CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

ECONSTOR
WWW.ECONSTOR.EU

Der Open-Access-Publikationsserver der ZBW – Leibniz-Informationszentrum Wirtschaft
The Open Access Publication Server of the ZBW – Leibniz Information Centre for Economics

Liesenfeld, Roman; Richard, Jean-François

Working Paper

Improving MCMC Using Efficient Importance Sampling

Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2006,05

Provided in cooperation with:
Christian-Albrechts-Universität Kiel (CAU)

Suggested citation: Liesenfeld, Roman; Richard, Jean-François (2006) : Improving MCMC Using Efficient Importance Sampling, Economics working paper / Christian-Albrechts-Universität Kiel, Department of Economics, No. 2006,05, http://hdl.handle.net/10419/22010

Nutzungsbedingungen:
Die ZBW räumt Ihnen als Nutzerin/Nutzer das unentgeltliche, räumlich unbeschränkte und zeitlich auf die Dauer des Schutzrechts beschränkte einfache Recht ein, das ausgewählte Werk im Rahmen der unter
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
nachzulesenden vollständigen Nutzungsbedingungen zu vervielfältigen, mit denen die Nutzerin/der Nutzer sich durch die erste Nutzung einverstanden erklärt.

Terms of use:
The ZBW grants you, the user, the non-exclusive right to use the selected work free of charge, territorially unrestricted and within the time limit of the term of the property rights according to the terms specified at
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
By the first use of the selected work the user agrees and declares to comply with these terms of use.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

Improving MCMC Using Efficient Importance Sampling

by Roman Liesenfeld, Jean-François Richard

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2006-05

Improving MCMC Using Efficient Importance Sampling

Roman Liesenfeld*
Department of Economics, Christian-Albrecht-Universität, Ohlshausenstr. 40-60, 24118 Kiel, Germany

Jean-François Richard
Department of Economics, University of Pittsburgh, 4S01 Wesley W. Posvar Hall, Pittsburgh, PA 15260, USA

May 15, 2006

Abstract

This paper develops a systematic Markov Chain Monte Carlo (MCMC) framework based upon Efficient Importance Sampling (EIS) which can be used for the analysis of a wide range of econometric models involving integrals without an analytical solution. EIS is a simple, generic and yet accurate Monte-Carlo integration procedure based on sampling densities which are chosen to be global approximations to the integrand. By embedding EIS within MCMC procedures based on Metropolis-Hastings (MH) one can significantly improve their numerical properties, essentially by providing a fully automated selection of critical MCMC components such as auxiliary sampling densities, normalizing constants and starting values. The potential of this integrated MCMC-EIS approach is illustrated with simple univariate integration problems and with the Bayesian posterior analysis of stochastic volatility models and stationary autoregressive processes.

Keywords: Autoregressive models, Bayesian posterior analysis, Dynamic latent variables; Gibbs sampling, Metropolis Hastings; Stochastic volatility

*Corresponding author. Tel.: +49-431-8803810; fax: +49-431-8807605. *E-mail address:* liesenfeld@stat-econ.uni-kiel.de (R. Liesenfeld)

1 Introduction

Monte Carlo (MC) simulation methods are widely used to analyze a broad range of econometric models involving integrals for which no analytical solution exist. Excellent surveys on MC based econometric methods are available (see, e.g., Geweke, 1999, Gilks et al., 1996, Gourieroux and Monfort, 1996 and Stern, 1997). Two methods dominate the field. Importance Sampling (IS) was introduced in econometrics by Kloek and van Dijk (1978) and widely used in the eighties. It became progressively supplanted by Markov Chain Monte Carlo (MCMC) in the nineties. Seminal papers are that of Gelfand and Smith (1990), proposing MCMC techniques for Bayesian computations, and that of Tanner and Wong (1987) who introduced data augmentation for the treatment of latent variables. The main argument against the use of IS is the potential non-existence of the MC sampling variance which could lead to disastrous properties of IS estimates (see, e.g., Geweke 1989, 1999 and more recently Koopman and Shephard, 2004). In the present paper we will argue that the IS and MCMC are more closely related than generally recognized and, in particular, that the argument of non-existing variance which contributed to the relative demise of IS also apply to Metropolis-Hastings (MH) procedures which represent, together with the Gibbs-sampling techniques, the most widely used MCMC algorithms.

The fundamental insight which motivates the present paper lies in the observation that the sampling properties of both IS and MCMC critically depend upon the adequacy of an auxiliary sampler m, meant to approximate (up to a proportionality constant c) a density kernel φ which needs to be numerically integrated (by itself and/or to compute expectations of functions of interest). Efficient Importance Sampling (EIS), as proposed by Richard and Zhang (2006), provides a generic and essentially automated Least Squares (LS) procedure to construct such (near) optimal approximations within a preassigned parametric class M of auxiliary samplers. Whence we should be able to facilitate the design as well as improve the sampling properties of MCMC by relying upon auxiliary EIS regressions to construct m. Most importantly, once it is recognized how closely related EIS and MCMC are, the key issue of deciding which approach to use for a particular application becomes one of objective comparison, of respective ease of implementation and statistical adequacy within the context of that application, not of subjective preference for one method or the other.

Typically, EIS can be expected to have comparative advantages at integrating out high-dimensional dynamic latent variables. The efficiency of EIS in such situations is illustrated by its application for the computation of the likelihood of various dynamic latent variable models (see, e.g., in Bauwens and Hautsch, 2003, Bauwens and Galli, 2005, Jung and Liesenfeld, 2001, and Liesenfeld and Richard

2003a, 2003b, 2005a). In particular, in the context of highly correlated latent variables, EIS can usefully be interpreted as a *single* block MCMC step, providing an efficient solution to the slow convergence of MCMC in such context. On the other hand, intricate Bayesian posterior densities are generally more amenable to MCMC integration. This is illustrated by the estimation of stochastic volatility (SV) models where MCMC and EIS are combined together along this lines (subsection 6.1).

Moreover, even in applications where MCMC is unequivocally to be preferred, we will illustrate ways of embedding EIS auxiliary steps within the MCMC algorithm in order to facilitate the construction of the MCMC sampler and, in the process, to improve its convergence properties. This will be illustrated in the context of the MCMC analysis of the roots of stationary AR models (subsection 6.2).

The rest of the paper is organized as follows. In section 2, we review the principles of the MC integration procedures under consideration. Section 3 and 4 contain a brief description of the EIS technique and MCMC approaches, respectively. In particular we, discuss how EIS can usefully be combined with MCMC. This is illustrated in section 5 and 6 with numerical examples. In particular, we discuss two simple univariate integration problems (subsections 5.1 and 5.2), a Bayesian analysis of a SV model (subsection 6.1), and a MCMC analysis of the roots of AR models (subsection 6.2). Section 7 concludes.

2 MC Integration

Consider a continuous random variable X with support $\Delta \subset \mathbb{R}^T$. We assume that its density function $f(x)$ is characterized by a density kernel $\varphi(x)$ whose integrating constant on Δ is unknown. That is to say

$$f(x) = c^{-1} \cdot \varphi(x), \quad \text{with} \quad c = \int_\Delta \varphi(x)dx. \tag{1}$$

Let $g(x)$ denote a φ–integrable function. Its expectation on f is given by

$$I_g = \mathrm{E}_f\left[g(X)\right] = \frac{\int_\Delta g(x)\varphi(x)dx}{\int_\Delta \varphi(x)dx}. \tag{2}$$

IS and MCMC (MH) are commonly used to evaluate such (ratios of) integrals. In order to motivate our paper, we first present both methods in a particular way which serves highlighting their intrinsically close relationship. Technical assumptions validating these methods are extensively discussed in the literature – see, e.g., Geweke (1989) and Robert and Casella (1999) – and are omitted here.

All MC integration techniques discussed in the present paper share the characteristic that they rely upon a sequence of 'primary' draws $\{\tilde{x}_i;\ i:1\to S\}$. While IS uses these draws as such, MH requires additional auxiliary steps based upon uniform draws in order to determine which of the primary draws are to be deleted, retained and/or repeated. For the purpose of the present discussion it proves convenient to reinterpret both IS and MCMC estimates of I_g as (randomized) weighted sums of functions of the primary draws, say

$$\bar{I}_g = \frac{\sum_{i=1}^{S} g(\tilde{x}_i)\cdot\alpha\left(\tilde{x}_{(i)},\tilde{v}_i\right)}{\sum_{i=1}^{S}\alpha\left(\tilde{x}_{(i)},\tilde{v}_i\right)}, \tag{3}$$

where $\tilde{x}_{(i)}=(\tilde{x}_j;\ j:1\to i)$ and $\tilde{v}_i$ denote additional auxiliary uniform draws as required for MCMC. Implicitly assuming convergence, let $m(x_i)$ denote the stationary margin of $\tilde{x}_i$. IS relies upon i.i.d. draws with weights

$$\alpha\left(x_{(i)},v_i\right)\equiv\omega(x_i)=\frac{\varphi(x_i)}{m(x_i)}. \tag{4}$$

There exists a wide variety of MH techniques, some of which are described below, which generally result in assigning integer values to the αs, which then represent the number of times a primary draw appears in the average. Let μ denote the (stationary) joint distribution of the auxiliary draws $(\tilde{x}_{(i)},\tilde{v}_i)$. Additional notations are:

$$\begin{aligned}
\bar{\omega} &= \mathrm{E}_m\left[\omega(\tilde{x}_i)\right]=c && (5)\\
a(x_i) &= \mathrm{E}_\mu\left[\alpha(\tilde{x}_{(i)},\tilde{v}_i)\mid\tilde{x}_i=x_i\right] && (6)\\
\bar{a} &= \mathrm{E}_m\left[a(\tilde{x}_i)\right]. && (7)
\end{aligned}$$

Next, we reinterpret $\bar{I}_g$ as an MC-estimator of I_g, i.e. a decision rule which maps a φ-integrable function g into a point estimate of its expectation on the density f. Under standard technical conditions consistency of $\bar{I}_g$ obtains under the following condition

$$\mathrm{E}_m\left[g(\tilde{x}_i)\cdot\frac{a(\tilde{x}_i)}{\bar{a}}\right]=\mathrm{E}_m\left[g(\tilde{x}_i)\cdot\frac{\omega(\tilde{x}_i)}{\bar{\omega}}\right]=I_g \tag{8}$$

for all φ-integrable g. This implies that there exists a constant $b>0$ such that

$$a(x_i)=b\cdot\omega(x_i),\quad m-\text{a.s.}\ , \tag{9}$$

where b accounts for different implicit normalization rules associated with the weights. It follows that

$$\begin{aligned}
\mathrm{Var}_\mu\left[\alpha\left(\tilde{x}_{(i)},\tilde{v}_i\right)\right] &= \mathrm{E}_m\left[\mathrm{Var}\left(\alpha\left(\tilde{x}_{(i)},\tilde{v}_i\right)\right)\mid\tilde{x}_i\right]+b^2\mathrm{Var}_m\left[\omega(\tilde{x}_i)\right] && (10)\\
&\geq b^2\mathrm{Var}_m\left[\omega\left(\tilde{x}_i\right)\right].
\end{aligned}$$

Due to the complex correlation structure of the α weights under MH schemes, the inequality (10) does not necessarily imply that IS is more efficient than MH. However, it has two fundamental implications which motivate the present paper: (i) The criticism commonly raised against IS that $\text{Var}[\omega(X)]$ might not exist also applies to MH, a fact which is often ignored in the literature on MCMC and cannot by itself justify the demise of IS procedures. (ii) More constructively, as we shall discuss further below, the MH α weights are produced by accept-reject steps based upon ratios of the form $\omega(x)/c$, where c is a calibration constant. Efficient MH algorithms are those for which the distribution of these ratios is tightly concentrated around one. Since, as discussed further below, EIS is designed with that objective, we should be able to facilitate the construction of MH algorithms and improve their efficiency by relying upon EIS auxiliary steps to construct the MH samplers. This is the objective of the present paper.

3 Efficient Importance Sampling (EIS)

The EIS procedure proposed by Richard and Zhang (2006) – hereafter RZ – provides a generic auxiliary LS algorithm to select an efficient sampler within a preassigned class M. For low-dimensional problems it approximates the integrand $\varphi(x)$ – itself a density kernel – by a kernel $k(x,a)$ for $a \in A$. The correspondence between k and m is given by

$$m(x;a) = \frac{k(x,a)}{\chi(a)}, \quad \text{with} \quad \chi(a) = \int_X k(x,a)dx. \tag{11}$$

A (near) optimal value of a obtains by solving the following auxiliary generalized LS (GLS) problem

$$\hat{a} = \arg\min_{a \in A} q(a),$$

$$q(a) = \int_X d^2(x,a) \cdot \omega(x,a) \cdot m(x;a)dx, \tag{12}$$

with

$$d(x,a) = \ln \varphi(x) - \gamma - \ln k(x,a), \tag{13}$$

$$\omega(x,a) = \frac{\varphi(x)}{m(x;a)}, \tag{14}$$

and $\gamma = \ln c$ is an intercept (calibrating constant), which is included in a for the ease of notation. For a tentative value $\hat{a}_j \in A$, an MC estimate of $q(a)$ is given by

$$\bar{q}_S(a \mid \hat{a}_j) = \frac{1}{S}\sum_{i=1}^{S} \left[\ln \varphi\left(\tilde{x}_i^j\right) - \gamma - \ln k\left(\tilde{x}_i^j, a\right)\right]^2 \cdot \omega\left(\tilde{x}_i^j, \hat{a}_j\right), \tag{15}$$

where $\{\tilde{x}_i^j;\ i:1\to S\}$ are i.i.d. draws from $m(x;\hat{a}_j)$.

Let $\hat{a}_{j+1}$ denote the value of a which minimizes $\bar{q}_S(a\mid\hat{a}_j)$. EIS consists of recursively solving the GLS problem associated with Equation (15) until a fixed point solution obtains whereby $\hat{a}_{j+1}\simeq\hat{a}_j$. In earlier iterations where the variance of ω can be large, it is recommended to set the ωs in Equation (15) equal to one in order to avoid imbalances due to large weights. Note that if k belongs to the exponential family of distribution, then it can be parameterized in such a way that $\ln k$ is linear in a and the search for $\hat{a}$ is reduced to a sequence of linear LS problems.

High-dimensional EIS requires that $\varphi(x)$ be partitioned into low-dimensional components in accordance with a natural (model based) sampling preordering (allowing for parallel sequences under appropriate conditional independence assumptions as for panels). Let x be partitioned conformably with such a preordering into $x=(x_1,\ldots,x_L)$, under the convention that x_ℓ is to be drawn conditionally on $x_{(\ell-1)}=(x_1,\ldots,x_{\ell-1})$. Both φ and m are factorized accordingly into

$$\varphi(x)=\prod_{\ell=1}^{L}\varphi_\ell\left(x_{(\ell)}\right) \tag{16}$$

$$m(x;a)=\prod_{\ell=1}^{L}m_\ell\left(x_\ell\mid x_{(\ell-1)},a_\ell\right). \tag{17}$$

Note that $\varphi_\ell(x_{(\ell)})$, which is given by the model specification, is typically not a kernel for the conditional density $f_\ell(x_\ell\mid x_{(\ell-1)})$. Kernels for f_ℓ obtain from the backward recursion

$$\varphi_\ell^*\left(x_{(\ell)}\right)=\varphi_\ell\left(x_{(\ell)}\right)\cdot\int\varphi_{\ell+1}^*(x_{(\ell+1)})\cdot dx_{\ell+1}, \tag{18}$$

together with $\varphi_L^*=\varphi_L$. Such integrals are generally intractable which is precisely why MC is used. However, a similar recursive procedure can be applied to a sequence of (operational) IS approximating kernels $k_\ell(x_{(\ell)},a_\ell)$. Specifically, the ratio φ/m can be rearranged as

$$\frac{\varphi(x)}{m(x;a)}=\chi_1(a_1)\cdot\prod_{\ell=1}^{L}\left[\frac{\varphi_\ell(x_{(\ell)})\cdot\chi_{\ell+1}(x_{(\ell)},a_{\ell+1})}{k_\ell(x_{(\ell)},a_\ell)}\right], \tag{19}$$

together with

$$\chi_{\ell+1}\left(x_{(\ell)},a_{\ell+1}\right)=\int k_{\ell+1}(x_{(\ell+1)},a_{\ell+1})dx_{\ell+1} \tag{20}$$

and $\chi_{L+1}(\cdot)\equiv 1$. Step j of the EIS fixed point search for $\hat{a}=(\hat{a}_\ell;\ \ell:1\to L)$ consists of drawing full trajectories $\{\tilde{x}_{i\ell}^j;\ \ell:1\to L;\ i:1\to S\}$ from the previous step sampler $m(x;\hat{a}_j)$ and solving a backward sequence of auxiliary GLS problems – of the form given by Equation (15) – whereby $\ln(\varphi_\ell\cdot\chi_{\ell+1})$ is approximated by $\ln k_\ell$ as a function of $x_{(\ell)}$.

High-dimensional EIS implicitly assumes that $\varphi_\ell \cdot \chi_{\ell+1}$ and, therefore, k_ℓ only depend on a sufficiently low-dimensional subvector of $x_{(\ell)}$ in accordance with conditional independence assumptions built into the model under consideration (see RZ, 2006 for full notation and implementation details).

It is important to emphasize that in order to secure smooth convergence towards a fixed point solution $\hat{a}$, all EIS iterations are to be run under Common Random Numbers (CRNs), in the sense that all auxiliary draws $\{\tilde{x}_i^j;\ i:1\rightarrow n\}$ are constructed as transformation of a single set of canonical draws $\{\tilde{u}_i;\ i:1\rightarrow n\}$, i.e. draws whose densities do not depend on a (for example, uniforms on $(0,1)$ from which the xs obtain by inversion of the cdf). As discussed in RZ (2006), the use of CRNs also robustifies EIS against outlier draws in the sense that outliers from $m(x;\hat{a}_j)$ will be highly influential in the auxiliary (G)LS problem, as given by Equation (15) and will generate an immediate adjustment in $\hat{a}_{j+1}$. While the use of CRNs thereby critically contributes reducing the estimated variance of $\bar{I}_g$, it obviously does not eliminate the possibility that this variance might actually not exist (whether under EIS or MCMC) as would be the case if the tails of k are thinner than those of φ (examples are provided below). RZ (2006) propose a simple and highly sensitive test to detect such imbalances. It consists of computing the ratio between two alternative MC estimates of the MC sampling variance of $\bar{I}_g$, one based upon draws from the EIS sampler $m(x;\hat{a})$ itself and the other from an alternative sampler $m(x;a_0)$ with artificially inflated variance (by a factor of 3 to 5). This ratio is given by

$$\Gamma_S^2\left(\hat{a},a_0\right)=\frac{\hat{\sigma}_S^2\left(\hat{a},a_0\right)}{\hat{\sigma}_S^2\left(\hat{a},\hat{a}\right)}, \tag{21}$$

where

$$\hat{\sigma}_S^2\left(\hat{a},a\right)=\frac{1}{S}\sum_{i=1}^{S}h\left[d^2\left(\tilde{x}_i^a,\hat{a}\right)\right]\cdot\frac{\varphi\left(\tilde{x}_i^a\right)}{m\left(\tilde{x}_i^a;a\right)}, \tag{22}$$

$$h(c)=e^{\sqrt{c}}+e^{-\sqrt{c}}-2, \tag{23}$$

and $\{\tilde{x}_i^a;\ i:1\rightarrow S\}$ denotes (CRN) draws from $m(x;a)$. (The MC sampling variance of $\bar{I}_g$ under a sampler $m(x;a)$ is given by $\{\int h[d^2(x,a)]\varphi(x)dx\}/S$, see RZ, 2006.) One might consider calibrating such a ratio by means of auxiliary MC simulations. Experience suggests that such calibration is hardly necessary as Γ_S^2 rapidly explodes in the presence of even mild imbalances between the tails of m and those of φ. Note that by inflating the variance of the sampler $m(x;a)$ used to compute $\hat{\sigma}_S^2\left(\hat{a},a\right)$, one increases the probability of drawing points in the potentially critical regions. Hence, in the presence of a thin-tail problem one would expect $\hat{\sigma}_S^2\left(\hat{a},a\right)$ to increase rapidly with the variance of $m(x;a)$. Illustrations are provided in section 5 below.

In conclusion of this brief description of EIS, we ought to mention that, relative to MH, EIS offers an additional degree of flexibility which can produce additional and occasionally important efficiency

gains in the computation of ratios of integrals such as I_g in Equation (2). Consider first the case where $g(x) > 0$ on X. It is then trivial to compute separate EIS samplers for the numerator and denominator of I_g (under a single set of CRNs). Specifically, let $m(x; \hat{a}_N)$ denote an EIS sampler for the product $g \cdot \varphi$ and $m(x; \hat{a}_D)$ one for φ alone. Let $(\tilde{x}_{Ni}; i : 1 \to S)$ and $(\tilde{x}_{Di}; i : 1 \to S)$ denote corresponding sets of draws obtained by transformation of a single set of CRNs. An EIS-2 samplers estimate of I_g is then given by

$$\bar{I}_{g2} = \frac{\sum_{i=1}^{S} \omega_N(\tilde{x}_{Ni})}{\sum_{i=1}^{S} \omega_D(\tilde{x}_{Di})}, \tag{24}$$

where $\omega_N = \varphi \cdot g / m(\cdot; \hat{a}_N)$ and $\omega_D = \varphi / m(\cdot; \hat{a}_D)$. Illustrations of the significant efficiency gains produced by EIS-2 samplers for $g(x) = x$ with $x > 0$ and $g(x) = x^2$ are provided in section 5 below. A similar idea applies to situations when $g(x)$ though not strictly positive on X can be decomposed into the product of a positive function and a remainder. Note that there is no MH counterpart to EIS-2 samplers since MH aims at producing draws from φ itself, from which expectations are evaluated as simple arithmetic means.

4 Markov Chain Monte Carlo (MCMC)

The object of MCMC is to construct an auxiliary Markov Chain whose stationary distribution is the one associated with the density kernel φ. Such chains include acceptance steps. Whence, in line with our discussion in section 2, we introduce an explicit distinction between 'primary' draws $(\tilde{x}_i)$ and 'accepted' draws $(\tilde{y}_i)$. As for EIS, we first discuss low-dimensional integration problems.

A baseline MH Markov Chain consists of two components:

(i) A transition probability density $m(x \mid y)$;

(ii) An acceptance probability $\rho(x \mid y)$ defined as:

$$\rho(x \mid y) = \min \left[\frac{\varphi(x)}{m(x \mid y)} \cdot \frac{m(y \mid x)}{\varphi(y)}, 1 \right]. \tag{25}$$

Assuming convergence after an initialization run, the MH algorithm proceeds as follows: Conditionally on the latest accepted draw $\tilde{y}_i$, draw $\tilde{x}_{i+1}$ from $m(x \mid \tilde{y}_i)$ and accept $\tilde{x}_{i+1}$, i.e. set $\tilde{y}_{i+1} = \tilde{x}_{i+1}$, with probability $\rho(\tilde{x}_{i+1} \mid \tilde{y}_i)$. Otherwise, set $\tilde{y}_{i+1} = \tilde{y}_i$. The corresponding MH estimate of I_g is a simple arithmetic mean over accepted draws

$$\bar{I}_g = \frac{1}{S} \sum_{i=1}^{S} g(\tilde{y}_i). \tag{26}$$

It trivially corresponds to formula (3) when reformulated in terms of the $\tilde{x}_i$s with $\alpha(\tilde{x}_{(i)}, v_i)$ representing the number of time a primary draw was counted in the sum. Note, in particular that a $\tilde{y}_i$ which obtains as the outcome of s successive rejections equals $\tilde{x}_{i-s}$, and that a rejected primary draw $\tilde{x}_i$ receives an α weight of zero. The MH algorithm as defined above is quite general in the sense that $\bar{I}_g$ converges almost surely toward I_g as long as the chain $\{\tilde{y}_i\}$ is ergodic (see, e.g. Chib and Greenberg, 1995 or Robert and Casella 1999).

A special version of MH is the independent MH (see Tierney, 1994), which is widely used in econometrics as a key component of MCMC algorithms. This method, which is most closely related to IS, relies upon auxiliary sampling densities which are independent from previous draws, in which case $m(x \mid y) = m(x)$. The corresponding acceptance probability ρ is then given by

$$\rho(x \mid y) = \min\left[\frac{\omega(x)}{\omega(y)}, 1\right], \tag{27}$$

with $\omega(x) = \varphi(x)/m(x)$. A necessary and sufficient condition for ergodicity of a chain from an independent MH (and hence for convergence) is that m is almost everywhere positive on the support Δ of φ, which is not particularly restrictive per se (see Robert and Casella, 1999, Chapter 6.2). This being said, the rate of convergence and the efficiency of MH critically depends upon how well m approximates φ – up to a multiplicative constant c which actually cancels out in Equations (25) and (27). Note, in particular, that the better the approximation of φ by m, the lower the variation in the ratio $(\min\{\omega(\tilde{x})/\omega(\tilde{y}), 1\})$, and hence, the higher the average acceptance rate. Furthermore, only for densities m, satisfying

$$\omega(x) < \infty \quad \forall x \in \Delta,$$

a fast exploration of the support Δ implied by uniform ergodicity of the resulting chain $\{\tilde{y}_i\}$ can be guaranteed (see Robert and Casella, 1999, Chap. 6.3). Without such a strong ergodicity property, which is rarely met in practical applications, the convergence can be very slow. In particular observe, that when m is lighter-tailed than φ, the algorithm can get stuck at points in the tails with very low acceptance rates for new candidate draws, leading to potentially very large α weights for such points. Finally, notice the close accordance between the condition for uniform ergodicity and the requirement for a finite MC variance of IS procedures.

It follows that in order to construct (independent) MH algorithms with fast convergence and high efficiency, one should select a density m which closely approximates φ and, in particular, such that the ratio φ/m is bounded. Therefore, subject to the usual thin-tail caveat, an EIS density m which, as discussed in section 2, provides a global LS approximation to φ should also result in high acceptance rates when used as an MH sampler. This even though there remains a conceptual

difference between the intrinsic objectives of optimizing an IS procedure and improving the rate of convergence of MH (through faster exploration of Δ). Nevertheless, as illustrated below, efficient IS samplers can significantly accelerate the convergence of MH.

MH is often used in combination with other (MC)MC techniques. In particular, the combination of independent MH with Accept-Reject (AR) sampling leads to the AR-MH algorithm proposed by Tierney (1994). In the initial AR step of AR-MH, candidates drawn from m are accepted with probability

$$p(x) = \min\left[\frac{1}{c}\omega(x), 1\right], \tag{28}$$

where $c > 0$ represents a calibration (tuning) parameter. MH is then applied to the accepted (primary) draws $\tilde{x}_i$ from this initial step, with a probability ρ modified as follows:

$$\rho(x \mid y) = \min\left[\frac{\varphi(x) \cdot \min\left[\varphi(y), c \cdot m(y)\right]}{\varphi(y) \cdot \min\left[\varphi(x), c \cdot m(x)\right]}, 1\right]. \tag{29}$$

The AR-MH algorithm is particularly useful in situations when the precondition for pure AR sampling – i.e. the existence of a constant c such that $\varphi(x) \leq c \cdot m(x), \forall\, x \in \Delta$ – either is hard to verify or leads to unacceptably high numbers of rejections. On the other hand, it is obvious that the convergence of an AR-MH algorithm also critically depends upon the constant c. In the absence of operational optimality criterion for the selection of c, rules of thumb are widely used in applications (see, e.g., Jacquier et al., 1994 or Chib and Greenberg, 1995). In contrast, the use of EIS to construct an AR-MH auxiliary sampler m offers the additional benefit that the tuning parameter c is a fully automated by-product obtained from the estimation of the intercept γ in the EIS regression (15). In particular, the estimate $\hat{\gamma}$ captures the average difference between $\ln\varphi$ and the EIS approximation $\ln k(x, \hat{a})$. Hence, by setting $c = \chi(\hat{a}) \cdot e^{\hat{\gamma}}$, one can expect a balanced trade-off between too frequent rejection in the initial AR-step from high c values and too frequent repetitions in the subsequent MH-step due to low c values. Accordingly, automated selection of c should be an important contribution of EIS within the AR-MH algorithm which, as illustrated below, can considerably accelerate convergence.

In higher dimensional problems, Gibbs sampling (itself an MCMC algorithm) is often used in combination with MH. Gibbs relies upon a partitioning of x into low-dimensional components, say $x = (x_1, \ldots, x_L)$ such that the conditional densities $f(x_\ell \mid x_{\setminus\ell})$, where $x_{\setminus\ell} = (x_i; i \neq \ell)$, are MC tractable, in the sense that they are either amenable to direct simulation, or can be sampled through MC approximations. Gibbs then repeatedly cycles across the complete set of such conditional densities until convergence obtains. While Gibbs offers the advantage that its low-dimensional components can be easy to draw from, it suffers from the fact that high correlations among xs can dramatically slow convergence and, therefore, produce highly inefficient estimates of I_g (see, e.g., Carter and Kohn, 1994

or Shephard and Pitt, 1997). Important examples of such problems are high-dimensional dynamic latent variable models such as SV models, an example of which is discussed below. Bayesian analysis of such models require integrating out a high-dimensional vector λ of latent variables in addition to a vector θ of unknown parameters, in which case $x = (\theta, \lambda)$. Partitioning λ into blocks instead of individual variables alleviates the high-correlation problem but also complicates approximating the corresponding (block) conditional densities. On the other hand, EIS has proved to be very efficient for (single block) high-dimensional integration of $\lambda \mid \theta$ (see, e.g., Liesenfeld and Richard, 2003b). We shall present below a highly efficient Gibbs-EIS-MH combination for the Bayesian analysis of SV models, whereby EIS is used to draw complete (single block) trajectories for $\lambda \mid \theta$, and conventional MH-Gibbs is used to draw $\theta \mid \lambda$.

5 Two Univariate Examples

In this section we discuss two simple univariate examples, illustrating how EIS can be used to automate the selection of an MH sampler (within a preassigned class). We show that the performance of such an EIS-MH combination is very close to that of EIS itself. The auxiliary statistic Γ_S defined in Equation (21) is used under both EIS-MH and EIS to detect thin-tail problems. We also illustrate the fact that neither EIS nor EIS-MH can match EIS-2 samplers for the estimation of (positive) moments.

5.1 Inverse Gaussian Density

Consider the computation of the mean $\mathrm{E}_f(x)$, when x follows an inverse Gaussian distribution f with a density kernel

$$\varphi(x, \theta) = x^{-3/2} e^{-\theta_1 x - \theta_2 / x}, \quad x > 0, \theta_1 > 0, \theta_2 > 0, \tag{30}$$

where $\theta = (\theta_1, \theta_2)'$. The analytical form of the mean is $\mathrm{E}_f(x) = \sqrt{\theta_2/\theta_1}$. Following Robert and Casella (1999, Chap. 6.4.1) – hereafter RC – let k be the following gamma kernel

$$k(x, a) = x^{\kappa - 1} e^{-x/\delta}, \quad x > 0, \kappa > 0, \delta > 0. \tag{31}$$

For MC estimation of $\mathrm{E}_f(x)$ based upon independent MH, RC (1999) propose to select a value of $a = (\kappa, \delta)'$ which (approximately) maximizes the MH acceptance rate subject to the simplifying assumption that the mean of the auxiliary sampler $m(x; a)$ coincides with that of f. This requires setting $\kappa = \sqrt{\theta_2/\theta_1}/\delta$. The acceptance rate to be maximized in δ is also subject to the constraint

that the ratio

$$\omega(x,a) = \frac{f(x)}{m(x;a)} \propto x^{-\kappa-1/2} e^{(1/\delta-\theta_1)x-\theta_2/x} \tag{32}$$

be bounded. The latter requires that $1/\theta_1 < \delta$ and ensures uniform ergodicity. Since the expected acceptance rate is impossible to compute it has to be approximated by simulation. For $\theta_1 = 1.5$ and $\theta_2 = 2$, the optimal value for δ subject to the constraint $1/1.5 < \delta$, as obtained by RC (1999), lies at the boundary $\delta^* = 1/1.5$. The corresponding value for a is given by $a^* = (\sqrt{\theta_2/\theta_1}/\delta^*, \delta^*)'$.

Note that the procedure described above is not easy to generalize to a wider range of applications. Optimization based upon estimated acceptance rates is a non-trivial and time consuming operation (with a discontinuous objective function). Nor would E_f be generally available in which case the RC approach would require iterations based upon intermediate estimates of E_f.

In contrast (unconstrained) EIS relies upon trivial bivariate linear auxiliary regression associated with the following expression for $d(x,a)$ as defined in Equation (13):

$$d(x,a) = \left[-\frac{3}{2}\ln x - \theta_1 x - \frac{\theta_2}{x}\right] - \gamma - \left[\alpha_1 \ln x + \alpha_2 x\right], \tag{33}$$

where $\alpha_1 = \kappa - 1$, $\alpha_2 = -1/\delta$, and γ denotes the regression intercept. It should be noted that the variance of $\omega(x,a)$ remains finite even when $1/\theta_1 > \delta$, in which case $\omega(x,a)$ is unbounded. Actually, the MC variance of $\omega(x,a)$ can be expressed in terms of Bessel functions of imaginary argument (see Gradshteyn and Ryzhik, 1979, 3.471.9). Whence we do not impose the uniform ergodicity condition and, unsurprisingly in view of the RC results, find that the EIS solution $\hat{a}$ violates that condition.

Results are reported in Table 1 for the RC approach based on the optimized acceptance rate and for the corresponding MH-EIS (using 20 EIS iterations to obtain the final EIS value $\hat{a}$). The results which are reported are sample means and standard deviations based upon 100 independent replications of the complete algorithms, providing a reliable measure of numerical accuracy. Individual runs are based upon $S = 5,000$ draws. The results of the experiment indicate that the optimized MH algorithm based on $m(x;a^*)$ and the MH-EIS procedure using $m(x;\hat{a})$ perform very similarly with respect to numerical efficiency: while the former has a MC standard deviation of 0.011, the latter has one of 0.013. The acceptance rates for the MH-EIS procedure (90%) is significantly larger than that for the optimized MH (71%). This indicates that the EIS kernel $k(x,\hat{a})$ based on an unconstraint GLS optimization associated with Equation (33) provides a significantly better global approximation to $\varphi(x,\theta)$ than the kernel $k(x,a^*)$, which is selected subject to the constrained that the ratio (32) remains bounded. For comparison, we considered the direct EIS estimate of $\mathrm{E}_f(x)$ using the same EIS-sampler as for MH-EIS. The MC standard deviation of this estimate (EIS-1) provided in the third row of Table 1 indicates a numerical efficiency which is close to that of the MH procedures.

In order to make sure that the unboundedness of $\omega(x, \hat{a})$ does not adversely impact the estimation of $\mathrm{E}_f(x)$ we computed the Γ_S statistic, as defined in Equation (21), under variance ratios varying from 5 to 30 (replacing $(\hat{\kappa}, \hat{\delta})$ by $(\hat{\kappa}/q, q\hat{\delta})$ amounts to multiplying the variance of $m(x; \hat{a})$ by q). We find that Γ_S remains tightly distributed around one in all cases (only the results for $q = 5$ are reported in Table 1), indicating the complete absence of adverse thin-tail effects. Clearly, MH-EIS provides a fully operational (and generic) alternative to the RC procedure as well as one which enables us to verify that we can safely relax the boundary condition in this application.

Last but not least, in order to illustrate the flexibility and full potential of EIS in the present case, we also computed an EIS-2 samplers estimate of $\mathrm{E}_f(x)$ whereby, as described in section 2, separate EIS samplers (under CRNs) are used to approximate $x \cdot \varphi$ and φ, respectively. The EIS for $x \cdot \varphi$ only requires adding $\ln x$ to $d(x, a)$ in Equation (33) (which represents the EIS for φ). The results for this EIS-2 are reported in the last row of Table 1. Note the large efficiency gain with MC-variance reduction by a factor of 150.

The results discussed above are also confirmed by Figure 1, where we reproduce plots of MC estimates of $\mathrm{E}_f(x)$ under increasing size S for MH based on optimized acceptance rate, MH-EIS, and EIS-2. Note that both MH procedures have similar convergence patterns with MH-EIS performing somewhat better. Clearly none of the MH approaches can remotely match the performance of EIS-2.

5.2 Student-t Density

In our second example, we consider the classical pathological problem of approximating a Student-t by a Normal density. In particular, we focus on the MC estimation of $\mathrm{E}_f(x^2)$ where f is a standardized t density with a kernel

$$\varphi(x, \nu) = \left(1 + \frac{x^2}{\nu - 2}\right)^{-(\nu+1)/2}, \tag{34}$$

such that $\mathrm{E}_f(x) = 0$, and $\mathrm{E}_f(x^2) = 1$ for all degrees of freedom $\nu > 2$. The normal sampling density kernel under consideration is parameterized as

$$k(x, a) = e^{-ax^2/2}, \quad a > 0, \tag{35}$$

such that $\mathrm{E}_m(x) = 0$ and $\mathrm{E}_m(x^2) = a^{-1}$. Notice that $\lim_{\nu \to \infty} f(x; \nu) = m(x; 1)$. The EIS univariate linear regression for φ corresponds to the following expression for $d(x, a)$

$$d(x, a) = \left[-\frac{1}{2}(\nu + 1) \cdot \ln\left(1 + \frac{x^2}{\nu - 2}\right)\right] - \gamma - \alpha x^2, \tag{36}$$

where $\alpha = -a/2$. Our initial sampler is a $\mathrm{N}(0, 1)$-density obtained by setting $\alpha_0 = -1/2$.

Table 2 and Figure 2 summarize the result for EIS and (independent) MH-EIS estimation of $\mathrm{E}_f(x^2)$. Since draws for MH-EIS need to be from f itself, we use first an EIS for φ as defined by Equation (36). The maximum number of EIS iterations is conservatively set equal to 100, as low degrees of freedom require more iterations. As in the example of subsection 5.1, we computed also direct EIS-1 estimates of $\mathrm{E}_f(x^2)$ using the same EIS sampler as for the MH-EIS as well as EIS-2 estimates using different EIS samplers for the numerator and denominator of $\mathrm{E}_f(x^2)$ though based upon a single set of CRNs (see, fifth and sixth row of Table 2). The EIS for the denominator only requires adding $\ln x^2$ to $d(x,a)$ in Equation (36). The results in Table 2 are means and standard deviations from 100 replications of the full estimation procedure, each based upon a sample size $S = 1,000$. The Γ_S statistic as defined in Equation (21), is computed using the EIS sampler and another sampler with variance equal to five times that of the EIS sampler (obtained by setting $a_0 = 5/\hat{a}$).

In order to interpret the results for MH-EIS and EIS-1 one needs to account for the fact that low degrees of freedom student t densities are significantly tighter around their mode than a $\mathrm{N}(0,1)$-density – in order to compensate for fatter tails. Therefore, an EIS sampler approximating φ will have smaller variance ($\hat{a} > 1$) which will result in greater downward bias when used to evaluate $\int x^2\varphi(x,\nu)dx$ in the numerator. Accordingly, decreasing the degrees of freedom ν (fatting the tails) increases rapidly the downward bias of MH-EIS and EIS-1 estimates of $\mathrm{E}_f(x^2)$. Under EIS-2 a different sampler is used for that numerator which explains the significantly lower biases in row 6 of Table 2 (see also the top line in Figure 2).

Note that, in sharp contrast with the previous example, the thin-tail issue matters greatly here as indicated by rapidly exploding values of Γ_S for low degrees of freedom. Figure 2 confirms our findings and indicates that the bias in $\mathrm{E}_f(x^2)$ estimates under EIS-MH and EIS-1 does not decrease much for $\nu = 2.5$ as sample size S increases up to 50,000. The occasional significant jumps are the results of CRN outlier draws affecting the integral of $x^2\varphi$. In contrast, the EIS estimate of the integral of φ remains extremely robust against outliers due to the use of CRNs, as explained in section 3. This is illustrated by the results for EIS estimation of $I = \int \varphi(x,\nu)dx$ reported in Table 3, where we artificially insert a single very large outlier ($x = 6$ corresponding to a p-value of 9.8e–10) in the first set of 1,000 $\mathrm{N}(0,1)$ draws. The exact value of I equals 1.2360. Note the minimal impact of that outlier, except for an increase from 0.017 to 0.027 in the MC standard deviation of $\bar{I}$. This is due to the fact that EIS adjusts each $\hat{a}$ to the particular set of CRNs being used. The third row of Table 3 indicates that if we prevent such adjustments by keeping $\hat{a}$ fixed, then $\bar{I}$ exhibits the traditional erratic behavior under outlier(s).

These two (classical) univariate examples have highlighted several key findings: (i) MH is subject to the same caveat as (E)IS relative to the thin-tail problem; (ii) Maximally efficient EIS estimation of positive moments obtains when numerator and denominator are treated separately (under a common set of CRNs) – note that EIS estimation of the integrals of φ and $g \cdot \varphi$ (for $g > 0$) only require minimal program changes; (iii) The reliance upon a single set of CRNs from the fixed point search for $\hat{a}$ to the final estimation of I_g is a critical component of EIS robustness against outliers and (iv) the Γ_S statistic defined in Equation (21) effectively discriminates between situations where the unboundedness of $\omega(x)$ is inconsequential (example 5.1) from those where it is critical (example 5.2).

6 Two Higher-dimensional Examples

We now consider two higher-dimensional full-fledged applications illustrating how EIS and MH can be combined together for greater numerical efficiency.

6.1 Bayesian Analysis of a Stochastic Volatility Model

SV models have received considerable attention in financial econometrics as specifications accounting for the dynamic behavior of the volatility of financial returns (see Ghysels et al., 1996 or Shephard, 2004). The standard univariate SV model has the form

$$y_t = \beta e^{\lambda_t/2} u_t, \qquad \lambda_t = \delta \lambda_{t-1} + \nu v_t, \qquad t : 1 \to T, \tag{37}$$

where y_t is the asset return observed in period t, λ_t is the unobservable log volatility of y_t, and (u_t, v_t) are i.i.d.N(0,1) variables which are mutually independent. The parameters to be estimated are $\theta = (\beta, \delta, \nu)'$ with the usual stationarity condition that $|\delta| < 1$.

Classical and Bayesian inference of the SV model are hindered by the fact that it has no closed-form likelihood. Various (E)IS procedures have been used to perform ML estimation of the SV model (37) and of numerous univariate and multivariate extensions (see, e.g., Sandmann and Koopman, 1998 and Liesenfeld and Richard, 2003b). Alternatively, MCMC algorithms have been proposed for a Bayesian analysis, which avoids the need to compute the likelihood (see, e.g, Jacquier et al., 1994 and Chib et al., 2002). Such an MCMC analysis is typically based on simulation of θ and of the vector of volatilities λ from their joint conditional distribution $f(\theta, \lambda|y)$ using the Gibbs factorization $f(\theta|\lambda, y)$ and $f(\lambda|\theta, y)$. The simulation of $f(\lambda|\theta, y)$ is most challenging. Most approaches are based on a 'single-move' Gibbs sampling scheme, whereby each element of λ is simulated individually from its full conditional posterior. For example, Shephard and Pitt (1997) consider an AR-MH

procedure to simulate $\lambda_t|\lambda_{\setminus t},\theta,y$ element by element. For this purpose, they apply a second-order (Taylor) expansion around the mean of $\lambda_t|\lambda_{\setminus t},\theta$ to approximate the log kernel of the target density $\ln\varphi_t(\lambda_t|\lambda_{\setminus t},\theta,y) = -[(\lambda_t-\mu_t)^2/\sigma_t^2 + \lambda_t + y_t^2\exp(-\lambda_t)/\beta^2]/2$, where $\mu_t = \delta(\lambda_{t+1}+\lambda_{t-1})/(1+\delta^2)$ and $\sigma_t^2 = \nu^2/(1+\delta^2)$. This yields for each period an auxiliary Gaussian sampler m_t which can be used to produce candidate draws for the AR-MH procedure. The mean and variance of m_t are

$$\mu_t^* = \sigma_t^{*^2}\left[\frac{\mu_t}{\sigma_t^2} + \frac{1}{2}\left(\frac{y_t^2}{\beta^2}e^{-\lambda_t}(1+\mu_t) - 1\right)\right], \qquad \sigma_t^{*^2} = \left(\frac{1}{\sigma_t^2} + \frac{y_t^2}{2\beta^2}e^{-\lambda_t}\right)^{-1}.$$

However, the typically high persistence in the λ_t-process creates a problem for the convergence of such single-move Gibbs sampling MCMC schemes. To improve the speed of convergence, Shephard and Pitt (1997) suggest a block-sampling technique, that is, a joint simulation of a smaller number of blocks of consecutive λ_ts.

In view of the numerical efficiency of EIS and of the fact that the EIS sampler can be generically be interpreted as a single block–sampler for λ, Liesenfeld and Richard (2005b) propose an MCMC-EIS Bayesian algorithm combining Gibbs for the parameters given the volatilities and EIS for the volatilities given the parameters. In particular, the EIS sampler is used as a proposal density for the AR-MH algorithm to simulate λ as a *single* block from $f(\lambda|\theta,y)$. The latter is proportional to

$$\varphi(\lambda;\theta,y) = \prod_{t=1}^{T}\exp\left\{-\frac{1}{2}\left[\lambda_t + \frac{y_t^2}{\beta^2}e^{-\lambda_t} + \frac{(\lambda_t-\delta\lambda_{t-1})^2}{\nu^2}\right]\right\}. \tag{38}$$

For any given θ, one can use EIS to construct the following sequential Gaussian sampler approximation to $\varphi(\lambda;\theta,y)$:

$$\varphi(\lambda;\theta,y) \simeq \hat{c}\cdot m(\lambda|y,\theta,\hat{a}(\theta)) = \hat{c}\cdot\prod_{t=1}^{T} m_t\left(\lambda_t|\lambda_{t-1},\theta,\hat{a}_t(\theta)\right), \tag{39}$$

where $\hat{a}_t(\theta) = (\hat{\alpha}_{1t},\hat{\alpha}_{2t})$ denotes the GLS coefficients of λ_t and λ_t^2 in the sequential EIS auxiliary regressions. Furthermore, $\ln\hat{c} = \sum_{t=1}^{T}\hat{\gamma}_t$, where the $\hat{\gamma}_t$s are the intercepts of the EIS regressions, and m_t is a univariate normal density with kernel

$$k_t(\lambda_t;\theta,\hat{a}_t(\theta)) = \exp\left\{-\frac{1}{2}\left[\frac{(\lambda_t-\delta\lambda_{t-1})^2}{\nu^2} - 2\hat{\alpha}_{1t}\lambda_t + \hat{\alpha}_{2t}\lambda_t^2\right]\right\}, \tag{40}$$

and with variance $\hat{\sigma}_t^2 = \nu^2/(1+\nu^2\hat{\alpha}_{2t})$ and mean $\hat{\mu}_t = \hat{\sigma}_t^2(\delta\lambda_{t-1}/\nu^2 + \hat{\alpha}_{1t})$. While m would be used as such for an EIS likelihood evaluation (see Liesenfeld and Richard 2003b), for a Bayesian MCMC analysis one need exact draws from $f(\lambda|\theta,y)$ for which one can use an AR-MH algorithm based upon the ratio $\varphi/(\hat{c}\cdot m)$.

Table 4 and Figure 3 summarizes results of a Bayesian MCMC analysis of the SV model (37) using this single-block EIS sampling procedure for λ, and for comparison, those obtained using the Shephard and Pitt's (1997) single-move Gibbs sampler described above. The model is applied to the daily exchange rate of the British Pound against the US Dollar from October 1, 1981 to June 28, 1985 ($T = 945$). (The data set and the prior specification are the same as in Shephard and Pitt, 1997.) For $\ln \beta$, we use a flat prior leading to an inverted chi-squared conditional posterior for β^2. An inverted chi-squared prior with a mean of 0.013 and a variance of 0.007 is used for ν^2. The resulting conditional posterior for ν^2 is also an inverted chi-squared distribution. Finally, a Beta prior is assumed for $(\delta + 1)/2$ with a prior mean for δ of 0.86 and a prior variance of 0.012. The resulting conditional posterior is non-conjugate and is sampled by an independent MH procedure based on a Gaussian proposal distribution.

Posterior moments are presented in Table 4 together with numerical MC standard errors which are computed with the Parzen based spectral estimator used by Shephard and Pitt (1997). For M draws of a parameter $\{\tilde{\theta}_i; i : 1 \to M\}$ the MC standard error is the square root of

$$\frac{1}{M}\Big[\Gamma_0 + \frac{2L_M}{L_M - 1}\sum_{\ell=1}^{L_M} K\Big(\frac{\ell}{L_M}\Big)\Gamma_\ell\Big], \qquad \text{where} \quad \Gamma_\ell = \frac{1}{M}\sum_{i=\ell+1}^{M}(\tilde{\theta}_i - \bar{\theta})(\tilde{\theta}_{i-\ell} - \bar{\theta}),$$

with $K(\cdot)$ the Parzen kernel and L_M the bandwidth. The results for both MCMC algorithms are based on 52,000 Gibbs draws of the parameters, where the first 2,000 are discarded. The EIS approximation to $f(\lambda|\theta, y)$ is obtained using for the EIS auxiliary regressions a MC sample size of $S = 50$ and 3 EIS iterations with an initial sampler for the λ_ts given by the sequence of $\mathrm{N}(\delta\lambda_{t-1}, \nu^2)$ densities (at the corresponding Gibbs draws of the parameters). The starting values of the parameters used for both MCMC algorithms are $\beta = 1$, $\delta = 0.9$, $\nu = 0.2$. The single-move Gibbs sampler is initialized by setting all λ_ts to zero and then iterating for the starting values of the parameters on the λs for 1000 iterations. For the MCMC-EIS procedure the first λ-draw is simulated from the EIS sampler at the starting values of the parameters.

The results in Table 4 show that the the single block MCMC-EIS algorithm performs notably better than the single-move MCMC procedure. In particular, the MC standard errors of the MCMC-EIS estimators are significantly lower than those of the single-move estimators. This improvement obtained by the single-block EIS sampler is not surprising since the posterior moments for the persistence parameter δ (with a mean close to unity) indicate a slowly mixing volatility process. Using a single-move sampler for the volatilities, this leads typically to a slowly mixing of the MCMC chains on the volatility parameters. This interpretation is confirmed by the autocorrelation function of the

Gibbs draws of parameters and the convergence diagrams, where the MCMC posterior means for the parameters are plotted against the Gibbs iteration (see Figure 3). Notice in particular the severe convergence problems of the MCMC posterior mean for β obtained from the single-move Gibbs sampler. It appears that it has not converged even after 50,000 iterations, while the MCMC-EIS posterior mean do not need more than 10,000 iterations to reach its convergence level. Finally we notice, that the acceptance rates of the AR-MH EIS algorithm for the simulation of $\lambda|\theta, y$ turn out to be 81% (initial AR step) and 80% (subsequent MH step). This reflects the close EIS approximation of the kernel of the target density φ by $\hat{c} \cdot m$ and indicates that $\hat{c}$ ensures a balanced trade-off between too frequent rejection in the initial AR-step from high c values and too frequent repetitions in the subsequent MH-step due to low c values.

6.2 Bayesian Analysis of a Stationary AR Model

The following last example illustrates the implementation of a fully automated MCMC-EIS algorithm in a situation where EIS alone is not operational for a nonlinear parametrization of interest. It consists of the Bayesian posterior analysis of a stationary AR process. Bayesian MCMC analysis of stationary AR processes are found e.g. in Chib and Greenberg (1994,1995). Our analysis differs from theirs in several critical ways. Firstly, it relies on a different (nonlinear) parametrization associated with the roots of the AR process, which are typically the key quantities of interest for such an analysis. Secondly, it makes use of an operational analytical expression for the inverse of the stationary covariance matrix and, relatedly, all draws belong by construction to the stationary region of the parameter space (while in Chib and Greenberg, 1995, primary draws are unconstrained and then accepted or rejected depending upon whether they satisfy or not the stationarity condition). Thirdly, it relies upon an initial EIS approximation to the likelihood function to construct a fully automated and numericlly efficient MCMC posterior analysis.

The model under consideration is

$$\phi(L)y_t = y_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} = \epsilon_t, \quad \epsilon_t \sim \text{i.i.d.N}(0, \sigma^2), \quad t : 1 \to T, \tag{41}$$

where $\phi(L)$ is a polynomial in the backshift operator L. For ease of presentation and without loss of generality we set $\sigma^2 = 1$. It is assumed that $\phi = (\phi_1, ..., \phi_p)'$ lies in the stationary region B_ϕ, that is characterized by roots of $\phi(L)$, which lie all outside the unit circle. Following Richard (1977), the joint stationary density of the initial variables $y_{[p]} = (y_1, ..., y_p)'$ can be written as

$$y_{[p]} \sim \text{N}_p(0, H_p^{-1}), \quad \text{with } H_p = S_p S_p' - R_p' R_p, \tag{42}$$

where S_p is a lower triangular band matrix with elements $s_{ij} = \phi_{i-j}$, $(j \leq i)$ and R_p an upper triangular band matrix with elements $r_{ij} = \phi_{p+i-j}$, $(j \geq i)$. The likelihood function is given by

$$\ell(\phi; y) \propto |H_p|^{1/2} \exp\Big\{ -\frac{1}{2} y'_{[p]} H_p y_{[p]} \Big\} \cdot \left[\exp\Big\{ -\frac{1}{2} \sum_{t=p+1}^{T} (y_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p})^2 \Big\} \right]. \quad (43)$$

Assuming a flat prior for $\phi \in B_\phi$, the posterior of the parameters $f(\phi|y)$ has a kernel of the form $\varphi(\phi; y) = \ell(\phi; y) I_{B_\phi}$, where I_{B_ϕ} is an indicator function of the set B_ϕ. In order to perform a Bayesian analysis, Chib and Greenberg (1995) simulate from such a posterior using the MH algorithm with an auxiliary sampler m, associated with the Gaussian density kernel for the observations $y_{t+p}, ..., y_T$ (the term in brackets of Equation, 43).

However, note that the ϕ coefficients are hard to interpret and one would typically have greater interest in the roots of the process. In order for the transformation from ϕ into the roots to be one-to-one, the latter need to be ordered in some appropriate way. An operational solution to this problem consists of factorizing the polynomial $\phi(L)$ into ordered binomials. Focusing for the moment on the case where the order p is even, say $p = 2r$, let

$$\phi(L) = \prod_{j=1}^{r} \psi_j(L) = \prod_{j=1}^{r} (1 + \beta_j L + \delta_j L^2), \quad (44)$$

where the β_js are ordered according to $-2 < \beta_1 < \cdots < \beta_r < 2$. The corresponding stationarity conditions for $\psi = (\beta_1, \delta_1, ..., \beta_r, \delta_r)'$ are given by $\beta_j + \delta_j > -1$, $\beta_j - \delta_j < 1$, and $\delta_j < 1$. It follows, that simulated draws of the ψ coefficients will automatically satisfy the stationarity restrictions if they are constrained in the following sequential way:

- For β_j given all other coefficients

$$\max\{\beta_{j-1}, -(1 + \delta_j)\} < \beta_j < \min\{\beta_{j+1}, (1 + \delta_j)\}, \quad \text{with} \quad \beta_0 = -2, \beta_{r+1} = 2; \quad (45)$$

- For δ_j given all other coefficients

$$|\beta_j| - 1 < \delta_j < 1. \quad (46)$$

Obviously, for $r > 1$, the ϕ coefficients are a nonlinear but trivial transformation of the ψ coefficients, e.g., for $r = 2$ we have

$$\phi_1 = \beta_1 + \beta_2, \quad \phi_2 = \delta_1 + \beta_1\beta_2 + \delta_2, \quad \phi_2 = \delta_1\beta_2 + \delta_2\beta_1, \quad \phi_4 = \delta_1\delta_2.$$

Furthermore, if we partition ψ into $\psi' = (\psi_\tau, \psi'_{\backslash\tau})$, where ψ_τ is a single parameter, we note that ϕ is a *linear* function of ψ_τ *given* $\psi_{\backslash\tau}$. This conditional linearity together with the iterated restrictions in

(45) and (46), provide the key to the proposed MCMC-EIS algorithm for a Bayesian analysis of the pairs (β_j, δ_j). It also allows for a trivial numerical evaluation of the Jacobian $\partial\phi/\partial\psi'$, whose τ-th column is obtained by increasing ψ_τ by 1 keeping $\psi_{\setminus\tau}$ fixed and computing the resulting difference in ϕ. Accordingly, the proposed MCMC-EIS can be implemented for priors on ϕ as well as priors on ψ.

Based on these preliminaries, the MCMC-EIS algorithm consists of the following key steps:

(i) In the first step, a Gaussian EIS approximation $m(\phi|y,a)$ to the posterior associated with the likelihood (43) is constructed. This requires an auxiliary GLS regression associated with the following expression for $d(\phi,a)$ as defined in Equation (13)

$$d(\phi,a) = \ln \ell(\phi;y) - \gamma - \alpha'\phi - \phi'\Sigma\phi, \tag{47}$$

where α is a p-dimensional vector and Σ a symmetric $p\times p$ matrix. In the application below with $p=6$, this regression includes 21 regressors plus one intercept. (For significantly higher ps, one could embed univariate EIS regressions within the MCMC algorithm instead of computing a single global initial EIS.) In our application we use the prior of ϕ as initial sampler. Beyond the sampler itself, this initial EIS step produces two other results which are useful for an efficient MCMC implementation. First, its mode can be used as initial value for the subsequent AR-MH steps, and second, the intercept of the EIS regression can be transformed into an efficient calibration constant c for the corresponding AR-MH ratios $\varphi/[c\cdot m]$.

(ii) In the next step, the AR-MH algorithm is used to sample individually $\psi_\tau|\psi_{\setminus\tau},y$ based on the Gibbs factorization of ψ. Specifically, under a uniform prior on ψ the posterior density of a single ψ_τ is given by $f(\psi_\tau|\psi_{\setminus\tau},y) \propto \varphi(\phi(\psi);y)$. The AR-MH sampling density for $\psi_\tau|\psi_{\setminus\tau},y$ is given by the Gaussian conditional density $m(\psi_\tau|\psi_{\setminus\tau},y)$ associated with the EIS sampler $m(\phi|y,a)$. Note that since ϕ is a linear function of ψ_τ given $\psi_{\setminus\tau}$, one only needs to evaluate $m(\phi|y,a)$ for *three* different values of ψ_τ (keeping $\psi_{\setminus\tau}$ fixed) to retrieve the mean and variance of $m(\psi_\tau|\psi_{\setminus\tau},y)$. The ψ_τ draws are truncated conformably with the stationarity and ordering conditions given by Equations (45) and (46). In order to accelerate draws of the truncated normals one can usefully rely upon interpolation tables for the c.d.f. and inverse c.d.f of the standardized normal.

In order to illustrate the performance of the proposed MCMC-EIS algorithm, we first reran the application of Chib and Greenberg (1995) for an AR(2) model based on simulated observations. In

particular, we simulated $T = 100$ observations from an AR(2) with $\phi_1 = -1$ and $\phi_2 = 0.5$. Since $p = 2$, the ϕ and ψ parameters are equivalent and we can benchmark a MCMC-EIS against an EIS posterior analysis. As in Chib and Greenberg, we use 5,500 MCMC draws of which the first 500 are discarded. Correspondingly, we use 5,000 draws for EIS. Posterior means and standard deviations are provided in Table 5, together with MC (numerical) standard errors for all relevant estimates. The latter are based upon 100 i.i.d. replications of the complete algorithm. We also provide the MC means and standard errors of the ratios $\varphi/[c \cdot m]$ for MCMC-EIS. The results show that the EIS-step has efficiently normalized these ratios. Furthermore, the EIS as well as the MCMC-EIS algorithm lead to posterior distributions which are centered around the true values of the data generating process. Since Chib and Greenberg (1995) implicitly use an EIS-type algorithm based upon the quadratic part of $\ell(\phi; y)$, their numerical accuracy is similar to the ones obtained here for the MCMC-EIS algorithm, while the computing time for the latter is absolutely competitive. The CPU times for the entire sampling process using the EIS and MCMC-EIS scheme on a 750 MHz UNIX server are 0.08 and 0.16 seconds, respectively, while Chib and Greenberg report CPU times of 8 minutes for a pure Accept-Reject algorithm and 2 minutes for the MH on a 50MHz PC. Moreover, preliminary investigations suggest that the EIS step has significantly improved the convergence properties of the MCMC algorithm. For example, we reduced the number of MCMC draws by a factor ten (only 550 draws of which the first 50 were dicarded) and produced virtually identical estimation results (not presented here) up to 2 or 3 decimals except for the obvious fact that numerical standard errors increased approximately by the factor $\sqrt{10}$.

Next, for a more stringent test, we repeated the numerical experiment for an AR(p) model with $p = 6$, using a pair of complex conjugate roots close to the unit circle ($\pm i\sqrt{0.95}$), a pair of real roots (-0.9,0.2) and a double real root (-0.7). The corresponding ordered (β_j, δ_j)-pairs in ψ are given by (0, 0.95), (0.7,-0.18) and (1.4,0.49). For $p > 1$, we can no longer offer an EIS benchmark on the ψ coefficients for comparison. The posterior means and standard deviations of ψ together with MC numerical standard errors and posterior probabilities for complex roots are provided in Table 6. The results are based on 5,500 MCMC draws, where the first 500 are discarded. CPU time for the entire sampling process is 0.7 seconds. The results indicate that the MCMC-EIS algorithm has accurately produced posterior distributions for the ψ parameters which are centered around the true parameter values of the data generating process. (Posterior moments of ϕ are immediate by-products of the MCMC-EIS analysis of ψ but are not reported here. When compared with the EIS posterior moments of ϕ, they were found to be near identical but numerically less efficient by factor ranging from 2 to 3. This is hardly surprising since the likelihood is nearly Gaussian in ϕ and EIS on ϕ is

therefore highly efficient.) Furthermore, a five fold reduction in the number of MCMC draws leads to virtually identical results (not presented here) except for the corresponding $\sqrt{5}$-fold increase in numerical standard errors. Hence, the results suggest that the convergence of MCMC-EIS is very fast and requires fewer draws than typically used in the literature. Again it appears that implementation of an EIS step improved the numerical properties of MCMC, essentially providing (near optimal) fully automated selection of critical MCMC components (starting values, normalizing constant, efficient univariate conditional samplers).

In conclusion of this subsection we briefly indicate how to handle odd orders of autoregression p. Direct *ordered* factorization remains feasible but gets more complicated as the isolated real root needs to be ordered relative to a random number of other pairs of real roots. This imposes further recursive constraints within MCMC. We have found it far easier to adopt a 'Bayesian' solution whereby one increases the order by one and specifies a corresponding prior which keeps the trailing coefficient in a region very near to zero.

7 Conclusions

This paper has shown how the Efficient Importance sampling (EIS) can be used to improve the numerical accuracy of Markov Chain Monte Carlo (MCMC) algorithms based on the Metropolis Hastings (MH) procedure. (E)IS and MH are two separate techniques which can be used to analyze econometric models involving integrals without analytical solutions. The MH procedure is a Markov-Chain method to simulate from a (unknown) distribution by 'weighting' draws from an auxiliary sampler according to an accept-reject mechanism. EIS is an algorithm for the construction of importance sampling densities which produce numerically efficient Monte Carlo (MC) estimates of integrals, provided that the corresponding MC sampling variance exists. It is based on a Least Squares approach to obtain a global approximation of the integrand, typically the posterior density of variables to be integrated out. As such the EIS technique can – beyond its use as a separate MC integration approch – be employed to systematically construct auxiliary samplers for the MH procedure which can be expected to have good MC-sampling properties. This approach of embedding EIS within MCMC is based on the key insight that there is a close relationship between the efficient selection of importance sampling densities and the construction of auxiliary sampling densities for the MH procedure. In both cases the numerical efficiency critically depends upon the approximation quality of the sampling densities (w.r.t. the integrand and the target density from which a simulated sample needs to be generated, respectively). Furthermore, the problem of possibly non-existing vari-

ance of EIS-MC estimates, leading to a large variation and slow convergence of the estimate, can also affect MH algorithms (in particular the independent MH). In order to reveal such convergence problems of EIS and MCMC-EIS estimates we propose a useful diagnostic statistic. Beyond the auxiliary sampler itself, EIS can also provide a fully automated selection of calibrating constants and starting values, which can also be critical elements of an numerically efficient implementation of MH procedures.

The potential of this integrated MCMC-EIS approach for the analysis of a broad range of econometric models is illustrated with numerical examples involving univariate as well as multivariate integration problems. The two 'textbook' univariate examples (integration of an inverse Gaussian and a Student-t) serve to illustrate the close relationship between (E)IS and MH, and the basic principle of the integrated MCMC-EIS approach.

The two multivariate examples illustrate the full potential of our proposed approach in the context of two important classes of models. In the Stochastic Volatility application, we fully exploit the comparative advantages of EIS (high-dimensional integration) and MCMC (Bayesian posterior factorization) to offer a numerically very efficient EIS-MH MCMC algorithm. In the Bayesian Autoregressive application, we focus our attention of a highly non-linear parametrization of intrinsic interest for which EIS is not operational but nevertheless are able to exploit EIS for the construction of an efficient and original MCMC algorithm.

References

Bauwens, L., Galli, F., 2005. EIS for the estimation of SCD models. Unpublished manuscript, CORE, Universite Catholique, Louvain-la-Neuve.

Bauwens, L., Hautsch, N., 2003. Dynamic latent factor models for intensity processes. Discussion Paper, 2003-103, CORE, Universite Catholique, Louvain-la-Neuve.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrica 81, 541-553.

Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA (p,q) errors. Journal of Econometrics 64, 183-206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistician 49, 327-335.

Chib, S., Nadari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics 108, 281-316.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398-409.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317-1339.

Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication. Econometric Reviews 18, 1-73.

Ghysels, E., Harvey, A.C., Renault, E., 1996. Stochastic volatility. In: Maddala, G., Rao, C.R., Handbook of statistics, Vol 14. Elsevier Sciences, Amsterdam.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in practice. Chapman & Hall, London.

Gourieroux, C., Monfort, A., 1996. Simulation-based econometric methods. Oxford University Press, Oxford.

Gradshteyn, I.S., Ryzhik, I.M., 1979. Table of integrals, series and products. Academic Press, San Diego.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal of Business & Economic Statistics 12, 371-389.

Jung, R.C., Liesenfeld, R., 2001. Estimating time series models for count data using efficient importance sampling. Allgemeines Statistisches Archiv 85, 387-407.

Kloek, T., van Dijk, H.K., 1978. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. Econometrica 46, 1-19.

Koopman, S.J., Shephard, N., 2004. Testing the assumptions behind the use of importance sampling. Unpublished manuscript, Free University Amsterdam, Dept. of Econometrics.

Liesenfeld, R., Richard, J.F., 2003a. Estimation of dynamic bivariate mixture models: comments on Watanabe (2000). The Journal of Business and Economic Statistics 21, 570-576.

Liesenfeld, R., Richard, J.F., 2003b. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505-531.

Liesenfeld, R., Richard, J.F., 2005a. Simulation techniques for panels: Efficient importance sampling. manuscript, University of Kiel , Dept. of Economics. (to appear in: Matyas, L., Sevestre, P., The Econometerics of Panel Data (3rd ed). Kluwer Academic Publishers.)

Liesenfeld, R., Richard, J.F., 2005b. Classical and Bayesian analysis of univariate and multivariate stochastic volatility models manuscript, University of Kiel , (to appear in: Econometric Reviews)

Richard, J.F., 1977. Bayesian analysis of the regression model when the disturbances are generated by an autoregressive process. In: Aykac, A., Brumat, C., New developments in the application of Bayesian methods. North Holland, Amsterdam.

Richard, J.F., Zhang, W., 2006. Efficient high-dimensional importance sampling. Unpublished manuscript, University of Pittsburgh, Dept. of Economics.

Robert, C.P, Casella, G., 1999. Monte Carlo statistical methods. Springer, New York.

Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of Econometrics 87, 271-301.

Shephard, N., 2004. Stochastic volatility: selected readings (editor). Oxford University Press, Oxford.

Shephard, N., Pitt, M.K., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653-667.

Stern, S., 1997. Simulation-based estimation. Journal of Economic Literature 35, 2006-2039.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions ba data augmentation. Journal of the American Statistical Association 82, 528-550.

Tierney, L., 1994. Markov chains for exploring posterior distributions. The Annals of Statistics 22, 1701-1762.

Table 1. MC Evaluation of the Mean Under an Inverse Gaussian

	$\hat{\mathrm{E}}_f(x)$	δ	κ	accept. rate	Γ_S
MH (opt. accept. rate)	1.1530	.6667	1.7320	.713	.9987
	(.0111)				(.0122)
MH-EIS	1.1537	.3158	3.6182	.904	1.0749
	(.0126)	(.0089)	(.0919)		(.0634)
EIS-1	1.1530	.3158	3.6182		
	(.0107)	(.0089)	(.0919)		
		numerator			
EIS-2	1.1551	.3876	3.8132		
	(.0008)	(.0071)	(.0628)		
		denominator			
		.3158	3.6182		
		(.0089)	(.0919)		

NOTE: MC estimation of $E_f(x)$ for the inverse Gaussian distribution (30) with $\theta_1 = 1.5$ and $\theta_2 = 2$ using a Gamma distribution (31) for simulation. The theoretical value is $E_f(x) = 1.155$. Simulation sample size is $S = 5,000$ and the number of EIS iterations is 20. The numbers in parentheses are MC standard deviations based upon 100 repeated estimates.

Table 2. MC Evaluation of the Variance Under a Standardized Student-t

ν		2.5	4.0	6.0	10.0	150.0
MH-EIS	$\hat{\mathrm{E}}(x^2)$	.4359	.8095	.9201	.9675	.9930
		(.1042)	(.1101)	(.0966)	(.0753)	(.0453)
	$\hat{a}$	2.0863	1.1487	1.0392	1.0080	1.0004
		(.5234)	(.1751)	(.0995)	(.0523)	(.0022)
	Γ_S	3.4e+4	1.5e+4	4.9e+3	7.1e+2	1.2363
		(1.3e+5)	(5.9e+4)	(1.9e+4)	(2.4e+3)	(.2710)
	accept. rate	.813	.879	.918	.952	.997
EIS-1	$\hat{\mathrm{E}}(x^2)$	.4420	.8119	.9170	.9653	.9929
		(.1158)	(.1164)	(.0898)	(.0682)	(.0459)
EIS-2	$\hat{\mathrm{E}}(x^2)$	.8619	.9826	.9961	.9990	.9991
		(.0531)	(.0149)	(.0162)	(.0177)	(.0195)

NOTE: MC estimation of $E_f(x^2)$ for the standardized student-t distribution with kernel (34) and ν degrees of freedom using a $\mathrm{N}(0, a^{-1})$ distribution (35) for simulation. The theoretical value is $E_f(x) = 1$. Simulation sample size is $S = 1,000$ and the number of EIS iterations is 100. The numbers in parentheses are MC standard deviations based upon 100 repeated estimates.

Table 3. (E)IS MC Evaluation of a Standardized Student-t Integral with an Injected Outlier

	$\hat{a}$	$\bar{I}$
EIS without injected outlier	2.0863	1.2159
	(.5234)	(.0173)
EIS with injected outlier	2.0567	1.2141
	(.5559)	(.0266)
IS with injected outlier and fixed a	2.0567	3.3893
		(21.6810)

NOTE: (E)IS MC estimation of $I = \int \varphi(x,\nu)dx = \sqrt{\pi(\nu-2)}\Gamma(\nu/2)/\Gamma([\nu+1]/2)$ for the standardized student-t distribution with kernel (34) and $\nu = 2.5$ degrees of freedom using a $\mathrm{N}(0, a^{-1})$ distribution (35) for simulation. The theoretical value of I is 1.2360. Simulation sample size is $S = 1,000$ and the number of EIS iterations is 100. The numbers in parentheses are MC standard deviations based upon 100 repeated estimates. The outlier injected into the draws of the initial $\mathrm{N}(0,1)$ sampler is $x = 6$.

Table 4. MCMC Posterior Analysis of the SV Model for the British Pound/US-Dollar Exchange Rate

		EIS	Single-move Gibbs
β	Mean	.654	.691
	Stand. Dev.	.111	.075
	MC S.E.	.00590	.01579
δ	Mean	.981	.982
	Stand. Dev.	.009	.008
	MC S.E.	.00027	.00062
ν	Mean	.144	.143
	Stand. Dev.	.027	.027
	MC S.E.	.00113	.002069

NOTE: The estimated model is given by Equation (37). Posterior moments are based on 52,000 Gibbs iterations (discarding the first 2,000 draws). MC standard errors are computed using a Parzen based spectral estimator with a bandwidth of 5000. The EIS approximation to the full conditional distribution of $\lambda|\theta, y$ is based on a MC sample size $N = 50$ (used to run the EIS auxiliary regressions) and three EIS iterations.

Table 5. Posterior Analysis of the AR(2) Model for Simulated Data

	EIS		MCMC-EIS	
	$\beta_1 = \phi_1$	$\delta_1 = \phi_2$	$\beta_1 = \phi_1$	$\delta_1 = \phi_2$
True	−1.000	.500	−1.000	.500
Mean	−.851	.383	−.851	.382
Stand. Dev.	.100	.101	.099	.100
MC S.E	.0015	.0011	.0020	.0020
MC Mean (S.E.) of $\varphi/(c \cdot m)$			.999 (.0470)	

NOTE: The estimated model is given by Equation (41) with $p = 2$. The sample size for the simulated data from that model is $T = 100$. The posterior moments are based on 5,500 parameter draws (discarding the first 500 draws). The EIS approximation to the likelihood $\ell(\phi; y)$ is based on 4 to 5 EIS iterations. The MC standard errors are based upon 100 replications of the complete algorithms.

Table 6. MCMC-EIS Posterior Analysis of the AR(6) Model for Simulated Data

	β_1	δ_1	β_2	δ_2	β_3	δ_3
True	.000	.950	.700	−.180	1.400	.490
Mean	−.004	.932	.841	.081	1.433	.510
Stand. Dev.	.040	.031	.204	.143	.188	.166
MC S.E	.0008	.0007	.0404	.0111	.0419	.0362
Prob.	1.000		.169		.426	
MC Mean (S.E.) of $\varphi/(c \cdot m)$			1.037 (1.680)			

NOTE: The estimated model is given by Equation (41) with $p = 6$ and re-parameterized according to (44). The sample size for the simulated data from that model is $T = 100$. The posterior moments are based on 5,500 parameter draws (discarding the first 500 draws). The EIS approximation to the likelihood $\ell(\phi; y)$ is based on 4 to 5 EIS iterations. The MC standard errors are based upon 100 replications of the complete algorithm.

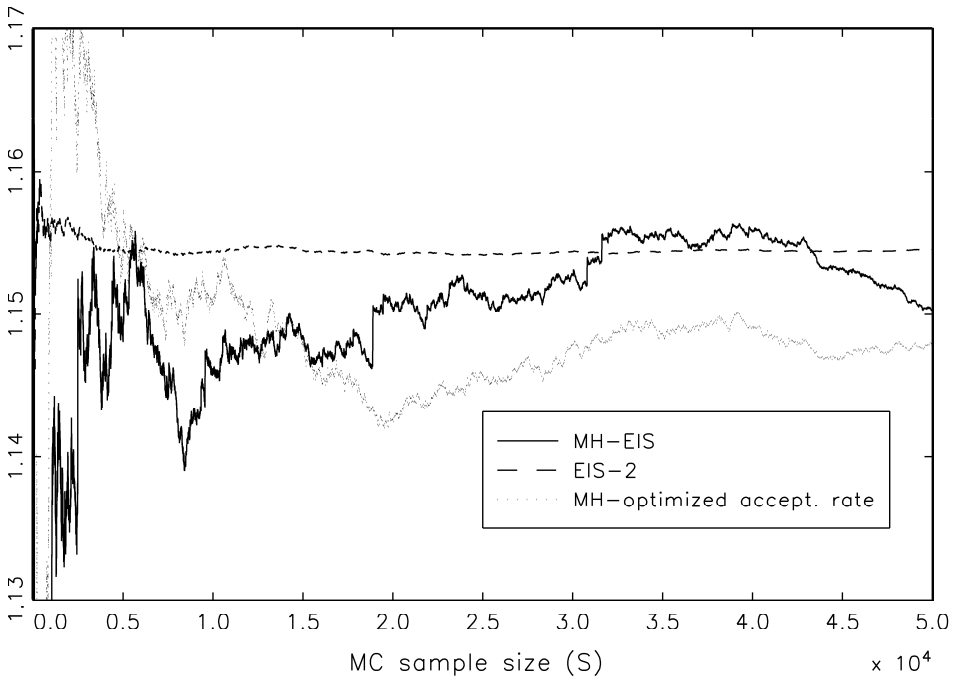

Fig. 1. Convergence of MC estimates of $E_f(x)$ where f is an inverse Gaussian distribution with kernel (30) and $\theta_1 = 1.5$ and $\theta_2 = 2$. The exact value is 1.155. The final estimates for a MC sample size 50,000 are 1.1503 (MH-EIS), 1.1545 (EIS-2), 1.1479 (MH-optimized acceptance rate).

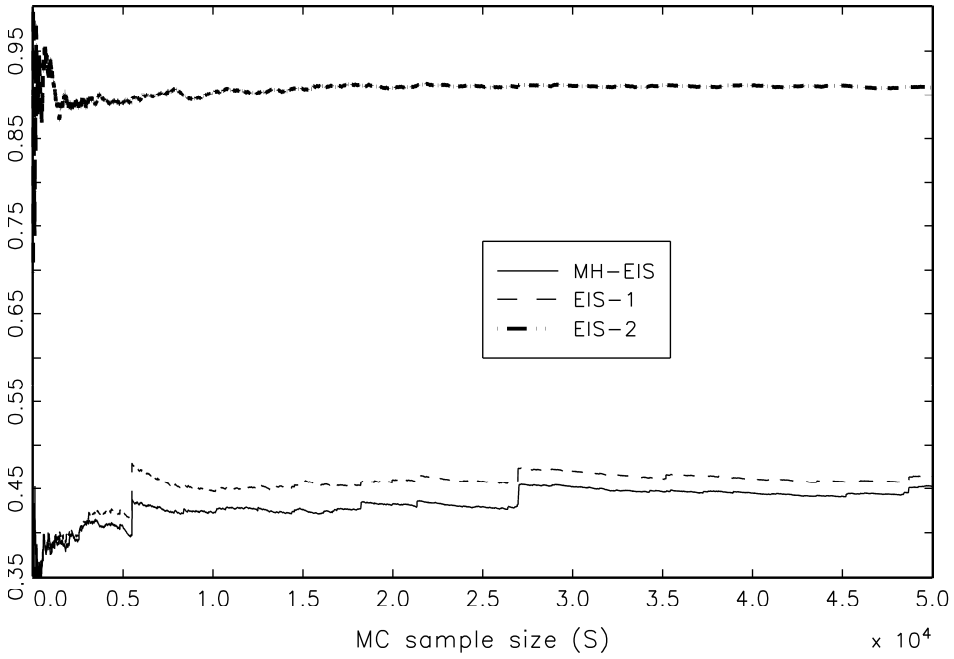

Fig. 2. Convergence of MC estimates of $E_f(x^2)$ where f is a standardized student-t with density kernel (34) and $\nu = 2.5$. The exact value is 1. The final estimates for a MC sample size 50,000 are 0.4520 (MH-EIS), 0.4648 (EIS-1), 0.9095 (EIS-2).

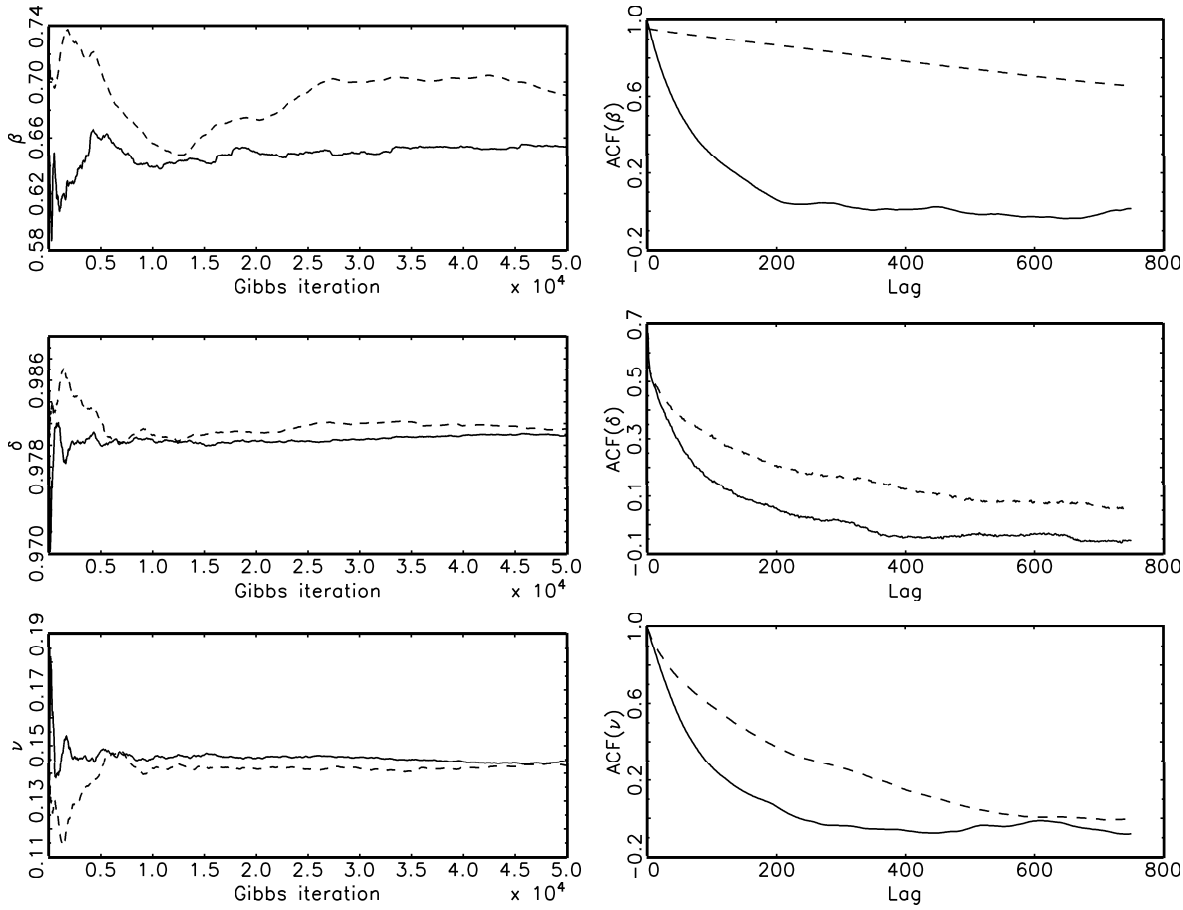

Fig. 3. Convergence of MCMC posterior means for the SV parameters (left) and autocorrelation functions of the Gibbs draws of the parameters (right). The solid lines represent the results for the MCMC algorithm based on the (single-block) EIS sampler for λ and the dashed lines those for the corresponding single-move Gibbs sampler.