_1 \sim \mathrm{N}(a_1, P_1)$, $\eta_t \sim \mathrm{N}(0, Q)$, where T is a diagonal matrix with elements ϕ_1, and ϕ_2 and Q is a diagonal matrix with elements $\sigma^2_{\eta,1}$, and $\sigma^2_{\eta,2}$.

	GE		JP Morgan		Coca-Cola		AT&T		Wal-Mart		Exxon	
	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS	NAIS	EIS
c	0.845	0.845	1.275	1.274	0.004	0.005	0.655	0.655	0.375	0.374	0.555	0.555
	[0.002]	[0.006]	[0.003]	[0.005]	[0.004]	[0.007]	[0.001]	[0.002]	[0.002]	[0.003]	[0.000]	[0.001]
	(0.690)		(0.458)		(0.393)		(0.652)		(0.413)		(0.193)	
ϕ_1	0.997	0.997	0.997	0.997	0.995	0.995	0.997	0.997	0.996	0.996	0.991	0.991
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)		(0.002)		(0.003)		(0.002)		(0.002)		(0.006)	
$\sigma^2_{1,\eta}$	0.007	0.007	0.008	0.008	0.006	0.006	0.005	0.005	0.005	0.005	0.008	0.008
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
	(0.003)		(0.003)		(0.003)		(0.003)		(0.002)		(0.008)	
ϕ_2	0.458	0.455	0.839	0.840	0.488	0.484	0.897	0.898	0.275	0.274	0.936	0.936
	[0.016]	[0.042]	[0.004]	[0.006]	[0.037]	[0.076]	[0.002]	[0.004]	[0.013]	[0.028]	[0.000]	[0.002]
	(0.271)		(0.065)		(0.164)		(0.081)		(0.118)		(0.048)	
$\sigma^2_{2,\eta}$	0.261	0.261	0.086	0.085	0.307	0.305	0.027	0.027	0.263	0.262	0.026	0.026
	[0.008]	[0.017]	[0.002]	[0.003]	[0.017]	[0.029]	[0.000]	[0.001]	[0.007]	[0.009]	[0.000]	[0.001]
	(0.091)		(0.032)		(0.077)		(0.018)		(0.054)		(0.013)	
Time (seconds)	86	244	67	196	69	146	74	212	36	73	46	86
Improvement	-65%		-66%		-52%		-65%		-51%		-47%	

ACCEPTED MANUSCRIPT

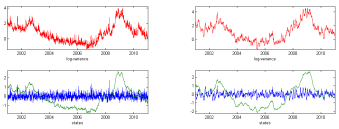
Table 6: Variance Reductions in Parameter Estimation.

The table presents the ratios between the Monte Carlo variances of the NAIS-JSDK method against the EIS-BF method in Table 4.

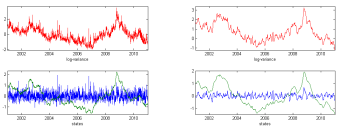
	GE	JPM	KO	T	WMT	XOM
c	0.127	0.482	0.293	0.482	0.160	0.559
ϕ_1	0.057	0.068	0.089	0.068	0.066	0.109
$\sigma^2_{1,\eta}$	0.051	0.052	0.090	0.052	0.042	0.118
ϕ_2	0.148	0.408	0.237	0.408	0.184	0.216
$\sigma^2_{2,\eta}$	0.207	0.492	0.324	0.492	0.266	0.629

Figure 1: NAIS algorithm for selecting the importance parameter set $\{b, C\}$.

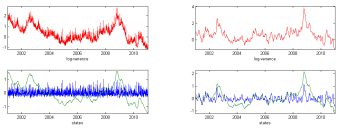
- Set value for M and obtain Gauss-Hermite nodes z_j with weights $h(z_j)$, $j = 1, \ldots, M$.
- Set $k = 0$ and select initial importance parameter set $\{b^{[0]}, C^{[0]}\}$.
- Repeat until convergence:
  - Set $k \leftarrow k + 1$
  - Compute mean $\widetilde{\theta}_t$ and variance V_t, for $t = 1, \ldots, n$, based on $\{b^{[k-1]}, C^{[k-1]}\}$ and linear state space model (9) using KFS or BF.
  - For each t :
    - Compute $\widetilde{\theta}_{tj} = \widetilde{\theta}_t + V_t^{1/2} z_j$ for $j = 1, \ldots, M$.
    - Weighted least squares regression for $\log p(y_t|\widetilde{\theta}_{tj}; \psi) = (1, \widetilde{\theta}_{tj}, -0.5\, \widetilde{\theta}_{tj}^2)\beta + \text{error}_{tj}$ with weight $\exp(\frac{1}{2} z_j^2)\, h(z_j)$ – fast version – or weight $\exp(\frac{1}{2} z_j^2)\, h(z_j) \omega(\widetilde{\theta}_{tj}, y_t; \psi)$.
    - Set $b_t^{[k]}$ and $C_t^{[k]}$ equal to second and third element of the regression estimate of β.
- Convergence when $n^{-1} \sum_{t=1}^{n} (b_t^{[k-1]} - b_t^{[k]})^2 < \epsilon$ and $n^{-1} \sum_{t=1}^{n} (C_t^{[k-1]} - C_t^{[k]})^2 < \epsilon$.



(a) General Electric (b) JP Morgan



(c) Coca-Cola (d) AT&T



(e) Wal-Mart (f) Exxon

Figure 2: Estimated log-variance (top) and states (bottom) for six Dow Jones stocks.