Testing the assumptions behind importance sampling

Siem Jan Koopman
Neil Shephard
Drew Creal

Testing the assumptions behind importance sampling

Siem Jan Koopman
Department of Econometrics, VU University Amsterdam
De Boelelaan 1105
NL-1081 HV Amsterdam
`s.j.koopman@feweb.vu.nl`

Neil Shephard
Oxford-Man Institute, University of Oxford
Blue Boar Court, 9 Alfred Street
Oxford OX1 4EH, UK
`neil.shephard@economics.ox.ac.uk`

Drew Creal
Department of Econometrics, VU University Amsterdam
De Boelelaan 1105
NL-1081 HV Amsterdam
`dcreal@feweb.vu.nl`

October 2007

Abstract

Importance sampling is used in many areas of modern 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 use extreme value theory to empirically assess the appropriateness of this assumption. Our main application is the stochastic volatility model, where importance sampling is commonly used for maximum likelihood estimation of the parameters of the model.

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

1 Introduction

This paper develops tools for evaluating the importance sampling techniques that are popular in econometrics. There are two situations where importance sampling is commonly applied. It is often used to numerically evaluate the likelihood function in missing data models, when an analytic expression for the likelihood function does not exist. This is of primary importance in the literature on maximum likelihood estimation of stochastic volatility (SV) models. Importance sampling is also popular in Bayesian applications when computing expectations of the posterior distribution.

In a fundamental contribution, Geweke (1989) argued that importance sampling should only be used in settings where the variance of the importance weights is known to exist. Failure of

this condition can lead to slow and unstable convergence of the estimator as the central limit theorem governing convergence fails to hold. Robert and Casella (2004, p.90) provide detailed examples of importance samplers that fail this condition and note that ignoring the problem can result in strongly biased estimators. While the variance conditions can be checked analytically in low dimensional problems, proving that they are met in high dimensions can be challenging. Many econometrics and statistics papers have recently been written which, in effect, *a priori* assume that this condition holds. This is the case in the stochastic volatility literature; see, e.g., Hendry and Richard (1991), Danielsson and Richard (1993), Sandmann and Koopman (1998), Elerian, Chib, and Shephard (2001), Durham and Gallant (2002), Durham (2006), and Asai and McAleer (2006). It is also assumed to hold in the macroeconomics literature on dynamic stochastic general equilibrium models; see, for example, An and Schorfheide (2007).

In this paper, we develop methods to check for the existence of the variance of the importance weights. This will be based on the application of extreme value theory.[1] Limit theorems from extreme value theory imply that we can learn about the variance of the weights by studying the behavior of their distribution in the right hand tail. Earlier work on testing whether the variance of the importance weights exists also using extreme value theory was conducted by Monahan (1993, 2001). Monahan examined the tail-behavior of the weights by building a test statistic from the Hill (1975) estimator. We take an alternative approach and develop three different test statistics using the well-known principles of the Wald, score and likelihood ratio tests. We also propose graphical diagnostics for a complete insight.

The new test procedures are carried out for two Monte Carlo studies and for the maximum likelihood analysis of two SV models. The latter models have been our main motivation to investigate the problem in detail. It will be shown that sometimes the variance of the importance weights does not exist when estimating parameters of SV models. Although our main focus will be on likelihood estimation, our methods apply generally to all applications of importance sampling including Bayesian approaches.

A review of importance sampling is provided in Section 2. In Section 3, we discuss the statistical estimation of tail indexes for extremes and introduce our test statistics. Section 4 presents and discusses results from some Monte Carlo experiments that are conducted to evaluate the power and size of the proposed tests. We apply the test statistics and graphical diagnostics to maximum likelihood estimation of the parameters of two SV models in Section 5. Section 6

---

[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. Furthermore, it is used to compute 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 in their context is harder since *i.i.d.* assumptions do not hold.

concludes.

2 Importance sampling

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. Consider an integral that cannot be solved analytically

$$\mathcal{I} = \int_{\mathcal{X}} h(x)f(x)dx, \tag{1}$$

where $h(x)$ is a function and $f(x)$ is a probability density. An importance sampling density $g(x)$ is selected from which it is easy to simulate and whose support is also $\mathcal{X}$. The integral $\mathcal{I}$ can be approximated by

$$\widehat{\mathcal{I}} = \frac{1}{N}\sum_{i=1}^{N} w_i, \qquad \text{where} \qquad w_i = w(x^i),$$

with

$$w(x) = \frac{h(x)f(x)}{g(x)} \quad \text{and} \quad x^i \sim g(x), \tag{2}$$

for $i = 1, \ldots, N$ and a large number of N draws. The strong law of large numbers implies that $\widehat{\mathcal{I}}$ will be a consistent estimator of $\mathcal{I}$ without any additional assumptions. By construction we know that the $\{w_i\}$ are $i.i.d.$ and that $\mathbb{E}[w(x)] = \mathcal{I}$. As a result, a simple application of Kolmogorov's strong law of large numbers (e.g. Geweke (1989) and Geweke (2005, p. 114)) shows that

$$\widehat{\mathcal{I}} \overset{a.s.}{\to} \mathcal{I}, \qquad \text{as } N \to \infty,$$

whatever importance sampler we design. In practice, the normalizing constant of $f(x)$ or $g(x)$ in (2) may not be known requiring a self-normalized importance sampling estimator.[2]

In order to easily measure the precision of $\widehat{\mathcal{I}}$ and to guarantee that the rate of convergence to $\mathcal{I}$ is $N^{1/2}$, it is helpful to have a Gaussian central limit theorem for $\widehat{\mathcal{I}}$. We know that a necessary condition for the application of the Lindeberg-Lévy central limit theorem is the existence of a variance for the importance weights. This would allow us to conclude that

$$\sqrt{N}\left(\widehat{\mathcal{I}} - \mathcal{I}\right) \overset{d}{\to} N(0, \gamma^2).$$

Expressions for the asymptotic variance γ^2 of the importance sampling estimator can be found in Geweke (2005, p. 114) and Cappé, Moulines, and Rydén (2005, p.287). However, the existence

---

[2]This self-normalized estimator induces a small bias but it remains consistent. Liu (2001, p. 35) and Robert and Casella (2004, p.95) note that the self-normalized estimator may be preferred as it typically has smaller mean square error than the unbiased estimator.

of the variance of the importance weights is by no means guaranteed. If it does not exist, the importance sampling estimator may convergence erratically making estimates unreliable. An approach for testing whether the variance exists, is considered in the next section.

3 Extreme value theory and inference

3.1 Extreme value theory

We begin this section by briefly reviewing the results from extreme value theory that are relevant for this paper; see Embrechts, Kluppelberg, and Mikosch (1997) for more detailed discussions. Assume we have a sequence $w_1, w_2, ..., w_N$ of $i.i.d$ random variables from a distribution function F. The $\{w_i\}$ need not represent importance weights, although it will be the case in our work below. Denote the endpoint of the distribution function F as $w_E = \sup\{w : F(w) < 1\} \leq +\infty$. Pickands (1975) defined a new sequence of random variables $Z_1, ..., Z_n$ created from those w_i that lie above a threshold u, i.e. $Z_i = w_i - u$. Pickands (1975) proved that as N and u get larger the conditional distribution of the excesses $\{Z_i\}$ is well approximated by the generalized Pareto distribution. More formally, he showed that

$$\lim_{u \to w_E} \sup_{0 < z < w_E - u} |F_u(z) - GPD(z; \xi, \beta)| = 0,$$

where $F_u(z)$ is the distribution function of the excesses conditional on $w > u$. Furthermore, GPD denotes the generalized Pareto distribution with density function

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

where

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

The threshold u plays a crucial role in this result and its selection will be discussed in detail in Section 4. We continue by assuming that u is a given threshold.

The shape parameter ξ of the generalized Pareto distribution characterizes the thickness of the distribution's tails. In the context of extreme value theory, the value of ξ is determined by the thickness of the tails of F. This means that we can extract information about the thickness of the tails of the importance weights by making inference on the shape parameter using the excesses $\{Z_i\}$. In fact, it can be shown that when $\xi > 0$ then $\mathbb{E}[w^j] = \infty$ for $j \geq 1/\xi$. Consequently, the number of moments that the distribution of the importance weights has can be determined by focusing on ξ. In particular, we are interested in establishing that the variance of the importance weights exists, $\xi \leq 1/2$.

4

We know by construction of the importance sampler that $\xi \leq 1$. The cases of $\xi < 0$ deal with situations where the $\{w_i\}$ have some upper bound. This is of some relevance in importance sampling in the case of the sampler being bounded, that is

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

for some finite choice of $M > 0$. Then if (4) is true for $x \in \mathcal{X}$, the support of the importance weights $\{w_i\}$ will be bounded and the existence of all moments is guaranteed. This is, in practice, quite unusual. In most cases of real interest, the sampler is not bounded, so $\xi \in [0, 1]$.

3.2   Previous work: Monahan's test

Monahan (1993, 2001) proposed a procedure for testing whether the importance sampling weights have a finite variance based on the Hill (1975) estimator. The Hill estimator is an alternative to the generalized Pareto distribution for estimating the tail parameter ξ. It is defined as

$$\xi^{Hill} = \frac{1}{k} \sum_{j=1}^{k} \log w_{(N-j+1)} - \log w_{(N-k)}, \tag{5}$$

where $w_{(1)} \leq w_{(2)} \leq \ldots \leq w_{(N)}$ are the order statistics from a sample of importance weights. The tuning parameter k determines the number of large order statistics used in (5) meaning that it plays a similar role as the threshold parameter u above.

The Hill estimator is largely regarded as a nonparametric estimator although it can be interpreted as a maximum likelihood estimator under appropriate conditions on the distribution function F. Asymptotic normality of the Hill estimator has been studied by Haeusler and Teugels (1985) who have shown that it depends on the rate of growth of k relative to N. Monahan used their result

$$\sqrt{k/\xi^2} \left(\widehat{\xi}^{Hill} - \xi \right) \xrightarrow{d} N(0, 1).$$

and chose the values $k = 2N^{1/3}$ and $k = 4N^{1/3}$ to build a test statistic. Monahan's test has as its null hypothesis that the variance of the importance weights is infinite making it a test of $H_0 : \xi \geq 1/2$ versus $H_1 : \xi < 1/2$. In Section 4, we compare Monahan's test with the tests we propose below. To make the comparison easier, we have reversed the null hypothesis of his test to make it consistent with ours.

3.3   Proposed methods

Our approach of testing for an infinite variance will be based on tests constructed from maximum likelihood estimates of the generalized Pareto distribution. Our paper builds on the work of Smith (1985, 1987) who studied the properties of the maximum likelihood estimator of $\theta = (\xi, \beta)'$

in this setting. We construct three test statistics using his results. Formally, the hypotheses we will be interested in are given by

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

The null hypothesis implies the existence of the variance while the alternative denies it. 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

$$H_0 : \xi = \frac{1}{2}, \qquad \text{and} \qquad H_1 : \xi > \frac{1}{2}. \tag{6}$$

As we have a parametric model for the weights over a threshold, we will use a likelihood function to carry out the testing. This approach to inference is often called the peaks over threshold method. We will study the behavior of the score and likelihood ratio tests of the hypothesis.

3.4 Estimation and Wald test

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

$$\log f(z; \theta) = -n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log \left(1 + \xi \beta^{-1} z_i\right). \tag{7}$$

Maximum likelihood estimation of the parameter vector $\theta = (\xi, \beta)'$ is discussed in Smith (1985, 1987). The asymptotic distribution of the maximum likelihood estimator $\hat{\theta}$ is given by

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \overset{d}{\to} N\left(0, \mathbf{I}(\theta)^{-1}\right) \qquad \text{where} \qquad \mathbf{I}(\theta)^{-1} = (1 + \xi)\begin{pmatrix} 2\beta^2 & \beta \\ \beta & 1 + \xi \end{pmatrix}, \tag{8}$$

where $n\mathbf{I}(\theta)$ denotes the expected information matrix. Smith (1985) has shown that likelihood inference is standard 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. Davison and Smith (1990) describe how to reduce maximum likelihood estimation of (7) to a one-dimensional parameter search. Substituting $\beta = \xi/\tau$ into (7) and differentiating with respect to ξ and τ produces

$$\xi = \xi(\tau) = \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + \tau z_i\right), \tag{9}$$

$$\frac{n}{\tau} = \left(1 + \frac{1}{\xi(\tau)}\right) \sum_{i=1}^{n} \frac{z_i}{1 + \tau z_i}. \tag{10}$$

Equation (9) can then be substituted into (10) and solved numerically for an estimate $\hat{\tau}$. Given $\hat{\tau}$, the maximum likelihood estimates $\hat{\xi}$ and $\hat{\beta}$ are easily obtained. A sensible starting value for $\tau = \xi/\beta$ can be found by imposing the null hypothesis $\xi = 0.5$ and noting that the expected value of Z is

$$\mathbb{E}[Z] = \frac{\beta}{\xi^2} \frac{\Gamma(\xi^{-1} - 1)}{\Gamma(1 + \xi^{-1})} = 2\beta, \qquad \text{for } \xi = 0.5. \tag{11}$$

Iterations can therefore start at $\tau = 1/\bar{z}$ where $\bar{z} = n^{-1} \sum_{i=1}^{n} z_i$.

Once the maximum likelihood estimator of θ is found the Wald test can be computed to test the null hypothesis (6). For our purposes it is appropriate to compute a t-test, that is

$$t = \sqrt{\frac{n}{3\hat{\beta}^2}} \left(\hat{\xi} - \frac{1}{2} \right), \tag{12}$$

where $\hat{\beta}$ is the maximum likelihood estimate of β. The null hypothesis is rejected when the t-test takes a large positive value compared to a standard normal random variable.

3.5 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

$$
\begin{aligned}
s_r^\beta &= -\frac{n}{\beta} + \frac{3}{2}\beta^{-2} \sum_{i=1}^{n} \frac{z_i}{1 + 0.5\beta^{-1}z_i} \\
&= n\beta^{-1} \left(\frac{3}{n} \sum_{i=1}^{n} \frac{z_i}{2\beta + z_i} - 1 \right).
\end{aligned}
$$

Using the results in (11) once again, we can take $0.5\bar{z}$ as the initial value for β. After convergence the restricted estimate of β is obtained which we denote by $\hat{\beta}_r$.

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

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

This score value for the null hypothesis (6) is a function of β and it can be evaluated when β is replaced by its (restricted) maximum likelihood estimator $\hat{\beta}_r$. We know from the results of Smith (1985, 1987) that

$$s_*^\xi = \frac{1}{\sqrt{2n}} s_r^\xi \xrightarrow{d} N(0,1), \tag{13}$$

which provides a simple test. We reject the null $H_0 : \xi = \frac{1}{2}$ when $\hat{s}_*^\xi$ is significantly positive and where $\hat{s}_*^\xi$ is defined as s_*^ξ with β replaced by $\hat{\beta}_r$.

3.6 Likelihood ratio test

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

7

the null

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

Of course, the null is on the boundary next to the alternative and so the limiting distribution of the statistic is

$$LR \xrightarrow{d} \frac{\chi_0^2 + \chi_1^2}{2}, \quad \text{under } H_0.$$

The χ_0^2 distribution is a degenerate distribution with all its probability mass at zero. There exists a 0.5 probability that the likelihood ratio statistic will be zero (as $\widehat{\xi}$ will be negative). The rest of the time the statistic will have a χ_1^2 distribution (e.g. Chernoff (1954) and Gourieroux, Holly, and Monfort (1981)).

4 Monte Carlo evidence

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. Our second experiment is the classic example of estimating a Student's t distribution using a normal candidate density.

4.1 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 a positive number. The function $h(x)$ is constant and equal to one. Thus for this problem it is implied that $\mathcal{I} = 1$ in (1). We consider how well our tests can correctly identify whether the variance of the importance weights exists. The weight function is given by

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

which obviously has $\mathbb{E}[w(x)] = 1$, while

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

which is bounded only if $\epsilon < 1$. The variance of the importance weights will not exist if $\epsilon \geq 1$, meaning that the central limit theorem will not hold for importance samplers built in that range of values.

In Table 1, we present the diagnostic tests for a large number of different tuning values k^* from which we determine the threshold values $u \approx w_{(N-k^*)}$ where $w_{(1)} \leq w_{(2)} \leq \ldots \leq w_{(N)}$ are the order statistics from a sample of N importance weights. The Monte Carlo results show that

	$N = 10,000$					$N = 100,000$				
	score	t	LR	M_{test}	M^*_{test}	score	t	LR	M_{test}	M^*_{test}
$k^* = 0.55N$										
$\epsilon = 0.5$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\epsilon = 0.8$	0.00	0.46	0.07	0.00	0.00	0.00	0.60	0.21	0.00	0.00
$\epsilon = 1.0$	0.13	0.97	0.76	0.01	0.02	1.00	1.00	1.00	0.00	0.00
$\epsilon = 1.2$	0.71	1.00	0.99	0.05	0.06	1.00	1.00	1.00	0.08	0.08
$\epsilon = 1.5$	0.99	1.00	1.00	0.23	0.19	1.00	1.00	1.00	0.46	0.33
$\epsilon = 3.0$	1.00	1.00	1.00	0.90	0.73	1.00	1.00	1.00	1.00	0.95
$\epsilon = 5.0$	1.00	1.00	1.00	0.98	0.89	1.00	1.00	1.00	1.00	1.00
$k^* = 0.50N$										
$\epsilon = 0.5$	0.00	0.00	0.00	-	-	0.00	0.00	0.00	-	-
$\epsilon = 0.8$	0.00	0.19	0.01	-	-	0.00	0.02	0.00	-	-
$\epsilon = 1.0$	0.03	0.87	0.47	-	-	0.93	1.00	0.99	-	-
$\epsilon = 1.2$	0.41	1.00	0.94	-	-	1.00	1.00	1.00	-	-
$\epsilon = 1.5$	0.95	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$\epsilon = 3.0$	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$\epsilon = 5.0$	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$k^* = 0.40N$										
$\epsilon = 0.8$	0.00	0.02	0.00	-	-	0.00	0.00	0.00	-	-
$\epsilon = 1.0$	0.00	0.48	0.11	-	-	0.02	0.83	0.42	-	-
$\epsilon = 1.2$	0.08	0.92	0.63	-	-	0.99	1.00	1.00	-	-
$\epsilon = 1.5$	0.64	1.00	0.98	-	-	1.00	1.00	1.00	-	-
$\epsilon = 3.0$	0.82	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$k^* = 0.30N$										
$\epsilon = 0.8$	0.00	0.01	0.00	-	-	0.00	0.00	0.00	-	-
$\epsilon = 1.0$	0.00	0.20	0.03	-	-	0.00	0.10	0.01	-	-
$\epsilon = 1.2$	0.01	0.68	0.30	-	-	0.55	0.99	0.97	-	-
$\epsilon = 1.5$	0.28	0.97	0.85	-	-	1.00	1.00	1.00	-	-
$\epsilon = 3.0$	0.91	1.00	0.98	-	-	1.00	1.00	1.00	-	-
$k^* = 0.10N$										
$\epsilon = 0.8$	0.00	0.01	0.00	-	-	0.00	0.00	0.00	-	-
$\epsilon = 1.0$	0.00	0.05	0.01	-	-	0.00	0.00	0.00	-	-
$\epsilon = 1.2$	0.00	0.19	0.08	-	-	0.01	0.37	0.22	-	-
$\epsilon = 1.5$	0.03	0.47	0.34	-	-	0.72	0.98	0.99	-	-
$\epsilon = 3.0$	0.75	0.98	0.76	-	-	1.00	1.00	1.00	-	-
$\epsilon = 5.0$	0.97	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$k^* = 0.01N$										
$\epsilon = 0.8$	0.00	0.01	0.01	-	-	0.00	0.00	0.00	-	-
$\epsilon = 1.0$	0.00	0.01	0.02	-	-	0.00	0.00	0.01	-	-
$\epsilon = 1.2$	0.00	0.00	0.05	-	-	0.00	0.00	0.08	-	-
$\epsilon = 1.5$	0.01	0.00	0.10	-	-	0.04	0.00	0.37	-	-
$\epsilon = 3.0$	0.08	0.00	0.34	-	-	0.81	0.01	0.99	-	-
$\epsilon = 5.0$	0.17	0.00	0.51	-	-	0.98	0.02	1.00	-	-

Table 1: *Normal density experiment. Monte Carlo study for power and size of tests for the existence of a variance for importance weights, with size $\alpha = 0.05$. Percentage of rejections of the null hypothesis (existence of a variance) computed from $10,000$ Monte Carlo replications. Results are presented for different tuning values k^* from which threshold values $u \approx w_{(N-k^*)}$ are determined. The M_{test} used $k = 4N^{1/3}$ while M^*_{test} used $k = 2N^{1/3}$.*

for small values of ϵ importance sampling is valid while for values of $\epsilon \geq 1$ the variance of the importance weights does not exist. For our test statistics to be effective, we require a simulation sample size N that is sufficiently large (say, $100,000$). The Wald test appears to have the most power. It can potentially be oversized if the threshold is selected too high. It can also perform poorly as the threshold gets extremely small. Tests that are slightly oversized are not necessarily problematic. Rejecting importance samplers whose weights are barely finite may not be a bad practice as these algorithms can still be unstable. The LR statistic has nearly the same power and it behaves more robustly for different values of k^* than the Wald statistic. This implies that computing more than one statistic is a good idea. Both of these statistics are considerably more powerful than the Monahan statistic computed using either $k = 4N^{1/3}$ or $k = 2N^{1/3}$. We note that the Monahan statistic performs better than our tests when the number of draws N is small (e.g. $N \leq 5,000$).

We have chosen the threshold u (implied by the choice of k^*) in an arguably *ad hoc* manner to both facilitate the Monte Carlo study and to understand how different values of u impact the tests. In practice, the threshold can be chosen in alternative ways. Methods that are common in the extreme value literature are a mean residual life plot or by minimizing the distance between the empirical distribution function and the generalized Pareto distribution evaluated at the maximum likelihood estimates; see, e.g. Embrechts, Kluppelberg, and Mikosch (1997) and Gonzalo and Olmo (2004). Graphing the maximum likelihood estimates of ξ along with confidence bands for a large number of thresholds is also informative. We use this strategy in Section 5.

Our proposed test statistics do not provide strict guarantees for a successful importance sampling procedure. We believe that graphical diagnostics, although informal, can play a complementary role. They are a simple way to detect severely deficient importance sampling algorithms. A graph of the largest 100 weights indicates the seriousness of the outliers. The remaining weights can be represented via a histogram. Plotting estimates of the variance of the weights recursively can show the impact that each weight has on the estimate. As the sample increases, the variance should converge to a constant if it exists. Plots like these do not always detect problems and we recommend viewing them for different sequences of random numbers.

Graphs for the normal density experiment with $\epsilon = 0.5, 1.0,$ and 1.2 are presented in the three rows of Figure 1. The values of ϵ were chosen to compare the output from importance samplers representing both the null and alternative hypotheses. When $\epsilon = 0.5$, the variance is finite and the recursive estimate of the variance convergences smoothly. The graphs for the cases $\epsilon = 1.0$ and 1.2 are clearly informative in this example. The large jumps in the recursive
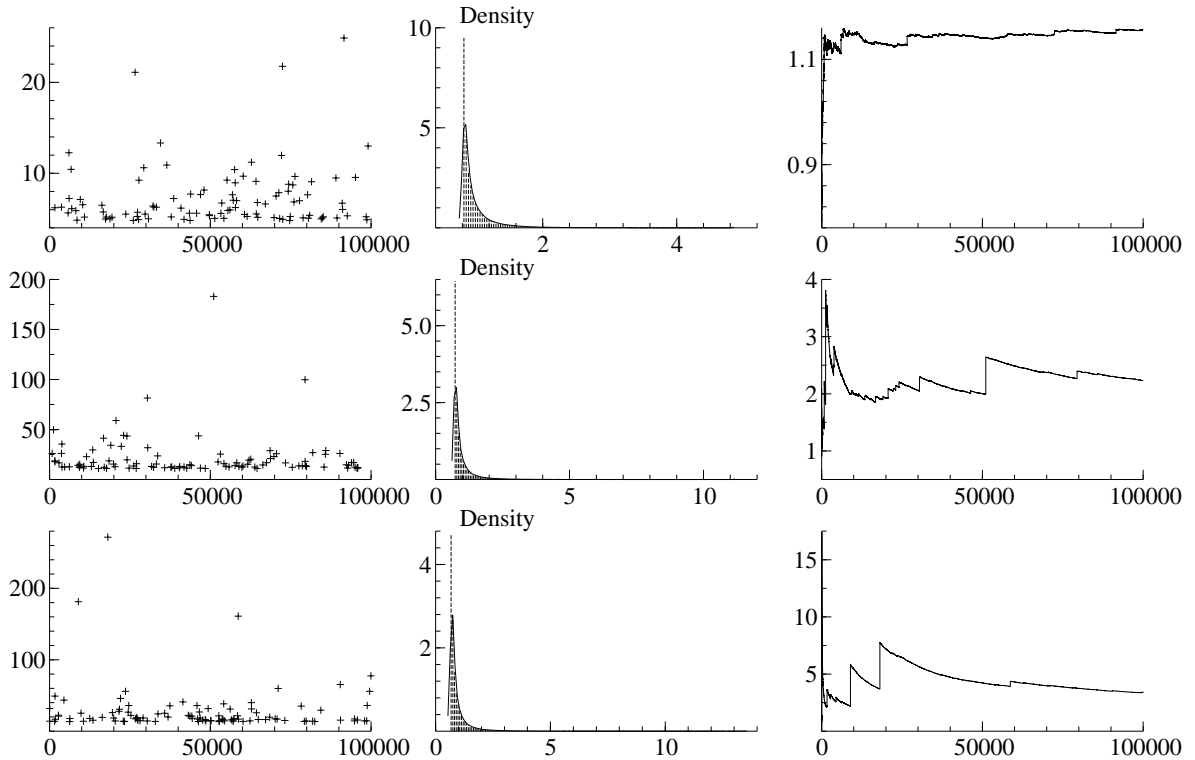
Figure 1: *Graphical diagnostics for the normal experiment. Top, middle, and bottom rows represent $\epsilon = 0.5, 1.0, 1.2$ respectively. The left hand side pictures are of the largest 100 weights, the middle pictures are a histogram of the remaining weights and the right hand side shows a recursive estimator of the variance of the weights. $N = 100,000$ draws.*

variance estimates correspond to the outliers in the left hand graphs of the middle and bottom rows.

4.2   Normal approximation to the Student's t distribution

Our second Monte Carlo experiment is the classic example of approximating a Student's t distribution with a normal distribution. This example can be used to evaluate the power of the tests. A similar experiment was also considered by Monahan (1993). We let $f(x)$ be a Student's t distribution with ν degrees of freedom and where $g(x)$ is a normal density with mean zero and variance $\nu/(\nu + 1)$.[3] Once again the function $h(x)$ is constant and equal to one. The weight function is simply given by

$$w(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{2}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu+1}} \frac{\left(1 + x^2/\nu\right)^{-(\nu+1)/2}}{\exp\left(-0.5(\frac{\nu+1}{\nu})x^2\right)}.$$

In this example, the quantity $\mathbb{E}[w(x)^2]$ does not exist.

---

[3]We believe we have implemented the same experiment as Monahan. A potential missprint in his paper may imply an alternative normal importance density with variance $(\nu + 1)/\nu$ rather than the importance density with

11

	$N = 10,000$					$N = 100,000$				
	score	t	LR	M_{test}	M^*_{test}	score	t	LR	M_{test}	M^*_{test}
$k^* = 0.50N$										
$\nu = 1$	1.00	1.00	1.00	0.97	0.88	1.00	1.00	1.00	1.00	1.00
$\nu = 2$	1.00	1.00	1.00	0.87	0.75	1.00	1.00	1.00	1.00	0.99
$\nu = 5$	1.00	1.00	1.00	0.19	0.26	1.00	1.00	1.00	0.84	0.76
$\nu = 20$	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00
$\nu = 100$	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00
$k^* = 0.10N$										
$\nu = 1$	0.98	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$\nu = 2$	0.89	1.00	1.00	-	-	1.00	1.00	1.00	-	-
$\nu = 5$	0.37	1.00	0.87	-	-	1.00	1.00	1.00	-	-
$\nu = 20$	0.00	0.50	0.05	-	-	0.00	0.46	0.06	-	-
$\nu = 100$	0.00	0.00	0.00	-	-	0.00	0.00	0.00	-	-
$k^* = 0.01N$										
$\nu = 1$	0.21	0.02	0.56	-	-	1.00	0.90	1.00	-	-
$\nu = 2$	0.15	0.05	0.47	-	-	0.97	0.84	1.00	-	-
$\nu = 5$	0.04	0.15	0.24	-	-	0.49	0.78	0.91	-	-
$\nu = 20$	0.00	0.14	0.01	-	-	0.00	0.02	0.00	-	-
$\nu = 100$	0.00	0.01	0.00	-	-	0.00	0.00	0.00	-	-

Table 2: *Normal approximation to the Student's t experiment. Monte Carlo study for power of tests for the existence of a variance for importance weights, with size $\alpha = 0.05$. Percentage of rejections of the null hypothesis (existence of a variance) computed from $10,000$ Monte Carlo replications. Results are presented for different tuning values k^* from which threshold values $u \approx w_{(N-k^*)}$ are determined. The M_{test} used $k = 4N^{1/3}$ while M^*_{test} used $k = 2N^{1/3}$.*
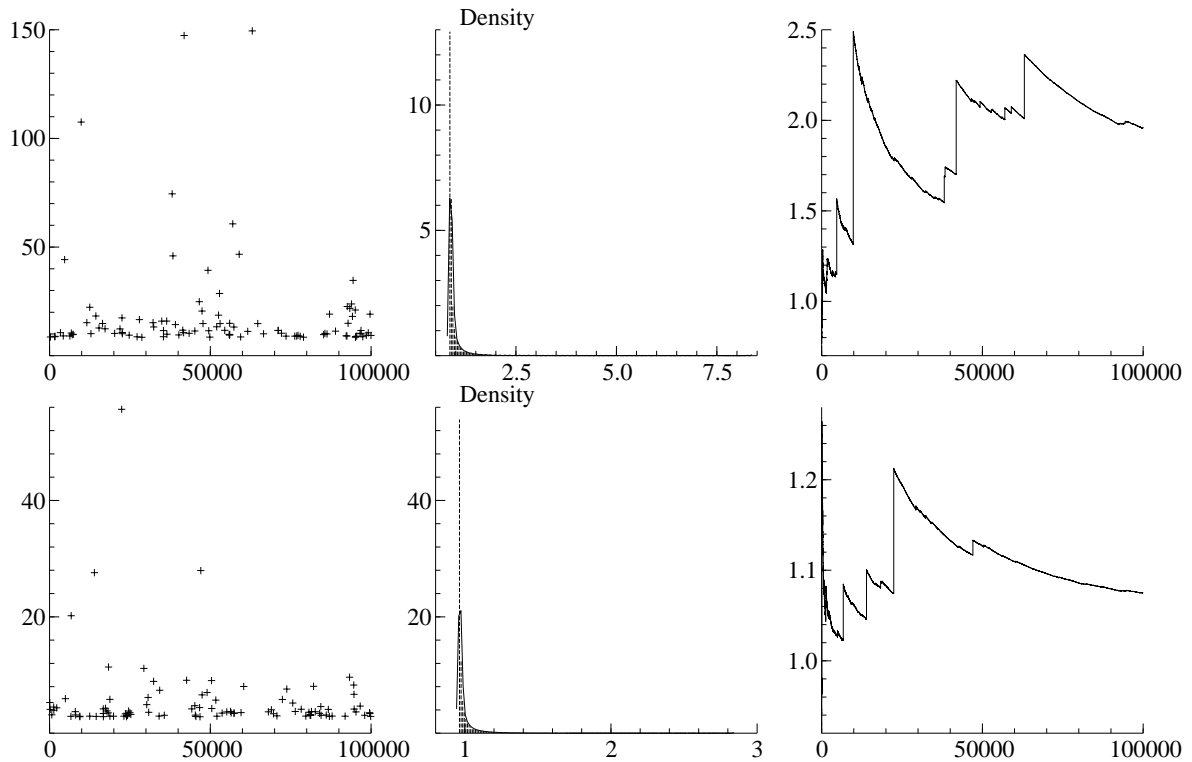
Figure 2: *Graphical diagnostics for the Student's t/normal experiment. Top and bottom rows represent $\nu = 5$ and 20, respectively. The left hand side pictures are of the largest 100 weights, the middle pictures are a histogram of the remaining weights and the right hand side shows a recursive estimator of the variance of the weights. $N = 100,000$ draws.*

Table 2 contains the diagnostic tests for different values of the degrees of freedom and the threshold. The performance of the tests are mostly in agreement with the results from the normal experiment above. The Wald test is the most powerful and performs the best when N is reasonably large. The difference in power between this test and the others is significant when $\nu \geq 20$. Figure 2 plots the series of graphical diagnostics for two values of the degrees of freedom. The recursive variance estimates easily detect problems when $\nu = 5$ while $\nu = 20$ also appears doubtful. Problems created by large values of ν can be hard to recognize with plots, making the combination of test statistics and graphical diagnostics valuable.

5 An illustration: stochastic volatility model

Stochastic volatility (SV) models have gained considerable interest in econometrics as a class of models which allows volatility to change stochastically through time. These models are often stated in continuous time and appear frequently in the option pricing literature. Ghysels,

---

variance $\nu/(\nu + 1)$ given here.

Harvey, and Renault (1996) and Shephard (2005) review the literature on this topic from an econometric viewpoint. The likelihood function for SV models is an integral that cannot be solved analytically but it can be evaluated by Monte Carlo simulation. Papers that carry out maximum likelihood analysis of SV models using importance sampling include Danielsson (1994), Shephard and Pitt (1997), Durbin and Koopman (1997), Sandmann and Koopman (1998), Durham and Gallant (2002), Durham (2003), Liesenfeld and Richard (2003), Brandt and Kang (2004), Durham (2006), Asai and McAleer (2006), Richard and Zhang (2007), and Jungbacker and Koopman (2007).

The SV model we use to illustrate our discussion will be a discrete time, log-normal, univariate SV model for a sequence of returns y_t for $t = 1, \ldots, T$. It can be thought of as an 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)). It is given by

$$
\begin{aligned}
y_t &= \mu_y + \exp(\alpha_t/2)\varepsilon_t, \\
\alpha_{t+1} &= \mu_\alpha + \phi(\alpha_t - \mu_\alpha) + \sigma_\eta \eta_t, \qquad \alpha_1 \sim N\{0, \sigma_\eta^2/(1-\phi^2)\},
\end{aligned}
\tag{14}
$$

where the disturbances ε_t and η_t are mutually and serially independent and independent of α_t for $t = 1, \ldots, T$. The unconditional mean of the returns is given by μ_y while the mean of the log-volatility is μ_α. The degree of volatility persistence is measured by the autoregressive parameter ϕ and the parameter σ_η is the standard deviation of the disturbances to the log-volatility process. We will consider two cases for ε_t. The first model, denoted by SV-N, assumes that $\varepsilon_t \sim N(0,1)$ while the second model, denoted SV-t, has $\varepsilon_t \sim t_\nu$ where ε_t is scaled to have unit variance.

Both SV models require the estimation of the parameter vector Θ. Inference for the unknown parameters will be based on $f(y; \Theta)$ where $y = (y_1, ..., y_T)'$. Clearly

$$
f(y; \Theta) = \int_{\mathcal{X}^T} f(y|\alpha; \Theta) f(\alpha; \Theta) d\alpha,
\tag{15}
$$

where $\alpha = (\alpha_1, ..., \alpha_T)'$. The analytic form of $f(y|\alpha; \Theta)$ and $f(\alpha; \Theta)$ can be deduced from (14), but the integral in (15) can only be solved numerically due to the nonlinear and, in case of the SV-t model, non-Gaussian nature of the model. Our approximation of $f(y; \Theta)$ by importance sampling is given by

$$
\widehat{f(y; \Theta)} = \frac{1}{N} \sum_{i=1}^{N} w_i, \qquad \text{where} \qquad w_i = w(\alpha^i),
\tag{16}
$$

with

$$
w(\alpha) = \frac{f(y|\alpha; \Theta) f(\alpha; \Theta)}{g(\alpha|y; \Theta)} \quad \text{and} \quad \alpha^i \sim g(\alpha|y; \Theta).
$$

It is assumed that $g(\alpha|y; \Theta)$ is strictly positive for all $\alpha \in \mathcal{X}^T$. We take the importance sampler $g(\alpha|y; \Theta)$ as the Laplace approximation to the posterior of $\alpha_1, \ldots, \alpha_T$ given the data $y_1, \ldots, y_T$;

14

Θ	full sample: 3/1/85 – 31/7/07		1/3 sample: 2/1/02 – 31/7/07	
	SV-N	SV-t	SV-N	SV-t
μ_y	0.0662	0.0636	0.0472	0.0481
	(0.0470 0.0854)	(0.0448 0.0825)	(0.0090 0.0854)	(0.0097 0.0864)
ϕ	0.983	0.993	0.993	0.994
	(0.978 0.987)	(0.988 0.996)	(0.981 0.997)	(0.981 0.998)
σ_η	0.0226	0.0077	0.0088	0.007
	(0.0183 0.0278)	(0.0050 0.0120)	(0.0044 0.0177)	(0.0033 0.0161)
μ_α	-0.367	-0.309	-0.434	-0.402
	(−0.600 −0.135)	(−0.663 0.0459)	(−1.125 0.258)	(−1.258 0.454)
ν	-	7.79	-	24.68
		(6.35 9.71)		(6.40 118.79)
log-lik	-7,270.0	-7,222.8	-1,743.7	-1,742.4

Table 3: *Simulated maximum likelihood estimates of the parameters in Θ for the SV-N model (14) with ε_t normally distributed and the SV-t model (14) with ε_t distributed as Student's t. Estimation results are presented for two samples. The values in parentheses give the asymptotic (asymmetric) 95% confidence intervals. The log-lik refers to the value of the likelihood function (16) in logs and evaluated at $\Theta = \widehat{\Theta}$.*

see, e.g., Shephard and Pitt (1997) and Robert and Casella (2004, p.107). The initial value of α_1 is a draw from the unconditional distribution of α_t. The maximum likelihood estimator of Θ is referred to as $\widehat{\Theta}$. When maximising the log-likelihood function, we use $N = 1,000$ draws where the same underlying uniform random variables are used for each likelihood evaluation.

In our empirical study we use the return series y_t of daily Standard & Poor's 500 stock index closures. The historical return series is for the period 3rd January 1985 to 31st July 2007 and was obtained from Yahoo Finance. Our sample consists of 5,693 daily observations. Importance sampling is being carried out over 5,693 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 500 index on day t.

5.1 Test statistics and graphical diagnostics

We start by considering the SV-N model whose estimated parameters are reported in Table 3 together with their 95% confidence intervals. We estimated both models on two different sample sizes, the full sample and a smaller sample corresponding to 2nd January 2002 through 31st July 2007. To illustrate our diagnostic procedure, we take the simulated maximum likelihood estimate $\Theta = \widehat{\Theta}$ as fixed. Conditional on this estimate, we increase the simulation size from
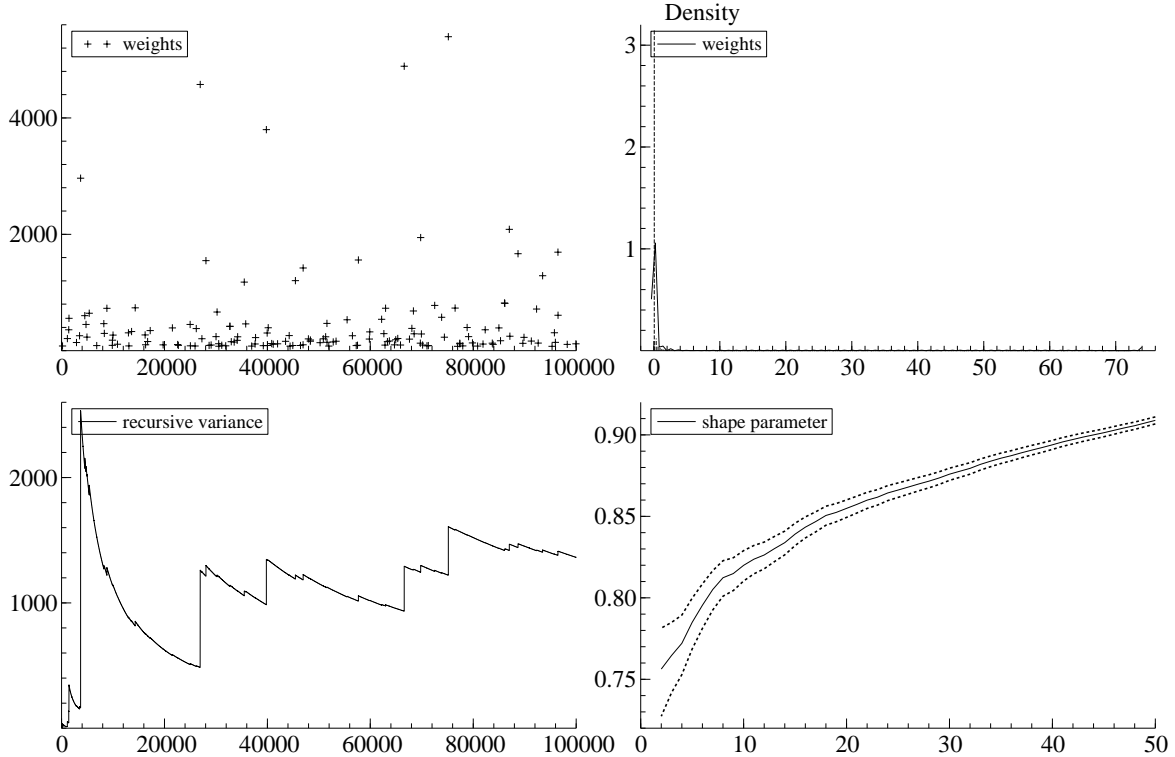
Figure 3: *Importance sampling diagnostic graphics for the SV-N model. (i) Largest* 100 *importance weights; (ii) histogram of the remaining weights; (iii) recursive variance estimator of the weights; (iv) estimated shape parameters with* 95% *confidence intervals (dotted lines) for* 50 *different thresholds going from the largest* 1% *to* 50% *of the* N^+ *weights, that is* $k^* = j \cdot N^+/100$ *for* $j = 1, \ldots, 50$.

$N = 1,000$ to $N^+ = 100,000$ and produce a set of diagnostic graphs. The question we ask is if the variance of the weights exists at $\Theta = \widehat{\Theta}$, basing the diagnostic graphics on these *i.i.d.* $100,000$ draws. The largest 100 importance weights are presented in Figure 3 together with a histogram of the remaining $99,900$ smaller weights. The third graph presents recursive estimates of the variance of the importance weights. The final graph presents estimates of ξ along with an asymmetric 95% confidence interval for different thresholds u ranging from the largest 1 percent to 50 percent of the N^+ importance weights.

These diagnostic graphs provide evidence that the importance sampler is unreliable as they indicate a maximum likelihood estimate of ξ which is larger than 0.5. The variance of the importance weights does not exist for the SV-N model. This result is in line with the more informal assessment given in Figure 3(iii). This records the recursive variance estimator, which contains several large jumps. Standard asymptotics to measure the uncertainty of the importance sampler's estimate of the log-likelihood function is problematic in this case.
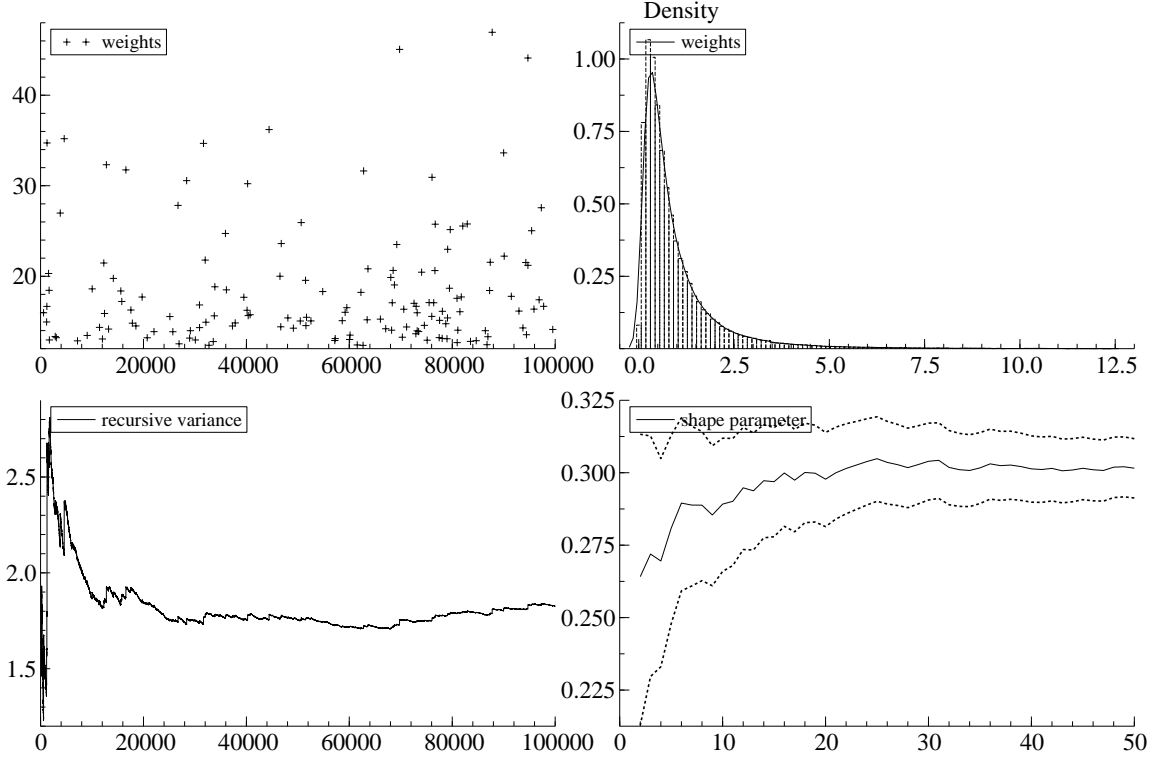
16

Figure 4: *Importance sampling diagnostic graphics for the SV-t model. (i) Largest 100 impor-*
tance weights; (ii) histogram of the remaining weights; (iii) recursive variance estimator of the
weights; (iv) estimated shape parameters with 95% confidence intervals (dotted lines) for 50
different thresholds going from the largest 1% to 50% of the N^+ weights, that is $k^* = j \cdot N^+/100$
for $j = 1, \ldots, 50$.

We now turn our attention to the empirically more interesting SV-t model and look at the
diagnostic graphs of the model with estimated parameters. The estimated parameter values are
provided in Table 3 and show a large improvement in the fit of the model over the full sample.
Figure 4 presents the same diagnostic graphs as in Figure 3 for the SV-t model. It becomes clear
by comparing these graphs that a better behaved importance sampler is obtained for the SV-t
model. In particular, the estimated shape parameters provide evidence that two moments exist
for the importance weights. When these diagnostic graphs are compared with the same graphs
for the estimated SV-N model, we conclude that the importance sampler for the SV-t model is
more reliable.

In Table 4, we present the test statistics for both SV models. The test statistics are computed
for two different thresholds where we have chosen u so that we have the 10,000 and 40,000 largest
weights, respectively. The score and t-value statistics are signed tests for the null hypothesis
$H_0 : \xi = \frac{1}{2}$. We suspect that no variance exists for the importance weights when these tests

have significant positive values. Both tests are asymptotically standard normally distributed, so having a 95% critical value of 1.64. The LR test follows an asymptotic $0.5\left(\chi_0^2 + \chi_1^2\right)$ distribution with the 95% critical value 2.69.

	Stochastic volatility model					
	full sample		1/3 sample		1/7 sample	
	SV-N	SV-t	SV-N	SV-t	SV-N	SV-t
$k^* = 0.4N$						
score	128.44	-10.42	17.09	-16.87	-0.252	-21.80
t	1810.0	-27.29	44.44	-60.42	-0.577	-118.5
LR	20464.1	0	1078.0	0	0	0
M_{test}	20.01	-3.83	0.061	-5.65	0.232	-8.15
$k^* = 0.1N$						
score	38.33	-5.52	3.68	-7.42	-0.825	-9.49
t	59.51	-9.28	3.22	-21.01	-0.930	-49.01
LR	2755.6	0	57.46	0	0	0

Table 4: *Likelihood based tests for the null of the existence of the variance of the importance weights for two SV models estimated by simulated maximum likelihood. The tests are presented for three different samples and for two tuning values k^* from which the treshold $u \approx w_{(N-k^*)}$ is determined. The M_{test} was computed using the tuning value $k = 4N^{1/3}$. The 95% critical values for the score, t, LR, and M_{test} are 1.65, 1.65, 2.69, and 1.65 respectively.*

The results given in Table 4 are consistent across the choice of statistic. For the SV-N model, the results are poor with all the statistics strongly rejecting the existence of a variance for the full sample. As we reduce the number of observations and the dimension of the integral decreases, the importance function for the SV-N begins to perform better. Table 4 also includes test statistics computed from models estimated on roughly one-seventh the sample size (818 observations). At this point, the test statistics fail to reject the null hypothesis. Conversely, the test statistics indicate that the importance function for the SV-t model has a finite variance for all sample sizes.

The point estimates in Table 3 appear reasonable for the SV-N model, despite the fact that the weight function has an infinite variance. We believe users should still proceed with caution. Robert and Casella (2004, p.95) emphasize that when the variance does not exist the range of Monte Carlo estimates computed using different sequences of random numbers will be enormous. For an SV model, this means that the estimated log-volatilities may be far from the true, unobserved values. Robert (2007) demonstrates how an importance sampler with a poorly chosen importance density can entirely miss-estimate the true log-volatilities for many sequences of random numbers.

6 Conclusion

Importance sampling is increasingly used in econometrics to estimate likelihoods and to compute moments of the posterior distribution in Bayesian analyses. Behind its use lies the assumption that the variance of the importance sampling weights exists. In this paper, we have proposed test statistics and graphical diagnostics to check this assumption in high dimensional settings when it is often difficult to obtain analytical results. We have applied our new tests on empirical problems and find this assumption is often overturned. This shows the relevance of the problem of which researchers should be aware. The tests and the graphical diagnostic procedures will provide researchers with a better understanding of when their importance samplers are reliable.

Acknowledgements

References

An, S. and F. Schorfheide (2007). Bayesian analysis of DSGE models. *Econometric Reviews 26*, 113–172.

Asai, M. and M. McAleer (2006). Asymmetric multivariate stochastic volatility. *Econometric Reviews 25*, 453–473.

Brandt, M. W. and Q. Kang (2004). On the relationship between the conditional mean and volatility of stock returns: a latent VAR approach. *Journal of Financial Economics 72*, 217–257.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models.* New York: Springer Press.

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.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society, Series B 52*, 393–442.

Doornik, J. A. (2006). *Ox: Object Oriented Matrix Programming, 4.0.* London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space models. *Biometrika 84*, 669–684.

Durham, G. (2003). Likelihood-based specification analysis of continuous-time models of the short-term interest rate. *Journal of Financial Economics 70*, 463–487.

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

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.

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

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics.* New Jersey: Wiley 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.

Gonzalo, J. and J. Olmo (2004). Which extreme values are really extreme? *Journal of Financial Econometrics 2*, 349–369.

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.

Haeusler, E. and J. L. Teugels (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics 13*, 743–756.

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*, pp. 3–17. Dordrecht: Kluwer.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics 3*, 1163–1174.

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

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*. Forthcoming.

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.

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

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

Marshall, A. W. (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.

Monahan, J. F. (1993). Testing the behaviour of importance sampling weights. Computer Science and Statistics: Proceedings of the 25th Annual Symposium on the Interface, 112–117.

Monahan, J. F. (2001). *Numerical Methods of Statistics*. Cambridge: Cambridge University Press.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

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

Robert, C. P. (2007). Stochastic volatility: an experimental approach. CREST ENSAE, Paris, France.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (Second ed.). New York: Springer Press.

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. (Ed.) (2005). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press.

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

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika 72*, 67–90.

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