Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 1385–1411
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Efficient high-dimensional importance sampling

Jean-Francois Richard[a,*], Wei Zhang[b]

[a]*Department of Economics, University of Pittsburgh, 4917, Posvar Hall, Pittsburgh, PA 15260, USA*
[b]*National University of Singapore, Singapore*

Available online 3 April 2007

Abstract

The paper describes a simple, generic and yet highly accurate efficient importance sampling (EIS) Monte Carlo (MC) procedure for the evaluation of high-dimensional numerical integrals. EIS is based upon a sequence of auxiliary weighted regressions which actually are linear under appropriate conditions. It can be used to evaluate likelihood functions and byproducts thereof, such as ML estimators, for models which depend upon unobservable variables. A dynamic stochastic volatility model and a logit panel data model with unobserved heterogeneity (random effects) in both dimensions are used to provide illustrations of EIS high numerical accuracy, even under small number of MC draws. MC simulations are used to characterize the finite sample numerical and statistical properties of EIS-based ML estimators.

JEL classification: C15

Keywords: Monte Carlo; Importance sampling; Marginalized likelihood; Stochastic volatility; Random effects

1. Introduction

Monte Carlo (hereafter MC) simulation techniques provide powerful tools to numerically evaluate expectations of functions of random variables for which no analytical expressions are available. See e.g. Fishman (1996) for an in-depth analysis of MC concepts and algorithms. One particular area where MC methods play a critical role is that of models which incorporate large numbers of unobserved random variables. Examples to be

*Corresponding author. Tel.: +1 412 648 1750.
E-mail address: fantin@pitt.edu (J.-F. Richard).
URL: http://www.econ.pitt.edu/fantin/

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.02.007

discussed further below are stochastic volatility models in finance and large panels with unobserved heterogeneity, where dimensions of integration can be in the thousands. It has long been recognized that the feasibility of evaluating such high-dimensional integrals by MC simulation critically depends on the selection of efficient MC samplers. Importance sampling relies upon an auxiliary sampler in combination with an appropriate probability redistribution scheme meant to compensate for the fact that the sampler is not the correct one. High variance of the redistribution ratio will produce highly inefficient MC estimates.

The object of the present paper is to propose a new algorithm to construct an efficient importance sampler (hereafter EIS), which is generic and particularly well adapted to (very) high-dimensional MC integration. In particular, under appropriate simplifying conditions it amounts to a simple recursive sequence of auxiliary least squares (LS) optimization problems which, as we shall illustrate below, can produce enormous efficiency gains at low cost of computation. The high efficiency of the proposed method is due to the fact that these EIS auxiliary regressions cover the full support of the integrand. Whence EIS samplers can be interpreted as global approximations of the integrand in sharp contrast with local approximation techniques such as those reviewed in Section 2.2.

Moreover, our EIS algorithm can also be used to facilitate the construction of auxiliary samplers for other MC techniques relying upon approximations of analytically intractable density kernels, such as Metropolis-Hastings (MH) and/or Markov Chain Monte Carlo (MCMC). A full investigation of the fundamental complementary between EIS and MCMC belongs to our current research agenda and is briefly discussed in the conclusion of the present paper.

The paper is organized as follows: Section 2 introduces importance sampling and briefly reviews relevant literature. The generic principle of EIS is introduced in Section 3 as a one-shot algorithm for low-dimensional integration. Section 3 also includes a discussion of key implementation details and provides two numerical illustrations including the now (in)famous Student-t/Gaussian example. High-dimensional sequential EIS is presented in Section 4. Pilot applications of EIS to stochastic volatility and panel data models are analyzed in Section 5. Section 6 concludes. Technical details are regrouped in Appendix A.

2. Importance sampling

2.1. Principle

Assume one has to evaluate a functional integral of the form

$$G(\delta) = \int_X \varphi(x;\delta)\,\mathrm{d}x \tag{1}$$

with $\varphi : X \times \Delta \to \mathbf{R}_+^*$. Whence φ denotes a density kernel in x given δ with support X and one needs to compute its integrating constant G as a function of δ. Important applications to be considered further below are those where x denotes a (high-dimensional) vector of latent variables and $G(\delta)$ a likelihood function.

It is often the case that the statistical formulation of the model under consideration produces an initial factorization of the form

$$\varphi(x;\delta) = g(x;\delta)\cdot p(x|\delta), \tag{2}$$

where p denotes a density for x given δ which is directly amenable to MC simulation. In the sequel of our analysis p will be referred to as an initial sampler. The corresponding MC estimator of $G(\delta)$ is given by

$$\overline{G_S}(\delta) = \frac{1}{S}\sum_{i=1}^{S} g(\tilde{x}_i;\delta), \tag{3}$$

where $\{\tilde{x}_i; i : 1 \to S\}$ denotes a set of S i.i.d. draws from p. If, however, the MC sampling variance of g on p is large, sufficiently accurate MC estimation of G may require a prohibitively large number of draws.

Importance sampling consists of replacing the initial sampler p by an auxiliary IS sampler $m(x|a)$, rewriting Eq. (1) as

$$G(\delta) = \int_X \omega(x;\delta,a)\cdot m(x|a)\,\mathrm{d}x, \tag{4}$$

where

$$\omega(x;\delta,a) = \frac{\varphi(x|\delta)}{m(x|a)}. \tag{5}$$

The corresponding IS estimator of G is given by

$$\overline{G_{S,m}}(\delta;a) = \frac{1}{S}\sum_{i=1}^{S} \omega(\tilde{x}_i;\delta,a), \tag{6}$$

where $\{\tilde{x}_i; i : 1 \to S\}$ now denotes a set of S i.i.d. draws from m. One's objective becomes that of selecting a class $M = \{m(x|a) : a \in A\}$ and a value $a(\delta) \in A$ which minimizes the MC sampling variance of the ratio φ/m on m. The latter is given by

$$\overline{V_{S,m}}(a;\delta) = \frac{1}{S}\int_X [\omega(x;\delta,a) - G(\delta)]^2 m(x|a)\,\mathrm{d}x. \tag{7}$$

Conditions for the finiteness of $\overline{V_{S,m}}$ and for a central limit theorem for $\sqrt{S}(\overline{G_{s,m}} - G)$ are discussed e.g. in Geweke (1996) or Stern (1997). It is now well recognized that large or even infinite values of $V(a;\delta)$ typically originate from thin tails of $m(x|a)$ which is precisely why IS pathologies can be empirically hard to detect. This critical issue will be addressed in Section 3.4 below.

Clearly, the two critical issues to be addressed in IS applications are (1) the selection of an appropriate class M of auxiliary samplers; and (2) the selection of an efficient sampler within M i.e. one for which $V(a;\delta)$ is as small as possible. The selection of M is bound to remain problem-specific, though our subsequent discussion will provide important guidelines for such selection. The EIS principle proposed in this paper specifically addresses the issue of selecting (near) optimal $a(\delta)$ in A.

It should be noted that our subsequent analysis applies to a specific functional of interest φ in Eq. (1). Obviously, there might be several functions of interest to be dealt with in the context of a particular application. In general, maximal numerical efficiency requires that EIS be run separately for each function of interest. This issue will be discussed further in the context of the applications presented below but will generally require only very minor modifications of our baseline EIS algorithm (essentially adjusting the definition of the dependent variables in the auxiliary EIS regressions). Functions of interest which are not

strictly positive on their support (e.g. first order moments) can be estimated under EIS applied to the positive part of the integrand or, when applicable, split into differences of integrals of positive functionals. The latter approach typically requires higher numerical accuracy in order to account for potential cancellation of significant digits (Antithetic MC, as discussed e.g. by Fishman, 1996; Geweke, 1996 could usefully be combined with EIS in this context but will not be discussed here).

2.2. Short literature review

Since the construction of importance samplers clearly constitutes the Achilles heel of IS, it has received much attention over the years. Let us review here some innovative proposals in this respect. While Tierney and Kadane (1986) do not specifically discuss IS, the concept of Laplace approximation they rely upon to evaluate posterior moments can also be used to construct importance samplers. It essentially consists of locally approximating $\varphi(x;\delta)$ around its modal value. Geweke (1989) explicitly discusses minimization of $V(a;\delta)$ within specific classes of fat-fail densities, typically multivariate Student-t densities and skewed generalizations thereof, labeled split-t densities. Evans (1991) relies upon adaptative methods whereby earlier draws of x are used to identify large values of the weight function ω in Eq. (4) and to revise accordingly the sampler. Owen and Zhou (2000) discuss various improvements of the *IS* technique which are well suited for low-dimensional applications.

Durbin and Koopman (1997) apply IS to evaluate the likelihood function of non-Gaussian state space models. Essentially, by constructing a Gaussian approximation to their model, they are able to express the ratio between the two likelihoods as an integral which is functionally similar to Eq. (4). The relevance of their method is twofold. First, it shows that the selection of an importance sampler can be approached via the construction of an operational approximation to a complex model and, in this respect, offers conceptual similarities with the EIS principle proposed below. Second, it is applicable in significantly higher dimensions than the alternative methods discussed above.

Another sophisticated implementation of IS is found in Madras and Piccioni (1999) where the authors use as *IS* the (implicit) equilibrium distribution associated with a MCMC simulator. The main advantage of their method lies in the flexibility of MCMC simulations but the convergence properties of their procedure are typically difficult to assess.

An important message which emerges from this brief literature overview is that importance samplers have to be carefully tailored to the problem under consideration. This has proved to be a significant obstacle to routine applications of IS. Another problem lies in the fact that, except for the specific problem addressed by Durbin and Koopman (1997), none of the existing IS methods appear to be applicable to (very) high-dimensional applications of the form of those considered in Section 5 below.

3. EIS baseline algorithm

EIS is fundamentally designed to achieve maximal efficiency in the context of high-dimensional sequential applications. In this section, however, we shall introduce EIS as a single shot fixed point minimization problem which while only applicable as such to low-dimensional problems, provides a convenient framework to discuss and illustrate EIS key features. Sequential implementations will be presented in Section 4 below.

Section 3 is organized as follows: in Section 3.1 we present the baseline LS fixed point version of EIS. A linear version thereof under sampling from the exponential family of distribution is introduced in Section 3.2. The (pre) selection of classes of IS samplers is discussed in Section 3.3. An operational diagnostic for the finiteness of the MC variance of EIS estimates is presented in Section 3.4. Measures of (stochastic) numerical accuracy are discussed in Section 3.5. Two univariate applications highlighting all key features of EIS are presented in Section 3.6. A more technical justification of EIS baseline algorithm is presented in Appendix A.

3.1. Baseline least squares EIS algorithm

In this section we momentarily assume that a class of auxiliary samplers $M : \{m(x|a); a \in A\}$ has already been selected. For any given δ, our objective then becomes that of selecting a value $a(\delta)$ which (approximately) minimizes the MC sampling variance of the ratio $\varphi(x;\delta)|/m(x|a)$ on draws from m. A11 factors in φ and/or m which do not depend on x are regrouped together in the form of a proportionality factor whose logarithm will serve as the (implicit) constant term of the auxiliary EIS regression to be estimated along with a. This is particularly relevant in the case of m since integrating constants of commonly used distributions are typically highly nonlinear functions of a. Whence EIS is conveniently reformulated in terms of the selection of an efficient density kernel $k(x;a)$ within a preassigned class $K = \{k(x;a), a \in A\}$. The relationship between k and m (K and M) is given by

$$m(x|a) = \frac{k(x;a)}{\chi(a)} \quad \text{with } \chi(a) = \int_X k(x;a)\,\mathrm{d}x. \tag{8}$$

We shall assume here that $\chi(a)$ is known analytically. (Extensions where χ can be numerically evaluated by non-stochastic quadrature rules are also feasible.) Obviously, if there existed $a_0(\delta) \in A$ and $\gamma_0(\delta) \in \mathbf{R}_+^*$ such that

$$\frac{\varphi(x;\delta)}{\gamma_0(\delta)\cdot k(x,a_0(\delta))} \equiv 1, \quad \forall x \in X, \tag{9}$$

then our problem would be trivially solved and $G(\delta)$ would be given by $\gamma_0(\delta)\cdot\chi(a_0(\delta))$. More generally, EIS aims at selecting a pair $(a_*(\delta), \gamma_*(\delta))$ which centers the ratio $\varphi/(\gamma\cdot k)$ around one and minimizes its MC sampling variance. As shown in Appendix A an operational (close) approximation to that problem is given by

$$(\hat{a}(\delta), \hat{c}(\delta)) = \operatorname{Arg} \operatorname*{Min}_{a\in A, c\in\mathbf{R}} Q(a,c;\delta), \tag{10}$$

where

$$Q(a,c;\delta) = \int_X d^2(x,a,c,\delta)\cdot\varphi(x;\delta)\,\mathrm{d}x, \tag{11}$$

$$d(x,a,c;\delta) = \ln\varphi(x;\delta) - c - \ln k(x;a). \tag{12}$$

Eqs. (2) and (10)–(12) can be interpreted as a (functional) generalized least squares (GLS) problem with x being distributed according to the initial sampler p and weight g. However, MC approximations of Q based upon i.i.d. draws from p would generally be highly inaccurate due to the typically (very) high MC sampling variance of g. In contrast, MC

approximations based upon an efficient sampler $m(x|a)$ would be expected to be far more accurate and numerically reliable. Therefore, rewrite Q as

$$Q(a,c;\delta) = \int_X d^2(x,a,c;\delta)\cdot\omega(x;\delta,a)\cdot m(x|a)\,\mathrm{d}x, \tag{13}$$

where $\omega(x;\delta,a)$ has been defined in (5).

The critical EIS step amounts to replacing Q by an MC approximation thereof. The complication arising from the fact that the auxiliary sampler itself depends upon a can be resolved by a standard fixed point argument. Specifically, our baseline EIS/GLS algorithm consists of computing a (converging) sequence of GLS estimates $\{\hat{a}_k(\delta)\}$ based upon the following MC recursion:

$$(\hat{a}_{k+1}(\delta),\hat{c}_{k+1}(\delta)) = \mathrm{Arg}\min_{a\in A,c\in\mathbf{R}} \overline{Q_S}(a,c;\delta|\hat{a}_k(\delta)), \tag{14}$$

where

$$\overline{Q_S}(a,c;\delta|\hat{a}_k(\delta)) = \frac{1}{S}\sum_{i=1}^{S} d^2(\tilde{x}_i^k,a,c,\delta)\cdot\omega(\tilde{x}_i^k;\delta,\hat{a}_k(\delta)), \tag{15}$$

and $\{\tilde{x}_i^k; i:1\to S\}$ denotes a set of i.i.d. draws from $m(x|\hat{a}_k(\delta))$. An initial value $\hat{a}_0(\delta)$ can be produced by conventional local approximations techniques or, more conveniently by p itself (since we can always add p to the class M and reinterpret it as $m(x|\hat{a}_0(\delta))$). It is generally advisable to set all weights ω equal to one during the initial iteration(s) to avoid numerical instability of GLS computations under high variance weights. Actually, for most problems, the OLS version of (15), whereby all weights remain set equal to one, is essentially as efficient as its GLS counterpart. At convergence the EIS estimate of $G(\delta)$ is trivially given by

$$\overline{G_S}(\delta) = \frac{1}{S}\sum_{i=1}^{S}\omega(\tilde{x}_i^k;\delta,\hat{a}_k(\delta)). \tag{16}$$

Note immediately that, in sharp contrast with local approximation techniques such as those reviewed in Section 2.2, EIS approximations are global in the sense that they are obtained by GLS on the full support X of φ.

In practice, we generally recommend against presetting a number R of EIS iterations. Instead, as illustrated in Section 3.6, we prefer using a stopping rule based upon a relative change threshold of the order of 10^{-3}–10^{-5}. Actually, we have found the speed of convergence to be a reliable rule-of-thumb indicator of the (in)adequacy of the class M of samplers. In particular, at an exploratory stage, occasional failure to converge is a clear indicator of potential pathologies (e.g. bimodality of the target density) which, in extreme cases, might require extensions of the class M of samplers. This being said, we have found EIS convergence to be very fast in well-behaved applications such as those presented in Section 5 where three iterations suffice.

Nevertheless, convergence remains subject to a critical implementation step which is familiar to users of numerical optimization procedures for MC functional approximations (see e.g. McFadden, 1989 in the context of the method of simulated moments). Specifically, whether for the convergence of the EIS–GLS recursion or for subsequent numerical optimization of $\overline{G_S}(\delta)$, it is critical that MC estimates of Q and G be smooth (continuous) functions of δ. Such smoothness is obtained by the application of a simulation technique

known as that of common random numbers (CRN), as described e.g. in Hendry (1984). Specifically, CRN requires that all successive sequences of i.i.d. draws $\{\tilde{x}_i^k; i: 1 \to S\}$ be generated as non-stochastic transformation of a single set $\{\tilde{u}_i; i: 1 \to S\}$ of canonical random draws, i.e. draws from a distribution which does not depend upon a and/or δ. The simplest application of CRN is that where $\tilde{x}_i^k$ is normally distributed with mean m_k and standard deviation s_k, in which case we have

$$\tilde{x}_i^k = m_k + s_k \cdot \tilde{u}_i \quad \text{with } \tilde{u}_i \sim \mathrm{N}(0,1). \tag{17}$$

More generally, one can rely upon the inversion technique as described e.g. in Devroye (1986). Let $F^{-1}(\cdot|a)$ denote the inverse cumulative distribution function (*cdf*) associated with $m(x|a)$. Then

$$\tilde{u}_i \sim \mathrm{U}(0,1) \quad \text{and} \quad \tilde{x}_i^k = F^{-1}(\tilde{u}_i|a) \Rightarrow \tilde{x}_i^k \sim m(\cdot|a). \tag{18}$$

While numerical inversion of F can be computationally demanding, it remains fully justified by the considerable efficiency gains resulting from the use of smooth MC functional approximations. Moreover, though this is not the focus of the present discussion, univariate EIS generally requires very low-dimensional auxiliary parametrization, in which case common interpolation techniques in (u, a) from an initial table of values of $F^{-1}(u|a)$ can prove numerically very fast. An example will be provided in Section 5.2 below.

Note finally that, as discussed further in Section 3.4 below, another major benefit of the CRN technique in the context of EIS is that it contributes by itself in a major way to reducing further the MC sampling variance of EIS estimates.

3.2. EIS from the exponential family of distributions

If m belongs to the exponential family of distributions, then $m(x|a)$ can be written as

$$m(x|a) = \chi(a) \cdot b(x) \cdot \exp(a' \cdot t(x)), \tag{19}$$

where $a \in A$ denotes a natural parametrization in the sense of Lehmann (1986, Section 2.7) and $t(x)$ a sufficient statistic of fixed dimension (for all sample sizes). The corresponding logkernels are

$$\ln k(x; a) = \ln(b(x)) + a' \cdot t(x). \tag{20}$$

Direct substitution in the expression of $\mathrm{d}(x, a, c; \delta)$ as defined by Eq. (12) indicates that the EIS auxiliary minimization problem is now linear in a.

Note that $a \in A$ typically has to satisfy inequality constraints for the integrability of k. For example, the variance of $x|a$ has to be strictly positive. It has been our experience that such constraints are typically non-binding in which case standard OLS (GLS) formulae provide an analytical solution to the minimization problem (14). Rare exceptions occur e.g. when the initial sampler $p(x|\delta)$ concentrates draws in a region where $\ln \varphi(x; \delta)$ is locally convex (such as a tail area, or an area squeezed between two local peaks). Efficient numerical solutions for such occasional pathologies consist of inflating the variance of $p(x|\delta)$ and/or setting the corresponding coefficient in a at some arbitrary value $\varepsilon > 0$.

Problems of extreme multicollinearity can also occur when the dimensionality of $t(x)$ is large (in relation to the number S of draws). A fully operational solution consists of implementing a trivial shrinkage version of OLS (GLS) which has proven to be

numerically far more reliable than increasing S itself. A spectacular implementation of shrinkage is provided in Section 5.2 below, where we run millions of EIS auxiliary regressions with 30 regressions and only $S = 100$ draws without any human intervention.

Formula (20) also implies that, subject to integrability restrictions, A is closed under addition in the sense that if $a_1 \in A$ and $a_2 \in A$ then $a_3 = a_1 + a_2$ is also in A. Equivalently, $K = \{k(x;a); a \in A\}$ is closed under multiplication. This property results in additional simplifications of the EIS auxiliary LS problem when there exists $a_\varphi(\delta) \in A$ such that $\varphi(x;\delta)$ can be factorized into

$$\varphi(x;\delta) = g_0(x;\delta) \cdot k(x; a_\varphi(\delta)). \tag{21}$$

In view of Eq. (2), $k(x; a_\varphi(\delta))$ could denote p itself when $p \in M$ but could also include additional factors from g. Under Eq. (21) we can take full advantage of the closure of K by redefining $k(x;a)$ as the following product:

$$k(x; a(\delta)) = k(x; a_1) \cdot k(x; a_\varphi(\delta)) \tag{22}$$

with $a(\delta) = a_1 + a_\varphi(\delta)$. It follows from Eqs. (12), (21) and (22) that $k(x; a_\varphi(\delta))$ cancels out in the auxiliary EIS regressions which simplify into LS regressions of $\ln[g_0(x;\delta)/b(x)]$ on $t(x)$. Let $\hat{a}_1(\delta)$ denote the corresponding LS estimate. Then $\hat{a}(\delta)$ is given by

$$\hat{a}(\delta) = \hat{a}_1(\delta) + a_\varphi(\delta). \tag{23}$$

Examples are provided in Section 5 below.

3.3. Selection of the class of auxiliary samplers

Preselection of a class M of auxiliary samplers is by no means specific to IS. It is equally critical for a much broader class of MC simulation techniques relying upon ratios of the form φ/m, including among others Metropolis-Hastings (MH) and related MCMC. Actually as outlined in Section 6 below, EIS techniques can be used to significantly improve the numerical performance of any such MC technique.

As discussed above, the selection of a class M of samplers within the exponential family of distributions considerably simplifies the implementation of EIS by reducing it to a simple linear LS auxiliary problem. In particular, Eqs. (21) and (22) provide useful guidelines for the selection of M within that family. Moreover, one can usefully rely upon a broad range of conventional regression diagnostic tests to examine whether $\ln k(x;a)$ provides a sufficiently close global approximation to $\ln \varphi(x;\delta)$. The EIS regression R^2 provides a direct measure of (global)goodness of fit and conventional residual analysis can help identifying lack of fit and suggesting extensions of k.

Obviously such extensions might require extending M beyond the exponential family of distributions, in which case the auxiliary EIS regressions in Eq. (14) would become nonlinear in a. Important examples of such extensions are the (thick tail) Student distribution for Bayesian posterior inference, and power transformation of the gamma distribution, such as the Weibull distribution, for empirical auction models. Even in such cases, one can take full advantage of the fact that EIS is based upon a recursive sequence of nonlinear LS estimates $\{\hat{a}_k(\delta)\}$ to efficiently implement nonlinear minimization. In particular, one only needs to implement a single minimization step (e.g. a single Newton–Raphson iteration) within each EIS iteration. Moreover, there exists a wide range

of numerically efficient estimation techniques for nonlinear LS problems which could usefully be applied to Eq. (14). Such extensions belong to our immediate research agenda.

3.4. The infinite variance problem

A common objection raised against IS is the possibility that its MC sampling variance as given in Eq. (7) might not exist. The classic example is that where M has thinner tails than φ in which case ω tends to infinity in the tails of m. This situation calls for some important qualifications:

1. It is by no means specific to IS and actually applies to all MC techniques which rely upon ratios of the form φ/m including therefore MH and MCMC.
2. The common suggestion that one should, therefore, test the stability of (E)IS estimates under increasing number of draws is highly unreliable since it depends upon the realization of extreme outliers. As illustrated below it is actually straightforward to artificially create such an outlier and directly examine its impact on the IS estimate. Moreover, as discussed below, direct diagnostics of variance stability obtain as immediate byproducts of our EIS algorithm.
3. Most importantly, our EIS implementation based upon a single set of CRN draws virtually eliminates the infinite variance problem for all practical purposes. The intuition behind this assertion is directly linked to the familiar concept of influential observations within a regression framework. Specifically, any CRN outlier draw will immediately prove highly influential within the EIS auxiliary regressions. Whence, the weight $\omega(\cdot)$ associated with that particular draw will typically remain close to one and will have no detectable impact on the variance of the EIS estimates. This will be dramatically illustrated in Section 3.6 below where we revisit the now (in)famous Student-t/Gaussian example. Obviously, the MC variance under such thin tails EIS samplers remains theoretically infinite though the actual probability of a variance explosion has been eliminated for all practical purposes (We have never faced one after several years of EIS applications such as those described in Section 6 below and literally billions of draws). It is up to the practitioner to decide whether the theoretical possibility of a MC variance explosion resulting from essentially zero probability extreme outliers remains a concern. Actually, we have found that a more relevant concern related to the use of thin tails EIS samplers is that they often effectively truncate the integrals in (1), as illustrated by the applications in Section 3.6 below.

Obviously, in order to carefully assess the empirical relevance of the potential problems associated with thin tails EIS samplers, one has to be able to detect their presence. We propose here a diagnostic which is an immediate byproduct of our EIS algorithm, does not require any additional draws and, as illustrated below, is highly effective detecting when the tails of an EIS sampler are thinner than those of the integrand φ. It consists of evaluating the MC sampling variance of the EIS estimate, as defined in Eq. (7) under two different auxiliary samplers, one being the EIS sampler itself and the other a sampler in M with inflated variance (as illustrated below, a factor of 3–5 works very effectively). The intuition underlying such a diagnostic is that of producing outliers in the far tails of the EIS samplers (where the weights ω are precisely expected to explode if the tails of that

sampler are too thin) and to measure their impact using the EIS variance metric. The MC estimates of the two variances to be compared are given by

$$\hat{V}_S(\cdot) = \frac{1}{S}\sum_{i=1}^{S} h[d^2(\tilde{x}_i, \hat{a}, \hat{c}, \delta)] \cdot \frac{\varphi(\tilde{x}_i; \delta)}{m(\tilde{x}_i|a)}, \tag{24}$$

where h is defined in Eq. (75) in Appendix A, $(\hat{a}, \hat{c})$ denote the EIS parameter values and $\{\tilde{x}_i; i: 1 \to S\}$ denotes S i.i.d. draws from an auxiliary sampler $m(x|a)$. See also Lee and Koopman (2004) for a test based upon extreme value theory.

3.5. Numerical accuracy of EIS estimate

MC (functional) estimates of $G(\delta)$ are typically not end products of an EIS analysis. For example $G(\delta)$ might represent a likelihood function for a model in which x denotes a set of latent variables to be integrated out. Let $\delta = (\theta, y)$ where θ denotes parameters to be estimated and y observable variables. $\overline{G_S}(\delta)$ might then be instrumental in the computation of a simulation point estimate $\hat{\theta}_S(y)$, to be interpreted as a MC estimate of the corresponding (infeasible) estimate $\hat{\theta}(y)$. Note immediately an important conceptual distinction between the numerical accuracy of $\hat{\theta}_S(y)$ as an MC estimate of $\hat{\theta}(y)$ and the conventional statistical accuracy of $\hat{\theta}(y)$ as an estimator of θ.

There now exists a vast literature on the statistical properties of simulation estimators—see e.g. Gourieroux and Monfort (1996). However, their results do not apply here since an EIS sampler is not a statistical component of the model itself but rather an artificial auxiliary sampler constructed to maximize numerical efficiency. Neither does formula (7) apply per se since the EIS parameter $\hat{a}(k)$ is itself random. Under such circumstances the most appropriate techniques for assessing the MC numerical accuracy of the final results consists of rerunning the complete application (computation of $\hat{a}(k)$, estimation of $\overline{G_S}(\delta)$, optimization in θ) under different seeds for the baseline CRN sequence. In the case of an estimation problem this will produce R i.i.d. MC estimates of $\hat{\theta}(y)$, say $\{\hat{\theta}^i_S(y); i: 1 \to R\}$ from which a final average MC estimate is obtained as

$$\bar{\hat{\theta}}^R_S(y) = \frac{1}{R}\sum_{i=1}^{R} \hat{\theta}^i_S(y), \tag{25}$$

together with its MC numerical standard deviation. Note that such an exercise is made possible by the fact that the EIS algorithm is fully automated once its user has specified $\ln\varphi$ and $\ln k$ and can, therefore, run without human intervention (calibration, restart,…). Liesenfeld and Richard (2007b) have recently proposed using auxiliary EIS steps to similarly automate the selection and calibration of the auxiliary samplers used for MCMC implementations.

Finally, we note a fundamental duality between the simulation treatment of the observation y and of the auxiliary CRN vector u. As described above, replications of u given y are used to compute numerical standard deviations for $\hat{\theta}_S(y)$ as an MC estimator of $\hat{\theta}(y)$. Equally trivially, replications of y given u and θ, can be used to compute statistical standard deviations for $\hat{\theta}(y)$ as an estimator of θ. In all applications which are presented in Section 5 below, we provide both numerical and statistical standard deviations for our EIS-ML estimates.

3.6. Two pilot univariate applications

First, let φ denote the following density kernel:

$$\varphi(x;\delta) = \exp(-x^{1/\delta}), \quad x>0, \ \delta>0 \tag{26}$$

in which case $G(\delta) = \Gamma(\delta+1)$. Let also

$$k(x;a) = \exp(-ax), \quad x>0, \ a>0 \tag{27}$$

denote an exponential density kernel. While k is not a particularly good choice of IS sampler for this problem, it enables us to illustrate several key points. Note, in particular, that k has thinner tails than φ for $\delta>1$ and that the variance of $\overline{G}(\delta)$ as defined in Eq. (7) is infinite for $\delta>2$. Under (27), the EIS auxiliary regression, as defined in Eqs. (10)–(12), amounts to a simple LS regression of $x^{1/\delta}$ on x. If the x's are drawn from $m(x|a)$, then the true value of the corresponding regression coefficient is given by

$$\hat{a}(\delta) = a^{1-1/\delta} \cdot \frac{1}{\delta} \cdot \Gamma\left(1+\frac{1}{\delta}\right) \tag{28}$$

with a fixed point solution

$$\hat{a}_f(\delta) = \left[\frac{1}{\delta} \cdot \Gamma\left(1+\frac{1}{\gamma}\right)\right]^{\delta}. \tag{29}$$

Formula (21) does not apply here so that we do not have a natural initial sampler $m(x|a_\varphi(\delta))$. But it is obvious from formulae (26) and (27) that a decreasing value of $a_\varphi(\delta)$ such as $a_\varphi(\delta) = 1/\delta$ should work well (other starting values such as $a_\varphi(\delta) = 1$ work equally well but may require a few additional EIS iterations). Results are presented in Table 1 for values of δ ranging from 0.6 to 2.6. The last column in Table 1 presents the diagnostic introduced in Section 3.4 in the form of a ratio between two MC estimates of the variance $\hat{V}_s$ as defined in Eq. (24), with a set equal to $5 \times \hat{a}(\delta)$ for the numerator and to $\hat{a}(\delta)$ for the denominator. Results are provided for $G(\delta)$ but also for $G_2(\delta) = \int x^2 \cdot \varphi(x;\delta)\,\mathrm{d}x$ and $E(x^2) = G_2(\delta)/G(\delta)$. While the exponential sampler is not particularly efficient for values of δ very different from 1, this pilot application illustrates several important points.

- The final EIS regression coefficients are accurate LS estimates of $\hat{a}(\delta)$. The (relative) convergence criterion for $\hat{a}(\delta)$ was set at 10^{-5} which is far too stringent. Using instead 10^{-2} produces results which are identical to three decimals and reduces the number of EIS iterations by more than half on average.
- Our convergence test is clearly very sensitive to the thin tails problem and rapidly explodes for $\delta>2$, when the MC variance of $\overline{G}(\delta)$ is infinite.
- The truncation effect alluded to in Section 3.4 translates into downward biased EIS estimates of $G(\delta)$ as soon as k has thinner tails than $\varphi(\delta>1)$—by two standard deviations for $\delta = 1.2$ and four or more for $\delta = 1.6$ and beyond.
- All results in Table 1 are derived under an EIS sampler for φ which explains why results for $G_2(\delta)$ are less accurate than those for $G(\delta)$. We ran the same computations under an EIS sampler for $x^2\varphi$ and, as expected the relative accuracy is reversed. Actually, if $E(x^2)$ were our primary object of interest, it would be more efficient to compute $G(\delta)$ and $G_2(\delta)$ under their own EIS using a common set of CRNs (to maximize positive correlation between G and G_2).

Table 1
EIS results for φ in Eq. (26)

δ		$\hat{a}(\delta)$	$G(\delta)$	$G_2(\delta)$	$E(X^2)$	EIS iter.	Var.[a] ratio
0.6	True	1.736	0.8935	0.5588	0.6254		
	EIS	1.688	0.8942	0.5745	0.6417	29.52	1.053
	St. dev.	0.013	0.0024	0.0083	0.0086		0.0012
0.8	True	1.321	0.9314	0.9937	1.067		
	EIS	1.304	0.9338	1.018	1.091	9.00	1.104
	St. dev.	0.004	0.0011	0.014	0.016		0.018
1.2	True	0.7466	1.102	4.460	4.048		
	EIS	0.7473	1.100	4.698	4.263	7.0	1.350
	St. dev.	0.0028	0.001	0.216	0.191		0.048
1.6	true	0.3959	1.430	28.54	19.96		
	EIS	0.4073	1.411	23.42	16.44	12.95	18.98
	St. dev.	0.0034	0.004	1.58	1.05		2.00
2.0	true	0.1963	2.000	240.0	120.0		
	EIS	0.2102	1.948	156.1	77.5	17.95	77.47
	St. dev.	0.0035	0.011	15.3	7.0		7.00
2.4	true	0.0916	2.981	2,521	845.6		
	EIS	0.0995	2.863	1,383	457.0	22.95	2,224
	St. dev.	0.0023	0.019	181	55.3		680

Notation and details:

$$G(\delta) = \int \varphi(x;\delta)\,\mathrm{d}x, \quad G_2(\delta) = \int x^2\varphi(x;\delta)\,\mathrm{d}x,$$

$E(x^2) = G_2(\delta)/G(\delta)$, $\hat{a}(\delta)$: EIS regression coefficient.

MC draws: $S = 100$. All standard deviations are based upon 100 replications of EIS under different CRN seeds and are for the averages reported here. Standard deviations for an individual EIS outcome are 10 times higher.

[a]See Section 3.6 for the precise definition of this MC variance ratio diagnostic.

Our second pilot application considers the classical (pathological) problem of approximating a standardized Student-t kernel by a Gaussian kernel. Let

$$\varphi(x;\delta) = \left(1 + \frac{x^2}{\delta - 2}\right)^{-(1/2)(\delta+1)}, \quad \delta > 2 \tag{30}$$

in which case $G(\delta) = B(\frac{1}{2}, \nu/2)\cdot\sqrt{\nu-2}$. The class of Gaussian kernels is parametrized as

$$k(x;a) = \exp\left(-\frac{a}{2}x^2\right), \quad a > 0 \tag{31}$$

with a variance equal to a^{-1}. Note that $\varphi(x;\delta)$ converges to $k(x|1)$ for $\delta \to \infty$. Results are presented in Table 2 and here again, highlight the high sensitivity of our diagnostic ratio. Note that in contrast the MC standard deviations for $\overline{G}(\delta)$ fail to detect the problem in the absence of a major outlier. Here again, it is the truncation effect which is dominant for low

Table 2
EIS results for student-t kernel in Eq. (30)

δ		$\hat{a}(\delta)$	$G(\delta)$	$G_2(\delta)$	$E(X^2)$	EIS iter.	Var. ratio
2.5	True	–	1.236	1.236	1.000		
	EIS	2.827	1.195	0.5748	0.4666	13.36	72,000
	St. dev.	0.092	0.005	0.0084	0.0600		68,460
3.0	True	–	1.571	1.571	1.000		
	EIS	1.929	1.522	0.8193	0.5346	10.77	713.2
	St. dev.	0.043	0.005	0.0312	0.0190		214.6
10.0	True	–	2.299	2.299	1.000		
	EIS	1.028	2.295	2.306	1.004	5.62	155.2
	St. dev.	0.014	0.003	0.044	0.018		47.8
150.0	True	–	2.494	2.494	1.000		
	EIS	1.003	2.493	2.457	0.9852	2.80	2.344
	St. dev.	0.001	0.000^{+}	0.038	0.0151		0.150
Outlier 1	True	–	1.236	–	–		
	EIS	2.787	1.198	–	–	14.25	2.44 E + 12
	St. dev.	0.095	0.006	–	–		2.43 E + 12
Outlier 2	True	–	1.236	–	–		
	EIS	2.787	32.06	–	–	–	3.99 E + 12
	St. dev.	0	30.67	–	–		3.97 E + 12

Notation and details are the same as for Table 1.

degrees of freedom ($\delta = 2.5$ and 3.0) with downward biases of the order of 10 MC standard deviations (but also very large values of the variance ratio diagnostic).

In order to illustrate the extreme robustness of EIS to outliers, we recalculated all results for $\delta = 2.5$ injecting a single outlier equal to -5.998 (corresponding to a p-value of 1.0E − 09 for the standardized Normal!) The results are reported in the bottom two lines of Table 2 for the following scenarios:

- *Outlier* 1: We run EIS under CRNs exactly as described in Section 3.4. The results for $G(\delta)$ hardly change as the EIS auxiliary regression captures this influential outlier. Foremost, note that the variance ratio test explodes which is highly desirable as the standard deviation of the EIS estimate of $G(\delta)$ provides no indication as to the presence of an outlier.
- *Outlier* 2: We run EIS under a fixed value of $\hat{a}(\delta)$, thereby preventing any adjustment to the outlier (mimicking the behavior of conventional IS). Note the immediate explosion in all results fully in line with conventional criticisms of standard IS procedures. The comparison between the last two runs dramatically illustrates the critical role of our use of a single set of CRNs for both the EIS search for $\hat{a}(\delta)$ and the subsequent estimation of $G(\delta)$. It also rationalizes the fact that over now several years of reliance upon EIS and billions of draws, we never faced a variance explosion due to outliers, which is also why we need a reliable diagnostic such as the one we described in Section 3.4.

4. High-dimensional sequential EIS

The EIS-LS algorithm introduced in Section 3 only applies to very low-dimensional x's. In higher dimensional set-ups, feasibility requires that it be decomposed into a sequence of low-dimensional optimization problems. We now present an operational sequential implementation of EIS which, as strikingly illustrated by the applications discussed in Section 6 below, is applicable in very high-dimensional integration problems. It exploits the fact that high-dimensional models are typically specified not as a single joint distribution but as a sequence of conditional distributions whereby latent and observable variables are generated individually (sequentially in the time dimension or in parallel in cross sections). Before we present sequential EIS in its full generality, let us introduce its heuristics justification through a simple example which we will revisit in Section 5 below.

4.1. Pilot example

Consider a latent variable model consisting of an AR(1) process for the latent variable λ_t and a (static) measurement process for an observable y_t given λ_t. This model is of the general form represented by Eq. (2) together with

$$p(x|\delta) = \prod_{t=1}^{T} p_t(\lambda_t|\lambda_{t-1}, \theta),$$

$$g(x;\delta) = \prod_{t=1}^{T} g_t(y_t|\lambda_t, \theta). \tag{32}$$

$\delta = (\theta; y_{(T)}), x = \lambda_{(T)}, y_{(t)} = (y_1, \ldots, y_t)$ and $\lambda_{(t)} = (\lambda_1 \cdots \lambda_t)$. Note that, as commonly the case with latent variable models, the state space equation represented here by the density $p(x|\delta)$ does not depend upon the observable $y_{(T)}$. This is precisely why p would be a highly inefficient sampler for $\lambda_{(T)}$. Interdependence between $\lambda_{(T)}$ and $y_{(T)}$ originates from the measurement equation which is represented here by the functional $g(x;\delta)$.

Our objective is that of constructing a sequential sampler which accounts for such interdependency by being a close approximation to the analytically intractable posterior density of $\lambda_{(T)}$ given $y_{(T)}$. For ease of notation it is assumed that λ_0 is given. Let

$$\varphi_t(\lambda_t, \lambda_{t-1}; \delta) = p_t(\lambda_t|\lambda_{t-1}, \theta) \cdot g_t(y_t|\lambda_t, \theta). \tag{33}$$

The full likelihood is given by

$$G(\theta; y_{(T)}) = \int \prod_{t=1}^{T} \varphi_t(\lambda_t, \lambda_{t-1}; \delta)\, \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_T. \tag{34}$$

In order to highlight the fundamental interdependency of such (potentially high-dimensional) integrals as well as to introduce the recursive concept of sequential EIS, we can rewrite the integral in $\lambda_{(T)}$ as a backward recursive sequence of integrals in λ_t, for $t : T \to 1$. This sequence is defined as

$$\ell_t(\lambda_{t-1}; \delta) = \int \varphi_t(\lambda_t, \lambda_{t-1}; \delta) \cdot \ell_{t+1}(\lambda_t; \delta)\, \mathrm{d}\lambda_t, \tag{35}$$

and is initialized by $\ell_{T+1}(\cdot) \equiv 1$. The full likelihood $G(\theta; y_{(T)})$ is then given by $\ell_1(\cdot)$. Note the transfer of ℓ_{t+1} from period $t+1$ to period t which is critical in order to properly account for the dynamics of the model. However except for linear Gaussian models, integrals such as these are bound to be analytically intractable.

In contrast, EIS will rely upon a sequence of approximating kernels $k_t(\lambda_t, \lambda_{t-1}; a_t)$ with known analytical (functional) integrating constants

$$\chi_t(\lambda_{t-1}; a_t) = \int k_t(\lambda_t, \lambda_{t-1}; a_t)\, \mathrm{d}\lambda_t \tag{36}$$

corresponding to the following sequential samplers:

$$m_t(\lambda_t|\lambda_{t-1}, a_t) = \frac{k_t(\lambda_t, \lambda_{t-1}; a_t)}{\chi_t(\lambda_{t-1}; a_t)}. \tag{37}$$

The $\chi_t s$ can trivially be transferred back from one period to another. Whence the likelihood integral can be rewritten as

$$\begin{aligned} G(\theta; y_{(T)}) &= \int \cdots \int \prod_{t=1}^{T} \frac{\varphi_t(\lambda_t, \lambda_{t-1}; \delta)}{m_t(\lambda_t|\lambda_{t-1}, a_t)} \cdot m_t(\lambda_t|\lambda_{t-1}, a_t) \cdot \mathrm{d}\lambda_T \cdots \mathrm{d}\lambda_1 \\ &= \chi_1(a_1) \cdot \int \cdots \int \prod_{t=1}^{T} \left[\frac{\varphi_t(\lambda_t, \lambda_{t-1}; \delta) \cdot \chi_{t+1}(\lambda_t; a_{t+1})}{k_t(\lambda_t, \lambda_{t-1}; a_t)} \right] m_t(\lambda_t|\lambda_{t-1}, a_t) \\ &\quad \cdot \mathrm{d}\lambda_T \cdots \mathrm{d}\lambda_1. \end{aligned} \tag{38}$$

Joint optimization with respect to $a' = (a_1 \cdots a_T)$ is clearly unfeasible for high-dimensional a. Instead we construct a (backward) recursive sequence of individual EIS minimizations whereby, conditionally on $\hat{a}^k_{t+1}(\delta), \hat{a}^k_t(\delta)$ (for $t : T \to 1$) is chosen so as to minimize the MC variance of the ratio

$$R^k_t(\lambda_t, \lambda_{t-1}; a_t, \delta) = \frac{\varphi_t(\lambda_t, \lambda_{t-1}; \delta) \cdot \chi_{t+1}(\lambda_t; \hat{a}^k_{t+1}(\delta))}{k_t(\lambda_t, \lambda_{t-1}; a_t)}. \tag{39}$$

EIS minimization with respect to a_t proceeds exactly as described in Section 3.1 under (full) trajectory draws $\{(\tilde{\lambda}^k_{t,i}, t : 1 \to T); i : 1 \to S\}$ from the (tentative) joint EIS sampler obtained at iteration k.

By comparison with formula (35) note that if the ratios $\varphi_t \cdot \chi_{t+1}/k_t$ were all constant then χ_t would be proportional to ℓ_t and EIS would produce an exact result. More generally, our sequential procedure implicitly assumes that the functional $\chi_{t+1}(\lambda_t; \hat{a}_{t+1}(\delta))$ are sufficiently close approximations for $\ell_{t+1}(\lambda_t; \delta)$, i.e. that the dynamic structure of the latent model is adequately captured by $\{\chi_t(\cdot), t : T \to 1\}$. This in turn requires that k_t closely approximates the product $\varphi_t \chi_{t+1}$ with respect to $(\lambda_t, \lambda_{t-1})$, not just λ_t since there is no revisiting of period t dynamics once χ_t has been produced.

Once the (backward) EIS sequence $\{(\hat{a}^k_t(\delta); t : T \to 1); k = 1, 2, \ldots\}$ has converged to a fixed point, the (forward) EIS estimate of $G(\theta; y_{(T)})$ obtains in the usual way and is given by

$$\bar{G}^k_S(\theta; y_{(T)}) = \frac{1}{S} \sum_{i=1}^{S} \left[\prod_{t=1}^{T} \frac{\varphi(\lambda_t, \lambda_{t-1}, \delta)}{m(\lambda_t|\lambda_{t-1}, \hat{a}^k_t(\delta))} \right]_{\{\lambda_t = \tilde{\lambda}^k_{t,i}\}}. \tag{40}$$

Full details under more general sampling preordering factorizations are presented in the next section. Note finally that violations of the implicit assumption that χ_t is near

proportional to ℓ_t would reduce the efficiency of the resulting EIS sampler but would not invalidate the consistency of the EIS-MC estimate in Eq. (40)—assuming the standard conditions discussed earlier are still valid. The two high- dimensional applications in Section 5.1 will highlight the exceptional numerical efficiency of sequential EIS in (very) high-dimensional important classes of latent variable models.

The key advantages of this procedure are its generic numerical tractability together with the fact that it implicitly produces an EIS sequential sampler for the λ_ts which is conditional upon the entire sample $y_{(T)}$. In contrast, alternative forward algorithms such as the widely used particle filters—see e.g. Pitt and Shephard (1999)—produces a sequence of filtered samplers, whereby λ_t is drawn conditionally on λ_{t-1} and $y_{(t)}$ only. Whence, whenever feasible, sequential EIS is intrinsically numerically more efficient than particle filters. The first author is currently investigating how to improve the numerical efficiency of particle filters by means of auxiliary (forward) EIS steps but this extension is outside the scope of the recent paper. We now describe sequential EIS in its full generality.

4.2. Sequential EIS

With reference to Eqs. (1) and (2) we now assume that there exists a natural sampling preordering partition of x into low-dimensional components, say $x = (x_1 \cdots x_L)$. A subscript ℓ is used to avoid focusing attention on time series models exclusively. The functionals φ, p and g are partitioned conformably with x into

$$\varphi(x;\delta) = \prod_{\ell=1}^{L} \varphi_\ell(X_\ell;\delta), \quad g(x;\delta) = \prod_{\ell=1}^{L} g_\ell(X_\ell;\delta), \tag{41}$$

$$p(x|\delta) = \prod_{\ell=1}^{L} p_\ell(x_\ell|X_{\ell-1},\delta), \tag{42}$$

where $X_\ell = (x_1 \cdots x_\ell)$. The use of a subscript ℓ for the individual components of φ, p and g allows in particular for a flexible treatment of initial conditions without explicit notational treatment. The EIS sampler $m(x|a)$ is partitioned conformably as

$$m(x|a) = \prod_{\ell=1}^{L} m_\ell(x_\ell|X_{\ell-1},a_\ell) \tag{43}$$

with $a = (a_1 \cdots a_L) \in A = \Pi_{\ell=1}^{L} A_\ell$. A kernel k_ℓ for m_ℓ is a function of (X_ℓ, a_ℓ) which is analytically integrable with respect to x_ℓ. The following notation applies to k_ℓ and m_ℓ:

$$m_\ell(x_\ell|X_{\ell-1},a_\ell) = \frac{k_\ell(X_\ell;a_\ell)}{\chi_\ell(X_{\ell-1};a_\ell)}, \tag{44}$$

$$\chi_\ell(X_{\ell-1};a_\ell) = \int k_\ell(X_\ell;a_\ell)\,\mathrm{d}x_\ell. \tag{45}$$

The generalization of formulae (12), (14) and (15) to this general sequential framework is conceptually straightforward but notationally somewhat tedious. Specifically, step

ℓ ($\ell : L \to 1$) of our sequential EIS is fully characterized by the following set of formulae:

$$(\hat{a}^{\ell}_{k+1}(\delta), \hat{c}^{\ell}_{k+1}(\delta)) = \text{Arg} \min_{a \in A_{\ell}, c \in \mathbf{R}} \overline{Q^{\ell}_{S}}(a, c; \delta | \hat{a}^{\ell}_{k}(\delta)), \tag{46}$$

$$\overline{Q^{\ell}_{S}}(a, c; \delta | \hat{a}^{\ell}_{k}(\delta)) = \frac{1}{S} \sum_{i=1}^{S} d^{2}_{\ell}(\tilde{X}^{k}_{\ell,i}, a, c; \delta) \cdot \omega_{\ell}(\tilde{X}^{k}_{\ell,i}; \delta, \hat{a}^{\ell}_{k}(\delta)), \tag{47}$$

$$d_{\ell}(\tilde{X}^{k}_{\ell,i} a, c; \delta) = \ln[\varphi_{\ell}(\tilde{X}^{k}_{\ell,i}; \delta) \cdot \chi_{\ell+1}(\tilde{X}^{k}_{\ell,i}; \hat{a}_{\ell+1}(\delta))] - c - \ln k_{\ell}(\tilde{X}^{k}_{\ell,i}; a), \tag{48}$$

$$\omega_{\ell}(\tilde{X}^{k}_{\ell,i}; \delta, \hat{a}^{\ell}_{k}(\delta)) = \frac{\varphi_{\ell}(\tilde{X}^{k}_{\ell,i}; \delta)}{m_{\ell}(\tilde{X}^{k}_{\ell,i} | \hat{a}^{\ell}_{k}(\delta))}, \tag{49}$$

where $\tilde{X}^{k}_{\ell,i} = (\tilde{x}^{k}_{1,i} \cdots \tilde{x}^{k}_{\ell,i})$ and $\{\tilde{x}^{k}_{\ell,i}; \ell : 1 \to L; i : 1 \to S\}$ denotes S i.i.d. trajectories drawn from the kth sequential sampler $\{m_{\ell}(x_{\ell} | \tilde{X}^{k}_{\ell-1,i}); \ell : 1 \to L\}$. Our notation highlights the fact that sequential EIS draws complete trajectories for each k until convergence in $\hat{a}^{\ell}_{k}(\delta)$ obtains. As previously discussed, it is critical that the individual trajectories $\{x^{k}_{\ell,i}; \ell : 1 \to L; i : 1 \to S\}$ be derived from a single set of CRNs $\{\tilde{u}_{\ell,i}; \ell : 1 \to L; i : 1 \to S\}$. We conclude the general presentation of sequential EIS with a few operational details.

- As already mentioned in Section 3.1, it is usually advisable to set the weights equal to one at least in the initial EIS iteration(s) in order to avoid severe imbalances due to inefficient choices of initial samplers.
- Feasibility of the individual LS regressions requires that a_{ℓ} remains of relatively low dimension relative to the number S of trajectories being used. This will generally be the case under (common) conditional independence assumptions whereby there exists a low-dimensional subvector z_{ℓ} of X_{ℓ} such that

  $$\varphi_{\ell}(X_{\ell}; \delta) \equiv \varphi_{\ell}(z_{\ell}, \delta) \tag{50}$$

  in which case k_{ℓ} itself only depends on z_{ℓ} (and a_{ℓ}). Higher (relative to S) dimensionality of a_{ℓ} may result in severe multicollinearity in the auxiliary EIS regressions and/or violations of the constraints which need to be satisfied by $\hat{a}_{\ell}$ in order to guarantee that k_{ℓ} is a proper density kernel. This being said, occasional such problems can often be addressed by implementing a modest amount of shrinkage toward admissible a_{ℓ}s. An illustration of successful shrinkage is provided in Section 5.2 below where a_{ℓ} is 30-dimensional and is estimated under $S = 100$ draws.
- It is useful to reiterate here that it is critical that k_{ℓ} (closely) approximates φ_{ℓ} with respect to z_{ℓ} (not simply x_{ℓ}) in order to adequately capture the dynamic interdependence of the sequential integrals.
- Reruns of EIS under added observations are expected to be very fast since one can use as initial sampler the EIS sampler previously computed augmented by an initial sampler for the added dimensions. Under standard mixing assumptions one would expect that the impact of added dimensions would fade out as the algorithm runs backwards so that only a relatively small number of component subsamplers would have to be significantly adjusted.

5. Two high dimensional pilot applications

We now present two high-dimensional applications highlighting the exceptional numerical accuracy of ML-EIS estimation under small numbers of draws for two important classes of latent variable models: stochastic volatility and panels with unobserved heterogeneity along both dimensions.

5.1. Stochastic volatility

Stochastic volatility models play a central role in finance. Pioneering contributions are Taylor (1986), Melino and Turnbull (1990) or Duffie and Singleton (1993). It has long been recognized that natural *IS*, as defined in Eq. (6), is hopelessly inefficient—see Danielsson and Richard (1993) for a dramatic illustration of such inefficiency. In order to illustrate the full flexibility of our EIS algorithm in this context we consider here four versions of the same baseline model. Let y_t denote the daily return of a financial asset and λ_t its unobserved variance. Stochastic volatility (hereafter SV) models typically consist of a stochastic equation for y_t given λ_t and another for λ_t given λ_{t-1} (which can trivially be generalized into a higher order autoregressive process). Two versions of the density $g(y_t|\lambda_t,\cdot)$ will be considered: a fat tail Student-t density (S) and a thin tail normal density (N). For the density $p(\lambda_t|\lambda_{t-1},\cdot)$, we shall consider in turn an inverted-gamma (IG) density and a lognormal density (L). These four densities are parametrized as follows:

$$g_t^1(y_t|\lambda_t,\theta) \propto \lambda_t^{1/2}\cdot\left[1+\frac{(y_t-\mu)^2)}{\lambda_t(w-2)}\right]^{-(1/2)(w+1)}, \tag{51}$$

$$g_t^2(y_t|\lambda_t,\theta) \propto \lambda_t^{-1/2}\exp\left[-\frac{1}{2}\cdot\frac{(y_t-\mu)^2}{\lambda_t}\right], \tag{52}$$

$$p_t^1(\lambda_t|\lambda_{t-1},\theta) \propto r_{t-1}^{(1/2)v}\cdot\lambda_t^{-(1/2)(v+2)}\exp\left(-\frac{r_{t-1}}{\lambda_t}\right), \tag{53}$$

$$\text{with } r_{t-1}=\tfrac{1}{2}(\gamma+\partial\lambda_{t-1}) \quad \text{and} \quad E(\lambda_t|\lambda_{t-1})=\frac{2}{v-2}\,r_{t-1}=q+r\lambda_{t-1},$$

$$p_t^2(v_t|v_{t-1},\theta) \propto \exp\left[-\frac{1}{2\sigma^2}(v_t-q-rv_{t-1})^2\right], \tag{54}$$

where $v_t=\ln\lambda_t$ and θ includes the corresponding subset of the parameters $(q,r,\sigma,\mu,\omega,v)$. We shall consider here the four pairwise combinations of q_t and p_t, respectively, denoted by SG, SL, NG and NL, but shall only detail the sequential EIS implementation for the SG version since, as hinted in Section 2.1, the other versions only require fairly obvious modifications of the auxiliary regressions (and are actually somewhat simpler).

Since p_t^1 belongs to the exponential family of distributions and is closed under multiplication, we can proceed as outlined in Section 3.2 and define accordingly the class K_t as consisting of products of two IG kernels, one regrouping all IG factors in $\varphi_t=g_t^1p_t^1$ and the other designed to best approximate the remainder. Since the latter only depends on

λ_t, we shall select the following forms for the EIS kernel k_t and its integrating constant χ_t:

$$k_t(\lambda_{(t)}, a_t) \propto \left[r_{t-1}^{(1/2)v} \cdot \lambda_t^{-(1/2)(v+3)} \exp\left(-\frac{r_{t-1}}{\lambda_t}\right)\right]\left[\lambda_t^{-b_t} \exp\left(-\frac{c_t}{\lambda_t}\right)\right], \tag{55}$$

$$\chi_t(\lambda_{t-1}, a_t) \propto r_{t-1}^{(1/2)v} \cdot (r_{t-1} + c_t)^{-(1/2)(v+1)-b_t} \tag{56}$$

with $a_t' = (b_t \; c_t)$. In line with the discussion in Section 4.2, note the inclusion of r_{t-1}, which only depends on λ_{t-1}, in the expression for k_t. By construction, all terms included between the first two brackets in Eq. (55) belong to both φ_t and k_t. Therefore, they cancel out in the auxiliary EIS regressions which simplify into the following OLS regressions:

$$\text{Dependent variables: } \ln \chi_{t+1}(\lambda_t, \hat{a}_{t+1}) - \frac{1}{2}(w+1)\ln\left[1 + \frac{(y_t - \mu)^2}{\lambda_t(w-2)}\right],$$

$$\text{Regressors: } \ln \lambda_t \quad \text{and} \quad \lambda_t^{-1} \quad \text{(plus intercept)}. \tag{57}$$

Eqs. (55)–(57)are all we need to apply the sequential EIS algorithm described in Section 4. As discussed above, we rerun EIS under a fixed set of CRNs for each evaluation of the likelihood function. The computing time required for a very large number of inversion of the IG distribution function is considerably reduced by the initial construction of a high-accuracy bivariate interpolation table (U and degrees of freedom w) for the standardized IG inverse distribution function. For the purpose of illustration we computed ML-EIS estimates for a sample consisting of 1,447 observations of IBM daily stock price changes for the period 1/9/82–3/31/87. As indicated by the results reported in the SG row of Table 3, ML-EIS estimates are numerically extremely accurate even with as little as three EIS iterations and $S = 10$ MC draws. Table 3 also reports results for the SL, NG and NL cases.

Total computing time for a full ML optimization (using a simplex algorithm which is extremely robust for this type of applications) is of the order of 45 s on a 750 MH UNIX

Table 3
Stochastic volatility; EIS-ML estimates (IBM data)

	$\hat{q}$	$\hat{r}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\omega}$	$\hat{v}$	$\ln L(\hat{\theta}, y) \times 10^{-4}$
SG	0.0526	0.9698	–	0.0518	21.63	226.55	−0.10869
	(0.0003)	(0.0002)		(0.0005)	(0.18)	(0.62)	(0.00001)
	[0.0275]	[0.0184]		[0.0216]	[13.91]	[61.89]	
NG	0.0742	0.9568	–	0.0503	–	133.83	−0.10887
	(0.0008)	(0.0004)		(0.0004)		(1.85)	(0.00001)
	[0.0280]	[0.0177]		[0.0205]		[62.13]	
SL	0.0116	0.9763	0.0788	0.0535	19.82	–	−0.10864
	(0.0001)	(0.0002)	(0.0005)	(0.0002)	(0.09)		(0.00000$^+$)
	[0.0110]	[0.0177]	[0.0261]	[0.0239]	[11.13]		
NL	0.0197	0.9574	0.1204	0.0514	–	–	−0.10883
	(0.0002)	(0.0004)	(0.0006)	(0.0005)			(0.00000$^+$)
	[0.0136]	[0.0315]	[0.0311]	[0.0234]			

Notes: Number of MC draws: $S = 10$; three EIS iterations; numerical ( ) and statistical [ ] standard deviations based upon 20 MC replications.

server (and about 15 s for the faster NL version). The high persistence with values of r in the 0.95–0.98 range is typical of SV models. We note that the lognormal density for λ_t is qualitatively better identified than its IG counterpart. The high numerical accuracy of EIS in the context of SV models results from the facts that: (i) high persistence of the latent process implies that the observations $\{y_t\}$ are very informative on the underlying latent process $\{\lambda_t\}$ and EIS is designed precisely to take full advantage of such situations; (ii) a total of $2T = 2,894$ auxiliary parameters were used to construct the EIS sampler. Such high accuracy has been fully confirmed by recent applications of EIS to a wide range of SV models (univariate, bivariate, two factors, semi-parametric versions) in dimensions up to $8,000^{+}$. See Liesenfeld and Richard (2003a, b). Comparisons with alternative numerical evaluations of SV models are found in Liesenfeld and Richard (2003c) and Bauwens and Hautsch (2006). The combination of high numerical accuracy and ease of implementation of EIS appears to be unmatched in that class of models.

5.2. Logit with unobserved heterogeneity

It is commonly held in the econometric literature that ML estimation of panel data models with unobserved random heterogeneity along both dimensions is unfeasible, see e.g. the comments in McFadden (1989). This has led to the development of alternative though statistically less efficient simulation-based estimation techniques. See Lerman and Manski (1981), McFadden (1989), Pakes and Pollard (1989) or Börsch-Supan and Hajivassiliou (1990). See also Gourieroux and Monfort (1993) for a survey or Gourieroux and Monfort (1996) for an in-depth analysis of simulation-based inference techniques.

In this section, we demonstrate that highly accurate numerical evaluation of the likelihood function of such panel data models is fully operational under EIS. Consider a model consisting of a logit for the observable $y_{it} \in \{0,1\}$ conditionally on the random effects α_i and λ_t, independent normal distributions for the α's and a multivariate normal distribution for the λ's. Let $\alpha' = (\alpha_1 \cdots \alpha_N), \lambda' = (\lambda_1 \cdots \lambda_T), y' = (y_{it}; i: 1 \to N, t: 1 \to T)$ and $\delta = (y, \theta)$ where θ regroups all unknown parameters in the model. Let x_{it} denote a vector of exogenous variables. The likelihood function is of the form given by Eqs. (1) and (2) together with

$$g(\alpha, \lambda, \delta) = g_0(\lambda, \delta) \cdot \prod_{i=1}^{N} g_i(\alpha_i, \lambda, \delta) \tag{58}$$

with $g_0(\lambda, \delta) \equiv 1$ and

$$g_i(\alpha_i, \lambda_i, \delta) = \prod_{t=1}^{T} \frac{[\exp(v_{it})]^{y_{it}}}{1 + \exp(v_{it})}, \tag{59}$$

$$v_{it} = \beta' x_{it} + \alpha_i + \lambda_t, \tag{60}$$

$$p(\alpha, \lambda|\delta) = p_0(\lambda|\theta) \cdot \prod_{i=1}^{N} p_i(\alpha_i|\theta), \tag{61}$$

$$p_0(\lambda|\theta) \propto |H_\theta|^{1/2} \exp -\tfrac{1}{2}\lambda' H_\theta \lambda, \tag{62}$$

$$p_i(\alpha_i|\theta) \propto \sigma_\alpha^{-1} \exp\left[-\frac{1}{2}\left(\frac{\alpha_i}{\sigma_\alpha}\right)^2\right]. \tag{63}$$

It is implicitly assumed here that $N \gg T$ as commonly the case. For $T \gg N$ we would permute the α's and λ's in all factorizations. Extensions such as exchangeable α's are trivially handled by conditioning the densities p_i in Eq. (60) on a common factor α_0 and adding an additional density for α_0. Notationally this would amount to incorporating α_0 into λ.

It is obvious from Eqs. (58) to (63) that we should partition EIS samplers conformably with p in Eq. (61), with the critical extension that the α_i's are now to be dependent upon λ in order to fully account for the (posterior) dependence between α and λ as induced by g. Specifically, we select for the α_i's conditionally independent kernels of the form

$$\ln k_i(\alpha_i, \lambda, a_i) = -\frac{1}{2}\left[2b_i' v_i + v_i' C_i v_i + \left(\frac{\alpha_i}{\sigma_\alpha}\right)^2\right] \tag{64}$$

with

$$v_i = X_i\beta + \lambda + \alpha_i e,$$

$$X_i' = (x_i, \ldots, x_{iT}), \quad \lambda' = (\lambda_1 \cdots \lambda_T), \quad e' = (1 \cdots 1), \quad b_i \in \mathbf{R}^{\mathrm{T}}, \quad C_i = \mathrm{Diag}(c_i),$$

$c_i = C_i e \in \mathbf{R}_+^{\mathrm{T}}$ and $a_i = (b_i\ c_i)$. It turns out that the constraints $c_i > 0$ never bind and can safely be ignored. Note that since the α's remain independent from one another *conditionally* on λ the integrating constant χ_i associated with k_i only depends upon (λ, a_i), not upon the other α_j's. In order to facilitate subsequent EIS integration w.r.t. $\lambda, \ln\ \chi_i$ is rewritten as a quadratic form in λ and is given by

$$\ln \chi_i(\lambda, a_i) \propto -\frac{1}{2}\left[\ell_i' C_i \ell_i + 2\ell_i' b_i - \left(\frac{\overline{\alpha}_i}{\sigma_i}\right)^2\right] \tag{65}$$

with

$$\ell_i = \lambda + X_i\beta, \quad \sigma_i^{-2} = \sigma_\alpha^{-2} + c_i' e \quad \text{and} \quad \overline{\alpha}_i = -\sigma_i^2(\ell_i' c_i + b_i' e). \tag{66}$$

In summary, the EIS auxiliary GLS regressions for the α_i's consist of regressing $\ln g_i$ on the $2T$ regressions $\{(v_{it}, v_{it}^2); t : 1 \to T\}$ with an intercept and weights g_{it} (OLS for the first EIS iteration). Under EIS sampling and conditionally on λ, the α_i's are independently normally distributed with means $\overline{\alpha}_i$ and variances σ_i^2.

As for λ, we note that $\varphi_0 = g_0 p_0$ and $\{\chi_i; i : 1 \to N\}$ all are in the forms of Gaussian kernels. Therefore, a perfect EIS sampler for λ obtains immediately by combining together the $N + 1$ quadratic forms in Eqs. (62) and (65) providing a remarkable example of perfect fit in the sense of Eq. (9). Rearranging terms in the usual way, we find that

$$m_0(\lambda|a_0) \propto |A|^{1/2} \exp\left[-\tfrac{1}{2}(\lambda - \mu)' A (\lambda - \mu)\right] \tag{67}$$

with $a_0' = (\mu',\ \mathrm{vec}'\, A)$ and

$$A = H_0 + \sum_{i=1}^{N} (C_i - \sigma_i^2 c_i c_i'), \tag{68}$$

Table 4
Logit: (a) EIS estimate of $L(\theta_0; y)$; (b) ML-EIS estimates

Method	Estimate	MC st. dev.
(a)		
NAT	0.8035D − 74	0.4458D − 74
EIS	0.9947D + 54	0.0044D + 54
TSE_0	0.9890D + 54	0.0832D + 54
TSE_1	0.9659D + 54	0.0558D + 54

(b) $\hat{\sigma}_\alpha$	$\hat{\sigma}_\lambda$	$\hat{\rho}$	$L(\hat{\theta}; y)$
0.2704	0.2972	0.2683	0.1610D + 55
(0.0004)	(0.0003)	(0.0013)	(0.0004D + 55)
[0.0482]	[0.0226]	[0.1828]	

Note for (a): Proportionality constants were ignored. *Notes* for (b): Sample size: $N = 1000$ and $T = 15$; three EIS iterations; number of MC draws: $S = 100$.

$$\mu = -A^{-1} \sum_{i=1}^{N} [C_i X_i \beta + b_i - \sigma_i^2 (c_i' X_i \beta + b_i' e) c_i]. \tag{69}$$

This completes the description of the EIS algorithm for this application. Note the very large number of auxiliary parameters (of the order of $2TN$) used to produce the EIS approximation to the actual posterior density of (α, λ). A caveat applies before we present numerical results. As usual we shall aim at running the auxiliary EIS regressions under relatively small numbers of MC draws (of the order of three times the number of regressors). As the v_{it}'s all depend on α_i, occasional bad draws can generate high multicollinearity in the auxiliary regressions and even crash the EIS algorithm (production of the results reported in Table 4 required several millions of EIS regressions!). We were able to completely eliminate the problem by introducing a small amount of shrinkage in the auxiliary EIS regressions. Specifically, we used a second order Taylor Series Expansion (hereafter TSE) of g_i around $\alpha_i = \lambda_t = 0$ in order to produce exact restrictions for a_i. Let these restrictions be written as

$$a_i = R_i \delta_i + q_i, \tag{70}$$

where $\delta_i \in \mathbf{R}^\ell$ with $\ell < 2T$ denotes free regression coefficients and $(R_i \; q_i)$ is a $2T \cdot (\ell + 1)$ matrix of known constants (which actually depend upon $\{x_{it}, y_{it}); t : 1 \to T\}$ and β which are all included in δ). An unconstrained EIS OLS estimator of the form $\hat{a}_i = G_i^{-1} h_i$ is then replaced by the shrinkage estimator

$$\tilde{a}_i = (G_i + \kappa M_i)^{-1} (R_i + \kappa M_i q_i) \tag{71}$$

with $M_i = I_{2T} - R_i (R_i' R_i)^{-1} R_i'$.

This shrinkage option completely eliminates unwanted interruptions of the EIS algorithm at virtually no loss of EIS efficiency even with very low values of κ (0.01 or less in the application which follows). In order to illustrate the impressive numerical performance of EIS within this class of models, we generated a fictitious sample of size $T = 15$ and $N = 1,000$, with no exogeneous variables and a stationary AR(1) process for λ

with autocorrelation coefficient ρ and stationary variance σ_λ^2. The parameters true values were set equal to $\sigma_\alpha = \sigma_\lambda = 0.3$ and $\rho = 0.5$, implying a relatively moderate amount of heterogeneity (results derived under different values produce similar qualitative results). The number of regressors in each EIS auxiliary regression equals 30 (plus one intercept). We used $S = 100$ MC draws and three EIS iterations.

In Table 4(a) we report MC estimates of the likelihood function at the parameter true values under the natural sampler (NAT), as defined by Eqs. (61)–(63), the EIS sampler (EIS), the sampler obtained by TSE of the g_i, s around $\alpha_i = \lambda_i = 0$ (TSE_0) and an (unfeasible) sampler obtained by TSE of the g_i, s around the true values of α_i and λ_t, which we had initially stored (TSE_1). TSE_1 is closest in spirit to the Laplace approximations proposed by Tierney and Kadane (1986). Obviously, in practice the α's and λ's would have to be estimated first which would result in increased numerical inefficiency. The results in Table 4(a) illustrate the clear superiority of our EIS global approximations relative to TSE local ones. The results also indicates that the problem associated with the natural sampler in this context is one of the enormous downward biased more than variance. Actually, this result is not surprising. Assume an individual α_i draw from p_i has a probability 0.75 of hitting the region of importance. The probability that $N = 1,000$ independent draws jointly hit the region of importance is then of the order of 10^{-124}.

Finally, in order to illustrate the performance of EIS within an inferential context, we computed ML-EIS estimates of θ. Full optimization using a simplex algorithm requires of the order of 50 EIS likelihood evaluations for a computing time of the order of 1 min on our 750 MH UNIX workstation. As in Section 5.1, numerical as well as statistical standard deviations were produced using 20 replications of the EIS-ML optimization in both cases. The results are reported in Table 4(b). Note here again the impressive numerical accuracy of the results. The particular sample we used appears to have produced a borderline value for $\hat{\rho}$ (note, however, that we only have $T = 15$ periods, that is to say only 15 latent λ_t's to identify ρ). In order to verify that the low value of $\hat{\rho}$ was due to that particular sample and not to an inherent EIS-ML problem, we ran 50 MC replications of our algorithm under the parameter true values. The corresponding statistical means and standard deviations equal (0.2922, 0.2965, 0.4726) and (0.0627, 0.0287, 0.1850), respectively. The (EIS) ML estimators are clearly statistically well-behaved.

This EIS-ML algorithm has recently been successfully applied by Liesenfeld and Richard (2007a) to a dynamic logit model for the union participation decision of young men. The data ($N = 545, T = 8$) were taken from Vella and Verbeek (1998) who estimated the model under random individual effects and fixed time effects. Randomizing both heterogeneity components enable Liesenfeld and Richard to qualify their relative impact on agent's decisions. They find that the dynamic of the union participation decision is dominated by individual heterogeneity.

6. Conclusion

We proposed an operational recursive LS algorithm to construct (very) high-dimensional importance samplers. In contrast with current procedures, which are mostly based upon local approximations of the integrand, our algorithm explicitly minimizes the variance of the MC-IS estimate in order to produce a global approximation to the

posterior density of the variables to be integrated out. Our algorithm's performance in high-dimensional latent variables models is unparalleled, as illustrated in the context of two important classes of models in the modern econometric literature. Its success appears to result from a combination of three factors: (1) the availability of full sequential factorizations which reduce the optimization problem to an operational sequence of low-dimensional LS problems; (2) the use of very large number of auxiliary parameters, typically a multiple of the sample size in order to produce very good global fit between the integrand and the importance sampler; and (3), the interdependence between the latent variables which implies that the information provided by the observables spills across the latent process and produces fairly well-behaved though highly interdependent (implicit) posterior densities for the latent variables.

The generic simplicity of EIS allows for fully automated applications within broad classes of models which are easily customized to one's particular application as only the selection of a class of auxiliary samplers and the definition of the variables entering the EIS auxiliary regressions are truly model specific. It also follows that it is a trivial matter to rerun EIS-based statistical applications under different seeds for the CRNs, thereby providing an effective and reliable measure of numerical (MC) accuracy for all results of interest.

EIS is not meant to substitute for other methods under all circumstances. In particular, Monte Carlo Markov Chain (MCMC) algorithms appear to be well adapted to Bayesian applications when posterior densities of the parameters are ill-behaved and/or cannot be conveniently sequentially factorized as required for EIS. But neither do we believe that MCMC should be indiscriminately applied across the board. Actually, there is more in common between IS and MCMC than generally recognized in that both critically rely upon ratios between component densities and auxiliary samplers and in particular both are exposed to the risk that the variance of such ratios might not exist.

Perhaps the most exciting future applications of EIS are linked to the recognition that auxiliary EIS steps can easily be embedded within any MC techniques which requires the construction of auxiliary samplers, thereby improving their numerical efficiency as well as facilitating their implementation and/or (automated) calibration. Preliminary investigations by the first author have already produced very promising results and are the object of an active ongoing (joint) research project beyond the scope of the present paper.

Acknowledgments

We are particularly indebted to Roman Liesenfeld who, together with the first author, has developed several highly successful applications of EIS. This experience has led to major improvements of the present paper. We also acknowledge numerous constructive comments from David DeJong and Hariharan Dharmarajan. Roman, Dave and Hari are currently engaged together with the first author developing applications of EIS to MCMC and particle filters. Last but not least, constructive comments from the referees and the editor have produced major clarifications of several key issues. Our work has been supported by the National Science Foundation (Grants SES-9223365 and SES-0516642).

Appendix A. Technical details

The MC sampling variance of $\overline{G}_s(\delta)$ in Eq. (16) is rewritten as

$$\overline{V}_{S,m}(a;\delta) = \frac{1}{S}\, G(\delta)\cdot V(a;\delta), \tag{72}$$

where

$$V(a;\delta) = \int_X \left[\frac{\omega(x;\delta,a)}{G(\delta)} - 1\right]^2 \cdot \frac{G(\delta)}{\omega(x;\delta,a)}\cdot \varphi(x;\delta)\,\mathrm{d}x. \tag{73}$$

Let

$$\mathrm{d}(x,a,c;\delta) = \ln\left[\frac{\omega(x;\delta,a)}{G(\delta)}\right]. \tag{74}$$

Eq. (12) follows together with

$$c = \ln G(\delta) - \ln \chi(a). \tag{75}$$

Next, $V(a;\delta)$ is rewritten as

$$V(a;\delta) = \int_X h[d^2(x,a,c,\delta)]\cdot \varphi(x;\delta)\,\mathrm{d}x, \tag{76}$$

where

$$h(r) = e^{\sqrt{r}} + e^{-\sqrt{r}} - 2 = 2\sum_{i=1}^{\infty}\frac{r^i}{(2i)!}. \tag{77}$$

For the ease of notation the argument c is ignored in the rest of this Appendix. Note that h is monotone and convex on $\mathbf{R}_+$. Since $G(\delta)$ is unknown, we treat c as an unknown intercept to be estimated along with a. An optimal choice for a is given by the solution of the nonlinear GLS problem

$$a_*(\delta) = \operatorname{Arg}\operatorname*{Min}_{a\in A}\ V(a;\delta). \tag{78}$$

Since, however, an efficient sampler is one for which d is expected to be small (on average) we can usefully consider replacing $h(r)$ by its leading term r, which implies solving the simpler GLS problem

$$\hat{a}(\delta) = \operatorname{Arg}\operatorname*{Min}_{a\in A}\ Q(a;\delta), \tag{79}$$

where Q is defined in Eq. (11). The following lemma provides an upper bound for the relative loss of efficiency resulting from the replacement of $h(r)$ by r.

Lemma A.1. *If, under conditions such as those proposed by* Geweke (1996), $V(a;\delta)$ *is finite, then*

$$V(\hat{a}(\delta);\delta) > V(a_*(\delta);\delta) > h[Q(\hat{a}(\delta);\delta)]. \tag{80}$$

Proof. The proof follows from Jensen's inequality, whereby

$$V(a;\delta) > h[Q(a;\delta)] \text{ on } \mathbf{R}_+,$$

together with Eqs. (78) and (79). □

Eq. (80) enables us to compute an upper bound for the relative loss of efficiency associated with using $\hat{a}(\delta)$ in place of $a_*(\delta)$. In all EIS applications we have run, that upper bound has never exceeded a few percents which is why we only discuss the simpler optimization problem in equations in the present paper.

References

Bauwens, L., Hautsch, N., 2006. Stochastic conditional intensity processes. Journal of Financial Econometrics 4, 450–493.

Börsch-Supan, A., Hajivassiliou, V., 1990. Smooth unbiased multivariate probability simulation for maximum likelihood estimation of limited dependent variable models. Cowles Foundation Discussion Paper No. 960, Yale University.

Danielsson, J., Richard, J.F., 1993. Accelerated Gaussian importance sampler with applications to dynamic latent variable models. Journal of Applied Econometrics 8, 153–173.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer, New York.

Duffie, D., Singleton, K.J., 1993. Simulated moment estimation of Markov models of asset prices. Econometrica 61, 929–952.

Durbin, J., Koopman, S.J., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.

Evans, M., 1991. Adaptative importance sampling and chaining. Contemporary Mathematics (Statistical Multiple Integration) 115, 137–142.

Fishman, G.S., 1996. Monte Carlo Concepts, Algorithms, and Applications. Springer, New York.

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

Geweke, J., 1996. Monte Carlo simulation and numerical integration. In: The Handbook of Computational Economics, vol. 1. North-Holland, Amsterdam.

Gourieroux, C., Monfort, A., 1993. Simulation based inference: a survey with special reference to panel data models. Journal of Econometrics 59, 5–33.

Gourieroux, C., Monfort, A., 1996. Simulation based econometric methods. CORE Lecture Series. Oxford University Press, Oxford.

Hendry, D.F., 1984. Monte Carlo experimentation in econometrics. In: The Handbook of Econometrics, North-Holland, New York.

Lee, K.M., Koopman, S.J., 2004. Estimating stochastic volatility models: a comparison of two importance samplers. Studies in Nonlinear Dynamics and Econometrics 8.

Lehmann, E.L., 1986. Testing Statistical Hypotheses. Wiley, New York.

Lerman, S., Manski, C., 1981. On the use of simulated frequencies to approximate choice probability. In: Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge (Chapter 7).

Liesenfeld, R., Richard, J.F., 2003a. Monte Carlo methods and Bayesian computation: importance sampling. In: The International Encyclopedia of the Social and Behavioral Sciences. Elsevier-Science, Oxford, pp. 10000–10004.

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

Liesenfeld, R., Richard, J.F., 2003c. Estimation of dynamic bivariate mixture models: comments on watanabe. Journal of Business and Economic Statistics 21, 570–576.

Liesenfeld, R., Richard, J.F., 2007a. Simulation techniques for panels: efficient importance sampling. Forthcoming In: Econometrics of Panel Data, A Handbook of the Theory, third ed. Kluwer, Boston.

Liesenfeld, R., Richard, J.F., 2007b. Improving MCMC using efficient importance sampling. University of Pittsburgh, Mimeo.

Madras, N., Piccioni, M., 1999. Importance sampling for families of distributions. Annals of Applied Probability 9, 1202–1225.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

Melino, A., Turnbull, S., 1990. Pricing foreign currency options with stochastic volatility. Journal of Econometrics 45, 239–265.

Owen, A., Zhou, Y., 2000. Safe and effective importance sampling. Journal of the American Statistical Association 95, 135–143.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association 94, 590–599.

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

Taylor, S., 1986. Modeling Financial Time Series. Wiley, Chichester.

Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81, 82–86.

Vella, F., Verbeek, M., 1998. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. Journal of Applied Econometrics 13, 163–283.