

Kleppe, Tore Selland; Liesenfeld, Roman

**Working Paper**

## Efficient high-dimensional importance sampling in mixture frameworks

Economics Working Paper, No. 2011-11

**Provided in Cooperation with:**

Christian-Albrechts-University of Kiel, Department of Economics

*Suggested Citation:* Kleppe, Tore Selland; Liesenfeld, Roman (2011) : Efficient high-dimensional importance sampling in mixture frameworks, Economics Working Paper, No. 2011-11, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/53125>

**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.*

# efficient high-dimensional importance sampling in mixture frameworks

by Tore Selland Kleppe and Roman Liesenfeld



# Efficient high-dimensional importance sampling in mixture frameworks

Tore Selland Kleppe\*      Roman Liesenfeld†

June 8, 2011

## Abstract

This paper provides high-dimensional and flexible importance sampling procedures for the likelihood evaluation of dynamic latent variable models involving finite or infinite mixtures leading to possibly heavy tailed and/or multi-modal target densities. Our approach is based upon the efficient importance sampling (EIS) approach of Richard and Zhang (2007) and exploits the mixture structure of the model when constructing importance sampling distributions as mixture of distributions. The proposed mixture EIS procedures are illustrated with ML estimation of a student- $t$  state space model for realized volatilities and a stochastic volatility model with leverage effects and jumps for asset returns.

**Keywords:** Dynamic latent variable model; Importance sampling; Marginalized likelihood; Mixture; Monte Carlo; Realized Volatility; Stochastic volatility.

**JEL classification codes:** C15

---

\*Department of Mathematics, University of Bergen, Bergen, Norway, (Corresponding author, email: [tore.kleppe@math.uib.no](mailto:tore.kleppe@math.uib.no), phone: +4795922364, fax: +4755589672, address: Universitetet i Bergen, Matematisk institutt, Postboks 7803, 5020 Bergen, Norway)

†Department of Economics, Christian Albrechts Universität, Kiel, Germany

# 1 Introduction

In recent decades Monte Carlo (MC) procedures based upon Importance Sampling (IS) have been successfully applied for the analysis of econometric models involving multiple integrals for which no analytical solutions exist. Important applications of IS are the evaluation of Bayesian a-posteriori expectations of functions of parameters of interest and that of likelihood functions in the presence of unobservable latent variables – see, e.g., Kloek and Dijk (1978); Geweke (1989); Durbin and Koopman (1997).

It is well-known that the reliable and efficient use of IS requires that the IS auxiliary density closely mimics the target density kernel which needs to be integrated, and exhibits tails that do not decay more quickly than the tails of the target density (see, Geweke (1989); Robert and Casella (2004)). This implies that IS implementations have to be tailored to the problem under consideration which has proved to be a significant obstacle to routine applications of IS. This is especially true for applications with ill behaved, and therefore, difficult to approximate target densities. A survey of IS approaches is found, e.g., in Liesenfeld and Richard (2001).

Another critical issue is that most of the existing IS approaches do not appear to be applicable to highly multidimensional integration problems. Prominent exceptions are the high-dimensional IS methods proposed by Shephard and Pitt (1997); Durbin and Koopman (1997, 2000), and the Efficient Importance Sampling (EIS) procedure of Richard and Zhang (2007). Existing implementations of those methods rely on IS densities from the exponential family of distributions, which, in the case of EIS, considerably simplifies the implementation. While the IS approaches of Shephard and Pitt (1997) and Durbin and Koopman (1997, 2000) use Gaussian IS densities constructed from local Taylor-series approximations to the target density, the IS densities of the EIS approach of Richard and

Zhang (2007) are based upon global approximations to the target obtained via a sequence of low-dimensional auxiliary least-squares regressions.

Those high-dimensional IS approaches have been successfully applied for the computation of the likelihood for a broad range of dynamic latent variable (DLV) models, where the target densities are reasonably well behaved such that they can be well approximated by IS densities from the exponential family of distributions (see, e.g., Sandmann and Koopman (1998); Liesenfeld and Richard (2003, 2010); Bauwens and Galli (2009)). However, for DLV models with pathological target densities, featuring e.g. multi-modality and/or fat tails, those high-dimensional procedures based upon IS densities from the exponential family might have severe convergence problems. In the context of DLV models, such ill-behaved target densities are often caused by finite or infinite mixture-of-distributions specifications assumed for some of the variables. Examples, to be discussed further below are diffusion models for stock prices with discrete jumps and state space models with student- $t$  measurement errors. Hence, there exists a need for high-dimensional IS procedures based upon flexible IS densities beyond the exponential family of distributions.

In the present paper, we extend the high-dimensional EIS approach of Richard and Zhang (2007), by introducing finite and infinite mixture of distributions as flexible classes of IS distributions allowing to approximate target densities which are possibly heavy-tailed and/or multi-modal. Our approach is particularly well adapted to the likelihood evaluations for DLV models involving variables characterized by a mixture of distributions, which can be exploited when constructing the IS densities. Under appropriate simplifying conditions our proposed mixture EIS procedures rely, similarly to EIS implementations for IS densities from the exponential family, on a simple sequence of auxiliary least-squares regressions used to obtain close approximations to the integrand.

Alternative IS procedures using flexible mixtures of distributions as IS

densities are the ‘split’-Student IS approach of Geweke (1989), the ‘defensive’ mixture technique proposed by Hesterberg (1995) and the adaptive method of Ardia et al. (2009) using mixture of Student-t distributions. While those approaches have been successfully applied to lower dimensional Bayesian integration problems, they do not appear to be applicable to very high-dimensional integrals, which needs to be approximated, e.g., for the likelihood evaluation of DLV models.

The rest of the paper is organized as follows. In Section 2, we briefly review the generic principles of EIS, and in Section 3 we introduce mixture EIS approaches based upon finite and infinite mixtures of distributions. Section 3 also provides two simple one-dimensional examples illustrating the benefits of the proposed mixture EIS procedures. Sequential implementation of mixture EIS procedures for high-dimensional integration required for a likelihood analysis of realistic models is illustrated in Section 4. In particular, we discuss the ML estimation of a student- $t$  state space model for realized volatilities (Section 4.1) and a stochastic volatility model with a jump component for asset returns (Section 4.2). Section 5 concludes.

## 2 Efficient importance sampling (EIS)

### 2.1 General principle

Consider the problem of evaluating an integral of the form

$$I = \int \varphi(z) dz, \quad (1)$$

where  $\varphi : \Delta \mapsto \mathbb{R}^+$  with  $\Delta \subseteq \mathbb{R}^T$ . Of special interest in this paper focusing on likelihood evaluations of DLV models is the case where the econometric model under consideration leads to an initial factorization of

the integrand of the form

$$\varphi(z) = g(z) \cdot p(z), \quad (2)$$

where  $p$  is a probability density function for  $z$  referred to as the initial or natural (model based) sampler, and  $g : \Delta \mapsto \mathbb{R}^+$  is a  $p$ -integrable function.

IS integration consists of selecting an IS density, say  $m(z)$ , and rewriting Equation (1) as

$$I = \int \left[ \frac{\varphi(z)}{m(z)} \right] \cdot m(z) dz. \quad (3)$$

The corresponding MC IS estimator of  $I$  is then given by

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M \frac{\varphi(z^{(j)})}{m(z^{(j)})}, \quad (4)$$

where  $\{z^{(j)}, j = 1, \dots, M\}$  denotes a set of  $M$  identically independently distributed draws from  $m$ .

The technical conditions, under which the IS estimator (4) converges almost surely to  $I$  and its variance is finite are discussed in Geweke (1989) and Robert and Casella (2004). A sufficient condition for the finiteness the variance of  $\hat{I}$  is that  $|\varphi(z)/m(z)|$  be bounded above on  $\Delta$ .

Efficient IS algorithms are those for which the variance of  $\hat{I}$  is as small as possible. This suggests to select an IS density  $m$ , which mimics the target integrand  $\varphi(z)$  as close as possible such that the ratio  $\varphi(z)/m(z)$  is almost constant on  $\Delta$ . The EIS approach of Richard and Zhang (2007) provides an algorithm to select such an efficient sampler within a preselected parametric class of densities, say  $\mathcal{M} = \{m(z; a), a \in A\}$ , indexed by the auxiliary parameter  $a$ . For lower dimensional problems, it approximates the target  $\varphi(z)$  by the density kernel  $k(z, a)$  associated with density  $m(z; a)$ , where the correspondence between  $k$  and  $m$  is given by

$$m(z; a) = \frac{k(z, a)}{\chi(a)} \quad \text{with} \quad \chi(a) = \int k(z, a) dz. \quad (5)$$

The near optimal value  $\hat{a}$  obtains as the solution to the least-squares problem

$$\{\hat{a}, \hat{c}\} = \arg \min_{a, c} \int [\ln \varphi(z) - c - \ln k(z, a)]^2 m(z; a) dz, \quad (6)$$

where  $c$  represents an intercept meant to calibrate the log-ratio  $\ln(\varphi/k)$ . Since the IS sampling density  $m$  itself depends upon  $a$ , the optimization problem (6) is reinterpreted as the search for a fixed-point solution with an operational MC version based upon the following step-wise recursion:

Given  $\hat{a}^{(\ell)} \in A$ , generate intermediate draws  $\{z^{(j, \ell)}, j = 1, \dots, M\}$  from the step- $\ell$  EIS sampler  $m(z; \hat{a}^{(\ell)})$  and solve

$$\{\hat{a}^{(\ell+1)}, \hat{c}^{(\ell+1)}\} = \arg \min_{a, c} \sum_{j=1}^M [\ln \varphi(z^{(j, \ell)}) - c - \ln k(z^{(j, \ell)}, a)]^2. \quad (7)$$

At convergence, whereby  $\hat{a} \simeq \hat{a}^{(\ell+1)} \simeq \hat{a}^{(\ell)}$ , the density  $m(z; \hat{a})$  is used to compute the EIS estimate of  $I$  according to Equation (4). An initial value  $\hat{a}^{(0)}$  can be obtained by a local approximation of  $\varphi$  or by the value of  $a$  associated with the natural sampler  $p$ . In order to secure smooth convergence to a fixed-point solution, all draws  $\{z^{(j, \ell)}\}$  must be obtained by transformation of a set of Common Random Numbers (CRNs)  $\{u^{(j)}\}$  from a canonical distribution associated with  $\mathcal{M}$ , i.e. one that does not dependent on  $a$  (e.g. a standardized Normal, when  $m$  is Gaussian).

Although the EIS approach is not restricted to any particular family of IS densities  $m$ , its existing implementations typically rely upon densities from the exponential family. In this case, the log kernel  $\ln k$  can be parameterized in such a way that it is linear in the auxiliary parameter  $a$  and the minimization problem (7) reduces to a computationally simple linear LS problem. If, in addition, the initial sampler  $p$  belongs to the exponential family and  $\mathcal{M}$  is selected to be a parametric class containing  $p$  as a particular member, one can exploit the property of the exponential family that a parametric class  $\mathcal{M}$  is closed under multiplication. This



allows one to define the EIS density kernel  $k$  as

$$k(z, a) = p(z)\zeta(z, a), \quad (8)$$

where  $\zeta(z, a)$  is itself a density kernel associated with  $m$  (for details, see Richard and Zhang (2007)). It follows that  $p(z)$  cancels out in the regression (7) which simplifies into a regression of  $\ln g(x)$  on a constant and  $\ln \zeta(z, a)$ .

## 2.2 Sequential high-dimensional EIS

High-dimensional EIS used, e.g., for the likelihood evaluation of DLV models, requires that the integrand  $\varphi(z) = g(z)p(z)$  in Equations (1) and (2) as well as the EIS density  $m(z; a)$  be factorized into low-dimensional components in accordance with a natural pre-ordering partition of  $z$  obtained from the model specification, say  $z_1, \dots, z_T$ , with an initial condition  $z_0$ . The factorizations of  $\varphi$  and  $m$  conformably with  $z$  are

$$\varphi(z) = \prod_{t=0}^T \varphi_t(z_{(t)}) = p_0(z_0) \prod_{t=1}^T g_t(z_{(t)}) p_t(z_t | z_{(t-1)}), \quad (9)$$

$$m(z; a) = \prod_{t=0}^T m_t(z_t | z_{(t-1)}; a_t), \quad (10)$$

where  $z_{(t)} = (z_0, \dots, z_t)$ ,  $a = (a_0, \dots, a_T)$  and  $m_t(z_0 | z_{(-1)}; a_0) = m_0(z_0; a_0)$ .

The natural model based sampling density of  $z$  is the product  $p(z) = p_0(z_0) \prod_{t=1}^T p_t(z_t | z_{(t-1)})$ . Denoting the approximating kernel of the IS density  $m_t(z_t | z_{(t-1)}; a_t)$  by  $k_t(z_{(t)}, a_t)$ , with

$$m_t(z_t | z_{(t-1)}; a_t) = \frac{k_t(z_{(t)}, a_t)}{\chi_t(z_{(t-1)}, a_t)}, \quad \text{and} \quad (11)$$

$$\chi_t(z_{(t-1)}, a_t) = \int k_t(z_{(t)}, a_t) dz_t,$$

the integral to be approximated can be rewritten as

$$I = \chi_0(a_0) \int \left[ \prod_{t=0}^T \frac{\varphi_t(z_{(t)}) \chi_{t+1}(z_{(t)}, a_{t+1})}{k_t(z_{(t)}, a_t)} \right] \prod_{t=0}^T m_t(z_t | z_{(t-1)}; a_t) dz_{(T)}, \quad (12)$$

where  $\chi_{T+1}(\cdot) \equiv 1$ .

EIS then aims at selecting values of the auxiliary parameters  $\{a_t\}$  that minimize factor by factor the variance of the ratios  $\varphi_t \cdot \chi_{t+1}/k_t$  as functions of  $z_{(t)}$  with respect to the  $m_t$ -distributions. An operational MC version of that fixed-point minimization problem generates at step  $\ell$  intermediate draws  $\{z_t^{(j,\ell)}; t = 0, \dots, T; j = 1, \dots, M\}$  from the step- $\ell$  EIS sampler  $m(z; \hat{a}^{(\ell)})$  and solves back recursively the following sequence of  $T + 1$  least-squares problems

$$\{\hat{a}_t^{(\ell+1)}, \hat{c}_t^{(\ell+1)}\} = \arg \min_{a_t, c_t} \sum_{j=1}^M \left[ \ln \{ \varphi_t(z_{(t)}^{(j,\ell)}) \chi_{t+1}(z_{(t)}^{(j,\ell)}, \hat{a}_{t+1}^{(\ell+1)}) \} \right. \\ \left. - c_t - \ln k_t(z_{(t)}^{(j,\ell)}, a_t) \right]^2, \quad t = T, T-1, \dots, 0. \quad (13)$$

The corresponding MC-EIS estimate of the integral  $I$  is given by

$$\hat{I} = \chi_0(a_0) \frac{1}{M} \sum_{j=1}^M \left[ \prod_{t=0}^T \frac{\varphi_t(z_{(t)}^{(j)}) \chi_{t+1}(z_{(t)}^{(j)}, \hat{a}_{t+1})}{k_t(z_{(t)}^{(j)}, \hat{a}_t)} \right], \quad (14)$$

where  $\{z_t^{(j)}; t = 0, \dots, T; j = 1, \dots, M\}$  are draws from the EIS density  $m(z; \hat{a})$ .

This sequential high-dimensional EIS approach has been successfully applied for likelihood evaluations for DLV models, where the natural sampling densities  $p_t$  in Equation (9) belong to the exponential family and the  $g_t$ s are well-behaved functions in  $z_t$ . In this case,  $\varphi_t \chi_{t+1}$  can be reasonably well approximated by exponential family density kernels  $k_t$  obtained as parametric extensions of the natural sampler  $p_t$  according to Equation (8). As mentioned above, the use of EIS densities from the exponential family considerably reduces the computational effort and simplifies the

implementation of EIS by reducing it to the solution of linear LS problems. Since the number of auxiliary LS problems is proportional to the dimension of the integral, this linearity becomes particularly desirable for very high-dimensional integration ( $1000^+$ ).

Applications of EIS in such well-behaved frameworks include the computation of the likelihood and ML estimates for stochastic volatility models (Liesenfeld and Richard (2003); Kleppe et al. (2010)), stochastic conditional intensity processes (Bauwens and Galli (2009)), multinomial-multiperiod Probit models (Liesenfeld and Richard (2010)) and dynamic factor models for multivariate count data (Jung et al. (2011)).

In many practical applications, however, DLV model specifications with natural sampling densities  $p$  within the exponential family appear to be too restrictive calling for more flexible alternatives. A straightforward extension of simple distributions from the exponential family is the class of mixture distributions providing a flexible environment for econometric modeling which enjoys great popularity. However, DLV models with natural sampling distributions  $p$  assumed to be finite or infinite mixtures of distributions lead to possibly multi-modality and/or fat tailed targets  $\varphi$ . In such cases, the performance of (E)IS based upon sampling densities  $m$  from the exponential family might be seriously hampered by a large or even infinite variance of the IS estimator  $\hat{I}$  resulting from poor global approximations of  $\varphi$  by exponential density kernels  $k$ .

In the following section, we present algorithms to construct flexible EIS densities for finite and infinite mixture frameworks generating possibly ill-behaved target integrands. The proposed algorithms exploit the mixture specification of the target density and use a data-augmentation step demarginalizing the mixture density to include the mixing variable as an additional latent variable. As we shall illustrate below, this artificial extension of the space of integration typically leads to better behaved target integrands (for the extended space), which can be well approximated

by standard distributions belonging to the exponential family.

### 3 EIS in mixture frameworks

In this section we introduce EIS in infinite mixture (subsection 3.1) as well as in finite mixture frameworks (subsection 3.2) focusing on simple one-dimensional integration problems. This provides a convenient setting to discuss and illustrate the key features of mixture EIS. Sequential implementations for high-dimensional integration required to analyze realistic models, will be discussed in Section 4.

#### 3.1 Infinite mixture EIS

Assume that the natural sampling density of the integral given by Equations (1) and (2) can be represented as an infinite mixture specification of the form

$$p(z) = \int p(z|w)p(w)dw, \quad (15)$$

where  $p(z|w)$  is a conditional density depending on a latent auxiliary variable  $w$  with density  $p(w)$ , referred to as the mixing density. Prominent examples for infinite mixture density functions include the student- $t$  and the negative binomial density.

Such a representation of the natural sampling density allows us to write the integral (1) as

$$I = \int g(z)p(z|w)p(w)dw dz. \quad (16)$$

Based upon this demarginalization which extends the space of integration, EIS can be implemented using a joint IS density for the augmented set of variables, say  $m(z, w; a)$ , and exploiting the factorization of the natural sampler given by  $p(z|w)p(w)$ . The factorization of the joint IS density conformably with that of the natural sampler is  $m(z, w; a) =$

$m(z|w; a_z)m(w; a_w)$ , with

$$m(z|w; a_z) = \frac{k(z, w; a_z)}{\chi(w, a_z)}, \quad m(w; a_w) = \frac{k(w; a_w)}{\chi(a_w)}, \quad (17)$$

where the  $k$ s and  $\chi$ s are the approximating EIS density kernels and the corresponding integrating factors, respectively. This allows us to rewrite the integral (16) as

$$I = \chi(a_w) \int \left[ \frac{g(z)p(z|w)}{k(z, w; a_z)} \right] \left[ \frac{\chi(w, a_z)p(w)}{k(w; a_w)} \right] m(z, w; a) dw dz. \quad (18)$$

Then the EIS values of the auxiliary parameters,  $\hat{a} = (\hat{a}_z^{(\ell+1)}, \hat{a}_w^{(\ell+1)})$ , obtain from the recursive solution of the following two LS problems

$$\begin{aligned} \min_{a_z, c_z} \sum_{j=1}^M \left[ \ln[g(z^{(j, \ell)})p(z^{(j, \ell)}|w^{(j, \ell)})] - c_z - \ln k(z^{(j, \ell)}, w^{(j, \ell)}, a_z) \right]^2 \\ \min_{a_w, c_w} \sum_{j=1}^M \left[ \ln[\chi(w^{(j, \ell)}, \hat{a}_z^{(\ell+1)})p(w^{(j, \ell)})] - c_w - \ln k(w^{(j, \ell)}, a_w) \right]^2, \end{aligned} \quad (19)$$

where  $\{(z^{(j, \ell)}, w^{(j, \ell)}), j = 1, \dots, M\}$  are draws from the step- $\ell$  intermediate sampler  $m(z, w; \hat{a}^{(\ell)})$ . The final EIS MC estimate of the augmented integral is then given by

$$\hat{I} = \frac{1}{M} \sum_{i=1}^m \frac{g(z^{(j)})p(z^{(j)}|w^{(j)})p(w^{(j)})}{m(z^{(j)}, w^{(j)}; \hat{a})}, \quad (20)$$

where  $(z^{(j)}, w^{(j)})$  represent draws from the final joint EIS sampler.

Hence, if both  $g(z)p(z|w)$  and  $\chi(w, \cdot)p(w)$  can be reasonably well approximated by density kernels from the exponential family, this approach allows us to retain the linearity of the EIS approximation problem, even for integrands  $\varphi(z)$  which are generated by infinite mixtures, and which cannot be well approximated by a univariate IS sampling density for  $z$  from the exponential family.

At a first glance, extending the dimension of integration seems to be counterintuitive since an increase of the dimension typically has an adverse

effect on the variance of the MC estimator. However, in the mixture framework considered here, the data-augmentation step can significantly increase the flexibility to construct efficient IS densities for ill-behaved integrands in infinite mixture frameworks, while retaining its simplicity.

The approach of augmenting the set of latent variables for the purpose of enhancing the numerical efficiency is well established in Bayesian MCMC literature. Examples include the approach of Geweke (1993) who exploits the equivalence of the student- $t$  and an appropriate scale mixture of normals for the design of a Gibbs sampling approach and Frühwirth-Schnatter and Wagner (2006) who propose a Gibbs sampler for a Poisson regression model based upon a data-augmentation introducing the latent inter-arrival times of a Poisson process.

In order to illustrate the mixture EIS approach, consider the computation of the integral  $\int g(z)p(z)dz$ , where  $g(z)$  is a  $N(z, \sigma^2)$ -density for a random variable  $y$  evaluated at  $y = 2, \sigma = 4$ , and  $p(z)$  is a standardized student- $t$  density with mean zero, unit variance and  $\nu > 2$  degrees of freedom, which obtains as an inverse Gamma scale mixture of normal densities. Hence, the augmented integrand in Equation (16) consists of

$$g(z) \propto \exp \left\{ -\frac{1}{2\sigma^2}(2-z)^2 \right\}, \quad (21)$$

$$p(z|w) \propto w^{\frac{1}{2}} \exp \left\{ -\frac{1}{2}z^2w \right\}, \quad p(w) \propto w^{\frac{\nu}{2}-1} \exp \left\{ -\frac{(\nu-2)}{2}w \right\}, \quad (22)$$

where  $p(z|w)$  is a  $N(0, 1/w)$  and  $p(w)$  is a  $\text{Gamma}(\frac{\nu}{2}, \frac{2}{\nu-2})$ -density.

Let the function  $k(z, w; a_z)$ , which is used to approximate the first factor  $g(z)p(z|w)$ , be a Gaussian kernel obtained as the following parametric extension of  $p(z|w)$ :

$$k(z, w; a_z) = p(z|w)\zeta_z(z; a_z), \quad \text{where} \quad (23)$$

$$\zeta_z(z; a_z) = \exp \left\{ -\frac{1}{2}(\alpha z^2 - 2\beta z) \right\},$$

with  $a_z = (\alpha, \beta)$ . It immediately follows that the conditional EIS sampler for  $z|w$  is given by  $m(z|w; a_z) \sim N[\beta/(\alpha + w), 1/(\alpha + w)]$ , and the integration of  $k(z, w; a_z)$  with respect to  $z$  yields

$$\chi(w; a_z) = \left( \frac{w}{w + \alpha} \right)^{\frac{1}{2}} \exp \left\{ \frac{\beta^2}{\alpha + w} \right\}. \quad (24)$$

Note that the factor  $p(z|w)$  appears on both sides of the first EIS-regression in Equation (19) and cancels out. It follows that the EIS-regression simplifies to an approximation of the Gaussian density  $g(z)$  by a Gaussian kernel  $\zeta_z(z; a_z)$  which leads to a perfect fit with EIS values of the auxiliary parameters obtained analytically as  $\hat{\alpha} = 1/\sigma^2$  and  $\hat{\beta} = 2/\sigma^2$ .

In order to approximate the second factor  $\chi(w; a_z)p(w)$ , we can use a Gamma kernel parameterized as

$$k(w, a_w) = p(w)\zeta_w(w; a_w), \text{ where } \zeta_w(w; a_w) = w^{\gamma-1} \exp \left\{ -\frac{1}{\delta}w \right\}, \quad (25)$$

with  $a_w = (\gamma, \delta)$ , so that the marginal EIS sampler for  $w$  is given by  $m(w; a_w) \sim \text{Gamma}[\frac{\nu}{2} + \gamma - 1, (\frac{\nu-2}{2} + \frac{1}{\delta})^{-1}]$ . Under this parametrization of  $k(w, a_w)$ , the density  $p(w)$  cancels out in the second EIS-regression in Equation (19), which simplifies into a linear regression of  $\{\ln \chi(w^{(j,\ell)}, \hat{a}_z)\}$  on  $\{\ln w^{(j,\ell)}, w^{(j,\ell)}\}$  and a constant.

As noted above, smooth convergence of the EIS-iterations to a fixed-point solution requires the use of CRNs. Hence, for the simulation from the Gamma EIS density  $m(w; \hat{a}_w^{(\ell)})$  we rely upon the cdf inversion technique (see, e.g. Robert and Casella (2004)), whereby the  $j$ th draw  $w^{(j,\ell)}$  obtain from solving

$$\int_{-\infty}^{w^{(j,\ell)}} m(w'; \hat{a}_w^{(\ell)}) dw' - u^{(j)} = 0, \quad (26)$$

where  $\{u^{(j)}, j = 1, \dots, M\}$  is a set of iid  $(0, 1)$  uniform CRNs<sup>1</sup>. Simula-

---

<sup>1</sup>Specifically, for the Gamma family  $m(w; a_w)$  we use inverse cdf `gaminv` in Matlab for the pilot example.

tions from the conditional Gaussian EIS density  $m(z|w; \hat{a}_z^{(\ell)})$  are based on transformations of a set of iid  $N(0, 1)$  CRNs.

Results for this mixture EIS estimator are reported in the last column of Table 1 for values of  $\nu$  ranging from 2.5 to 15. The results reported are the mean bias, the standard deviation and Geweke (1989) relative numerical efficiency (RNE) factor based upon 10000 independent replications of the complete algorithm using a simulation sample size of  $M = 32$  and 7 EIS iterations. The RNE of an IS density obtains as the ratio between the variance of the direct MC estimates obtained by using the natural sampler  $p(z)$  and the IS MC estimates. High RNE values are preferable. For comparison, we also considered the IS estimators based upon a univariate Gaussian sampler for  $z$  resulting from a local Laplace approximation of the target integrand  $g(z)p(z)$ , as well as from a global Gaussian EIS approximation as described in Section 2. The Laplace-IS and Gaussian-EIS algorithm are implemented using the same sample size  $M$  as for the mixture EIS.

The results of the experiment indicate that the mixture EIS sampler clearly outperforms the Gaussian EIS and the Laplace sampler with respect to numerical accuracy: For all degrees of freedom  $\nu$ , the former has a substantially smaller standard deviation and a significantly higher RNE factor than the latter. In fact, the RNE for the mixture EIS exceeds in all cases 1, which is the benchmark value for the natural sampler  $p(z)$ , while the RNE for the univariate Gaussian samplers is smaller than 1. This indicates that the joint bivariate EIS sampler  $m(z, w; \hat{a})$  provides a better approximation to the augmented integral  $g(z)p(z|w)p(w)$  than the univariate Gaussian EIS and Laplace samplers to the marginalized integral  $g(z)p(z)$ . This is confirmed by Figure 1 which shows the target integrand  $g(z)p(z)$  for  $\nu = 2.5$  (normalized such that it integrates to one) together with the Gaussian EIS and Laplace sampling densities as well as the marginal density for  $z$  obtained from the joint EIS sampler  $m(z, w; \hat{a})$ .



Figure 1 also reveals that for low degrees of freedom  $\nu$ , the Gaussian Laplace sampler is subject to a thin-tail problem since its tails decay more quickly than those of the target integrand. For  $\nu = \{2.5, 3.5, 6\}$  this translates into significantly downward biased MC estimates with a bias ranging from 3.9 to 10.9 standard deviations (see Table 1).

### 3.2 Finite mixture EIS

Here we consider the case where the natural sampler  $p(z)$  of the integral given by Equations (1) and (2) is a finite mixture of the form

$$p(z) = \sum_{i=1}^l p(z|w_i)p_i, \quad (27)$$

where  $p_i$  denotes the mixing proportion for the  $i$ th mixture component with  $\sum_{i=1}^l p_i = 1$ , and  $p(z|w_i)$  is a conditional density depending on the parameter  $w_i$ . We assume that all mixture components  $p(z|w_i)$  belong to the same exponential family of distribution, which is essential for simplifications of the calculations.

The integral (1) to be approximated has the augmented form

$$I = \int \sum_{i=1}^l g(z)p(z|w_i)p_i dz. \quad (28)$$

EIS may then be implemented using a joint mixed EIS sampler  $m(z, i; a) = m(z|w_i, a_z) \cdot m_i$  with

$$m(z|w_i, a_z) = \frac{k(z, i; a_z)}{\chi_z(i, a_z)}, \quad \chi_z(i, a_z) = \int k(z, i; a_z) dz, \quad (29)$$

and EIS mixing proportions

$$m_i = \frac{k_i}{\bar{\chi}}, \quad \bar{\chi} = \sum_{i=1}^l k_i, \quad i = 1, \dots, l, \quad (30)$$

where  $a = (a_z, k_1, \dots, k_l)$ . This allows us to rewrite the integral (28) as

$$I = \bar{\chi} \int \sum_{i=1}^l \left[ \frac{g(z)p(z|w_i)}{k(z, i; a_z)} \right] \left[ \frac{\chi_z(i, a_z)p_i}{k_i} \right] m(z, i; a) dz. \quad (31)$$

The density kernel  $k(z, i; a_z)$  used to approximate  $g(z)p(z|w_i)$  can be specified as a parametric extension of  $p(z|w_i)$  as

$$k(z, i; a_z) = p(z|w_i)\zeta(z, a_z), \quad (32)$$

where  $\zeta$  is a kernel for a density belonging to the same parametric class as  $p(z|w_i)$ . Since  $p(z|w_i)$  is assumed to be a member of the exponential family,  $k(z, i; a_z)$  itself defines a kernel for a density of the same parametric class as  $p(z|w_i)$ . From the factor in the rightmost bracket of (31), we see that by setting the non-normalized mixing proportions of the importance density to

$$\hat{k}_i^{(\ell+1)} = \chi_z(i, \hat{a}_z^{(\ell+1)}) \cdot p_i \quad i = 1, \dots, l, \quad (33)$$

we obtain a further simplified expression for (28):

$$I = \bar{\chi} \int \left[ \frac{g(z)}{\zeta(z, a_z)} \right] \sum_{i=1}^l m(z, i; a) dz. \quad (34)$$

Under this specification of  $k(z, i; a_z)$  and  $k_i$ ,  $i = 1, \dots, l$ , the EIS auxiliary parameters  $a$  are obtained from the solution of the LS regression

$$\min_{a_z, c_z} \sum_{j=1}^M \left[ \ln g(z^{(j, \ell)}) - c_z - \ln \zeta(z^{(j, \ell)}, a_z) \right]^2. \quad (35)$$

Here  $\{z^{(j, \ell)}, j = 1, \dots, M\}$  are draws from the step- $\ell$  intermediate marginal sampling density  $m(z; \hat{a}^{(\ell)})$  associated with the joint density  $m(z, i; \hat{a}^{(\ell)}) = m(z|w_i, \hat{a}_z^{(\ell)}) \cdot \hat{m}_i^{(\ell)}$ . In order to simulate from the marginal EIS density  $m(z; \hat{a}^{(\ell)})$  using CRNs, we can rely upon the cdf inversion technique<sup>2</sup>.

---

<sup>2</sup>When  $m(z, w_i; a)$  is Gaussian, as it is the case in the examples below, we use the approximation `erfcc` given in Press et al. (2007) to the complementary error function involved in the relevant Gaussian cumulative distribution functions.

Since the marginal EIS density  $m(z; \hat{a})$  is analytically available, we can directly use it to obtain the MC-EIS estimate of the integral  $I$  as

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M \frac{g(z^{(j)})p(z^{(j)})}{m(z^{(j)}; \hat{a})}, \quad (36)$$

and do not need to rely upon the joint EIS sampler for the augmented integral.

In order to illustrate this finite mixture EIS approach, consider the MC estimation of  $\int g(z)p(z)dz$ , where  $g(z)$  is a  $N(0, \exp\{z\})$  density for  $y$  evaluated at  $y = 3$ , and  $p(z)$  is a zero mean and unit variance Gaussian mixture of the form  $p(z) \sim N(0, \sigma_1^2)p_1 + N(0, \sigma_2^2)p_2$  with  $p_1 = p_2 = 0.5$  and  $\sigma_2^2 = 2 - \sigma_1^2$ . In this case, the augmented integrand (28) consists of

$$g(z) \propto \exp\left\{-\frac{1}{2}(3^2 e^{-z} + z)\right\} \quad (37)$$

$$p(z|w_i) \propto (\sigma_i^2)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{z}{\sigma_i}\right)^2\right\}, \quad i = 1, 2. \quad (38)$$

The joint mixed density kernel meant to approximate  $g(z)p(z|w_i)$  is specified as

$$k(z, i; a_z) = p(z|w_i)\zeta_z(z; a_z), \quad \text{where} \quad (39)$$

$$\zeta_z(z; a_z) = \exp\left\{-\frac{1}{2}(\alpha z^2 - 2\beta z)\right\},$$

with  $a_z = (\alpha, \beta)$ . The corresponding EIS density for  $z|w_i$  is given by  $m(z|w_i, a_z) \sim N(\beta\sigma_i^2/[1 + \alpha\sigma_i^2], \sigma_i^2/[1 + \alpha\sigma_i^2])$ , and the EIS values of the auxiliary parameters  $(\alpha, \beta)$  are obtained from solving the LS problem (35) where  $\{\ln g(z^{(j, \ell)})\}$  is regressed on  $\{[z^{(j, \ell)}]^2, z^{(j, \ell)}\}$  and a constant. The expression for the integrating factor of  $k(z, i; a_z)$ , which is used to construct the EIS mixing proportions  $m_i$  defined in Equations (30) and (33), is

$$\chi_z(i, a_z) = (1 + \alpha\sigma_i^2)^{-1/2} \exp\left\{\frac{1}{2} \frac{\sigma_i^2 \beta^2}{1 + \alpha\sigma_i^2}\right\}. \quad (40)$$

Table 2 summarizes the results for the finite mixture EIS estimates of the integral for values of  $\sigma_1$  ranging from 0.1 to 0.75. Note that for  $\sigma_1 \leq 0.25$  the target integrand is bimodal and for  $\sigma_1 \leq 0.4$  it is not log-concave. As in the example in section 3.1, we also computed the MC estimates of the integral using the Gaussian EIS sampler and, in cases where the integrand is uni-modal, the Laplace sampler. The results in Table 2 are the mean biases, standard deviations and RNEs from 10,000 replications of the full procedure each based upon a simulation sample size of  $M = 32$  and 7 EIS iterations. The results of the experiment reveal that in all cases the mixture EIS sampler performs well leading to numerically very accurate estimates. As expected, the Gaussian samplers perform poorly when the integrand is bimodal and/or not log-concave. In those cases, the EIS iterations for the Gaussian EIS sampler often failed to converge indicating a severe mismatch between the target integrand and the class of Gaussian densities. In sharp contrast, we have found that the EIS convergence for the mixture EIS is fast indicating the adequacy of the class of Gaussian mixture samplers. This is confirmed by Figure 2 which shows the target integrand for  $\sigma_1 = 0.1$  together with the mixture EIS density and the Gaussian EIS density.

## 4 High-dimensional mixture EIS

In this section we discuss operational implementations of the mixture EIS approach for high-dimensional integration required for a likelihood analysis of realistic models. Those high-dimensional implementations combine data augmentation steps as discussed for univariate problems in the previous section with the generic sequential EIS approach for high-dimensional problems introduced in Section 2.1.

## 4.1 Example: student- $t$ state-space model

We start to analyze a state space model with student- $t$  measurement errors, which represents a useful extension of the linear Gaussian specification used, e.g. by Barndorff-Nielsen and Shephard (2002) and Bach and Christensen (2011) to model time series of realized volatilities of asset returns.

The log of the daily realized volatility denoted by  $y_t$  is assumed to follow the process

$$y_t = \lambda_t + \sigma_y u_t \quad (41)$$

$$\lambda_t = \mu + \phi(\lambda_{t-1} - \mu) + \sigma_\lambda \epsilon_t, \quad (42)$$

where  $(u_t, \epsilon_t)$  are mutually independent iid variables with zero mean and unit variance. For the measurement error  $u_t$  we assume a standardized student- $t$  distribution with  $\nu > 2$  degrees of freedom, while the shock in the latent state equation  $\epsilon_t$  is normally distributed.

The conditional distribution of  $y_t | \lambda_t$  defines a student- $t$  density kernel for  $\lambda_t$ . In light of the first pilot example, the use of (E)IS procedures for the likelihood evaluation based on Gaussian sampling densities for the  $\lambda_t$ s appears to be inadequate. However, the interpretation of the fat-tailed student- $t$  distribution as a scale mixture of Normals allows us to augment the state space model (41) and (42) in such a way that we can evaluate the likelihood via the sequential EIS outlined in Section 2.1 using convenient sampling densities from the exponential class. The first step of data augmentation consists in representing the student- $t$  error as  $u_t \sim N(0, 1/\eta_{t-1})$ , where  $\eta_{t-1} \sim \text{Gamma}(\nu/2, 2/(\nu - 2))$ . In the second step, we use the inverse cdf of the  $\text{Gamma}(\nu/2, 2/(\nu - 2))$ -distribution,  $F_\Gamma^{-1}$ , and the cdf of a standardized normal,  $\Phi$ , to rewrite  $\eta_t$  as a map of

a standard normal variable

$$\eta_t = F_\Gamma^{-1}[\Phi(w_t)] \equiv h_\nu(w_t), \quad w_t \sim N(0, 1). \quad (43)$$

This augmentation scheme allows us to recast the non-Gaussian linear state space model (41) and (42) into a non-linear one with a Gaussian measurement density given by

$$g(y_t | \lambda_t, w_{t-1}) \propto h_\nu(w_{t-1})^{1/2} \exp \left\{ -\frac{h_\nu(w_{t-1})}{2\sigma_y^2} (y_t - \lambda_t)^2 \right\}, \quad (44)$$

and two Gaussian state-transition densities

$$p(\lambda_t | \lambda_{t-1}) \propto \exp \left\{ -\frac{1}{2\sigma_\lambda^2} [\lambda_t - \mu(\lambda_{t-1})]^2 \right\} \quad (45)$$

$$p(w_{t-1}) \propto \exp \left\{ -\frac{1}{2} w_{t-1}^2 \right\}, \quad (46)$$

where  $\mu(\lambda_{t-1}) = \mu + \phi(\lambda_{t-1} - \mu)$ . As we shall see below, the time shift of the auxiliary variable  $w_t$  relative to  $u_t$  and  $y_t$  together with the representation of  $\eta_t$  as a function of the Gaussian variable  $w_t$  allows us to factorize the integrand of the likelihood integral period-by-period into simple bivariate Gaussian density kernels for the latent state variables  $(\lambda_t, w_t)$ .

Let  $z'_t = (\lambda_t, w_t)$ ,  $z_T = \lambda_T$ , and  $z'_{(t)} = (z'_0, \dots, z'_t)$  and assume for the initial condition  $\lambda_0$  the stationary distribution  $\lambda_0 \sim N(\mu, \sigma_\lambda^2/(1 - \phi^2))$ . Then the likelihood for the augmented state space representation (44)-(46) is given by

$$L(\psi) = \int \prod_{t=0}^T \varphi_t(z_{(t)}) dz_{(T)}, \quad (47)$$

with

$$\varphi_t(z_{(t)}) = \begin{cases} g(y_T | \lambda_T, w_{T-1}) p(\lambda_T | \lambda_{T-1}), & t = T \\ g(y_t | \lambda_t, w_{t-1}) p(\lambda_t | \lambda_{t-1}) p(w_t), & t = 1, \dots, T-1 \\ p(\lambda_0) p(w_0), & t = 0, \end{cases} \quad (48)$$

and  $\psi = (\mu, \sigma_y, \nu, \phi, \sigma_z)'$ .

In order to apply the sequential EIS of Section (2.1) to this augmented likelihood, we first note that the factor  $g(y_t|\lambda_t, w_{t-1})p(\lambda_t|\lambda_{t-1})$  of the period- $t$  integrand (48) defines a kernel of a  $N(m_t^*, v_t^*)$  distribution for  $\lambda_t$  with

$$m_t^* = \frac{\sigma_y^2 \mu(\lambda_{t-1}) + h_\nu(w_{t-1}) \sigma_\lambda^2 y_t}{h_\nu(w_{t-1}) \sigma_\lambda^2 + \sigma_y^2}, \quad v_t^* = \frac{\sigma_y^2 \sigma_\lambda^2}{h_\nu(w_{t-1}) \sigma_\lambda^2 + \sigma_y^2}, \quad (49)$$

and that  $p(\lambda_0)$  and  $p(w_t)$  present Gaussian densities. It follows that for  $t < T$  the period- $t$  integrands  $\varphi_t$  define bivariate Gaussian kernels in  $z_t$ , while the period- $T$  integrand  $\varphi_T$  is a univariate Gaussian kernel for  $z_T = \lambda_T$ .

This immediately suggests to specify the EIS density kernels  $k_t$  of the period- $t$  importance samplers  $m_t$  in Equation (10) as the following parametric extensions of the Gaussian kernels  $\varphi_t$ :

$$k_t(z_{(t)}, a_t) = \varphi_t(z_{(t)}) \zeta_t(z_t, a_t), \quad (50)$$

with

$$\zeta_t(z_t, a_t) = \exp \left\{ -\frac{1}{2} (z_t' P_t z_t - 2 R_t' z_t) \right\}, \quad t = 0, \dots, T-1, \quad (51)$$

and  $\zeta_T(\cdot) \equiv 1$ . Here  $P_t$  is a symmetric  $(2 \times 2)$  matrix and  $R_t$  a  $(2 \times 1)$  vector representing the EIS auxiliary parameters, i.e.  $a_t = (\text{vech}(P_t)', R_t')'$ . Since  $\varphi_t$  and  $\zeta_t$  represent Gaussian density kernel for  $z_t$ , it follows the EIS samplers  $m_t$  associated with  $k_t$  are Gaussian distributions for  $z_t|z_{(t-1)}$ .

Using this parametrization of  $k_t$  and rewriting the likelihood integral (47) according to Equation (12) leads to

$$L(\psi) = \chi_0(a_0) \int \prod_{t=0}^{T-1} \frac{\chi_{t+1}(z_{(t)}, a_{t+1})}{\zeta_t(z_t, a_t)} \prod_{t=0}^{T-1} m_t(z_t|z_{(t-1)}, a_t) dz_{(T-1)}. \quad (52)$$

Note that since  $\chi_{T+1} \equiv 1$  and  $\zeta_T \equiv 1$ , the integration w.r.t. the variable

$z_T = \lambda_T$  can be done analytically and amounts to computing the integral  $\int m_T(z_T|z_{T-1}, a_T) dz_T$ , which is equal to 1. It immediately follows that the EIS auxiliary regressions in Equation (13) simplify into a sequence of  $T$  linear LS regressions of  $\{\ln \chi_{t+1}(z_{(t)}^{(j,\ell)}, \hat{a}_{t+1}^{(\ell+1)})\}$  on a constant and  $\{[\lambda_t^{(j,\ell)}]^2, [w_t^{(j,\ell)}]^2, [\lambda_t^{(j,\ell)} \cdot w_t^{(j,\ell)}], \lambda_t^{(j,\ell)}, w_t^{(j,\ell)}\}$  for  $t = T-1, \dots, 0$ . As initial samplers  $\{m_t(z_t|z_{t-1}, \hat{a}_t^{(0)})\}$  we use the Gaussian distributions associated with the Normal density kernels given in Equation (48). Details of the implementation of this sequential mixture EIS application are provided in Appendix A.

The data we use to estimate the student- $t$  state space model (41) and (42) by ML based upon sequential mixture EIS consists of the daily realized variances computed for the returns of the IBM stock traded at the New York Stock Exchange. We use  $T = 2156$  daily observations from January 1, 2000 until July 30, 2008. The daily realized variances can be computed as  $rv_t = \sum_{\tau=1}^J x_{t,\tau}^2$  where  $x_{t,\tau}$  is the return computed for the  $\tau$ s 5-minute interval of trading day  $t$  (This data set is discussed in more details in Chiriac and Voev (2011)).

The ML EIS results based upon a simulation sample size  $M = 64$  and 4 EIS iterations are summarized in Table 3. The parameter estimates and their MC (numerical) standard deviations are computed as the means and standard deviations of 100 ML estimates conducted under different sets of CRNs. The statistical standard deviations are obtained from 500 ML estimates for artificial data generated from the model. Total computing time for a full ML optimization using a BFGS optimizer is of the order of 1 min on a Dell PowerEdgeR200 computed with 2.66 GHz Quad core processor for a code written in Fortran 90. For comparison, we also report the ML estimation results for the state space model (41) and (42) with Gaussian measurement errors  $u_t$  obtained as the limit for  $\nu \rightarrow \infty$ , in which case we can use the standard Kalman filter to evaluate the likelihood. The MC standard deviations indicate that the ML estimates based on



the mixture EIS approach are numerically very accurate. The estimated degrees of freedom  $\nu$  is 6.4, indicating a large deviation from normality. The substitution of the Normal for the student- $t$  distribution decreases the value of the maximized likelihood function by 25, indicative of a significant deterioration in fit. Finally note that the strong persistence with a value of  $\phi$  close to one is typical for volatility models for asset returns.

## 4.2 Example: stochastic volatility model with jumps

In order to illustrate the mixture EIS approach for high-dimensional integration when the target integrand involves finite mixtures, we consider the ML estimation of a discrete-time stochastic volatility (SV) model with leverage effects and jumps (see, e.g., Yu (2005); Durham (2006) for a discussion of the discrete-time SV model and Eraker et al. (2003); Malik and Pitt (2011) for SV models involving jumps).

The stock log-returns  $y_t$  are assumed to follow the process

$$y_t = \gamma + \sigma_y \exp\{z_{t-1}/2\}(\rho\eta_t + \sqrt{1-\rho^2}\xi_t) + J_t\sigma_u u_t \quad (53)$$

and the latent volatility factor

$$z_t = \phi z_{t-1} + \sigma_z \eta_t. \quad (54)$$

Here  $J_t$  represents the time- $t$  jump arrival, which follows a Bernoulli process with state probabilities  $P(J_t = j) = p_j$  with  $j \in \{0, 1\}$ . The jump size is given by  $\sigma_u u_t$ , where  $u_t \sim N(0, 1)$ . The innovations  $\eta_t$  and  $\xi_t$  are assumed to be independent from  $u_t$  and are normally distributed as  $(\eta_t, \xi_t) \sim N(0, I)$ , where  $I$  is the identity matrix. The parameter  $\rho$  measures the leverage effect, while  $\sigma_u$  and  $p_1$  represent the average jump size and jump intensity respectively.

Note that under the SV model given in Equations (53) and (54) the

conditional density of  $z_t$  and  $y_t$  given  $z_{t-1}$  and  $J_t$  is a bivariate Gaussian density, which can be factorized as

$$p(z_t, y_t | z_{t-1}, J_t) = p(z_t | y_t, z_{t-1}, J_t) p(y_t | z_{t-1}, J_t), \quad (55)$$

where  $p(z_t | y_t, z_{t-1}, J_t)$  and  $p(y_t | z_{t-1}, J_t)$  are conditional Gaussian densities for  $z_t$  and  $y_t$  given by

$$p(z_t | y_t, z_{t-1}, J_t) \sim N(\mu_{z_t}, \sigma_{z_t}^2) \quad (56)$$

$$p(y_t | z_{t-1}, J_t) \sim N(\gamma, \sigma_{y_t}^2), \quad (57)$$

with

$$\mu_{z_t} = \phi z_{t-1} + (y_t - \gamma) \frac{\rho \sigma_y \sigma_z \exp\{z_{t-1}/2\}}{\sigma_y^2 \exp\{z_{t-1}\} + J_t \sigma_u^2} \quad (58)$$

$$\sigma_{z_t}^2 = \sigma_z^2 \left(1 - \frac{\sigma_y^2 \exp\{z_{t-1}\}}{\sigma_y^2 \exp\{z_{t-1}\} + J_t \sigma_u^2}\right) \quad (59)$$

$$\sigma_{y_t}^2 = \sigma_y^2 \exp\{z_{t-1}\} + J_t \sigma_u^2. \quad (60)$$

Based on this factorization of  $p(z_t, y_t | z_{t-1}, J_t)$ , the likelihood function obtains as  $L(\psi) = \int \prod_{t=0}^T \varphi_t(z_{(t)}) dz_{(T)}$ , where

$$\varphi_t(z_{(t)}) = \begin{cases} \sum_{j \in \{0,1\}} p(z_t | y_t, z_{t-1}, J_t) p(y_t | z_{t-1}, J_t) p_j, & t = 1, \dots, T \\ p(z_0), & t = 0, \end{cases} \quad (61)$$

and  $\psi = (\gamma, \sigma_y, \sigma_u, \sigma_z, \rho, \phi, p_1)'$ . The initial condition  $z_0$  is assumed be generated by the stationary distribution as  $z_0 \sim N(0, \sigma_z^2/(1 - \phi^2))$ .

In order to apply sequential EIS outlined in Section 2.1 to the evaluation of this likelihood function, we first note that for  $t \geq 1$  the factor  $\varphi_t$  in the likelihood integral defines a density kernel for a two-component normal mixture distribution for  $z_t | z_{t-1}$  with non-normalized mixing proportions  $p(y_t | z_{t-1}, J_t) p_j$ . The corresponding proper density for this two-component

normal mixture distribution obtains as

$$p^*(z_t|z_{t-1}) = \frac{\varphi_t(z_t)}{p(y_t|z_{t-1})} = \sum_{j \in \{0,1\}} p(z_t|y_t, z_{t-1}, J_t) \pi_{J_t}, \quad (62)$$

where the normalized mixing proportions  $\pi_{J_t}$  are

$$\pi_{J_t} = \frac{p(y_t|z_{t-1}, J_t) p_j}{p(y_t|z_{t-1})} \quad \text{with} \quad p(y_t|z_{t-1}) = \sum_{j \in \{0,1\}} p(y_t|z_{t-1}, J_t) p_j. \quad (63)$$

Using the normalized factors of the likelihood integral in Equation (62) we can rewrite the likelihood as

$$L(\psi) = \int \prod_{t=0}^{T-1} p(y_{t+1}|z_t) p^*(z_t|z_{t-1}) dz_{(T-1)}, \quad (64)$$

with  $p^*(z_0|z_{-1}) = p(z_0)$ . The integration w.r.t. the variable  $z_T$  is done analytically and amounts to computing  $\int p^*(z_T|z_{T-1}) dz_T$  which is equal to 1.

Clearly, (E)IS approaches based on (E)IS densities for the  $z_t$ s from the exponential family are not adequate to approximate this likelihood function involving finite mixtures leading to a possibly bimodal target integrand. Instead we shall specify the EIS density kernel  $k_t$  in Equation (11) as the following parametric extension of the natural Gaussian mixture sampler  $p^*$  given in Equation (62):

$$\begin{aligned} k_t(z_t, a_t) &= p^*(z_t|z_{t-1}) \zeta_t(z_t, a_t) \\ &= \begin{cases} \sum_{j \in \{0,1\}} p(z_t|y_t, z_{t-1}, J_t) \zeta_t(z_t, a_t) \pi_{J_t}, & t \geq 1 \\ p(z_0) \zeta_0(z_0, a_0), & t = 0, \end{cases} \end{aligned} \quad (65)$$

where

$$\zeta_t(z_t, a_t) = \exp \left\{ -\frac{1}{2} (\alpha_t z_t^2 - 2\beta_t z_t) \right\}, \quad t = 0, \dots, T-1, \quad (66)$$

and  $a_t = (\alpha_t, \beta_t)$ . This selection of the class of kernels implies that

for period  $t = 0$  the EIS sampler  $m_0(z_0; a_0) = k_0(z_0, a_0)/\chi_0(a_0)$  is a simple Gaussian density, while for  $t \geq 1$  the corresponding EIS densities  $m_t(z_t|z_{(t-1)}, a_t)$  are two-component Gaussian mixtures associated with the joint mixed density kernel  $p(z_t|y_t, z_{t-1}, J_t)\zeta_t(z_t, a_t)\pi_{J_t}$  for  $z_t$  and  $J_t$ . The corresponding EIS densities for  $z_t$  obtain as  $m_t(z_t|z_{(t-1)}; a_t) = \sum_{j \in \{0,1\}} m_t(z_t|z_{(t-1)}, J_t; a_t)m_{J_t}$ , where the Gaussian EIS mixture components  $m(z_t|z_{(t-1)}, J_t; a_t)$  are

$$m_t(z_t|z_{(t-1)}, J_t; a_t) = \frac{p(z_t|y_t, z_{t-1}, J_t)\zeta_t(z_t, a_t)}{\chi_t(z_{(t-1)}, J_t, a_t)}, \quad (67)$$

$$\chi_t(z_{(t-1)}, J_t, a_t) = \int p(z_t|y_t, z_{t-1}, J_t)\zeta_t(z_t, a_t)dz_t, \quad (68)$$

and the resulting EIS mixing proportions  $m_{J_t}$  obtain as

$$m_{J_t} = \frac{\chi_t(z_{(t-1)}, J_t, a_t)\pi_{J_t}}{\chi_t(z_{(t-1)}, a_t)}, \quad (69)$$

$$\chi_t(z_{(t-1)}, a_t) = \sum_{j \in \{0,1\}} \chi_t(z_{(t-1)}, J_t, a_t)\pi_{J_t}. \quad (70)$$

It immediately follows that  $\chi_t(z_{(t-1)}, a_t)$  as given in Equation (70) represents the integrating factor for the EIS density kernel  $k_t(z_t, a_t)$  in Equation (65) such that the EIS density can be represented as  $m_t(z_t|z_{(t-1)}; a_t) = k_t(z_t, a_t) / \chi_t(z_{(t-1)}, a_t)$ .

Finally note that using the parametrization of the density kernel  $k_t$  for the EIS density given in Equations (65) and (66), the likelihood function (64) can be rewritten according to Equation (12) as

$$L(\psi) = \chi_0(a_0) \int \left[ \prod_{t=0}^{T-1} \frac{\chi_{t+1}(z_t, a_{t+1})p(y_{t+1}|z_t)}{\zeta_t(z_t, a_t)} \right] \quad (71)$$

$$\times \prod_{t=0}^{T-1} m_t(z_t|z_{(t-1)}, a_t)dz_{(T-1)},$$

with  $\chi_T \equiv 1$ . Hence, the EIS auxiliary regressions in Equation (7) simplify into a sequence of  $T$  linear LS problems where  $\{\ln[\chi_{t+1}(z_t^{(j,\ell)}, \hat{a}_{t+1}^{(\ell+1)})]$

$p(y_{t+1}|z_t^{(j,\ell)})\}$  is regressed on a constant and  $\{[z_t^{(j,\ell)}]^2, z_t^{(j,\ell)}\}$  for  $t = T - 1, \dots, 0$ . As initial samplers  $\{m_t(z_t|z_{(t-1)}, \hat{a}_t^{(0)})\}$  we let  $\log \zeta_t$  match the two first derivatives of  $\log p(y_{t+1}|z_t, J_{t+1} = 0)$  around the mode for  $t = 0, \dots, T - 1$ . This amounts to setting  $\alpha_t = 1$  and  $\beta_t = \log\{[(y_{t+1} - \gamma)/\sigma_y]^2\}$ . Details of the implementation of this sequential finite mixture EIS application are provided in Appendix B. Before presenting the empirical results, an important remark is in order. As described above, the EIS implementation requires to approximate  $\chi_{t+1}(z_{(t)}, a_{t+1})p(y_{t+1}|z_t)$  by a Gaussian kernel  $\zeta_t(z_t, a_t)$ , where both  $\chi_{t+1}(z_{(t)}, a_{t+1})$  and  $p(y_{t+1}|z_t)$  are weighted sums of two functions in  $z_t$  (see Equations 63 and 70). Hence, theoretically there is no guarantee that their product will define a well-behaved function in  $z_t$ . However, we have found that those products are very accurately approximated by the Gaussian kernels  $\zeta_t(z_t, a_t)$ . In fact, the  $R^2$ s of the EIS auxiliary regressions are typically larger than 0.99.

The data we use to estimate the SV model (53) and (54) by ML based upon sequential finite mixture EIS consists of daily continuously compounded log-returns on the S&P500 stock price index from January 3, 2000 to December 31, 2007 with  $T = 2008$ . The ML-EIS results based upon a simulation sample size  $M = 32$  and 10 EIS iterations are summarized in Table 4. The parameter estimates and their MC standard deviations are computed as the means and standard deviations of 100 ML estimates obtained under different sets of CRNs. The statistical standard deviations are based upon 500 ML estimates for artificial data from the model. Total computing time for a full ML estimation using a BFGS optimizer is 142 seconds for a Fortran 90 code on a laptop computer with an Intel core i7 processor with 8 Gb of memory using 2 cores in parallel.

For comparison, we also estimate the corresponding SV model without jumps (obtained by letting either  $\sigma_u \rightarrow 0$  or  $p_1 \rightarrow 0$ ) using the same code and same simulation sample size as above. In this case, the mixture-EIS procedure boils down to a standard EIS algorithm based on Gaussian IS

densities. The MC standard deviations reported in Table 4 reveal that the ML parameter estimates for the SV model with jumps based on finite mixture EIS are numerically very accurate. In fact, the MC standard deviations for the SV jump model are of the same order of magnitudes as those obtained by the standard EIS procedure for the ML estimation of the model without jumps. The value of the likelihood ratio statistic for the hypothesis  $H_0 : \sigma_u = 0, p_1 = 0$  is 10.2, indicating a rejection of the model without jumps against the SV jump specification at the 1% level. The estimate of the jump intensity parameter  $p_1$  implies that jumps occur at a fairly low frequency with an average of 173 trading days between two jumps. This result is consistent with the findings of Eraker et al. (2003) and Malik and Pitt (2011). The high persistence with a value of  $\phi$  close to 1 and the strong leverage effect with a significantly negative value of  $\rho$  is typical of SV models.

## 5 Discussion

In this paper, we propose efficient importance sampling (EIS) procedures for the evaluation of likelihood functions involving finite or infinite mixtures leading to possibly ill-behaved target densities with multiple modes or fat tails. Our approach exploits the mixture specification of the target density and uses a data augmentation step, demarginalizing the mixture density to include the mixing variable as an additional latent variable.

In the examples used to illustrate our mixture EIS approach, we have found that this extension of the space of integration can lead to well behaved target integrands for the extended space which can be well globally approximated by standard distributions belonging to the exponential family.

The set of examples considered here do by no means deplete the potential area of application of the mixture EIS. Further examples where we

have also successfully applied this methodology include the model of Cox et al. (1985) observed with Gaussian measurement noise (exploiting that the transition law of the CIR-diffusion is a Poisson-mixture of Gammas) and a time-discretized jump-diffusion model observed with measurement noise.

As for most other (E)IS methods, mixture EIS involves a high degree of tailor-making for each particular model. However, many latent variable models today involve implicitly or explicitly some form of mixing or hierarchical structures, generating non-Gaussian and dependent target densities. In so far, the demarginalization step underlying the mixture EIS can in many cases be quite automatic.

As shown in the two high-dimensional examples, complex latent variable models are generally needed to capture the salient features of economic data. However, such complex models often give rise to integration problems involving high-dimensional and complex target densities. Thus may the required flexibility come at the cost of making the models difficult to fit to the data. We think that the mixture EIS provides a useful and efficient procedure for likelihood-based inference for such non-standard, but generally more appropriate models. The estimation can be done without imposing severe restrictions on the model building, while retaining the sparsely parameterized local importance densities and computationally simple LS regressions of the exponential family EIS.

## Appendix A

This appendix details the implementation of the sequential mixture EIS application for the MC likelihood evaluation of the student- $t$  state space model (41) and (42) outlined in Section 4.1.

The sequence of EIS density kernels  $k_t$  ( $t = 0, \dots, T$ ) as defined by Equations (48), (50), and (51) and the corresponding integrating factors  $\chi_t$  have the following functional forms:

For  $t = T$  the univariate Gaussian kernel  $k_T = \varphi_T$  for  $Z_T = \lambda_T$  is given by

$$k_T(z_{(T)}) = \frac{\varphi_T^*(z_{T-1})}{2\pi\sigma_y\sigma_\lambda} \exp\left\{-\frac{1}{2}\left(\frac{\lambda_T^2}{v_T^*} - 2\frac{m_T^*}{v_T^*}\lambda_T\right)\right\}, \quad (72)$$

where

$$\varphi_T^*(z_{T-1}) = h_\nu(w_{T-1})^{1/2} \exp\left\{-\frac{1}{2}\left(\frac{h_\nu(w_{T-1})y_T^2}{\sigma_y^2} + \frac{\mu(\lambda_{T-1})}{\sigma_\lambda^2}\right)\right\}, \quad (73)$$

and  $m_T^*$  and  $v_T^*$  are defined by Equation (49). Integrating the Gaussian kernel  $k_T$  w.r.t.  $\lambda_T$  and omitting irrelevant constants leads to

$$\chi_T(z_{(T-1)}) \propto \varphi_T^*(z_{T-1})v_T^{*1/2} \exp\left\{\frac{1}{2}\frac{m_T^{*2}}{v_T^*}\right\}. \quad (74)$$

For  $t = 1, \dots, T-1$  the bivariate Gaussian kernel  $k_t = \varphi_t \cdot \zeta_t$  for  $z_t$  has the form

$$k_t(z_{(t)}, a_t) = \frac{\varphi_t^*(z_{t-1})}{(2\pi)^{3/2}\sigma_y\sigma_\lambda} \exp\left\{-\frac{1}{2}\left(z_t'(\Sigma_t^{-1} + P_t)z_t - 2(\mu_t'\Sigma_t^{-1} + R_t')z_t\right)\right\}, \quad (75)$$

where

$$\varphi_t^*(z_{t-1}) = h_\nu(w_{t-1})^{1/2} \exp\left\{-\frac{1}{2}\left(\frac{h_\nu(w_{t-1})y_t^2}{\sigma_y^2} + \frac{\mu(\lambda_{t-1})}{\sigma_\lambda^2}\right)\right\} \quad (76)$$



$$-\frac{m_t^{*2}}{v_t^*} + \mu_t' \Sigma_t^{-1} \mu_t \Big) \Big\}$$

and the mean and variance associated with the density kernel  $\varphi_t$  are given as

$$\mu_t = (m_t^*, 0)', \quad \Sigma_t = \begin{pmatrix} v_t^* & 0 \\ 0 & 1 \end{pmatrix}. \quad (77)$$

It follows that the EIS sampler  $m_t$  for  $z_t|z_{(t-1)}$  associated with the density kernel (75) is a bivariate Gaussian distribution with mean and variance given by

$$\mu_t^* = \Sigma_t^* (\Sigma_t^{-1} \mu_t + R_t), \quad \Sigma_t^* = (\Sigma_t^{-1} + P_t)^{-1}, \quad (78)$$

Integrating the Gaussian kernel  $k_t$  w.r.t.  $z_t$  leads to

$$\chi_t(z_{(t-1)}, a_t) \propto \varphi_t^*(z_{t-1}) |\Sigma_t^*|^{1/2} \exp \left\{ \frac{1}{2} \mu_t^{*'} \Sigma_t^{*-1} \mu_t^* \right\}. \quad (79)$$

For  $t = 0$  the bivariate Gaussian kernel  $k_0 = \varphi_0 \cdot \zeta_0$  for  $z_0$  has the form

$$k_0(z_0, a_0) = \varphi_0^* \exp \left\{ -\frac{1}{2} \left( z_0' (\Sigma_0^{-1} + P_0) z_0 - 2(\mu_0' \Sigma_0^{-1} + R_0') z_0 \right) \right\}, \quad (80)$$

where

$$\varphi_0^* = \frac{(1 - \phi^2)^{1/2}}{2\pi\sigma_\lambda} \exp \left\{ -\frac{1}{2} \mu_0' \Sigma_0^{-1} \mu_0 \right\}, \quad (81)$$

$$\mu_0 = (\mu, 0)', \quad \Sigma_0 = \begin{pmatrix} \sigma_\lambda^2 / (1 - \phi^2) & 0 \\ 0 & 1 \end{pmatrix}. \quad (82)$$

It follows that the EIS sampler  $m_0$  for  $z_0$  associated with the density kernel (80) is a bivariate Gaussian distribution with mean and variance given by

$$\mu_0^* = \Sigma_0^* (\Sigma_0^{-1} \mu_0 + R_0), \quad \Sigma_0^* = (\Sigma_0^{-1} + P_0)^{-1}. \quad (83)$$

Integrating the Gaussian kernel  $k_0$  w.r.t.  $z_0$  leads to

$$\chi_0(a_0) = 2\pi\varphi_0^* |\Sigma_0^*|^{1/2} \exp \left\{ \frac{1}{2} \mu_0^{*'} \Sigma_0^{*-1} \mu_0^* \right\}. \quad (84)$$

## Appendix B

In this appendix we provide details for the implementation of the sequential finite mixture EIS application for the MC likelihood evaluation of the SV model with jumps (53) and (54) outlined in Section 4.2.

The sequence of EIS densities  $m_t(z_t|z_{(t-1)}, a_t)$  associated with the density kernels  $k_t(z_t, a_t)$  defined in Equations (65) and (66) and the corresponding integrating factors  $\chi_t(z_{(t-1)}, a_t)$  such that  $m_t(z_t|z_{(t-1)}; a_t) = k_t(z_t, a_t) / \chi_t(z_{(t-1)}, a_t)$  obtain as follows:

For  $t = 1, \dots, T - 1$  the EIS density is a two-component mixture,

$$m_t(z_t|z_{(t-1)}; a_t) = \sum_{j \in \{0,1\}} m_t(z_t|z_{(t-1)}, J_t; a_t) m_{J_t}, \quad (85)$$

where the mixture component  $m_t(z_t|z_{(t-1)}, J_t; a_t)$  and the mixing proportions  $m_{J_t}$  are defined by Equations (67) and (69), respectively. The density kernel for the mixture component  $m_t(z_t|z_{(t-1)}, J_t; a_t)$  has the form (see Equation 67)

$$p(z_t|y_t, z_{t-1}, J_t) \zeta_t(z_t, a_t) = \frac{1}{\sqrt{2\pi}\sigma_{z_t}} \exp \left\{ -\frac{1}{2} \left( \left[ \frac{1}{\sigma_{z_t}^2} + \alpha_t \right] z_t^2 - 2 \left[ \frac{\mu_{z_t}}{\sigma_{z_t}^2} + \beta_t \right] z_t + \frac{\mu_{z_t}^2}{\sigma_{z_t}^2} \right) \right\}, \quad (86)$$

where  $\mu_{z_t}$  and  $\sigma_{z_t}^2$  are defined in Equations (58) and (59). It follows that the EIS mixture component  $m_t(z_t|z_{(t-1)}, J_t; a_t)$  is a Gaussian density with mean and variance given by

$$\mu_t^* = \sigma_t^{*2} \left( \frac{\mu_{z_t}}{\sigma_{z_t}^2} + \beta_t \right), \quad \sigma_t^{*2} = \frac{\sigma_{z_t}^2}{1 + \alpha_t \sigma_{z_t}^2}. \quad (87)$$

Integrating the Gaussian kernel (86) w.r.t.  $z_t$  leads to the following form of the integrating factor given by Equation (68):

$$\chi_t(z_{(t-1)}, J_t, a_t) = \frac{\sigma_t^*}{\sigma_{z_t}} \exp \left\{ \frac{1}{2} \left( \frac{\mu_{z_t}^2}{\sigma_{z_t}^2} - \frac{\mu_t^{*2}}{\sigma_t^{*2}} \right) \right\}. \quad (88)$$

Based upon this integrating factor for the mixture component, we can compute the mixing proportions  $m_{J_t}$  and the integrating factor  $\chi_t(z_{(t-1)}, a_t)$  associated with the EIS density kernel according to Equations (69) and (70).

For  $t = 0$  the EIS density kernel  $k_t(z_0; a_0)$  defined in Equations (65) and (66) has the form

$$p(z_0)\zeta_0(z_0, a_0) = \frac{(1 - \phi^2)^{1/2}}{\sqrt{2\pi}\sigma_z} \exp \left\{ -\frac{1}{2} \left( \left[ \frac{1 - \phi^2}{\sigma_z^2} + \alpha_0 \right] z_0^2 - 2\beta_0 z_0 \right) \right\}. \quad (89)$$

It follows that the EIS sampler  $m_0$  for  $z_0$  is a Gaussian distribution with mean and variance given by

$$\mu_0^* = \sigma_0^{*2} \beta_0, \quad \sigma_0^{*2} = \frac{\sigma_z^2 / (1 - \phi^2)}{1 + \alpha_0 \sigma_z^2 / (1 - \phi^2)}. \quad (90)$$

Integrating the Gaussian kernel (89) w.r.t.  $z_0$  leads to the following form of the integrating factor

$$\chi_0(a_0) = \frac{(1 - \phi^2)^{1/2} \sigma_0^*}{\sigma_z} \exp \left\{ -\frac{1}{2} \left( \frac{\mu_0^{*2}}{\sigma_0^{*2}} \right) \right\}. \quad (91)$$

For the this model, we slow down the convergence of the EIS iterations by choosing shorter step-lengths. This is done in order to reduce the probability of getting an  $a^{(\ell)}$  corresponding to zero variance or some other pathology in the first iterations. In practice, this done by setting  $a_t^{(\ell+1)} \leftarrow s a_t^{(\ell+1)} + (1 - s) a_t^{(\ell)}$  with  $s = \min(1, \exp(-0.6 + 0.1\ell))$  immediately after each regression.

## References

- Ardia, D., L. F. Hoogerheide, and H. K. van Dijk (2009, 1). Adaptive mixture of student-t distributions as a flexible candidate distribution for efficient simulation: The r package admit. *Journal of Statistical Software* 29(3), 1–32.
- Bach, C. and B. J. Christensen (2011). Latent integrated stochastic volatility, realized volatility, and implied volatility: A state space approach. CREATES Research Papers 2010-61, School of Economics and Management, University of Aarhus.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64(2), pp. 253–280.
- Bauwens, L. and F. Galli (2009). Efficient importance sampling for ML estimation of SCD models. *Computational Statistics and Data Analysis* 53, 1974–1992.
- Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, Forthcoming.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53(2), 385–407.
- Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* 84(3), 669–684.
- Durbin, J. and S. J. Koopman (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society Series B* 62, 3–56.

- Durham, G. B. (2006). Monte carlo methods for estimating, smoothing, and filtering one and two-factor stochastic volatility models. *Journal of Econometrics* 133, 273–305.
- Eraker, B., M. Johannes, and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance* 58(3), 1269–1300.
- Frühwirth-Schnatter, S. and H. Wagner (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika* 93(4), 827–841.
- Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica* 57(6), pp. 1317–1339.
- Geweke, J. (1993). Bayesian treatment of the independent student- t linear model. *Journal of Applied Econometrics* 8(S), S19–S40.
- Hesterberg, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37(2), pp. 185–194.
- Jung, R. C., 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.
- Kleppe, T. S., J. Yu, and H. J. Skaug (2010). Simulated maximum likelihood estimation of continuous time stochastic volatility models. In W. Greene and R. C. Hill (Eds.), *Maximum Simulated Likelihood Methods and Applications (Advances in Econometrics, Volume 26)*, pp. 137–161. Emerald Group Publishing Limited.
- Kloek, T. and H. K. v. Dijk (1978). Bayesian estimates of equation system parameters: An application of integration by monte carlo. *Econometrica* 46, pp. 1–19.

- Liesenfeld, R. and J.-F. Richard (2001). Monte carlo methods and bayesian computation: importance sampling. In N. Smelser and P. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier Science, Oxford.
- Liesenfeld, R. and J.-F. Richard (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* 10, 505–531.
- Liesenfeld, R. and J.-F. Richard (2010). Efficient estimation of probit models with correlated errors. *Journal of Econometrics* 156(2), 367 – 376.
- Malik, S. and M. K. Pitt (2011). Modelling stochastic volatility with leverage and jumps: A simulated maximum likelihood approach via particle filtering. Documents de Travail 318, Banque de France.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.
- Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics* 127(2), 1385–1411.
- Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods*. Springer.
- Sandmann, G. and S. Koopman (1998). Maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289–306.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika* 84, 653–667.
- Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics* 127, 165–178.

Table 1. IS integration for a student- $t$  density.

	$\nu$	Laplace	Gaussian EIS	Mixture EIS
Bias	2.5	-3.27e-03	-7.66e-04	-4.08e-06
Standard dev.		0.03	0.07	4.48e-04
RNE		2.76e-03	5.17e-04	12.3
# fail			2	0
Bias	3.5	-1.45e-03	-2.62e-04	-2.64e-06
Standard dev.		0.02	0.04	2.87e-04
RNE		7.84e-03	2.00e-03	38.4
# fail			0	0
Bias	6.0	-3.77e-04	-5.60e-05	-1.03e-06
Standard dev.		9.77e-03	0.01	1.10e-04
RNE		0.03	0.02	284.5
# fail			0	0
Bias	15.0	-2.82e-05	-1.93e-06	-2.20e-07
Standard dev.		2.86e-03	2.34e-03	2.04e-05
RNE		0.42	0.63	8.26e+03
# fail			0	0

Note: IS-MC estimation of  $\int g(z)p(z)dz$  where  $p$  is a standardized student- $t$  density with  $\nu$  degrees of freedom, and  $g$  is a  $N[z, 4^2]$ -density. IS simulation sample size is  $M = 32$  and the number of EIS iterations is 7. Mean biases relative to the true value, standard deviations and RNEs are based upon 10000 replications under different CRN seeds. # fail denotes the number of replications where the EIS approximation failed to converge. The true values of the integrals were computed using high precision numerical integration.

Table 2. IS integration for a two-component Gaussian mixture density.

	$\sigma_1$	Laplace	Gaussian EIS	Mixture EIS
Bias	0.10		-6.02e-05	-9.81e-06
Standard dev.			6.71e-03	8.99e-04
RNE			0.338	18.8
# fail			1457	0
Bias	0.25		-1.85e-04	-1.01e-05
Standard dev.			6.58e-03	1.02e-03
RNE			0.348	14.6
# fail			1493	0
Bias	0.40	-1.06e-03	-1.52e-04	-9.93e-06
Standard dev.		0.02	0.02	1.08e-03
RNE		0.05	0.031	13.1
# fail			335	0
Bias	0.75	-8.03e-05	-3.02e-06	-6.53e-06
Standard dev.		1.92e-03	1.83e-03	8.96e-04
RNE		4.62	5.09	21.1
# fail			0	0

Note: IS-MC estimation of  $\int g(z)p(z)dz$  where  $p$  is a Gaussian mixture of the form  $N(0, \sigma_1^2)p_1 + N(0, \sigma_2^2)p_2$  with  $p_1 = p_2 = 0.5$  and  $\sigma_2^2 = 2 - \sigma_1^2$ , and  $g$  is a  $N[0, \exp(z)]$ -density.

IS simulation sample size is  $M = 32$  and the number of EIS iterations is 7. Mean biases relative to the true value, standard deviations and RNEs are based upon 10000 replications under different CRN seeds. # fail denotes the number of replications where the EIS approximation failed to converge. The true values of the integrals were computed using high precision numerical integration.



Table 3. ML estimation results for  
state space models for realized volatility.

	estimate	MC-std.	statistical std.
student- $t$ measurement error			
$\phi$	0.983	7.82e-05	0.005
$\mu$	0.383	2.07e-04	0.195
$\sigma_\lambda$	0.154	3.71e-04	0.008
$\sigma_y$	0.372	3.79e-04	0.009
$\nu$	6.435	0.105	1.382
log-likelihood	-1328.2	0.472	
Gaussian measurement error			
$\phi$	0.982		0.005
$\mu$	0.382		0.185
$\sigma_\lambda$	0.162		0.010
$\sigma_y$	0.366		0.008
log-likelihood	-1353.1		

Note: The estimated model is given by Equations (41) and (42).

Table 4. *ML estimation results for the SV model with leverage effects and jumps for returns.*

	estimate	MC-std.	statistical std.
SV model with jumps			
$\gamma$	-1.80e-04	4.65e-05	0.018
$\sigma_y$	0.910	2.68e-04	0.097
$\sigma_z$	0.147	1.46e-04	0.015
$\rho$	-0.889	5.21e-04	0.039
$\phi$	0.985	2.43e-05	0.003
$\sigma_u$	1.484	0.020	0.760
$\log(p_1)$	-5.156	0.046	1.658
log-likelihood	-2718.7	0.060	
Akaike	5451.4		
SV model without jumps			
$\gamma$	-0.003	1.78e-05	0.019
$\sigma_y$	0.920	7.80e-05	0.091
$\sigma_z$	0.148	1.83e-04	0.013
$\rho$	-0.875	4.70e-04	0.052
$\phi$	0.985	3.01e-05	0.003
log-likelihood	-2723.8	0.075	
Akaike	5457.6		

Note: The estimated model is given by Equations (53) and (54).

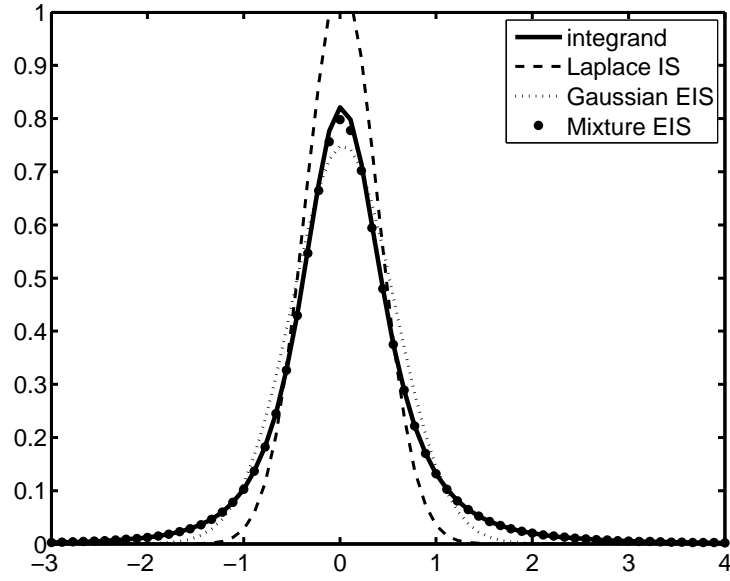


Figure 1. Target integrand  $g(z)p(z)$  (normalized such that it integrates to 1) where  $p$  is a standardized student- $t$  density with  $\nu = 2.5$  degrees of freedom and  $g$  is a  $N[z, 4^2]$ -density, Gaussian Laplace IS sampling density, Gaussian EIS sampling density, and marginal density for  $z$  associated with the joint mixture EIS sampling density  $m(z, w; \hat{a})$ .

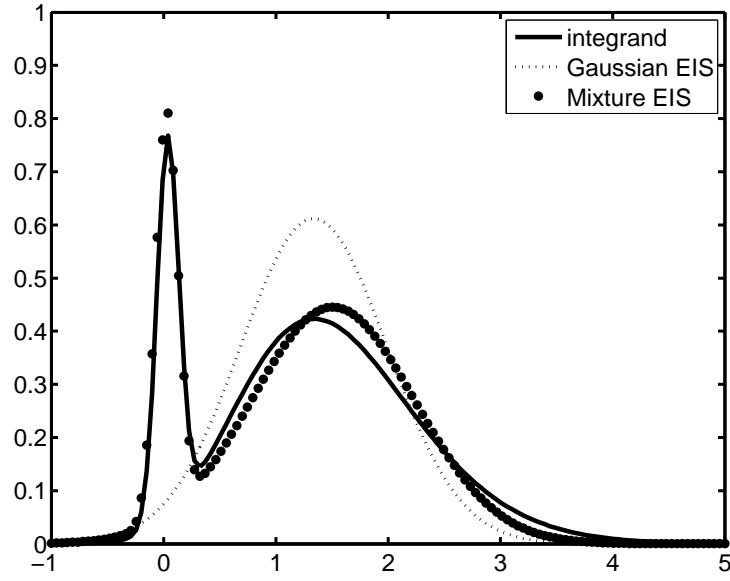


Figure 2. Target integrand  $g(z)p(z)$  (normalized such that it integrates to 1) where  $p(z) \sim 0.5 \cdot N(0, \sigma_1^2) + 0.5 \cdot N(0, \sigma_2^2)$  with  $\sigma_2^2 = 2 - \sigma_1^2$ ,  $\sigma_1 = 0.1$  and  $g$  is a  $N[0, \exp(z)]$ -density, Gaussian EIS sampling density, and mixture EIS sampling density.