Testing the assumptions behind the use of importance sampling

Siem Jan Koopman
Free University Amsterdam, Department of Econometrics,
NL-1081 HV Amsterdam, The Netherlands
s.j.koopman@feweb.vu.nl

Neil Shephard
Nuffield College, Oxford OX1 1NF, UK
neil.shephard@nuf.ox.ac.uk

May 2003

Abstract

Importance sampling is used in many aspects of modern statistics and econometrics to approximate unsolvable integrals. Its reliable use requires the sampler to possess a variance, for this guarantees a square root speed of convergence and asymptotic normality of the estimator of the integral. However, this assumption is seldom checked. In this paper we propose to use extreme value theory to empirically assess the appropriateness of this assumption. We illustrate this method in the context of a maximum simulated likelihood analysis of the stochastic volatility model.

Keywords: Extreme value theory; Importance sampling; Simulation; Stochastic Volatility.

1 Introduction

One of the most important recent developments in statistics and econometrics has been the use of simulation methods to estimate models (see, for example, the reviews in Ripley (1987), Hajivassiliou and Ruud (1994), Gourieroux and Monfort (1996), Geweke (1997) and Liu (2001)). A basic tool in much of this literature is importance sampling, which approximates the solution to integrals via averages of simulations. In order to assess the accuracy of the importance sampler the Lindeberg-Lévy central limit theory is used. However, this assumes the existence of the variance of the importance sampler. Checking the validity of this assumption is often difficult. In this note we provide a simple empirical check on this assumption based on extreme value theory.

Importance sampling was discussed as early as Kahn and Marshall (1953) and Marshall (1956), while it was popularised in the influential monograph by Hammersley and Handscomb (1964, Section 5.4). It was first used in econometrics by Kloek and Van Dijk (1978) in their work on computing posterior densities. Further significant developments on this topic were reported by Geweke (1989). In particular we wish to evaluate the integral

$$c = \int_{\Omega} f(x)dx, \tag{1}$$

where we know how to calculate $f(x)$, but cannot solve the integral analytically. To deal with this we introduce an importance sampling density $g(x)$ which is easy to evaluate and simulate from and whose support is also Ω. We then approximate c by

$$\widehat{c} = \frac{1}{R}\sum_{j=1}^{R} w_j, \qquad \text{where} \qquad w_j = w(x^j), \tag{2}$$

where

$$w(x) = \frac{f(x)}{g(x)} \quad \text{and} \quad x^j \overset{i.i.d.}{\sim} g(x),$$

with $g(x)$ assumed to be strictly positive for all $x \in \Omega$. By construction we know that $\{w_j\}$ are i.i.d. and that $E(w) = c$. As a result, a simple application of Kolmogorov's strong law of large numbers (e.g. Geweke (1989, p. 1320) and Gallant (1997, p. 132)) shows that

$$\widehat{c} \overset{a.s.}{\to} c, \qquad \text{as } R \to \infty,$$

whatever importance sampler we design. However, in order to easily measure the precision of $\widehat{c}$ and to guarantee that the rate of convergence to c is $R^{1/2}$, it is helpful to have a Gaussian central limit theorem for $\widehat{c}$. We know from the Lindeberg-Lévy central limit theorem that a necessary and sufficient condition for this is that $Var(w)$ exists, which would allow us to conclude that

$$\sqrt{R}\,(\widehat{c} - c) \overset{d}{\to} N(0, Var(w)).$$

However, the existence of this quantity is by no means guaranteed.

In a fundamental contribution, Geweke (1989) argues that we should only use importance sampling in cases where we can prove that $Var(w)$ exists. However, in practice this is actually quite difficult to check in large dimensional problems and so many econometrics and statistics papers have recently been written which, in effect, *a priori* assume that this condition holds. See, for example, Hendry and Richard (1991), Danielsson and Richard (1993), Danielsson (1994), Sandmann and Koopman (1998), Durbin and Koopman (2000), Stephens and Donnelly (2000), Elerian, Chib, and Shephard (2001) and Durham and Gallant (2002).

In this paper we present a simple diagnostic check for the existence of $Var(w)$. This will be based on the application of extreme value theory[1]. We will apply it to the analysis of financial econometric models and show that sometimes this variance does not exist.

[1]We should note that extreme value theory has been applied in the context of financial economics and insurance in order to determine the thickness of the tail of financial returns. This can have potential application in computing various measures of risk. References to this literature include Embrechts, Kluppelberg, and Mikosch (1997, Section 6.5.2) and Danielsson and de Vries (1997). The application of the theory is much harder in that context for the *i.i.d.* assumption certainly does not hold. The inference problem we face here is much simpler.

In Section 2 we review the extreme value theory that is appropriate for checking the assumption that $Var(w)$ exists, while we go on to discuss the statistical estimation of tail indexes for extremes, while Section 3 reports the results from some experiments we have conducted based on this approach. Section 4 illustrates the approach we are advocating with an application from financial econometrics. Section 5 concludes.

2 Extreme value theory and inference

2.1 The model and hypothesis

By construction we know from (2) that the weights $\{w_j\}$ are *i.i.d.* and that $E(w) = c$. The key issue is whether the variance exists, which really means we need to know the behaviour of the weights in their right hand tail. To study the tail behaviour we use extreme value theory.

Smith (1987) argues that if we have an *i.i.d.* $\{w_j\}$ population then as a threshold value $u > 0$ increases, the limit distributions of the random variables over this threshold will be generalised Pareto. In particular, define these (unordered) large weights minus the threshold u as $Z_1, ..., Z_n$, then the asymptotic density of these excesses is

$$f(z) = \frac{1}{\beta}\left(1 + \xi\frac{z}{\beta}\right)^{-\frac{1}{\xi}-1}, \qquad z \in D(\xi, \beta) > u, \quad \beta > 0. \tag{3}$$

Here

$$D(\xi, \beta) = \begin{cases} [0, \infty), & \xi \geq 0, \\ [0, -\beta/\xi], & \xi < 0. \end{cases}$$

Importantly for this model only $1/\xi$ moments exist. This implies that we can determine the number of moments that the weights have by focusing on $1/\xi$. Thus we must be interested in knowing if $\xi \leq 1/2$, while we know by construction of the importance sampler that $\xi \leq 1$. The cases of $\xi < 0$ deals with situations where the $\{w_j\}$ have some upper bound. This is of some relevance in importance sampling in the case of the sampler being bounded, that is

$$f(x) \leq kg(x), \tag{4}$$

for some finite choice of $k > 0$. Then if (4) is true for $x \in \Omega$, the support of the $\{w_j\}$ will be bounded and the existence of all moments is guaranteed. This is, in practice, quite unusual. Most of the cases of real interest are where the sampler is not bounding, so $\xi \in [0, 1]$.

Formally the hypotheses we will be interested in deciding between is

$$H_0 : \xi \leq \frac{1}{2}, \qquad \text{and} \qquad H_1 : \xi > \frac{1}{2}.$$

The null will imply the existence of the variance, the alternative will deny this[2]. In practice it is helpful to make the null a point hypothesis (see, for example, Cox and Hinkley (1974, pp. 331–334)), so making the comparison between

$$H_0 : \xi = \frac{1}{2}, \qquad \text{and} \qquad H_1 : \xi > \frac{1}{2}. \tag{5}$$

As we have a parametric model for the weights over a threshold we will use a likelihood function to carry out the testing[3]. This approach to inference is often called the peak over threshold method (see, for example, Embrechts, Kluppelberg, and Mikosch (1997)). We will study the behaviour of the score and likelihood ratio tests of the hypothesis. Their behaviour is regular due to the standard asymptotics of the maximum likelihood estimators of parameters ξ and β.

2.2 Estimation and Wald test

The log-likelihood for a sample $z_1, ..., z_n$ over the threshold u equals

$$\log f(z; \lambda) = -n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log x_i, \tag{6}$$

where $x_i = 1 + \xi\beta^{-1} z_i$. Maximum likelihood estimation of parameter vector $\lambda = (\xi, \beta)'$ is discussed in Smith (1987) and can be based on the standard method of Fisher scoring which relies on the score vector

$$s = \begin{pmatrix} s_\xi \\ \\ s_\beta \end{pmatrix} = \frac{\partial \log f(z;\lambda)}{\partial \lambda} = \begin{pmatrix} \xi^{-2} \sum_{i=1}^{n} \log x_i - (1 + \xi^{-1})\beta^{-1} \sum_{i=1}^{n} z_i / x_i \\ \\ -n\beta^{-1} + (1+\xi)\beta^{-2} \sum_{i=1}^{n} z_i / x_i \end{pmatrix}, \tag{7}$$

and the expected information matrix $n\mathcal{I}$ where

$$\mathcal{I} = \frac{1}{(1+2\xi)(1+\xi)} \begin{pmatrix} 1+\xi & -1 \\ -1 & 2 \end{pmatrix}. \tag{8}$$

A sensible starting value for ξ is 0.5, the one specified by H_0, while for β it can be constructed using the expected value of z given by

$$E(z) = \frac{\beta}{\xi^2} \frac{\Gamma(\xi^{-1} - 1)}{\Gamma(1 + \xi^{-1})} = 2\beta, \qquad \text{for } \xi = 0.5.$$

So iterations can start at $\lambda = 0.5(1, \bar{z})'$ where $\bar{z} = n^{-1} \sum_{i=1}^{n} z_i$.

The asymptotic distribution of the maximum likelihood estimator $\hat{\lambda}$ is given by

$$\sqrt{n}\left(\widehat{\lambda} - \lambda\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}\right) \qquad \text{where} \qquad \mathcal{I}^{-1} = (1+\xi) \begin{pmatrix} 2 & 1 \\ 1 & 1+\xi \end{pmatrix}. \tag{9}$$

[2] It is also possible to setup the hypothesis in the opposite manner, with the null that $\xi > 1/2$ and the alternative that $\xi \leq 1/2$. In our treatment this will not lead to different diagnostic test statistics, although clearly the critical values will change.

[3] An alternative to this approach is to use the Hill estimator, which has the advantage that it has a closed form. However, it is well known that this estimator behaves in an unreliable manner in some important situations.

Smith (1987) has shown that likelihood inference is regular for this problem as long as $\xi > -1/2$. This covers the null hypothesis value $H_0 : \xi = \frac{1}{2}$ and all the values under the alternative.

Once the maximum likelihood estimator of λ is found the Wald test can be computed to test the null hypothesis (5). For our purposes it is appropriate to compute an asymptotic signed t-test, that is

$$t = \sqrt{\frac{n}{3}}\left(\hat{\xi} - \frac{1}{2}\right). \tag{10}$$

The null hypothesis is rejected when the t-test takes a large positive value compared to a standard normal.

2.3 Estimation under null hypothesis and score test

The maximum likelihood estimator of β under the null hypothesis of $\xi = 0.5$ can be found by univariate Fisher scoring. The score value of β is given by

$$\begin{aligned} s_\beta^0 &= -\frac{n}{\beta} + \frac{3}{2}\beta^{-2}\sum_{i=1}^{n} z_i/x_i, \\ &= n\beta^{-1}\left(3\bar{z}^* - 1\right), \end{aligned}$$

where

$$\sum_{i=1}^{n} z_i/x_i = 2\beta n\bar{z}^*, \qquad \bar{z}^* = \frac{1}{n}\sum_{i=1}^{n}\frac{z_i}{2\beta + z_i}, \qquad \text{for } \xi = 0.5.$$

The expected information in the sample is $2n/3$, under the null hypothesis. As in the previous section, we can take $0.5\bar{z}$ as the initial value for β. After convergence the restricted estimate of β is obtained which we will denote by $\hat{\beta}^0$.

The one-sided score statistic will be used for testing the null hypothesis as this will be computationally simple. It is based on the score value of ξ under the null hypothesis and is given by

$$\begin{aligned} s_\xi^0 &= 4\sum_{i=1}^{n}\log x_i - 3\beta^{-1}\sum_{i=1}^{n} z_i/x_i \\ &= 4\sum_{i=1}^{n}\log\left(1 + \frac{z_i}{2\beta}\right) - 6n\bar{z}^*, \qquad \text{for } \xi = 0.5. \end{aligned}$$

This score value for the null hypothesis (5) is a function of β and it can be evaluated when β is replaced by its (restricted) maximum likelihood estimator $\hat{\beta}^0$. We know from the results of Smith (1987) that

$$s_\xi^* = \frac{1}{\sqrt{2n}} s_\xi^0 \xrightarrow{d} N(0,1), \tag{11}$$

which gives a very simple test. We reject the $H_0 : \xi = \frac{1}{2}$ when $\hat{s}_\xi^*$ is significantly positive and where $\hat{s}_\xi^*$ is s_ξ^* with β replaced by $\hat{\beta}^0$.

2.4 Likelihood ratio test

When estimation of λ has taken place under the constraint that $\xi \geq 0.5$, to deliver $\widetilde{\lambda}$, together with estimation under the null hypothesis, the likelihood ratio statistic can also be used to test the null:

$$LR = 2\left\{\log f(z;\widetilde{\lambda}) - \log f(z;\hat{\beta}^0, \xi = 0.5)\right\}.$$

Of course, the null is on the boundary next to the alternative and so the

$$LR \xrightarrow{d} 0.5\left(\chi_0^2 + \chi_1^2\right), \quad \text{under } H_0.$$

That is there is a 0.5 probability that the likelihood ratio statistic will be zero (as $\widehat{\xi}$ will be negative) and the rest of the time the statistic will have a χ_1^2 distribution (e.g. Chernoff (1954) and Gourieroux, Holly, and Monfort (1981)).

3 Some experiments

To check the effectiveness of extreme value theory in practice for testing the assumptions behind importance sampling we consider two experiments. The first is based on sampling from a normal density to approximate a normal density with a different variance. Another experiment is derived from the binomial example considered by Geweke (1989, §5.1).

3.1 Normal density experiment

We take $f(x)$ to be the normal density $N(0,1)$ and the importance sampling density $g(x)$ is $N\{0,(1+\epsilon)^{-1}\}$ where ϵ is positive number. Thus for this problem $c = 1$ in (1). The question is whether the variance of the weight function exists? The weight function is given by

$$w(x) = \frac{f(x)}{g(x)} = \frac{1}{\sqrt{1+\epsilon}}\exp\left(\frac{\epsilon}{2}x^2\right),$$

which obviously has $Ew(x) = 1$, while

$$\begin{aligned} Ew(x)^2 &= \frac{1}{\sqrt{2\pi}}\frac{\sqrt{1+\epsilon}}{1+\epsilon}\int_{-\infty}^{\infty}\exp\left(\epsilon x^2\right)\exp\left(-\frac{1}{2}x^2(1+\epsilon)\right)dx \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1+\epsilon}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}x^2(1-\epsilon)\right)dx, \end{aligned}$$

which is bounded only if $\epsilon < 1$. Hence the variance of the weight function of the importance sampler will not exists if $\epsilon \geq 1$, which implies the central limit theory for importance samplers will not hold in that range of cases.

In Table 1 we present the diagnostic tests and they show that for small values of ϵ importance sampling is valid while for larger values higher moments do not exist for the importance weights.

	$R=1,000$				$R=10,000$			
	score	t	LR	$\sqrt{\widehat{Var(w)}}$	score	t	LR	$\sqrt{\widehat{Var(w)}}$
10% of w								
$\epsilon=0.1$	-0.6	-1.6	0	0.110	-2.6	-6.4	0	0.113
$\epsilon=0.5$	0.3	0.5	0.3	1.81	0.1	0.1	0	2.36
$\epsilon=0.8$	1.2	2.1	5.0	8.73	2.9	4.9	30.1	18.4
$\epsilon=1.0$	1.9	2.9	11.0	24.6	5.2	8.1	80.7	80.5
$\epsilon=1.2$	2.7	2.9	18.1	70.1	7.8	9.1	150	369
5% of w								
$\epsilon=0.1$	-0.8	-1.9	0	0.115	-2.2	-5.4	0	0.120
$\epsilon=0.5$	-0.2	-0.4	0	2.22	-0.4	-0.8	0	3.05
$\epsilon=0.8$	0.4	0.6	0.6	11.4	1.5	2.7	9.1	25.5
$\epsilon=1.0$	0.9	1.3	2.5	33.0	3.1	5.0	30.8	113
$\epsilon=1.2$	1.5	2.0	5.6	95.6	4.9	6.5	64.0	520
1% of w								
$\epsilon=0.1$	-0.5	-0.9	0	0.114	-1.0	-2.4	0	0.132
$\epsilon=0.5$	-0.3	-0.9	0	2.99	-0.3	-0.5	0	5.31
$\epsilon=0.8$	-0.1	-0.5	0	17.9	1.2	0.9	1.2	52.4
$\epsilon=1.0$	-0.0	-0.1	0	56.0	1.3	1.9	4.7	242
$\epsilon=1.2$	-0.1	0.3	0.1	171	2.0	2.8	10.4	1137

Table 1: *Likelihood based tests for the null of the existence of the variance of the importance sampler for normal densities. Based on 1,000 and 10,000 samples using 10%, 5% and 1% of these in the extreme value analysis. Score and t-tests are both standard normal, with the test being rejected for large positive values. LR test has a* $0.5\left(\chi_0^2+\chi_1^2\right)$ *null. It is rejected for large values. The 95% critical values are 1.65, 1.65 and 2.69, respectively.*

The Table shows that for the test statistics to be effective, we require a simulation sample size R that is sufficiently large (say, $10,000$) and from which a small fraction (say, 1%) is used in the extreme value testing.

3.2 Geweke's binomial experiment

Here we reanalyse a simple example constructed by Geweke (1989) which illustrates the possibility that importance samplers behave poorly. Suppose we have a sample of size n of Bernoulli random variables $y=(y_1,...,y_n)'$ with $\Pr(y_i=1)=\theta$. Assume a uniform prior for θ and write $y_\bullet=\sum_{i=1}^n y_i$, then the posterior for θ is

$$f(\theta|y)=\frac{\Gamma\left(n+2\right)}{\Gamma\left(y_\bullet+1\right)\Gamma\left(n-y_\bullet+1\right)}\theta^{y_\bullet}\left(1-\theta\right)^{n-y_\bullet},\quad \theta\in[0,1],$$

which implies the marginal likelihood is

$$f(y)=\int f(y|\theta)f(\theta)d\theta=\frac{f(y|\theta)f(\theta)}{f(\theta|y)}=\frac{1}{n+1}.$$

The maximum likelihood estimator of θ is $\widehat{\theta} = y_\bullet/n$ while $\sqrt{n}\left(\widehat{\theta} - \theta\right) \overset{d}{\to} N(0, \theta(1-\theta))$ as $n \to \infty$. A direct application of this theory suggests the important sampler:

$$g(\theta|y) = \frac{1}{\sqrt{2\pi\widehat{\theta}(1-\widehat{\theta})/n}} \exp\left[-\frac{n}{2}\left(\widehat{\theta} - \theta\right)^2 / \left\{\widehat{\theta}(1-\widehat{\theta})\right\}\right].$$

The corresponding weight function is

$$w(\theta) = \frac{f(\theta|y)}{g(\theta|y)},$$

as we are going to sample $\theta^j \overset{i.i.d.}{\sim} g(\theta|y)$.

In Table 2 we report the diagnostics for case A with $\theta = 0.5$ and for case B with $\theta = 0.7$. The test statistics confirm the findings of Geweke and show that case B with $\theta = 0.7$ lead to an importance density that generates weights for which a variance exists. However, for case A with $\theta = 0.5$ the test statistics give evidence that the variance does not exist.

				$R = 1,000$				$R = 10,000$			
	n	$y_\bullet$	θ	score	t	LR	$\sqrt{\widehat{Var}(w)}$ $\times 10^{-5}$	score	t	LR	$\sqrt{\widehat{Var}(w)}$ $\times 10^{-5}$
10% of w											
Case A	69	6	.5	6.9	2.9	49.6	2.6	22.5	9.1	545	3.4
Case B	71	54	.7	-2.1	-2.9	0	1160	-7.1	-9.1	0	1168
5% of w											
Case A	69	6	.5	5.5	2.0	31.5	3.6	14.7	6.5	259	4.7
Case B	71	54	.7	-1.6	-2.0	0	317	-5.0	-6.5	0	290
1% of w											
Case A	69	6	.5	1.7	0.9	3.87	7.4	3.0	2.9	21.0	10.1
Case B	71	54	.7	-0.8	-0.9	0	5.4	-2.4	-2.9	0	11.8

Table 2: *Likelihood based tests for the null of the existence of the variance of the importance sampler for Geweke's binomial problem. Score and t-tests are both standard normal, with the test being rejected for large positive values. Note the standard deviation of w is given $\times 10^{-5}$. LR test has a $0.5\left(\chi_0^2 + \chi_1^2\right)$ null. It is rejected for large values. The 95% critical values are 1.65, 1.65 and 2.69, respectively.*

3.3 Graphical diagnostics

Our proposed test statistics do not necessarily give the complete picture and do not provide strict guarantees for a successful importance sampling procedure. Graphical diagnostics play a complementary role in this respect. We will illustrate this using the two experiments of this Section.

A graph of the largest values of the weights gives an indication of the seriousness of the outliers that indicates the non-existence of a variance. The non-largest weights can be represented via a histogram. This is an effective way to graphically present all weights. Further the recursive plot of the standard deviation of weights can show the impact of large weights on the variance estimate. As the sample increases the variance should converge to a constant if it exists. The three graphs for the normal density experiment with $\epsilon = 0.1$ are presented in Figure 1 together with the graphs for cases A and B of Geweke's binomial experiment. The diagnostic plots are based on an importance simulation sample of $R = 100,000$. The first row-panel of Figure 1

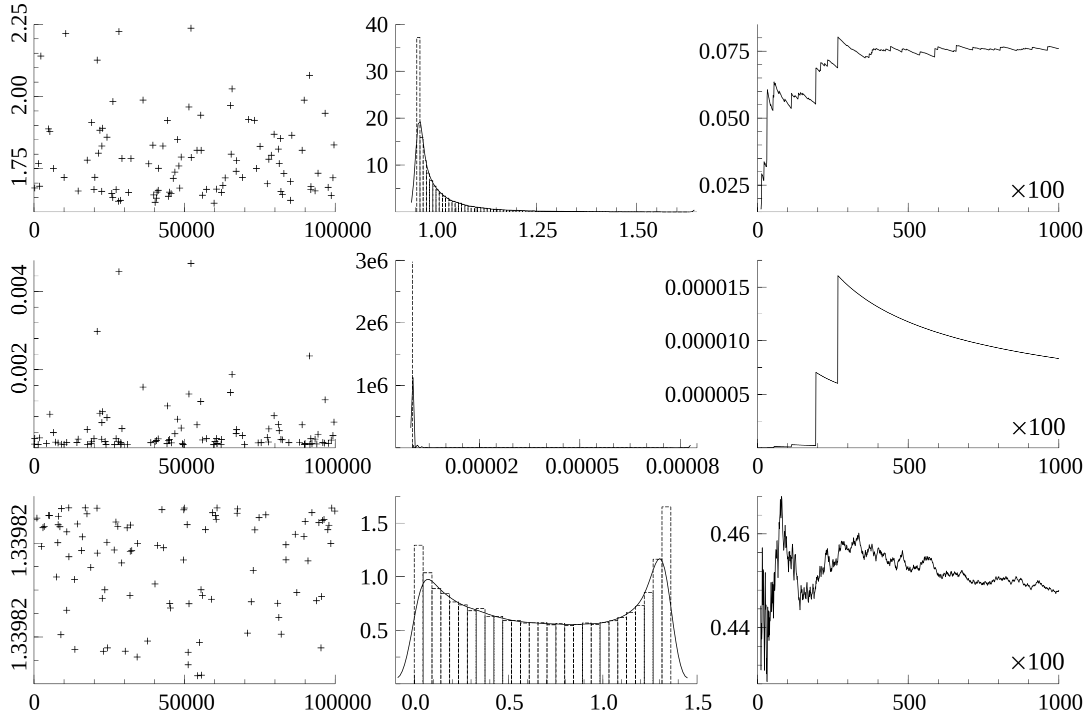

Figure 1: *Top row corresponds to importance sample weights for the Gaussian problem with* $\epsilon = 0.1$*. Middle row has corresponding results for Case A from the Geweke experiment, while the bottom row has the results for Case B. The left hand side pictures are of the largest 100 weights, the middle pictures are a histogram of all the other weights and the right hand side shows a recursive estimator of the variance of the weights.*

provides further evidence that the importance sampler behaves well for the normal experiment with $\epsilon = 0.1$. The second row-panel present plots that are consistent with the test statistics for case A of the binomial experiment. Some weights can be regarded as outliers and they have a huge effect on the computation of the sample variance indicating that the importance sampling procedure does not behave well. The bottom row-panel also confirm the conclusion for case B that a variance for the weight function exists. However, the weight function is badly behaved as we can see from the bi-modal histogram and from the recursive plot of standard deviation that is not converging smoothly. This last experiment makes the case that graphical procedures should also be considered for the task of diagnosing importance sampling.

4 Illustration: stochastic volatility

We investigate the effectiveness of our proposed diagnostic procedures using a class of stochastic volatility (SV) models (see Taylor (1994), Shephard (1996) and Ghysels, Harvey, and Renault (1996) for reviews of the associated literature). Various approaches exist for applying importance sampling to SV models such as the ones of Danielsson (1994), Sandmann and Koopman (1998) and Durham and Gallant (2002). Here we take the approach of Shephard and Pitt (1997) and Durbin and Koopman (1997).

4.1 Importance sampling for stochastic volatility models

A flexible discrete time stochastic volatility model for an univariate time series of returns y_t is given by

$$\begin{aligned} y_t &= \mu + \sigma_\varepsilon \exp(\alpha_t/2)(\rho\eta_t + \sqrt{1-\rho^2}\varepsilon_t), \\ \alpha_{t+1} &= \phi\alpha_t + \sigma_\eta \eta_t, \qquad \alpha_1 \sim N\{0, \sigma_\eta^2/(1-\phi^2)\}, \end{aligned} \tag{12}$$

for $t = 1, \ldots, T$, where the disturbances ε_t and η_t are serially and mutually independent series of disturbances with zero means and unit variances, independent of α_1. The disturbances η_t are assumed normally distributed for $t = 1, \ldots, T$. The mean of the returns is given by μ. The inclusion of the log-volatility disturbance η_t in the measurement equation introduces correlation between the two equations of model (12) allowing for a leverage effect in return series (e.g. Black (1976) and Nelson (1991)). This particular formulation ensures that $|\rho| < 1$. The degree of volatility persistence is measured by the autoregressive parameter ϕ and the parameter σ_η is the standard deviation of the disturbances of the autoregressive process for α_t. The initial value of α_1 is a draw from the unconditional distribution of α_t.

The model given in (12) can be thought of as Euler discretisation of a continuous time SV model where the spot volatility follows a log-normal Ornstein-Uhlenbeck process (e.g. Hull and White (1987)).

The likelihood function $f(y)$ of the SV model is not available analytically. Instead it is in the form of the unsolved integral

$$f(y) = \int f(y|\alpha)f(\alpha)d\alpha,$$

where y is playing an inactive role in the integral. Thus it is of the form of the usual importance sampling problem (1). It is computed as indicated below (1) with the importance sampler $g(\alpha)$

being the Laplace approximation to the posterior of $\alpha_1, \ldots, \alpha_T$ given the data $y_1, \ldots, y_T$; see, for example, Gelman, Carlin, Stern, and Rubin (1995, p.306). Then the weight function will be

$$w(\alpha) = \frac{f(y|\alpha)f(\alpha)}{g(\alpha)}.$$

In the Appendix we give the details of how such an approximation for SV type models is obtained quickly. We will now proceed to use extreme value theory to check the validity of a central limit theory based on this importance sampler.

4.2 Data and design of empirical study

In our empirical study we use the return series y_t of daily Standard & Poor's 100 stock index closures. The historical return series is for the period 2nd January 1990 to 31st December 1999 and was obtained from Datastream. After adjusting the series for holidays, our sample consists of $2,516$ daily observations. Hence the importance sampling is being carried out over $2,516$ dimensions in this case. The continuously compounded (raw) returns on the stock index are not adjusted for dividends and they are expressed in percentage terms and therefore given by $y_t = 100(\ln P_t - \ln P_{t-1})$ where P_t denotes the closing price of the Standard & Poor's 100 index at day t.

The stochastic volatility model requires the estimation of parameter vector

$$\psi = (\mu, \phi, \sigma_\eta, \sigma_\varepsilon, \rho)'.$$

Maximum likelihood estimation is based on numerically maximising the log of the estimated likelihood function, where the estimation is based on importance sampling. To start the iterative process of the search for the maximum likelihood estimator, initial values for the parameters are needed and we have chosen them realistically as

$$\mu = \rho = 0, \qquad \phi = 0.98, \qquad \sigma_\eta = 0.2, \qquad \sigma_\varepsilon = 0.6,$$

The initial parameter vector is referred to as ψ^0 and the maximum likelihood estimator of ψ is referred to as $\hat{\psi}$. Within the process of maximising the simulated log-likelihood value, we compute the log-likelihood via importance sampling using $R = 1,000$ samples where the same underlying uniform random variables are used for each likelihood evaluation. The approximate maximum likelihood estimates and the maximum log-likelihood values are given in Table 3.

4.3 Empirical results: graphical diagnostics

To illustrate our diagnostic procedure, we boost the simulation size to $M = 100,000$ and produce a set of diagnostic graphs. The question we ask is if the variance of the weights exists for $\psi = \widehat{\psi}$, basing the diagnostic graphics on these i.i.d. $100,000$ samples.

	Stochastic volatility model without leverage	with leverage
μ	0.0713	0.0515
	(0.0427 0.0999)	(0.0461 0.0960)
ϕ	0.985	0.980
	(0.971 0.992)	(0.963 0.989)
σ_η	0.134	0.160
	(0.103 0.174)	(0.120 0.211)
σ_ε	0.798	0.799
	(0.667 1.047)	(0.681 0.936)
ρ		-0.485
		(-0.612 -0.318)
sim. log lik.	-3090.71	-3076.30

Table 3: *Maximum simulated likelihood estimates of stochastic volatility parameters. The values in parentheses give the asymptotic (asymmetric)* 95% *confidence intervals.*

We start by considering the stochastic volatility model without leverage whose estimated parameters are reported in Table 3 together with their 95% confidence intervals. By taking the maximum simulated likelihood estimate $\hat{\psi}$ as fixed, we compute the M importance weights $\{w_j\}$. The largest 100 weights are presented in Figure 2 together with a histogram of the remaining $99,900$ smaller weights. The third graph presents the recursive estimation of the standard deviation of the weights $w_1, \ldots, w_j$ for $j = R+1, \ldots, M$ with $R = 1000$ (the number of weights used for maximum likelihood estimation). The final graph presents estimates of ξ, together with an asymmetric 95% confidence interval, based on thresholds u which excludes the smallest $98 + 0.05j$ percent of the weights for $j = 1, \ldots, 30$.

These diagnostic graphs provide evidence that the importance sampler is unreliable in this case, for they indicate a maximum likelihood estimate of ξ which is larger than 0.5. Thus the variance of the weights is either not existing or close to not existing. This result is in line with the more informal assessment given in the jumpy plot of Figure 2(iii). This records the recursive variance estimator. Taken together, these results point towards the conclusion that the use of standard asymptotics to measure the uncertainty of the importance sampler's estimate of the log-likelihood function is problematic in this case.

We now turn our attention to the empirically more interesting stochastic volatility model with leverage and look at the diagnostic graphs of the model with estimated parameters. The estimated parameter values are shown in Table 3 and show a large improvement in the fit of

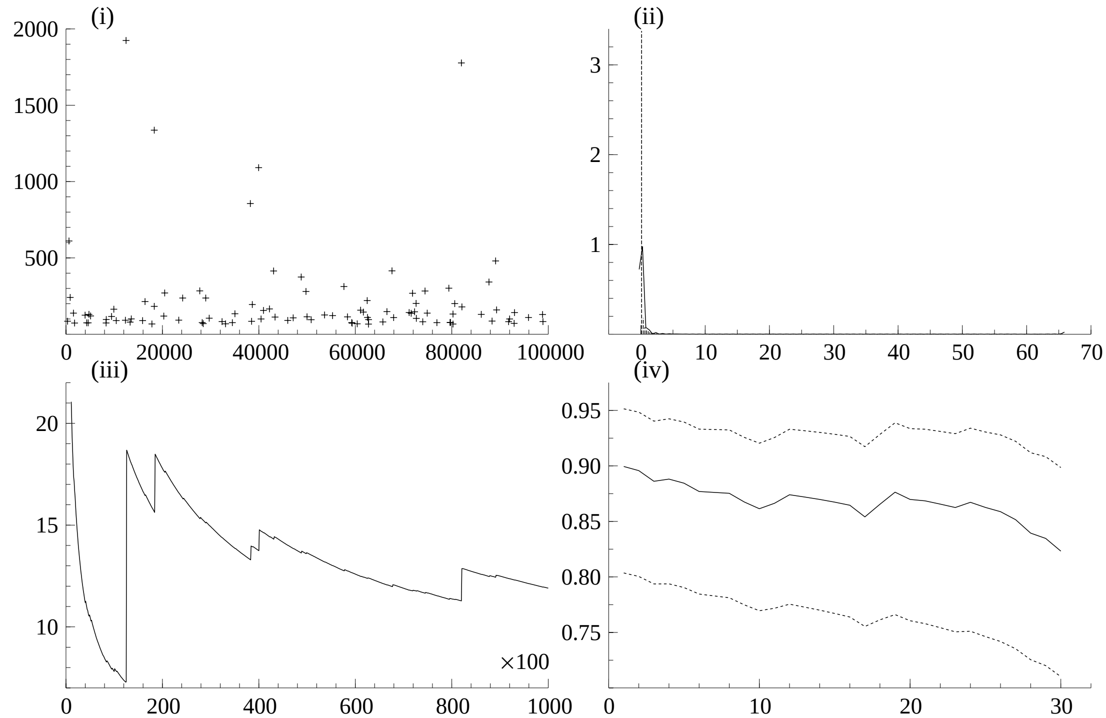

Figure 2: *Importance sampling diagnostic graphics for SV model without leverage. (i) Largest hundred weights from an importance sample of* $M = 100,000$*; (ii) histogram of weights excluding the largest hundred; (iii) recursive standard deviation of weights for* $j = R+1, \ldots, M$*; (iv) estimated shape parameters (solid line) with* 95% *confidence intervals (dotted lines) for thirty different tresholds.*

the model. Figure 3 presents the same types of diagnostic graphs as in Figure 2. It is clear that these graphs indicate a marginally better behaved importance sampler. In particular, the estimated shapes may provide some evidence that two moments exist in the weight series. When these diagnostic graphs are compared with the same graphs for the estimated model without leverage we conclude that the importance sampler for a stochastic volatility model with leverage is more reliable than the sampler for the model without leverage.

4.4 Empirical results: likelihood-based tests

In Table 4 we present the likelihood tests for the two classes of SV models. The threshold value u for constructing large weights z_i, for $i = 1, \ldots, n$, as defined in §2 is chosen so that we have $n = 1000$ large values. The score and t-value statistics are signed tests for the null hypothesis $H_0 : \xi = \frac{1}{2}$; it means that we suspect that no variance exists for the importance weights when these tests have significant positive values. Both tests are asymptotically standard

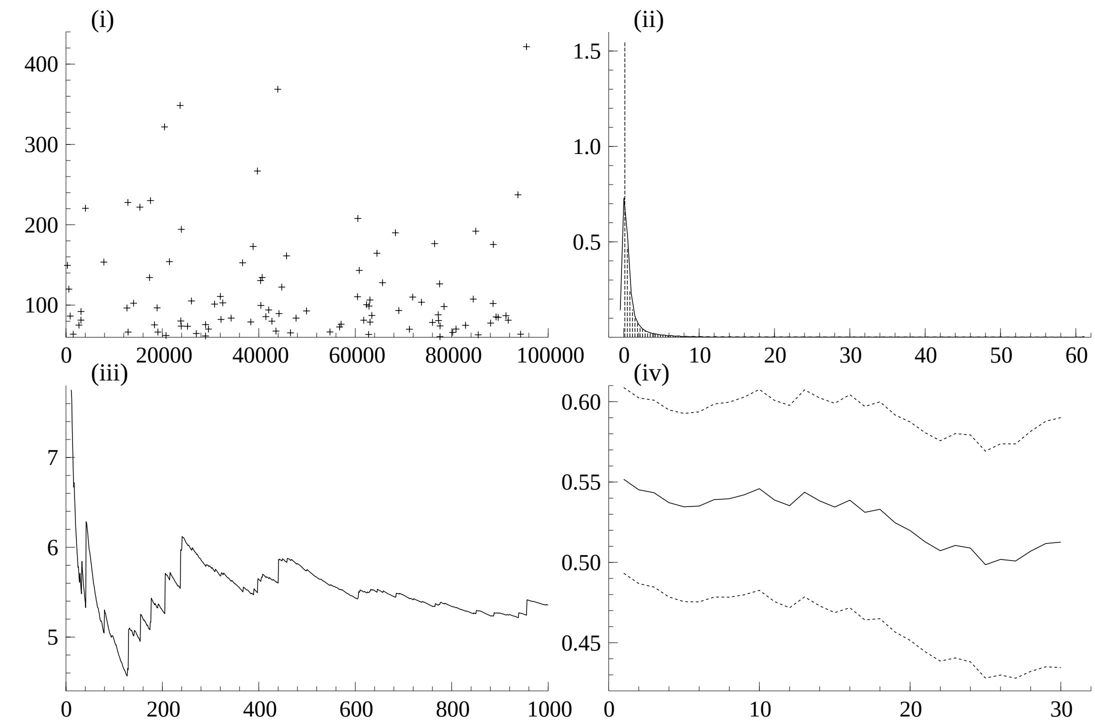

Figure 3: *Importance sampling diagnostic graphics for SV model with leverage. (i) Largest 100 weights; (ii) histogram of weights excluding the largest hundred; (iii) recursive standard deviation of weights for $j = R+1, \ldots, M$; (iv) estimated shape parameters with* 95% *confidence intervals for thirty different thresholds.*

normal distributed, so having a 95% critical value of 1.64. The LR test follows an asymptotic $0.5\left(\chi_0^2 + \chi_1^2\right)$ behaviour with the 95% critical value 2.69.

The results given in Table 4 are more or less consistent across the choice of statistic. For the stochastic volatility model without leverage the results are poor with all the statistics strongly rejecting the existence of a variance. This difficulty is somewhat reduced when the stochastic volatility model is taken with leverage. However the statistics suggest that the procedure is still problematic.

5 Conclusion

In this note we have suggested using extreme value theory to assess the validity of assuming the existence of a variance in the importance sampling weights. This is relatively easy to carry through in practice, providing a formal justification for the use of importance sampling in wider situations than those derived by Geweke (1989) who required the researcher to prove the existence of $Var(w)$. We have illustrated the methods on the problem of estimating stochastic

	Stochastic volatility model	
	without leverage	with leverage
score	3.627	1.174
t-value	3.378	2.124
LR	17.134	5.807

Table 4: *Likelihood based tests for the null of the existence of the variance of the importance sampler for stochastic volatility models estimated by maximum simulated likelihood. Score and t-tests are both standard normal, with the test being rejected for large positive values. LR test has a* $0.5\left(\chi_0^2+\chi_1^2\right)$ *null. It is rejected for large values. The 95% critical values are 1.65, 1.65 and 2.69, respectively.*

volatility models. This involved integrating in many thousands of dimensions.

Acknowledgements

Neil Shephard's research is supported by the UK's ESRC through the grants "Econometrics of trade-by-trade price dynamics" and "High frequency financial econometrics based upon power variation". All the computations reported in this paper were carried out using the Ox 3.0 programming language of Doornik (2001). The programs are available from Siem Jan Koopman by request.

Appendix

The normal importance density $g(y|\alpha)$ for model (12), with

$$y=(y_1,\ldots,y_T)' \quad \text{and} \quad \alpha=(\alpha_1,\ldots,\alpha_T)',$$

is obtained by equating the first and second derivatives of the conditional log-densities $\log f(y,\alpha)$ and $\log g(y,\alpha)$ with respect to α as in Shephard and Pitt (1997). Thus g will be a Laplace approximation to f. The joint density of the stochastic volatility model with leverage can be written $f(y,\alpha)=f(\alpha)f(y|\alpha)$ with $\log f(y|\alpha)=\sum_{t=1}^n f_t$ and

$$f_t=-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma_*^2-\frac{\alpha_t}{2}-\exp(-\alpha_t)\frac{\{y_t-\mu-\rho\exp(\alpha_t/2)\sigma_\varepsilon\eta_t\}^2}{2\sigma_*^2},$$

where $\sigma_*^2=\sigma_\varepsilon^2(1-\rho^2)$ for $t=1,\ldots,T$.

The importance joint density $g(y,\alpha)$ can be represented in terms of the additive model

$$y=m+\alpha+\omega, \qquad \alpha\sim N(0,\Omega), \qquad \omega\sim N(0,\Sigma), \tag{13}$$

where variance matrix Ω has the well-known band structure implied by an autoregressive process of order one which is the model for α_t. The $T\times 1$ vector m and $T\times T$ variance matrix Σ are chosen such that

$$\frac{\partial \log f(y|\alpha)}{\partial\alpha} = \frac{\partial \log g(y|\alpha)}{\partial\alpha}, \qquad \frac{\partial^2 \log f(y|\alpha)}{\partial\alpha\partial\alpha'} = \frac{\partial^2 \log g(y|\alpha)}{\partial\alpha\partial\alpha'}. \tag{14}$$

The set of equations for m and Σ in terms of α are solved iteratively. In the case of the stochastic volatility model with no leverage, the variance matrix Σ is diagonal whereas it is tridiagonal in the case of the model with leverage since η_t in the first equation of (12) is a function of α_{t+1} and α_t. It turns out that the densities are log-concave in α and this ensures that Σ is always positive definite.

The resulting equations for m and Σ can be solved out using an iterative procedure for which new 'trial' values for α are used. In particular, for each iteration we compute $E(\alpha|y)$ for which an expression is obtained by exploiting the properties of the multivariate normal distribution applied to model (13). We have

$$\hat{\alpha} = E(\alpha|y) = \Omega\Sigma_y^{-1}(y-m), \quad \text{where} \quad \Sigma_y = Var(y) = \Omega + \Sigma,$$

and, given the special structures of both matrices Ω and Σ, this can be computed efficiently by an order T computation. Note that Ω^{-1} is a tridiagonal matrix.

Importance sampling requires draws from the multivariate conditional density $g(\alpha|y)$ which in our case is also normal. Thus we can draw directly from

$$\alpha^j \sim N(\hat{\alpha}, V), \quad \text{where} \quad V = Var(\alpha|y) = \Omega - \Omega\Sigma_y^{-1}\Omega.$$

Such draws can be generated by scaling independent $N(0,1)$ realisations. The special structures of the matrices involved ensure that the generation of importance simulations can be implemented as an order T computation.

References

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Annals of Mathematical Statistics 25*, 573–8.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. London: Chapman & Hall.

Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics 61*, 375–400.

Danielsson, J. and C. G. de Vries (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance 4*, 241–257.

Danielsson, J. and J. F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics 8*, S153–S174.

Doornik, J. A. (2001). *Ox: Object Oriented Matrix Programming, 3.0*. London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika 84*, 669–84.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of the Royal Statistical Society, Series B 62*, 3–56.

Durham, G. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *Journal of Business and Economic Statistics 20*, 297–338.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69*, 959–993.

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Gallant, A. R. (1997). *An Introduction to Econometric Theory*. Princeton: Princeton University Press.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Geweke, J. (1997). Posterior simulators in econometrics. In D. M. Kreps and K. F. Wallis (Eds.), *Advances in Econometrics: Seventh World Congress*, pp. 128–165. Cambridge: Cambridge University Press.

Ghysels, E., A. C. Harvey, and E. Renault (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 119–191. Amsterdam: North-Holland.

Gourieroux, C., A. Holly, and A. Monfort (1981). Kuhn-Tucker, likelihood ratio and Wald tests for nonlinear models with inequality constraints on the parameters. Harvard Institute of Economic Research Discussion Paper 770.

Gourieroux, C. and A. Monfort (1996). *Simulation Based Econometric Methods*. Oxford: Oxford University Press.

Hajivassiliou, V. A. and P. A. Ruud (1994). Classical estimation methods for LDV models using simulation. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, pp. 2383–2441. Amsterdam: North-Holland.

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods*. London: Methuen and Co.

Hendry, D. F. and J. F. Richard (1991). Likelihood evaluation for dynamic latent variable models. In H. M. Amman, D. A. Delsley, and L. F. Pau (Eds.), *Computational Economics and Econometrics*. Dordrecht: Kluwer.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*, 281–300.

Kahn, H. and A. W. Marshall (1953). Methods of reducing sample size in Monte Carlo computations. *Journal of the Operational Research Society of America 1*, 263–271.

Kloek, T. and H. K. Van Dijk (1978). Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica 46*, 1–20.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Marshall, A. (1956). The use of multi-stage sampling schemes in Monte Carlo computations. In M. Meyer (Ed.), *Symposium on Monte Carlo Methods*, pp. 123–140. New York: Wiley.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica 59*, 347–370.

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

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics 87*, 271–301.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics 15*, 1174–1207.

Stephens, M. and P. Donnelly (2000). Inference in molecular population genetics (with discussion). *Journal of the Royal Statistical Society, Series B 62*, 605–655.

Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.