ANALYSIS OF AN IMPORTANCE SAMPLING IN A STOCHASTIC VOLATILITY MODEL

by

Qiang Sun

MS in Management, Zhejiang University, 1998

Submitted to the Graduate Faculty of

the the School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Qiang Sun

It was defended on

May 19th 2010

and approved by

Xinfu Chen, Dept. Mathematics

Ivan Yotov, Dept. Mathematics

Gunduz Caginalp, Dept. Mathematics

John Chadam, Dept. Mathematics

Steven Husted, Dept. Economics

Dissertation Director: Xinfu Chen, Dept. Mathematics

ANALYSIS OF AN IMPORTANCE SAMPLING IN A STOCHASTIC VOLATILITY MODEL

Qiang Sun, PhD

University of Pittsburgh, 2010

This thesis analyzes an importance sampling method whose effectiveness relies in many cases on the selection of sampler's parameters. In its typical application of a Taylor's stochastic volatility model, a new approach, referred to as 'universal importance sampling', was designed and shown to be much more efficient than those in the literature, such as the sequential importance sampling. One obvious advantage of the universal sampling is that the parameters selected do not rely on the sampling process, so that Monte Carlo simulations can be done on different computers with a final averaging.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am very grateful to my advisor Professor Xinfu Chen. He gave me the guidance and advice throughout the period of completion of my thesis. He also gave me the wonderful years of researching and teaching at the University of Pittsburgh. Here, I also want to thank to some other professors, Professor Ivan Yotov, Professor Gunduz Caginalp, Professor John Chadam and Professor Steven Husted for joining in the committee for my defense and overview. Many thanks go to my wife Ting Ting Huang for fruitful and encouraging discussions. I am also grateful to my department colleagues for the valuable comments on my research. Finally, I appreciated the support and encouragement from my parents throughout the entire years of working on these papers.

1.0 INTRODUCTION

Stochastic volatility (SV) models get much concerned in mathematical finance and econometrics. In econometrics, researchers have more interests in GARCH or ARCH-type models to investigate volatility. In mathematical finance, discreet and continuous stochastic volatility models have been playing more outstanding roles. Since 90's, a great number of researchers concerned about the behavior of the stochastic volatility, and its application to the pricing of the financial derivatives or the other financial assets. One of the difficulties is to how to capture the different stochastic process of volatilities and do the estimations.

In this paper, like many other related researchers, we consider Taylor's stochastic volatility model. The volatility is modeled by a latent stochastic process. We review the joint density function here to find the basic definition and the total integral over T-folds. Then we will evaluate the log-likelihood function numerically by integrating out the latent volatility variables. Because it is impossible to compute the high-dimensional integral by conventional quadrature, we will use Importance Sampling to solve the problem. Finally parameter estimations are followed by numerically maximization of the log-likelihood functions. Let's examine Taylor's SV model as the following:

$$r_t = \beta e^{v_t/2}\varepsilon_t, \quad \varepsilon_t \sim N(0,1), \tag{1.0.1}$$

$$v_t = \delta v_{t-1} + \upsilon\eta_t, \quad \eta_t \sim N(0,1), \tag{1.0.2}$$

where r_t is observable time series of financial return, v_t a is latent factor following a AR(1) process, and ε_t and η_t are serially and mutually independent Gaussian random variables, which follows [6, 5].

In Taylor's SV model, the observable variable $\mathbf{r} = \{r_t\}_{t=1}^T$ and auxiliary unobservable (artificial or latent) variables $\mathbf{V} = \{v_t\}_{t=1}^T$ are introduced, along with parameters Λ. Let T be the number of period of interest and being observed, $\widetilde{\rho}(v_1, ..., v_T, r_1, ..., r_T|\Lambda)$ be the joint distribution of $(v_1, ..., v_T, r_1, ..., r_T)$. Then the marginal distribution density of $(r_1, ..., r_T)$ is given by

$$\rho(\mathbf{r}|\Lambda) \equiv \int_{\mathbb{R}^T} \widetilde{\rho}(v_1, ..., v_T, r_1, ..., r_T|\Lambda) dv_1 \cdots dv_T, \tag{1.0.3}$$

Given an observation value $(r_1, ..., r_T)$, for the model will then be obtained, the parameter, Λ^*, if maximum likelihood method is used, by

$$\Lambda^* = \arg\max_{\Lambda} \log \rho(r_1, ..., r_T|\Lambda).$$

In many applications, analytic evaluation of the integral in (1.0.3) is impossible, so numerical evaluation, with given values of $(r_1, ..., r_T|\Lambda)$, are needed here. We will investigate and clearly derive the proofs and results of the Monte-Carlo approximation for high dimensional integrals. For notational simplicity, we consider

$$\text{numerical evaluation of } \iint_{\mathbb{R}^m} \varphi(\mathbf{v}) d\mathbf{v}$$

where φ is a known function. In terms of (1.0.3), we also have

$$m = T, \varphi(v) = \widetilde{\rho}(v, r, \Lambda), (r, \Lambda) \text{ are numerically given.}$$

We use m instead of T to indicate that time is irrelevant, as far as only integration is concerned. Later on when sequential integration are considered, we change m back to T .

The Monte Carlo method is to construct estimator to evaluate numerically the

integral $\int\int_{\mathbb{R}^m} f(\eta)d\eta$ by simulations. Let's consider a sequence of $i.i.d$ random draws from random variable (r.v.) η with the probability density $\rho(\cdot)$. After a sequence $\eta^{(1)}, \eta^{(2)}, ...$ of $i.i.d$ draws with $\rho(\cdot)$ are generated, by Law of Large numbers, we have

$$\int\int_{\mathbb{R}^m} f(\eta)\rho(\eta)d\eta \cong \frac{1}{N}\sum_{i=0}^{N} f(\eta^{(i)}),$$

with the convergence rate of $O(\frac{1}{\sqrt{N}})$.

Importance Sampling is one of the Monte Carlo methods to solve the above problem of numerically evaluating $\int\int_{\mathbb{R}^m} f(\eta)d\eta$. First, we design some probability density $\rho(., a)$ with some parameter a; Second, we do the transformation as

$$\int\int_{\mathbb{R}^m} f(\eta)d\eta = \int\int_{\mathbb{R}^m} \left\{\frac{f(\eta)}{\rho(\eta, a)}\right\} \rho(\eta, a)d\eta,$$

with the optimal parameter a is selected according to some criteria;Finally use the estimator to evaluate the likelihood as

$$\int\int_{\mathbb{R}^m} f(\eta)dx \cong \frac{1}{N}\sum_{i=0}^{N} \frac{f(\eta^{(i)})}{\rho(\eta^{(i)}, a)},$$

With respect to Importance Sampling applications to the estimation of SV models, the following is the partial list of the reference including Geweke(1989), Danielsson and Rechard(1993), Shephard and Pitt(1997), Durbin and Koopman(1997), Liesenfeld and Richard(2003), Richard and Zhang(2007). Geweke (1989) firstly discusses minimization of Variance of $\log \rho(r_1, ..., r_T|\Lambda)$ by explicit procedures within specific classes of fat-tail densities, typically multivariate student -t densities and skewed generalizations labeled split-t densities. Durbin and Koopmans (1997) apply IS to evaluate the likelihood function of non-Gaussian state space models. They showed that the selection of an importance sampler can be approached via the construction of an operational approximation to a complex model. They verified that this method is applicable in significantly higher dimensions than the other alternative methods. Owens and Zhou (2000) discuss various improvements of the IS technique which are

well fit for low-dimensional applications. They extend the theoretical results to more general multiple and mixture samplers and describe conditions under which estimated coefficients approach the true ones. Richard and Zhang (2007) describes a sequential and efficient Importance Sampling Monte Carlo (MC) procedure for the evaluation of high-dimensional numerical integrals, based upon a sequence of auxiliary weighted regressions which actually are linear under appropriate conditions. Their method can be used to evaluate likelihood functions and ML estimators for models which have unobservable variables.

The purpose of this paper is to develop a new importance sampling method, Universal Importance Sampling. Its computational cost and efficiency are much improved relative to that of the methods in the previous literatures. Our method is built on the works of Durbin and Koopmans(1997), Shephard and Pitt(1997), Richard and Zhang(2007). Comparing to those previous methods, UIS has the following advantages,

1. The estimator from the UIS is unbiased;
2. The "universal importance sampler" is obtained by solving an algebraic system which by the Newton's iteration only takes a fraction of second of computing time; on the other hand, the previous method, sequential importance sampling depends on random draws and takes a multiple of the time needed for an actual Monte-Carlo integration based on a fixed sampler;
3. The universal importance sampler can be used in parallel computation since the sampler is universal; The SIS may not be able to do that since the "optimal sampler" depends on the collection of all the common random numbers;
4. Numerical simulation shows that the UIS is at least 10 times faster than SIS; for the particular example at hand, the total random draws can be as large as 1,000,000,000 for UIS whereas for SIS, the maximum number is about 100,000, partially due to the limitation of virtual memory (2GB).

The rest of the paper is organized as follows. Section 2 reviews Monte-Carlo

integration methods and sequential importance sampling. Section 3 gives a general analysis of how to use previous sequential importance sampling method to evaluate the likelihood function in the stochastic volatility model .Section 4 presents how the new importance sampling method UIS are designed based on the analysis. Then we do the comparisons of the results from two methods. Section 5 concludes this paper.

2.0 MONTE–CARLO INTEGRATION AND SEQUENTIAL IMPORTANCE SAMPLING

Monte–Carlo Integration (MCI) is a numerical algorithm used to evaluate

$$\int_{\mathbb{R}^T} f(x)\,dx =: \mu \quad . \tag{2.0.1}$$

In this section, we review the basic MCI and its recent development by Geweke (1989) [6], Fishman (1996) [5], Owen and Zhou (2000) [13], and Richard and Zhang (2007) [15], etc..

2.0.1 Monte–Carlo Integration

The basic idea of MCI is to introduce a T-dimensional random variable X with a designed probability density function ρ and to write the integral in (2.0.1) as

$$\mu = \int_{\mathbb{R}^T} \frac{f(x)}{\rho(x)}\,[\rho(x)\,dx] =: \mathbf{E}[\varphi(X)], \qquad \varphi(x) := \frac{f(x)}{\rho(x)}. \tag{2.0.2}$$

According to the law of large numbers, the expectation μ of $\varphi(X)$ can be approximated by the sample mean

$$\mu_n(X^1, \cdots, X^n) := \frac{1}{n}\sum_{i=1}^{n} \varphi(X^i) \tag{2.0.3}$$

where $\{X^i\}_{i=1}^n$ is a set of random draws of X. We highlight the method as follows:

Monte–Carlo Integration (MCI)

0. Design a probability density ρ and a random number generator for X that has density ρ.
1. Generate random draws $\{X^i\}_{i=1}^n$ from the random number generator for X.
2. Use the sample mean $\mu_n := \mu_n(X^1, \cdots, X^n)$ in (2.0.3) as an approximation of μ.

Remark 2.0.1. *Theoretical and Numerical Technicalities.*

(1) One criterion in choosing a good probability density function ρ is the closeness of the function $\varphi(x) := f(x)/\rho(x)$ to a constant. Another criterion is the convenience of producing random numbers with the chosen density ρ, from available software packages. For the second criteria, we recall that one dimensional random variables can be converted each other via a standard transformation. For example, if X is a 1-D random variable with a cumulative distribution function (cdf) $F(x)$, then $U := F(X)$ is uniformly distributed on $[0,1]$. In other words, if U is uniformly distributed on $[0,1]$, then $X = F^{-1}(U)$ is a random variable with density $\rho := F'$.

(2) For consistency, the law of large number [18, p323] states that

$$\mathbf{E}[|\varphi(X)|] = \int_{\mathbb{R}} |f(x)|dx < \infty \quad \Longrightarrow \quad \lim_{n\to\infty} \mu_n(X^1, \cdots, X^n) = \mu \quad \textit{in probability.}$$

For accuracy, let σ be the standard deviation and σ^2 the variance of $\varphi(X)$:

$$\sigma^2 := \mathbf{V}[\varphi(X)] = \int_{\mathbb{R}^T} \frac{f^2(x)}{\rho(x)} dx - \Big(\int_{\mathbb{R}^T} f(x)dx \Big)^2.$$

Assume $\sigma < \infty$ and regard $X^1, \cdots, X^n$ as i.i.d random variables. Then for μ_n in (2.0.3),

$$\mathbf{E}[\mu_n] = \mu, \quad \mathbf{V}[\mu_n] = \frac{\sigma^2}{n}, \quad \mathrm{Std}[\mu_n] = \sqrt{\mathbf{V}[\mu_n]} := \frac{\sigma}{\sqrt{n}}.$$

Indeed, the central limit theorem [18, p324] states that for each $z > 0$,

$$\lim_{n\to\infty} Prob\Big(\big|\, \mu_n - \mu \big| > \frac{\sigma z}{\sqrt{n}} \Big) = \frac{2}{\sqrt{2\pi}} \int_z^\infty e^{-s^2/2} ds.$$

Note that σ can be approximated by the sample standard derivation of $\{\varphi(X^i)\}_{i=1}^n$:

$$\sigma_n = \sigma_n(X^1, \cdots, X^n) := \Big\{ \frac{1}{n-1} \sum_{i=1}^{n} \Big(\varphi(X^i) - \mu_n\Big)^2 \Big\}^{1/2}.$$

Hence, to describe the accuracy of the MCI, it is informative to write the output as

$$\mu = \mu_n \pm \frac{\sigma_n}{\sqrt{n}} \quad .$$

(3) In certain extensions of MCI, $\varphi(\cdot)$ depends on the random draw $\{X_i\}_{i=1}^n$ so we write it as $\varphi[X^1, \cdots, X^n](\cdot)$. Analytical investigation on the variance of μ_n may be quite complicated. To estimate the error, we may repeat a certain number of MCIs to generate needed statistics. In other words, instead of performing one MCI using n samples to produce one approximation, μ_n, for μ, we perform n_e ($\geqslant 2$) number of MCIs each of which using n_s ($= n/n_e \geqslant 1$) samples to produce n_e number of approximations, $\{\mu_{n_s}^{(j)}\}_{j=1}^{n_e}$, of which the sample mean and sample standard deviation can be used to generate an informative approximation for μ. More precisely, we compute

$$\begin{cases} \mu_{n_s}^{(j)} := \dfrac{1}{n_s} \displaystyle\sum_{i=1}^{n_s} \varphi[X^{(j-1)n_s+1}, \cdots, X^{jn_s}](X^{i+(j-1)n_s}), \qquad j = 1, \cdots, n_e, \\ \\ \mu_{n_s,n_e} := \dfrac{1}{n_e} \displaystyle\sum_{j=1}^{n_e} \mu_{n_s}^{(j)}, \quad \sigma_{n_s,n_e} := \sqrt{n_s} \Big\{ \dfrac{1}{n_e - 1} \displaystyle\sum_{j=1}^{n_e} \Big(\mu_{n_s,n_e} - \mu_{n_s}^{(j)}\Big)^2 \Big\}^{1/2}, \end{cases}$$

and express an informative numerical approximation of μ as

$$\mu = \mu_{n_s,n_e} \pm \frac{\sigma_{n_s,n_e}}{\sqrt{n_s n_e}}.$$

Theoretically, if $X^1, \cdots, X^{n_s n_e}$ are regarded as i.i.d, then $\mu_{n_s}^{(1)}, \cdots, \mu_{n_s}^{(n_e)}$ are i.i.d and

$$\mathbf{V}[\mu_{n_s,n_e}] = \frac{1}{n_e}\, \mathbf{V}[\mu_{n_s}^{(j)}] = \frac{\mathbf{E}[\sigma_{n_s,n_e}^2]}{n_e\, n_s} \quad \forall\, j \ .$$

Example 1. *Let $X \sim N(0,1)$ and $Y = \ln X^2$. Using MCI evaluate*

$$A := \mathbf{E}[Y] = \int_{\mathbb{R}} \ln x^2 \, \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx, \qquad B := \mathbf{V}[Y] = \int_{\mathbb{R}} \Big(\ln x^2 - A \Big)^2 \, \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

Solution. Choose integers $n_s \geqslant 1$ and $n_e \geqslant 2$ and set $n = n_s n_e$. Let $\{X^i\}_{i=1}^n$ be independent random draws from $N(0,1)$. Set $\varphi^i = \ln[\varepsilon^2 + (X^i)^2]$ where ε is a small number, say $\varepsilon = 10^{-10}$, introduced to avoid possible overflows. Evaluate the following quantities:

$$A_n := \sum_{i=1}^{n} \frac{\varphi^i}{n}, \qquad B^i := \frac{n[\varphi^i - A_n]^2}{n-1}, \qquad B_n := \sum_{i=1}^{n} \frac{B^i}{n}, \qquad A_{n_s}^{(j)} := \sum_{i=1}^{n_s} \frac{\varphi^{i+(j-1)n_s}}{n_s},$$

$$\sigma_n^A := \sqrt{B_n}, \qquad \tilde{\sigma}_n^B := \Big\{ \sum_{i=1}^{n} \frac{(B^i - B_n)^2}{n-1} \Big\}^{1/2}, \qquad B_{n_s}^{(j)} := \sum_{i=1}^{n_s} \frac{(\varphi^{i+(j-1)n_s} - A_{n_s}^{(j)})^2}{n_s - 1},$$

$$A_{n_s,n_e} := \sum_{j=1}^{n_e} \frac{A_{n_s}^j}{n_e} \Big(= A_n \Big), \qquad B_{n_s,n_e} := \frac{1}{n_e} \sum_{i=1}^{n_e} B_{n_s}^{(j)},$$

$$\sigma_{n_s,n_e}^A := \Big\{ \sum_{j=1}^{n_e} \frac{(A_{n_s,n_e} - A_{n_s}^{(j)})^2}{n_e - 1} \Big\}^{1/2}, \qquad \sigma_{n_s,n_e}^B := \Big\{ \sum_{j=1}^{n_e} \frac{(B_{n_s,n_e} - B_{n_s}^{(j)})^2}{n_e - 1} \Big\}^{1/2}.$$

The output of a numerical experiment is summarized in Table 1.

Table 1: Numerical Simulation for Example 1

		A			B				
N	A_n	$\sigma_n^A/\sqrt{n}$	$\hat{\sigma}_{n_s}^A/\sqrt{n_e}$	$\lvert A - A_n\rvert$	B_n	$\hat{B}_n$	$\sigma_n^B/\sqrt{n}$	$\hat{\sigma}_{n_s}^B/\sqrt{n_e}$	$\lvert B - \hat{B}_n\rvert$
4	−1.5650	0.75421	0.79256	0.29464	2.2753	2.1567	0.75421	2.15640	2.65950
16	−2.1935	0.65658	0.43916	0.92311	6.8975	7.8505	0.65658	2.19390	1.96270
64	−2.0869	0.30118	0.33556	0.81653	5.8055	5.6303	0.30118	1.66820	0.87066
256	−1.2627	0.13508	0.17154	0.00765	4.6708	4.4919	0.13508	0.46736	0.26400
1024	−1.3621	0.07214	0.07120	0.09174	5.3297	5.3340	0.07214	0.41793	0.39488
4096	−1.2566	0.03431	0.03845	0.01381	4.8221	4.8029	0.03431	0.17003	0.11267
16384	−1.3012	0.01758	0.01764	0.03085	5.0611	5.0608	0.01758	0.09168	0.12629
65536	−1.2704	0.00862	0.00858	0.00007	4.8728	4.8730	0.00862	0.04630	0.06196
262144	−1.2779	0.00435	0.00437	0.00755	4.9698	4.9697	0.00435	0.02357	0.03497
1048576	−1.2704	0.00217	0.00223	0.00006	4.9447	4.9445	0.00217	0.01173	0.00994
∞		−1.2703628454614781700					4.9348022005446793094		

Note: The last row is the true value of $B = \pi^2/2$ and $A = -\gamma - \ln 2$ where γ is the Euler constant. Note the $\sigma_n^A/\sqrt{n}$ and $\hat{\sigma}_{n_s}^A/\sqrt{n_e}$ approach to be the same, and B_n and $\hat{B}_n$ approach to be the same as N gets larger. We can see that the choice of n_e does not affect the asymptotic results, so, for simplicity, we can choose $n_e = \sqrt{N}$.

From Table 1, one may observe the following:

(1) The MCI relies crucially on the quality of the software that generates random numbers. Here we are "fortunate" to catch a "rare" event to demonstrate clearly the

statistical nature of the Monte-Carlo simulation: when $n = 64$, $B_n - B$ *is "unusually" large.*

(2) When n_e *changes from small to large (keeping* $n = n_s n_e$ *fixed), there is no fundamental change in* $\sigma_{n_s,n_e}/\sqrt{n}$, *which represents one standard deviation of the numerical approximation to the true value. That is, when there is no good way to estimate* σ_n, *one can, instead of performing one MCI using* n *samples, perform* n_e *MCIs each of which using* n_s *samples, and use the scaled sample standard deviation* $\sigma_{n_s,n_e}/\sqrt{n_s n_e}$ *as an estimation for the true one.*

2.0.2 Importance Sampling

The effectiveness of MCI for (2.0.1) depends on the statistical behavior of the random variable $\varphi(X) := f(X)/\rho(X)$ where X is a random variable with density ρ. In practice quite often one chooses the density from a family $\{\rho(\cdot, a)\}_{a\in\mathcal{A}}$ of density functions. Here A is a parameter set and for each parameter a in the set A, $\rho(\cdot, a)$ is a probability density function. Note that

$$\mu := \int_{\mathbb{R}^T} f(x)\,dx = \int_{\mathbb{R}^T} \frac{f(x)}{\rho(x,a)}\,[\rho(x,a)\,dx] = \mathbf{E}[\varphi(X_a, a)], \qquad \varphi(x,a) := \frac{f(x)}{\rho(x,a)}.$$

Here and in the sequel, X_a denotes a random variable with probability density $\rho(\cdot, a)$. If a parameter $a \in A$ is chosen, μ can be approximated by

$$\mu_n(a, X_a^1, \cdots, X_a^n) := \frac{1}{n}\sum_{i=1}^{n} \varphi(X_a^i, a)$$

where $\{X_a^i\}_{i=1}^n$ are random draws from a random number generator for X_a.

Importance Sampling (IS) is a special technique used for Monte–Carlo integration which selects an "optimal" probability density function $\rho(\cdot, a^*)$ from a carefully designed family $\{\rho(\cdot, a)\}_{a\in A}$.

While designing a good probability density family is important and in general very hard, setting up a criterion for optimality can be quite delicate. Quite often one uses the following [6, 15, 16]:

$$a^* := \operatorname*{argmin}_a \ \min_c \int_{\mathbb{R}^T} \Big[Q(\varphi(x,a)) - c\Big]^2 \omega(x,a)\rho(x,a)dx \qquad (\mathbf{C})$$

where $\omega > 0$ is a weight function. Typical choices of Q and ω are

$$Q(t) = t, \qquad Q(t) = \ln t, \qquad Q(t) = \cosh(\lambda\, t), \qquad \omega(x,a) \equiv 1, \qquad \omega(x,a) = \varphi(x,a).$$

In general, analytical evaluation of a^* from (C) is quite difficult since it may involve integrals that are more complicated than the original integral (2.0.1). One common practice is to use MCI to evaluat the integrals in (C) and perform an appropriate minimization via an iteration process. For an optimization that involves random draws (needed by the MCI), usually it is necessary to use common random numbers (CRNs), explained as follows.

Let U be a random variable whose random draws can be generated from a standard software package. Assume that there exists a smooth function $\Phi(\cdot,\cdot)$ such that for each $a \in A$, the random variable $X_a := \Phi(U,a)$ has the distribution density $\rho(\cdot,a)$. Now let $\{U^i\}_{i=1}^n$ be n random draws of U. Consider the family $\{\{X_a^i\}_{i=1}^n\}_{a\in\mathcal{A}}$ defined by

$$X_a^i := \Phi(U^i,a) \quad \forall\, i = 1,\cdots,n,\ a \in \mathcal{A}. \qquad (2.0.4)$$

For each $a \in A$, $\{X_a\}_{i=1}^n$ can be considered as random draws of $X_a := \Phi(U,a)$ and it varies continuously with respect to the parameter a. For a functional that depends only on $\{X_a^i\}_{i=1}^n$, numerical optimizations with respect to a can be performed, quite often, with stability.

When random draws of every X_a in a family $\{X_a\}_{a\in\mathcal{A}}$ are produced from a single set of random draws $\{U^i\}_{i=1}^n$, we call $\{U^i\}_{i=1}^n$ and $\{\{X_a^i\}_{i=1}^n\}_{a\in\mathcal{A}}$ the common random numbers (CRNs).

Once CRNs are generated, the criterion (C) can be implemented numerically as

$$a^* := \operatorname*{argmin}_a \min_c \sum_{i=1}^{n} \Big(Q(\varphi(X_a^i, a)) - c \Big)^2 \omega(X_a^i, a) \, . \tag{C1}$$

In [15], Researchers introduced the following (not necessarily equivalent to the above):

$$a^* := \lim_{k\to\infty} a^{(k)}, \qquad a^{(k+1)} := \operatorname*{argmin}_a \min_c \sum_{i=1}^{n} \big\{ Q(\varphi(X_{a^{(k)}}^i, a)) - c \big\}^2 \omega(X_{a^{(k)}}^i, a^{(k)}). \tag{C2}$$

Using importance sampling, the Monte-Carlo integration can be described as follows:

Importance Sampling Monte–Carlo Integration (ISMCI)

0. Pick a random variable U and design a family $\{\rho(\cdot, a)\}_{a\in A}$ of probability densities with a function Φ such that $X_a := \Phi(U, a)$ has density $\rho(\cdot, a)$. Also design Q and ω.

1. Generate independent random draws $\{U^i\}_{i=1}^n$ of U.

2. Find an "optimal" value a^* from (**C**), (**C1**) or (**C2**), where $\{X_a^i\}$ are produced from (2.0.4).

3. The output of an informative numerical approximation for μ is $\mu_n \pm \sigma_n/\sqrt{n}$ where

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \varphi(X_{a^*}^i, a^*), \quad \sigma_n := \Big\{ \frac{1}{n-1} \sum_{i=1}^{n} [\varphi(X_{a^*}^i, a^*) - \mu_n]^2 \Big\}^{1/2}.$$

Remark 2.0.2. *(1) The "optimal" parameter a^* obtained from criterion (C1) or (C2) depends on the CRNs used for MCI and hence is local or non-universal. The "optimal" parameter a^* from the criterion (C), on the other hand, is universal in the sense that it does not depend on the CRNs. A universal parameter, albeit can be found, has the obvious advantages: (i) it is efficient for repeated MCIs, say, for various different sizes of samplings; (ii) in error analysis, $\{\varphi(X_{a^*}^i, a^*)\}_{i=1}^n$ can be regarded as i.i.d random variables so central limit theorem applies.*

(2) When a^* depends on $\{U^i\}_{i=1}^n$, the quantity $\sigma_n/\sqrt{n}$ may not be a good approx-

imation of the true standard deviation of μ_n. If this is the case, one may consider to use the sample standard deviation of a set of MCI approximations; see Remark 2.0.1 (3).

Example 2 ([15]). *Let $\delta > 0$ and consider the use of the probability density family $\{\rho(x,a)\}_{a\in\mathcal{A}}$ with $\rho(x,a) = ae^{-ax}1_{\{x\geqslant 0\}}$ and $A = (0,\infty)$ to evaluate $\mu = \int_0^\infty e^{-x^\delta}dx$.*

(1) The MCI Method. Let U be a random variable that is uniformly distributed on $[0,1]$ and

$$\varphi(x,a) = \frac{e^{-x^\delta}}{a\,e^{-ax}}, \qquad X_a = -\frac{\ln U}{a} \qquad \forall\, a > 0.$$

Then X_a is a random variable with density $\rho(\cdot,a)$ and $\mu = E[\varphi(X_a,a)]$. If a is a selected parameter and $\{U^i\}_{i=1}^n$ are random draws of U, μ is approximated by $\mu_n(a) \pm \sigma_n(a)/\sqrt{n}$ where

$$\mu_n(a) := \sum_{i=1}^n \frac{\varphi^i(a)}{n}, \quad \sigma_n(a)^2 := \sum_{i=1}^n \frac{[\varphi^i(a)-\mu_n(a)]^2}{n-1}, \quad \varphi^i(a) := \frac{\exp\Big(-[\frac{1}{a}\ln\frac{1}{U^i}]^\delta\Big)}{aU^i}.$$

(2) The Criteria. We list a number of criteria for the "optimal" parameter a^*.

1. $a^* = a_1^* := \underset{a}{\operatorname{argmin}}\, V[\varphi(X_a,a)] = \underset{a}{\operatorname{argmin}} \int_0^\infty e^{ax-2x^\delta-\ln a}dx$, *for $\delta > 1$.*

   When $\delta \in (0,1)$, the variance of $\varphi(X_a,a)$ is infinite so we may consider

   $$a^* = \hat{a}_1^* := \underset{a}{\operatorname{argmin}}\, \mathbf{E}[|\varphi(X_a,a)-\mu|] \qquad \forall\, \delta > 0.$$

2. $a^* = a_2^* := \underset{a}{\operatorname{argmin}}\, V[\ln\varphi(X_a,a)] = \left[\dfrac{\Gamma(1+2\delta)-\Gamma^2(1+\delta)}{\delta\Gamma(1+\delta)}\right]^{\frac{1}{\delta}}$, *for $\delta > 0$.*

3. $a^* = a_3^* := \{\delta\Gamma(1+\delta)\}^{1/\delta} = \lim_{k\to\infty} a^{(k)}$ *where* $a^{(k+1)} := \underset{a}{\operatorname{argmin}}\, V[\ln\varphi(X_{a^{(k)}},a),a)]$.

   Indeed, $a^{(k+1)} = (a^{(k)})^{1-\delta}\delta\Gamma(1+\delta)$ so $\lim_{k\to\infty} a^{(k)}$ exists if and only if $\delta \in (0,2)$.

4. $a^* = a_4^*(U^1,\cdots,U^n) := \lim_{k\to\infty} a^{(k)}$ *where* $a^{(k+1)} = \underset{a}{\operatorname{argmin}} \min_c \sum_{i=1}^n \{\ln\varphi(X^i_{a^{(k)}},a) -$

$c\}^2$ *One can show that* a_4^* *exists if and only if* $\delta \in (0,2)$ *and in such a case,*

$$a_4^*(U^1,\cdots,U^n) := \left\{ \frac{\sum_{i=1}^n (X_1^i)^\delta \,[X_1^i - c]}{\sum_{i=1}^n [X_1^i - c]^2} \right\}^{1/\delta}, \qquad c := \frac{1}{n}\sum_{i=1}^n X_1^i, \quad X_1^i := \ln\frac{1}{U^i}.$$

Asymptotically, one can verify that $\lim_{n\to\infty} a_4^*(U^1,\cdots,U^n) = a_3^*$.

(3) Numerics. To demonstrate the effectiveness of choosing the optimal parameter a*, we present our numerical results in Table 2 and Table 3, and Figure 1–2.*

Figure 1: The sample standard deviation curve on $a-\sigma$ coordinate system.

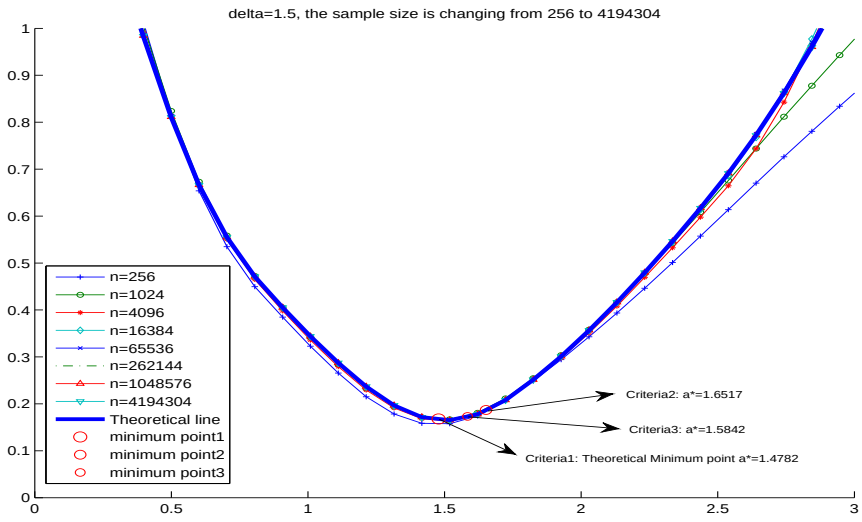


Figure 2: The sample standard deviation curve on $a - \sigma$ coordinate system.

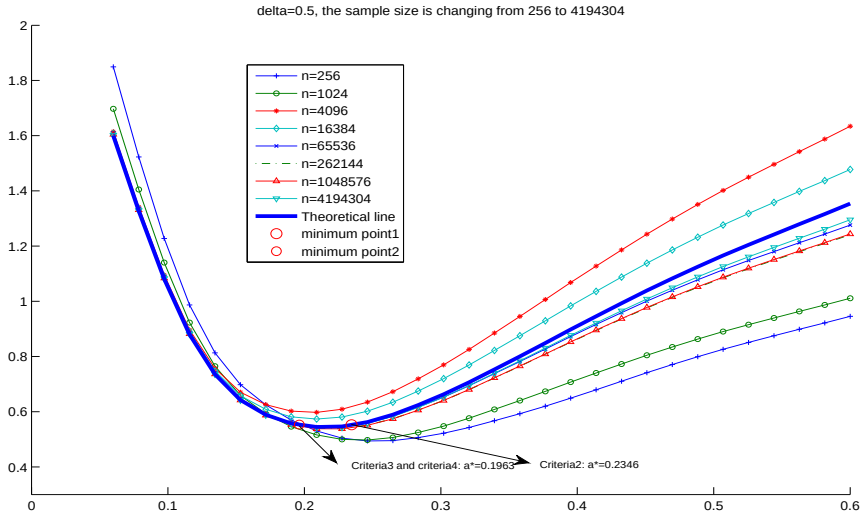


Table 2: Numerical Simulation for Example 2($\delta = 1.5$)

	$\delta = 1.5$								
	(1) $a^* = 1.4782$		(2) $a^* = 1.6517$		(3) $a^* = 1.5842$			(4)	
N	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	a^*	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$
4	0.9811 ± 0.04916	0.07831	0.8154 ± 0.05062	0.08733	0.8031 ± 0.06649	0.09960	1.8734	0.7656 ± 0.13599	0.13713
16	0.9089 ± 0.00573	0.00611	0.8513 ± 0.03409	0.05145	0.9300 ± 0.01774	0.02722	1.2718	0.9769 ± 0.01111	0.07418
64	0.9225 ± 0.00518	0.01978	0.8921 ± 0.00890	0.01060	0.8980 ± 0.00598	0.00470	1.5979	0.8844 ± 0.01236	0.01838
256	0.8996 ± 0.00273	0.00310	0.9077 ± 0.00350	0.00493	0.9057 ± 0.00318	0.00296	1.5666	0.8946 ± 0.00211	0.00814
1024	0.9081 ± 0.00098	0.00533	0.8983 ± 0.00098	0.00447	0.9105 ± 0.00096	0.00773	1.5454	0.9042 ± 0.00095	0.00146
4096	0.9015 ± 0.00035	0.00127	0.9040 ± 0.00038	0.00122	0.9032 ± 0.00031	0.00050	1.5762	0.9023 ± 0.00030	0.00049
16384	0.8988 ± 0.00011	0.00392	0.9009 ± 0.00013	0.00189	0.9032 ± 0.00011	0.00050	1.6011	0.9016 ± 0.00013	0.00118
65536	0.9030 ± 0.00004	0.00023	0.9040 ± 0.00005	0.00129	0.9021 ± 0.00004	0.00061	1.5761	0.9038 ± 0.00004	0.00103
262144	0.9030 ± 0.00001	0.00021	0.9035 ± 0.00002	0.00072	0.9029 ± 0.00001	0.00016	1.5858	0.9030 ± 0.00002	0.00030
1048576	0.9026 ± 0.00001	0.00016	0.9028 ± 0.00001	0.00005	0.9028 ± 0.00001	0.00002	1.5829	0.9030 ± 0.00001	0.00024
∞	0.90274529295093361129685								

*Note:*The last row is the true value of the pilot integral $\mu = \Gamma(1 + 1/\delta)$. Compared with all four criterium, all the methods are good for numerical approximation, as N approaching infinity, and the interval are really small. The choice of a^* does not affect the asymptotic results, and all the them approach to the true value. As extending to 3 S.D., the fourth method could be a good alternative to the third one, but it should be noticed that the optimal parameter is associated with the sample, so the sample variance is bigger than that of the third one, and the performance of variance is worse even N is big enough.

Table 3: Numerical Simulation for Example 2($\delta = 0.5$)

	$\delta = 0.5$								
	(1) $a^* = 1$		(2) $a^* = 0.2346$		(3) $a^* = 0.1963$			(4)	
N	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$	a^*	$\mu_n \pm \sigma_{n_s}/\sqrt{n_e}$	$\lvert\mu - \mu_n\rvert$
4	1.1985 ± 0.26627	0.80154	1.5422 ± 0.04345	0.45781	1.8801 ± 0.12308	0.11994	0.0890	2.5611 ± 0.08101	0.56108
16	1.4018 ± 0.17452	0.59824	1.8433 ± 0.02751	0.15667	1.8650 ± 0.05139	0.13497	0.2653	1.7632 ± 0.01510	0.23678
64	3.7060 ± 0.75878	1.70600	2.1222 ± 0.05349	0.12217	2.2109 ± 0.11678	0.21089	0.1893	1.9520 ± 0.03843	0.04802
256	1.6850 ± 0.04718	0.31504	1.9498 ± 0.01023	0.05017	2.0843 ± 0.01616	0.08431	0.1886	2.0168 ± 0.01652	0.01675
1024	2.0015 ± 0.08368	0.00153	2.0057 ± 0.00799	0.00572	1.9757 ± 0.00466	0.02427	0.1850	1.9678 ± 0.00597	0.03225
4096	1.5445 ± 0.00737	0.45548	1.9822 ± 0.00217	0.01783	2.0179 ± 0.00331	0.01789	0.1874	2.0183 ± 0.00334	0.01825
16384	1.6634 ± 0.00642	0.33658	1.9795 ± 0.00135	0.02045	2.0078 ± 0.00122	0.00779	0.1965	1.9868 ± 0.00070	0.01323
65536	1.7935 ± 0.00786	0.20647	1.9746 ± 0.00062	0.02536	1.9903 ± 0.00044	0.00966	0.1963	2.0002 ± 0.00060	0.00017
262144	1.7954 ± 0.00216	0.20458	1.9881 ± 0.00024	0.01195	2.0042 ± 0.00025	0.00421	0.1947	2.0051 ± 0.00030	0.00508
1048576	1.7614 ± 0.00075	0.23864	1.9994 ± 0.00014	0.00065	1.9994 ± 0.00006	0.00064	0.1965	1.9980 ± 0.00010	0.00196
∞	2.000000000000000000000000								

*Note:*The last row is the true value of the pilot integral $\mu = \Gamma(1 + 1/\delta)$. The equation in first method has no solution, so we arbitrarily let $a^* = 1$, so, explicitly, the first method performs poorly, when $\delta = 0.5$. Compared with all other four criterium, we can see that all of them are not good indeed, since the true value does not fall into any confidence interval of each criteria, as N approaching infinity, although the third is better for the sample variance is smaller than others.

In Figure 1 , $\delta = 1.5$. We plot the sample standard derivation curve $\sigma = \sigma_n(a)$ on the a-σ coordinate system. Curves displayed are for $n = 2^k$ with $k = 8, \cdots, 22$. All curves are smooth since CRNs are used for different a's. The thick dashing curve corresponds to the curve $\sigma = \sqrt{\mathbf{V}[\varphi(X_a, a)]} = \sqrt{\mathbf{E}[\sigma_n(a)^2]} = \sqrt{n\mathbf{E}[|\mu_n(a) - \mu|^2]}$ for all n.

In Figure 2 , $\delta = 0.5$ so $E[|\mu_n(a) - \mu|^2] = \infty$ and a_1^* is undefined. We plot the scaled L^1 norm $d = E[|\mu_n(a) - \mu|]/d_n^*$ with $d_n^* = \min_a E[|\mu_n(a) - \mu|]$ on the a-d plane. It seems that

$$a_{\text{opt}} := \operatorname*{argmin}_a \mathbf{E}[|\mu_n(a) - \mu|] \approx \frac{\hat{a}_1^*}{\ln(\ln n)} \ (n \geqslant 2^4), \qquad d_n^* := \min \mathbf{E}[|\mu_n(a) - \mu|] \approx \frac{\ln(\log_2 n)}{\sqrt{n}}.$$

Here the values $E[|\mu_n(a) - \mu|]$ displayed are indeed numerical approximations obtained from an average of $n_e = 32$ Monte-Carlo simulations. The dashing curve represents the function $d = E[|\varphi(X_a, a) - \mu|]/d_1^*$ which attains its minimum at $\hat{a}_1^* \approx 0.088$.

(4) Conclusion. It is quite clear that when $\delta > 1$, the MCI is not sensitive to the choice of a^*; for example, when $\delta = 1.5$, for any a between 1 and 2, the standard deviation of the resulting MCI is no bigger than twice of the optimal one. Nevertheless, when $0 < \delta < 1$, the performance of MCI is very sensitive to the choice of a ; namely, importance sampling is truly the key for efficiency.

For discussions from other points of view, see Owen and Zhou (2000) [13], and Richard and Zhang (2007) [15].

Remark 2.0.3. *The MCI is typically used for high space dimensions, i.e. for T large. When T is small, it is better to use integration quadrature rules. For instance,*

$$\int_0^\infty e^{-x^\delta} dx \overset{x=-\frac{1}{a}\ln\frac{1}{u}}{=\!=\!=} \int_0^1 \frac{e^{-[\frac{1}{a}\ln\frac{1}{u}]^\delta}}{au} du \approx \mu_R^n := \frac{1}{n}\sum_{i=1}^n \frac{e^{-[-\frac{1}{a}U_i]^\delta}}{aU_i}\Big|_{U_i=\frac{i}{n}}.$$

While error estimates from numerical quadrature rules are absolute (confidence level $= 100\%$), error estimates in MCI are statistical and confidence levels can never reach 100%.

2.0.3 Sequential Importance Sampling

Sequential Importance Sampling (SIS), a variation of IS, is a powerful technique developed by Geweke (1989) [6], and Fishman (1996) [5], for the Monte-Carlo integration of (2.0.1) where $T \gg 1$. In this method, a family $\{\rho(\cdot, a)\}_{a\in\mathcal{A}}$ of probability density functions with parameter $a = (a_1, \cdots, a_T)$ is to be designed and when a parameter $a^* \in A$ is selected, the integral μ in (2.0.1) is approximated by

$$\mu_n(a^*, U^1, \cdots, U^n) := \frac{1}{n}\sum_{i=1}^{n} \varphi(X^i_{a^*}, a^*), \qquad \varphi(x, a) := \frac{f(x)}{\rho(x, a)}, \quad X^i_a := \Phi(U^i, a),$$

where $\{U_i\}_{i=1}^n$ is a set of common random numbers drawn independently from U and Φ is a smooth function having the property that for each $a \in A$, $X_a = \Phi(U, a)$ has density $\rho(\cdot, a)$.

For the importance sampling, the "optimal" a^* is selected by solving a minimization problem having the same space dimension as that of A. In the sequential importance sampling, a fairly large dimension of the set A is introduced and a "sub-optimal", that is, not necessarily "optimal", parameter a^* is obtained by solving a sequence of minimization problems each of which involves only a very small parameter space dimension. The introduction of a large set A allows a good approximation of $f(x)$ by a constant multiple of certain density functions in the family $\{\rho(\cdot, a)\}_{a\in\mathcal{A}}$, and the sequential minimization reduces significantly both the theoretical and the numerical complexity. A good balance between the size of A and the degree of sub-optimality improves significantly the performance of SISMCI.

To transform the global minimization of the importance sampling to a sequential minimization of the sequential importance sampling, one designs a special decomposition of the form

$$\psi(x, a) := \ln \frac{f(x)}{\rho(x, a)} = \psi_0(a) + \sum_{t=1}^{T} \psi_t(x, a).$$

While the importance sampling uses optimal a that minimizes certain variance of ψ,

the sequential importance sampling only requires a_t to minimize certain variance of ψ_t, for each $t = 1, \cdots, T$. One version of a "suboptimal" $a^* = (a_1^*, \cdots, a_T^*)$ can be defined as the solution of

$$a_t^* = \underset{z}{\operatorname{argmin}}\, \mathbf{V}[\psi_t(X_{a^*}, a_1^*, \cdots, a_{t-1}^*, z, a_{t+1}^*, \cdots, a_T^*)] \quad \forall\, t = 1, \cdots, T.$$

Quite often analytical evaluation of $V[\psi_t]$ is very difficulty so a Monte–Carlo integration with CRNs is used to evaluate $V[\psi_t]$. To illustrate this aspect, let $\{U^i\}_{i=1}^n$ be a set of CRNs drawn independently from a random number generator for U. Set $X_a^i = \Phi(U^i, a)$ where $X_a = \Phi(U, a)$ is a random variable with density $\rho(x, a)$. Then $V[\psi_t(X_a, a)]$ is in certain sense propositional to

$$\min_c \sum_{i=1}^n \Big(\psi_t(X_a^i, a) - c\Big)^2 . \tag{2.0.5}$$

Hence, (2.0.5) can be implemented numerically as the solution of

$$a_t^* = \underset{z}{\operatorname{argmin}}\ \min_c \sum_{i=1}^n \Big\{\psi_t(X_{a^*}^i, a_1^*, ..., a_{t-1}^*, z, a_{t+1}^*, ..., a_T^*) - c\Big\}^2 \omega(X_{a^*}^i, a^*) \quad \forall\, t. \tag{2.0.6}$$

Here ω is a weight function introduced as an added feature.

Problem (2.0.6) is usually solved numerically by an iteration process: $a^* = \lim_{k\to\infty} a^{(k)}$ where

$$a_t^{(k+1)} = \underset{z}{\operatorname{argmin}}\ \min_c \sum_{i=1}^n \Big\{\psi_t(X_{a^{(k)}}^{(i)}, a_1^{(k)}, ..., a_{t-1}^{(k)}, z, a_{t+1}^{(k)}, ..., a_T^{(k)}) - c\Big\}^2 \omega(X_{a^{(k)}}, a^{(k)}).$$

Although it not guaranteed that as $k \to \infty$, $a^{(k)}$ converges, one can argue that the effectiveness of the method should be related to the rate of the convergence of $\{a^{(k)}\}$; fast convergence provides a hard evidence towards the effectiveness (accuracy) of the method; poor convergence warns that it is most probably the family of density functions, instead of the numerical iteration scheme, that needs to be redesigned.

An example of the application of the sequential importance sampling Monte–Carlo integration will be presented in the next section.

3.0 AN APPLICATION OF THE SISMCI

In this section, we apply the sequential importance sampling Monte–Carlo integration method to evaluate a likelihood function of a stochastic volatility model. This particular application has been presented by Geweke (1989) [6], Fishman (1996) [5], Owen and Zhou (2000) [13], and Richard and Zhang (2007) [15], etc., in a context of very general setting. Here we present a simplified version of their derivation. Additional analysis will be given in the subsequent sections.

3.0.4 The Problem

1. The Stochastic Volatility Model. We consider a stochastic volatility model [17]

$$r_t = \beta e^{v_t}\,\varepsilon_t, \qquad v_t = \delta v_{t-1} + \nu\,\eta_t$$

where $\varepsilon_t, \eta_t, t = 0, \pm 1, \pm 2, \cdots,$ are i.i.d $N(0,1)$ distributed random variables. Here $\{r_t\}_{t=-\infty}^{\infty}$ models an observable time series, say stock returns with mean subtracted, βe^{v_t} is the conditional volatility of r_t. The condition of knowing v_t cannot be materialized since v_t is assumed to be a latent factor, i.e., a non-observable stochastic process. In this model, we have three parameters:

$$\Lambda := (\beta, \delta, \nu) \;\in\; (0,\infty) \times (-1,1) \times [0,\infty).$$

Under the new variable $x_t = v_t/\nu$, the stochastic volatility model can be reformulated

as

$$r_t = \beta e^{\nu x_t/2}\varepsilon_t, \qquad x_t = \delta x_{t-1} + \eta_t.$$

2. The Marginal, Conditional, and Joint Density Functions. For notational simplicity, in this section we omit most of the dependence of functions on the parameters $\Lambda = (\beta, \delta, \nu)$. By abusing the notation we denote by $\rho(x_t)$ the marginal density functions of the random variables x_t, by $\rho(x_t|x_{t-1})$ the conditional density of x_t under the condition of known x_{t-1}, and by $\rho(r_t|x_t)$ the conditional density of r_t under the condition of known x_t. We denote by $\rho(r_1, \cdots, r_T, x_1, \cdots, x_T; \Lambda)$ the joint density of $r_1, \cdots, r_T, x_1, \cdots, x_T$ with the given parameter Λ. Finally, we denote by $\rho(r; \Lambda)$ the joint density of $r = (r_1, \cdots, r_T)$. Now we use (1.0.2) to derive these density functions.

Since $|\delta| < 1$, one derives from the recursion $x_t = \delta x_{t-1} + \eta_t$ and the independency and normality of $\{\eta_s\}_{s=-\infty}^{\infty}$ that

$$\rho(x_t|x_{t-1}) = \frac{1}{\sqrt{2\pi}} e^{-(x_t - \delta x_{t-1})^2/2}, \quad x_t = \sum_{i=0}^{\infty} \delta^i \eta_{t-i}, \qquad \rho(x_t) = \frac{\sqrt{1-\delta^2}}{\sqrt{2\pi}} e^{-\frac{1-\delta^2}{2} x_t^2}.$$

The equation $r_t = \beta e^{\nu x_t/2}\varepsilon_t$ and the independency of ε_t and $x_t = \sum_{i=0}^{\infty} \delta^i \eta_{t-i}$ imply that

$$\rho(r_t|x_t) = \frac{e^{-r_t^2/(2\beta^2 e^{\nu x_t})}}{\sqrt{2\pi\beta^2 e^{\nu x_t}}} = \frac{e^{-\nu x_t/2 - r_t^2 e^{-\nu x_t}/(2\beta^2)}}{\sqrt{2\pi\beta^2}}.$$

We then derive from these conditional densities that

$$\begin{aligned}\rho(r_1, \cdots, r_T, x_1, \cdots, x_T; \Lambda) &= \rho(x_1)\rho(r_1|x_1) \prod_{t=2}^{T} \Big[\rho(x_t|x_{t-1})\rho(r_t|x_t)\Big] \\ &= \frac{\sqrt{1-\delta^2}}{(2\pi\beta)^T} \exp\left[-\sum_{t=1}^{T}\Big(\frac{\nu x_t}{2} + \frac{r_t^2 e^{-\nu x_t}}{2\beta^2}\Big) - \frac{(1-\delta^2)x_1^2}{2} - \sum_{t=2}^{T} \frac{(x_t - \delta x_{t-1})^2}{2}\right].\end{aligned}$$

Finally the marginal distribution density $\rho(r; \Lambda)$ of $r = (r_1, \cdots, r_T)$ under given parameter Λ can be calculated from the above joint distribution by

$$\rho(r; \Lambda) = \int_{\mathbb{R}^T} \rho(r_1, \cdots, r_T, x_1, \cdots, x_T; \Lambda) dx_1 \cdots dx_T. \tag{3.0.1}$$

3. The Maximum Likelihood Estimator (MLE). Suppose $R = (R_1, \cdots, R_T)$ is a

set of observations of $r = (r_1, \cdots, r_T)$. Then the maximum likelihood estimator $\Lambda^* = (\beta^*, \delta^*, \nu^*)$ for the parameters in the model (1.0.2) is defined as

$$\Lambda^* = \Lambda^*(R) := \underset{\Lambda \in (0,\infty)\times(-1,1)\times[0,\infty)}{\operatorname{argmax}} \rho(R;\Lambda)\ .$$

To find Λ^*, one has to evaluate the T-dimensional integral in (3.0.1) which does not seem to have an analytical closed form. Since T is usually quite large, efficient numerical algorithm is needed to perform the integration. Indeed this is our current focus of the attention. For definiteness, we formulate our problem as follows.

Problem: Given constants $\beta > 0, \delta \in (-1,1)$, $\nu \geqslant 0$, and $(R_1, \cdots . R_T) \in R^T$, numerically evaluate the integral $\int_{\mathbb{R}^T} f(x)dx$ where $x = (x_1, \cdots, x_T)$ and

$$f(x) = \frac{\sqrt{1-\delta^2}}{(2\pi\beta)^T}\ \exp\Big[-\sum_{t=1}^{T}\Big(\frac{\nu x_t}{2} + \frac{R_t^2 e^{-\nu x_t}}{2\beta^2}\Big) - \frac{(1-\delta^2)x_1^2}{2} - \sum_{t=2}^{T}\frac{(x_t - \delta x_{t-1})^2}{2}\Big]. \tag{3.0.2}$$

3.0.5 Sequential Importance Sampling Monte–Carlo Integration

Here we present a simplified derivation of the algorithm used for the numerical evaluation of (2.0.1) with f given by (3.0.2).

1. The family of probability density functions. The special form of f suggests that we use the Gaussian distribution family. More specifically, we define $\{\rho(x,a)\}_{a\in\mathcal{A}}$ by

$$\begin{cases} a & := & (a_1, \cdots, a_T, b_1, \cdots, b_T) \in \mathcal{A} := (0,\infty)^T \times \mathbb{R}^T, \\ \rho(x,a) & := & \displaystyle\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi a_t^2}} e^{-[x_t - a_t^2(\delta x_{t-1} - b_t)]^2/(2a_t^2)} \quad (x_0 \equiv 0). \end{cases} \tag{3.0.3}$$

For any $a \in A$, a random variable X_a that has density $\rho(\cdot, a)$ can be obtained from a single T-dimensional random variable η as follows. Let $\eta_1, \cdots, \eta_T$ be i.i.d. $N(0,1)$ distributed random variables. Set $\eta = (\eta_1, \cdots, \eta_T)$ and $\Phi(\eta, a) = (\Phi_1, \cdots, \Phi_T)$ where

$\{\Phi_t\}_{t=0}^T$ are defined by

$$\Phi_0 \equiv 0, \qquad \Phi_t = a_t^2[\delta\Phi_{t-1} - b_t] + a_t\eta_t, \quad t = 1, \cdots, T. \tag{3.0.4}$$

It is easy to verify that $X_a := \Phi(\eta, a)$ is a random variable with density $\rho(\cdot, a)$. Given random draws $\{\eta^i\}_{i=1}^n = \{(\eta_1^i, \cdots, \eta_T^i)\}_{i=1}^n$ of η, we denote by $\{X_a^i\}_{i=1}^n$ the corresponding CRNs defined by $X_a^i = \Phi(\eta^i, a)$. The components of X_a^i are denoted as $(X_{a,1}^i, \cdots, X_{a,T}^i)$.

For convenience, in the sequel we use the following "artificial" values:

$$x_0 := 0, \quad a_{T+1} := 1, \quad b_{T+1} := 0, \qquad \delta_{t\tau} := \begin{cases} 1 & \text{if } t = \tau, \\ 0 & \text{if } t \neq \tau. \end{cases}$$

2. The function $f(x)/\rho(x, a)$. We can calculate, for any $a = (a_1, \cdots, a_T, b_1, \cdots, b_T)$ in A,

$$\psi(x, a) := \ln \frac{f(x)}{\rho(x, a)} = \psi_0(a) + \sum_{t=1}^T \psi_t\big(x_t, p_t(a), q_t(a)\big),$$

where

$$\begin{aligned} \psi_0(a) := \quad & \ln \frac{\sqrt{1-\delta^2}}{(\sqrt{2\pi}\,\beta)^T} + \sum_{t=1}^T \left(\ln a_t + \frac{a_t^2 b_t^2}{2}\right), \\ \psi_t(z, \kappa, \ell) := \quad & \kappa\, z^2 + \ell\, z - \Big(\frac{R_t^2 e^{-\nu z}}{2\beta^2} + \frac{\nu z}{2}\Big), \end{aligned} \tag{3.0.5}$$

$$\begin{cases} p_t(a) & := \quad \dfrac{1}{2a_t^2} - \dfrac{1 + \delta^2[1 - a_{t+1}^2 - \delta_{t1}]}{2}, \\ q_t(a) & := \quad b_t - \delta a_{t+1}^2 b_{t+1}. \end{cases} \tag{3.0.6}$$

Hence, for each $a \in A$,

$$\int_{\mathbb{R}^T} f(x)dx = e^{\psi_0(a)} \int_{\mathbb{R}^T} e^{\sum_{t=1}^T \psi_t(x_t, p_t(a), q_t(a))}[\rho(x, a)dx].$$

The simple expression of $\psi_t(z, \kappa, \ell)$ in (3.0.5) suggests that it is convenient to use the parameters $p = (p_1, \cdots, p_T, q_1, \cdots, q_T)$ as independent parameters and use $a =$

$(a_1, \cdots, a_T, b_1, \cdots, b_T)$ as dependent parameters. The following result then becomes very helpful.

Sequential Importance Sampling Monte-Carlo Integration

0. Load input $\{\beta, \delta, \nu, R_1, \cdots, R_T\}$. Assign $\varepsilon =$"tolerance" a value, say, 10^{-4}. Set $a_t = 1$ and $b_t = 0$ for all $t = 1, \cdots, T+1$, and $X_0^i = 0$ for all $i = 1, \cdots, n$. Generate independent $N(0,1)$ distributed random numbers $\{ \{\eta_t^i\}_{t=1}^T \}_{i=1}^n$.

1. Define $\{\{X_t^i\}_{t=1}^T\}_{i=1}^n$ by $X_t^i = a_t^2[\delta X_{t-1}^i - b_t] + a_t\eta_t^i, \quad t = 1, \cdots, T, \quad i = 1, \cdots, n.$

2. For each $t = 1, \cdots, T$, find

$$(p_t, q_t, c_t) = \underset{(\kappa,\ell,c)}{\operatorname{argmin}} \sum_{i=1}^n \Big(\psi_t(X_t^i, \kappa, \ell) + c\Big)^2.$$

3. Set $a^{\text{old}} = (a_1, \cdots, a_T, b_1, \cdots, b_T)$, define $a = (a_1, \cdots, a_T, b_1, \cdots, b_T)$ by (3.0.7), and make a switch as follows:

$$\boxed{|a - a^{\text{old}}| < \varepsilon\ ?} \xrightarrow{\text{No}} \boxed{\begin{array}{l}\text{Max Itera-}\\ \text{tion?}\end{array}} \xrightarrow{\text{No}} \boxed{\text{goto } \mathbf{1}}$$

$$\downarrow \text{Yes} \qquad\qquad \downarrow \text{Yes}$$

$$\boxed{\text{goto } \mathbf{4}} \qquad\qquad \boxed{\text{Abort!}}$$

4. Produce the output

$$\ln \int_T f(x)dx \approx \psi_0(a) - \sum_{t=1}^T c_t + \ln\Big(\frac{1}{n}\sum_{i=1}^n \exp\Big(\sum_{t=1}^T [\psi_t(X_t^i, p_t, q_t) + c_t]\Big)\Big).$$

Table 4: Sensitivity Analysis for SIS

	(1)						(2)						(3)					
n	L_n	Err	$\frac{S_n}{\sqrt{n}}$	S_n	C.T.	It	L_n	Err	$\frac{S_n}{\sqrt{n}}$	S_n	C.T.	It	L_n	Err	$\frac{S_n}{\sqrt{n}}$	S_n	C.T.	It
4	-69.0406	-0.0214	0.01698	0.03396	0.06	13	-68.6818	0.3374	0.01692	0.03384	0.06	11	-69.0618	-0.0426	0.01696	0.03391	0.08	13
8	-68.8293	0.1899	0.06678	0.18888	0.06	14	-68.4719	0.5474	0.06709	0.18977	0.06	13	-68.8508	0.1684	0.06671	0.18869	0.05	13
16	-68.9993	0.0199	0.05604	0.22415	0.06	14	-68.6455	0.3737	0.05631	0.22525	0.05	13	-69.0200	-0.0008	0.05598	0.22393	0.03	14
32	-69.0594	-0.0402	0.05742	0.32481	0.05	14	-68.7060	0.3132	0.05745	0.32497	0.03	13	-69.0799	-0.0607	0.05741	0.32478	0.06	14
64	-69.0010	0.0182	0.05464	0.43713	0.05	15	-68.6460	0.3732	0.05465	0.43719	0.06	12	-69.0213	-0.0021	0.05463	0.43702	0.05	15
128	-69.0365	-0.0173	0.03300	0.37336	0.03	14	-68.6824	0.3368	0.03302	0.37362	0.03	11	-69.0570	-0.0377	0.03299	0.37323	0.03	14
256	-69.0377	-0.0185	0.02564	0.41025	0.08	13	-68.6841	0.3351	0.02565	0.41044	0.06	11	-69.0581	-0.0389	0.02562	0.40997	0.08	13
512	-69.0348	-0.0156	0.01933	0.43745	0.08	13	-68.6811	0.3381	0.01933	0.43743	0.11	11	-69.0552	-0.0360	0.01932	0.43721	0.08	13
1024	-69.0111	0.0081	0.01318	0.42180	0.13	13	-68.6574	0.3618	0.01319	0.42204	0.13	11	-69.0316	-0.0123	0.01317	0.42149	0.14	13
2048	-68.9974	0.0218	0.01003	0.45387	0.23	14	-68.6438	0.3754	0.01003	0.45383	0.19	11	-69.0179	0.0013	0.01002	0.45350	0.20	13
4096	-69.0010	0.0183	0.00694	0.44386	0.45	14	-68.6473	0.3719	0.00694	0.44403	0.38	11	-69.0214	-0.0022	0.00693	0.44334	0.44	13
8192	-68.9993	0.0199	0.00487	0.44047	0.94	14	-68.6457	0.3735	0.00487	0.44069	0.81	11	-69.0198	-0.0006	0.00486	0.43994	1.00	14
16384	-68.9982	0.0210	0.00343	0.43894	2.23	14	-68.6446	0.3746	0.00343	0.43919	2.00	11	-69.0188	0.0005	0.00342	0.43838	2.20	14
32768	-68.9974	0.0218	0.00239	0.43196	5.69	14	-68.6438	0.3754	0.00239	0.43211	4.72	11	-69.0179	0.0013	0.00238	0.43141	5.67	14
65536	-68.9981	0.0211	0.00171	0.43880	15.56	14	-68.6445	0.3747	0.00171	0.43882	13.44	11	-69.0186	0.0006	0.00171	0.43823	15.41	14
131072	-68.9968	0.0224	0.00121	0.43847	98.36	14	-68.6433	0.3760	0.00121	0.43858	93.75	11	-69.0173	0.0019	0.00121	0.43786	97.73	14
262144	-68.9970	0.0222	0.00086	0.43844	437.09	14	-68.6435	0.3757	0.00086	0.43855	453.06	11	-69.0176	0.0017	0.00086	0.43784	455.80	14

Lemma 3.0.1. *The map* $a \in A \to p(a) := (p_1(a), \cdots, p_T(a), q_1(a), \cdots, q_T(a))$ *defined by (3.0.6) is one-to-one. Its inverse can be calculated iteratively as follows (recalling* $a_{T+1} = 1$ *and* $b_{T+1} = 0$*):*

$$\begin{cases} a_t &= (\ 2p_t + 1 + \delta^2[1 - a_{t+1}^2 - \delta_{t1}]\ )^{-1/2}, \\ b_t &= q_t + \delta a_{t+1}^2 b_{t+1}, \end{cases} \qquad t = T,\ T-1,\ \cdots, 1. \tag{3.0.7}$$

In addition, this inverse p^{-1} *maps* $[0,\infty)^T \times R^T$ *to* $\left(0, \dfrac{1}{\sqrt{1-\delta^2}}\right] \times (0,1]^{T-1} \times R^T$.

The proof is a straightforward verification and hence is omitted.

3. Criteria for "Optimal" Parameters. Based on (2.0.5) and (2.0.6), we present two criteria.

When True (Theoretical) variance are used, we propose the following: the "optimal" parameter $a^* \in A$ to be used for SISMCI is the solution of

$$\Big(p_t(a^*), q_t(a^*)\Big) = \underset{(\kappa,\ell)}{\operatorname{argmin}}\ \mathbf{V}\left[\psi(X_{a^*,t},\ \kappa, \ell)\right] \quad \forall\, t = 1, \cdots, T. \tag{T}$$

Here $X_a := (X_{a,1}, \cdots, X_{a,T}) := \Phi(\eta, a)$ where $\Phi(\eta, a) := (\Phi_1, \cdots, \Phi_T)$ is defined in (3.0.4). We shall investigate this criterion in detail in the next section.

In rest of this section, we consider the criterion used by previous literature, in which the variance in (T) is replaced by Sample variance; that is, after the generation of CRNs, $\{\eta^i\}_{i=1}^n$, the optimal parameter a^* is the solution of

$$\Big(p_t(a^*), q_t(a^*)\Big) = \underset{(\kappa,\ell)}{\operatorname{argmin}}\ \min_c \sum_{i=1}^n \Big(\psi_t(X_{a^*,t}^i,\ \kappa, \ell) + c\Big)^2 \quad \forall\, t = 1, \cdots, T, \tag{S}$$

where $(X_{a,1}^i, \cdots, X_{a,T}^i) := \Phi(\eta^i, a)$.

4. The Sequential Minimization. To solve problem (S), first we consider a simple minimization where a^* on the right–hand side of (S) is replaced by a generic $a \in A$. That is, we consider the map $a \in A \to p^*(a) := (p_1^*(a), \cdots, p_T^*(a), q_1^*(a), \cdots, q_T^*(a))$

where

$$\Big(p_t^*(a), q_t^*(a)\Big) := \underset{(\kappa,\ell)}{\operatorname{argmin}} \ \min_{c} \ \sum_{i=1}^{n} \Big(\psi_t(X_{a,t}^i, \kappa, \ell) \ + \ c\Big)^2 \quad \forall\, t = 1, \cdots, T. \tag{3.0.8}$$

From the explicit expression of $\psi_t(z, \kappa, \ell)$ in (3.0.5), we see that the minimization problem on the right-hand side is a standard linear regression whose geometric interpretation is the orthogonal projection of the vector F onto the space expanded by $1, L$ and B where

$$F = \frac{R_t^2}{2\beta^2}(e^{-\nu \mathbf{x}_{a,t}^1}, \cdots, e^{-\nu X_{a,t}^n}) + \frac{\nu}{2} L, \qquad B = ([X_{a,t}^1]^2, \cdots, [X_{a,t}^n]^2),$$
$$L = (X_{a,t}^1, \cdots, X_{a,t}^n), \qquad \mathbf{1} = (1, \cdots, 1)_{n\times 1}.$$

Thus, $(p_t^*(a), q_t^*(a))$ is indeed the solution of the following linear system (with $\varepsilon = 0$)

$$\begin{pmatrix} B\cdot B + \varepsilon & L\cdot B & \mathbf{1}\cdot B \\ B\cdot L & L\cdot L + \varepsilon & \mathbf{1}\cdot L \\ B\cdot \mathbf{1} & L\cdot \mathbf{1} & \mathbf{1}\cdot \mathbf{1} + \varepsilon \end{pmatrix} \begin{pmatrix} p_t^*(a) \\ q_t^*(a) \\ c_t(a) \end{pmatrix} = \begin{pmatrix} F\cdot B \\ F\cdot L \\ F\cdot \mathbf{1} \end{pmatrix}.$$

When $n \geqslant 3$, the theoretical probability that the n-dimensional vectors $1, L$, and B are linearly dependent is zero. Hence, we can assume that the above system (with $\varepsilon = 0$) always admits a unique solution so that $p^*(a)$ is well-defined by (3.0.8). Numerically we can take a tiny positive ε to ensure the robustness of the program.

It is easy to see now that a^* solves (S) if and only if a^* is a fixed point of the composite map

$$a \in \mathcal{A} \quad \xrightarrow[(3.0.8)]{\mathbf{p}^*} \quad p = \mathbf{p}^*(a) \quad \xrightarrow[(3.0.7)]{\mathbf{p}^{-1}} \quad \tilde{a} = \mathbf{p}^{-1}(p) = \mathbf{p}^{-1} \circ \mathbf{p}^*(a).$$

One may be happy to notice the following. The minimization problem in (3.0.8) is in certain sense meant to approximate the exponential function $he^{-\nu s}$ ($h = R_t^2/(2\beta^2) \geqslant 0$) by a quadratic function $\kappa s^2 + [\ell - \nu/2]s - c$. Since the function $s \in R \to he^{-\nu s}$ is convex, it should be true in general that the solution of problem (3.0.8) satisfies $p_t^*(a) \geqslant 0$, i.e., $p^*(a) \in [0, \infty)^T \times R^T$. Hence, by Lemma 3.0.1, $p^{-1} \circ p^*(a)$ is

well-defined.

Numerically, the fixed point a^* of the map $p^{-1} \circ p^*$ can be obtained by an iteration process:

$$a^* = \lim_{k \to \infty} a^{(k)}, \qquad a^{(k+1)} := \mathbf{p}^{-1} \circ \mathbf{p}^*(a^{(k)}).$$

4. A Numerical Scheme. Based on the above discussion, if we use (S) as the criterion for the "optimal" parameter, we can design a numerical algorithm for sequential importance sampling (SIS) as follows:
Here c_t is introduced for the necessity of numerical implementation that avoids overflow of the floating-point arithmetic for exponentially large or exponentially small numbers.

3.0.6 A Numerical Experiment.

In the sequel, a point $x = (x_1, \cdots, x_T)$ in R^T will be referred to as a path since we visualize it as a curve $\{(t, x_t) \in R^2 \mid 0 \leqslant t \leqslant T\}$.
1. The Input. As an illustration, we use artificial data generated by a Monte-Carlo simulation for (1.0.2) with $(\beta, \delta, \nu) = (1,\ 0.9,\ 0.5)$. For the convenience of numerical analysis, we take a small size $T = 50$. The data $\{R_t\}_{t=1}^T$ are plotted as the thin curve in Figure 3 (a).

2. The Optimal Parameters. The optimal parameter, denoted by $a^*(n)$ where n is the number of sample paths, depends on the CRNs $\{\{\eta_t^i\}_{t=1}^T\}_{i=1}^n$ used in the MCI. One of such parameters is plotted in Figure 3 (a). The theoretical mean, denoted by a^{**}, of the optimal parameters $a^*(n)$ is the solution of (T). For the particular CRNs used in our simulation, we list the deviation $|a^*(n) - a^{**}|$ in the last column in Table 3. It is quite easy to see the tendency that

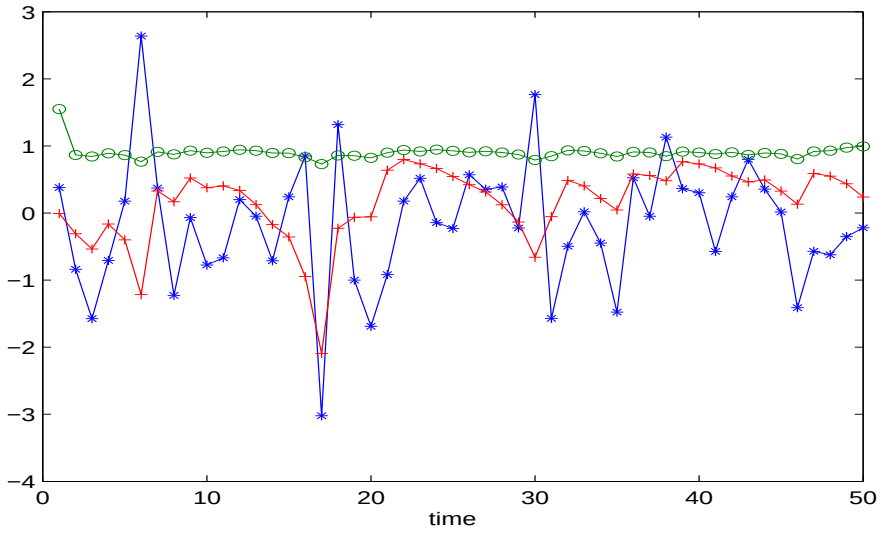
$$\max_{1 \leqslant t \leqslant T} |a_t^*(n) - a_t^{**}| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty.$$

Based on this observation, it is then quite natural to take the initial value $a^{(0)}$ of

$a^* = \lim_{k\to\infty} a^{(k)}$ as a^{**}, which will be fully analyzed in the next section. In Table 3, comparisons of the number of iterations and the CPU time (in seconds) are made between the case of setting $a^{(0)} = a^{**}$ and the default case of setting $a^{(0)} \equiv 1$.

When $n = 4$, the sequence $a^{(k)}$ oscillates and does not seem to converge. This is not a generic phenomenon and it came to our attention by chance.

Figure 3: SIS Curves (a)



The thin curve represents $\{R_t\}_{t=1}^{50}$ which has mean -0.16, Std 0.95, skewness -0.05, and kurtosis 4.47. The two thick curves are $\{a_t^*\}_{t=1}^{50}$ and $\{b_t^*\}_{t=1}^{50}$.

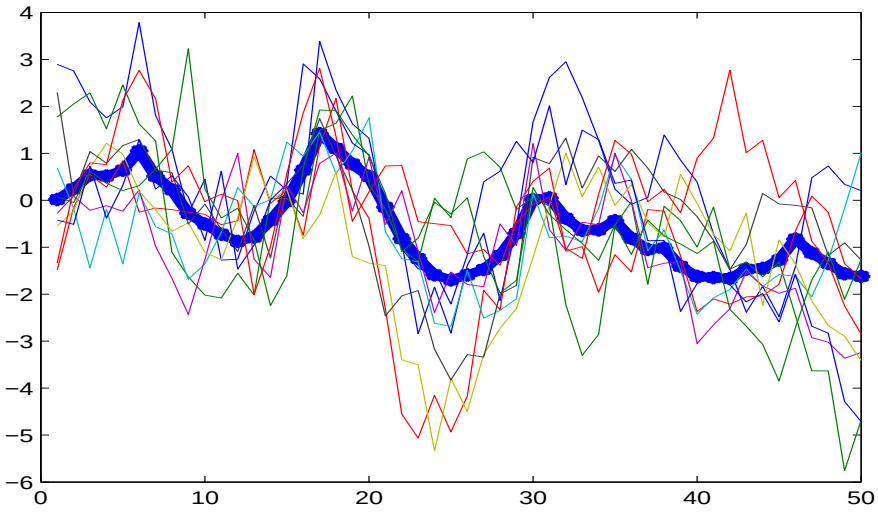
3. Importance Sample Paths. In Figure 3 (b) we plot two of the important sample paths $\{X_t^i\}_{t=1}^T$ ($i = 1, 2, \cdots, n$) defined by $X_t^i = a_t^{*2}[\delta X_t - b_t^*] + a_t^* \eta_t^i$. Also plotted are the mean path (setting $\eta_t^i = 0$) with optimal parameter $a = a^*(n)$ for different n. These mean paths are almost the same since $a^*(n) \approx a^{**}$. As $(\eta_1, \cdots, \eta_T) \sim N(0, I)$, all important sample paths oscillate around the mean paths.

4. Error Analysis. Let $\mu := \int_{\mathbb{R}^T} f(x)dx$ and μ_n be its numerical approximation. Note that

$$\ln \mu_n = \ln \mu + \ln\left(1 + \frac{\mu_n - \mu}{\mu}\right) \approx \ln \mu + \left(\frac{\mu_n}{\mu} - 1\right).$$

An informative approximation of $\ln \mu$ can be written as $\ln \mu = \ln \mu_n + \varepsilon_n/\sqrt{n} \pm \sigma/\sqrt{n}$ where $\varepsilon_n/\sqrt{n} = E[\ln \mu_n] - \ln \mu$ is the bias and σ is a constant. We propose the following for the estimation of σ.

Figure 4: SIS Curves (b)



The curves clustered in the middle are the mean paths for different n. The thin curves are two of the important sample paths used in one of the MCIs.

(1) As an approximation, the optimal parameter $a^*(n)$ can be regarded as a constant vector, since its dependence on the CRNs used in the MCI is very weak. Hence we can approximate σ by the sample relative standard deviation

$$\sigma_n := \left\{ \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{e^{\psi^i}}{\overline{e^{\psi}}} - 1 \right)^2 \right\}^{1/2}, \qquad \overline{e^{\psi}} := \frac{1}{n} \sum_{i=1}^{n} e^{\psi^i}, \qquad \psi^i := \sum_{t=1}^{T} \psi_t(X_t^i, p_t, q_t).$$

From Table 3, one sees that $\sigma_n/\sqrt{n}$ does provide a basic size of error of the MCI output for the Log likelihood. In addition, one does see a tendency that $\sigma_n \to \sigma \approx 0.5\cdots$ as $n \to \infty$. One may also notice that there does exist a negative bias of size comparable to $\sigma_n/\sqrt{n}$.

(2) We use a simple version of ANOVA (analysis of variance). Suppose we perform

n_e number of SISMCI experiments, each of which takes n_s samples. We denote the approximations and sample relative standard-deviations from these SISMCIs by $\mu_{n_s}^{(j)}, \sigma_{n_s}^{(j)},\ j = 1, \cdots, n_e$.

(i) The relative variance of each $\mu_{n_s}^{(j)},\ j = 1, \cdots, n_e$, is approximately σ^2/n_s so the sample relative variance of $\{\mu_{n_s}^{(j)}\}_{j=1}^{n_e}$ also approximates σ^2/n_s. Hence σ can be approximated by

$$\sigma_{n_s,n_e} := \sqrt{n_s}\left\{\frac{1}{n_e-1}\sum_{j=1}^{n_e}\Big(\frac{\mu_{n_s}^{(j)}}{\mu_{n_sn_e}}-1\Big)^2\right\}^{1/2}, \qquad \mu_{n_sn_e} := \frac{1}{n_e}\sum_{j=1}^{n_e}\mu_{n_s}^{(j)}.$$

(ii) Since each $\sigma_{n_s}^{(j)},\ j = 1, \cdots, n_e$, approximates σ, so does their average. This average can be finely tuned by the following. Set $\bar{\mu}_{\cdot,j}$ the average of μ_{ij} over $i = 1, \cdots, n_s$ and $\bar{\mu}$ the overall average. Then

$$\begin{aligned}(n_sn_e-1)\sigma^2 &\approx \sum_{j=1}^{n_e}\sum_{i=1}^{n_s}\Big(\frac{\mu_{ij}}{\bar{\mu}}-1\Big)^2 = \frac{1}{\bar{\mu}^2}\sum_{j=1}^{n_e}\Big\{\sum_{i=1}^{n_s}\Big(\mu_{ij}-\overline{\mu_{\cdot j}}\Big)^2 + n_s\Big(\overline{\mu_{\cdot j}}-\bar{\mu}\Big)^2\Big\}\\ &\approx (n_s-1)\sum_{j=1}^{n_e}\sigma_{n_s}^{(j)2} + (n_e-1)\sigma_{n_e,n_s}^2\end{aligned}$$

Here the second term is of the order $n_e\sigma^2 = [n_en_s\sigma^2]/n_s$ and can be neglected when $n_s \gg 1$. Anyway, we can approximate σ by

$$\overline{\sigma_{n_sn_e}} := \frac{1}{\sqrt{n_en_s-1}}\Big\{(n_s-1)\sum_{j=1}^{n_e}\sigma_{n_s}^{(j)2} + (n_e-1)\,\sigma_{n_sn_e}^2\Big\}^{1/2}.$$

In summary, denoting by $n = n_sn_e$ the total number of sample points used, we expect

$$\ln\mu = \ln\mu_{n_s,n_e} + \frac{\varepsilon_n}{\sqrt{n}} \pm \frac{\sigma}{\sqrt{n}}, \qquad \sigma \approx \sigma_n \approx \overline{\sigma_{n_sn_e}} \approx \sigma_{n_s,n_e}.$$

In Tables 3 and 4, we list these sample standard deviations $\sigma_n, \sigma_{n_s\,n_e}$ and $\overline{\sigma_{n_sn_e}}$ from numerical simulations. From these listed data, one can obtain a basic size $\sigma \approx 0.5$. That is to say, a log likelihood estimation from an SISMCI has an error of size $0.5/\sqrt{n}$.

5. Biased Estimator. Since the parameter a^* depends on the CRNs, fully theoretical analysis of the statistical behavior of the random variable $\mu_n = \mu_n(\{\{\eta_t^i\}_{t=1}^T\}_{i=1}^n)$ is out of our reach for the moment. Nevertheless, from Table 4, one sees that most of $\ln \mu_n$ underestimate $\ln \mu$. Indeed, one can run a statistical analysis to show that it is statistically significant to reject the hypothesis $E[\ln \mu_n] - \ln \mu = 0$. For example, in $n_e = 128$ experiments, the average of $\ln \mu_{128} - \ln \mu$ is -0.0182 which is more than 5 sample standard-deviation ($= 0.00354$) away from zero. This gives us an extremely high level of confidence to accept the hypothesis that $\varepsilon_{128} > \sigma$. We tend to believe that

$$\lim_{n\to\infty} \varepsilon_n := \lim_{n\to\infty} \sqrt{n}\Big(\ln \mu - \mathbf{E}[\ln \mu_n]) \propto \sigma.$$

Remark 3.0.4. *1. The CPU time listed is only for reference and it is not always proportional to the amount of calculation needed. We use Matlab on Dell's PC which seems to be efficient in handling certain particular dimensions of vectors so it may take shorter time to do larger job; see the amount of CPU time for the comparison of* $(n_e, n_s) = (32, 128)$ *and* $(n_e, n_s) = (32, 256)$ *and also the comparison of* $(n_e, n_s) = (64, 256)$ *and* $(n_e, n_s) = (128, 128)$.

2. If an estimator is biased by more than one standard-deviation, then taking the average of a large number of the values produced by Monte-Carlo simulation won't improve the approximation. If $\varepsilon_n > \sigma$ *were rigorously verified, then taking for example,* $n_e = 100$*, of the Monte-Carolo experiments with fixed number of sample points, say* $n_s = 100$ *, does not guarantee to obtain an approximation with error size* $O(1/\sqrt{n_s n_e})$*. Indeed, when* $n_e \to \infty$*, one obtains a limit value* $\ln \mu - \varepsilon_{n_s}/\sqrt{n_s}$.

4.0 A NEW APPROACH OF SISMCI FOR SVM

In this section, we investigate the SISMCI for the likelihood function defined in (3.0.1), using the criterion (T) which is new in the literature. To be specific, we assume that $\Lambda = (\beta, \delta, \nu) \in (0, \infty) \times (-1, 1) \times [0, \infty)$ and $R = (R_1, \cdots, R_T) \in R^T$ are given and fixed, and consider the integral

$$\ell := \ln \int_{\mathbb{R}^T} e^{-\phi(x)} dx \tag{4.0.1}$$

where $\phi(x)$ is given by

$$\phi(x) := \ln \frac{(2\pi\beta)^T}{\sqrt{1-\delta^2}} + \sum_{t=1}^{T} \left\{ \frac{\nu x_t}{2} + \frac{R_t^2}{2\beta^2} e^{-\nu x_t} \right\} + \frac{(1-\delta^2)x_1^2}{2} + \sum_{t=2}^{T} \frac{(x_t - \delta x_{t-1})^2}{2}. \tag{4.0.2}$$

4.0.7 Normalization

Note that ϕ is a sum of convex functions, so ϕ is convex. Indeed, it is strictly convex so it has a unique point of minimum. Hence, the major contribution toward the integral (4.0.1) comes from the integration of the integrand in a certain vicinity of the point of local minimum of ϕ. Clearly, in Monte-Carlo integration, of importance are those samples that are near the point of minimum of ϕ. It is therefore useful, at least from numerical point of view, that we perform a normalization by shifting the origin to the point of minimum of ϕ, so important samples in MCI are near the origin in the new coordinate system.

Lemma 4.0.2. *The function $\phi(\cdot)$ defined in (4.0.2) has the following properties:*

1. *ϕ is strictly convex in R^T; that is, for every $x \in R^T$, the matrix $D^2\phi(x)$ of all second order derivatives of ϕ at x is positive definite.*
2. *There exists a unique $x^* = (x_1^*, \cdots, x_T^*) \in R^T$ such that ϕ attains its global minimum:*

$$\phi(x^*) < \phi(x) \qquad \forall\, x \neq x^*.$$

3. *Let $\zeta(x) = \phi(x^* + x) - \phi(x^*)$. Then*

$$\zeta(\mathbf{0}) = 0 < \zeta(x) \quad \forall\, x \neq \mathbf{0}, \quad D\zeta(\mathbf{0}) = \mathbf{0}, \qquad D^2\zeta(x) > (1 - |\delta|)^2\, \mathbf{I} \quad \forall\, x \in \mathbb{R}^T,$$

$$\zeta(x) = \frac{(1-\delta^2)x_1^2}{2} + \sum_{t=2}^{T} \frac{(x_t - \delta x_{t-1})^2}{2} + \sum_{t=1}^{T} h_t\Big(e^{-\nu x_t} - 1 + \nu x_t\Big) \tag{4.0.3}$$

where I is the $T \times T$ identity matrix, and

$$h_t \equiv \frac{R_t^2}{2\beta^2} e^{-\nu x_t^*}, \quad t = 1, \cdots, T.$$

4. *In terms of ζ, the integral (4.0.1) can be written as*

$$\rho(R; \Lambda) = \int_{\mathbb{R}^T} e^{-\phi(x)} dx = e^{-\phi(x_*)} \int_{\mathbb{R}^T} e^{-\zeta(x)} dx. \tag{4.0.4}$$

Note that the function $e^{-\nu z} - 1 + \nu z$ is convex and attains its global minimum zero at $z = 0$.

Proof. Since $e^{-\nu z}$ is a convex function, ϕ is the sum of convex functions so it is also convex. Notice the identity

$$(1-\delta^2)x_1^2 + \sum_{t=2}^{T}(x_t - \delta x_{t-1})^2 = (1-|\delta|)(x_1^2 + x_T^2) + (1-|\delta|)^2 \sum_{t=2}^{T-1} x_t^2 + |\delta| \sum_{t=2}^{T} \left(x_t - \frac{\delta}{|\delta|} x_{t-1}\right)^2.$$

We see that the minimum eigenvalue of the Hessian $D^2\phi$ is no smaller than $(1 - |\delta|)^2$ so $D^2\phi$ is positive definite and ϕ is strictly convex.

The above identity shows that ϕ grows to ∞ as $|x| \to \infty$ so that ϕ admits a unique point of local minimum. We denote this point by $x^* = (x_1^*, \cdots, x_T^*)$. It is easy

to check that $\zeta(x) := \phi(x^* + x) - \phi(x^*)$ can be expressed as

$$\zeta(x) = \frac{(1-\delta^2)x_1^2}{2} + \sum_{t=2}^{T} \frac{(x_t + \delta x_{t-1})^2}{2} + \sum_{t=1}^{T} \left(h_t e^{-\nu x_t} + g_t x_t \right) - c$$

where $c, g_1, \cdots, g_T$ are constants. Using $\zeta(0) = 0$ we have $c = \sum_{t=1}^{T} h_t$. Also from $0 = D\zeta(0) = (g_1 - \nu h_1, \cdots, g_T - \nu h_T)$, we see that $g_t = \nu h_t$ for all t. Hence, $\zeta(x)$ has the form (4.0.3). □

4.0.8 The New Criterion for the Sequential Importance Sampling

We now investigate a numerical evaluation of (4.0.4) by using the SISMCI. Same as before, we use the probability density family $\{\rho(x,a)\}_{a\in\mathcal{A}}$ defined in (3.0.3). One finds that

$$\psi(x,a) := \ln \frac{e^{-\phi(x^*+x)}}{\rho(x,a)} = \psi_0(a) + \sum_{t=1}^{T} \psi_t(x_t, p_t(a), q_t(a))$$

where $\{(p_t, q_t)\}$ are given by (3.0.6) and

$$\psi_0(a) := \sum_{t=1}^{T} \left(\log \sqrt{2\pi a_t^2} + \frac{a_t^2 b_t^2}{2} \right) - \phi(x^*), \tag{4.0.5}$$

$$\psi_t(z,\kappa,\ell) := \kappa\, z^2 + \ell\, z - h_t[e^{-\nu z} - 1 + \nu z]. \tag{4.0.6}$$

We shall use the following

Criterion For Universal Optimal Parameters

The universal optimal parameter a^* is the solution $a \in (0,\infty)^T \times^T$ of the system

$$(p_t(a), q_t(a)) = \underset{(\kappa,\ell)}{\operatorname{argmin}} \operatorname{Var}\left[\psi_t(X_{a,t}, \kappa, \ell)\right] \quad \forall\, t = 1, \cdots, T. \tag{T}$$

Here $\psi_t(z,\kappa,\ell)$ and $(p_t(a), q_t(a))$ are defined in (4.0.5) and (3.0.6) respectively, whereas the random variable $X_{a,t}$ is defined by $X_{a,t} = \Phi_t(\eta, a)$ with Φ_t given by (3.0.4) where $\eta \sim N(\mathbf{0}, \mathbf{I}_{T\times T})$.

Table 5: Sensitivity Analysis for UIS

	UIS					SIS with initial parameters from natural sampler						SIS with initial parameters from UIS					
n	L_n	Error	$\frac{S_n}{\sqrt{n}}$	S_n	CPU T	L_n	Error	$\frac{S_n}{\sqrt{n}}$	S_n	CPU T	It	L_n	Error	$\frac{S_n}{\sqrt{n}}$	S_n	CPU T	It
4	−69.00257	0.01664	0.14193	0.28	0.000	-69.1500	-0.1307	0.0182	0.03	0.046	6	-69.1500	-0.1297	0.0182	0.03	0.015	5
8	−68.98750	0.03171	0.13134	0.37	0.000	-69.1441	-0.1249	0.0464	0.13	0.062	8	-69.1441	-0.1247	0.0464	0.13	0.015	5
16	−68.92605	0.09316	0.09218	0.36	0.000	-69.0129	0.0062	0.0579	0.23	0.046	8	-69.0129	0.0064	0.0579	0.23	0.031	5
32	−69.06370	−0.04448	0.10403	0.58	0.000	-69.0921	-0.0729	0.0661	0.37	0.046	8	-69.0921	-0.0734	0.0661	0.37	0.000	4
64	−68.93594	0.08327	0.07131	0.57	0.000	-68.9857	0.0334	0.0466	0.37	0.031	9	-68.9857	0.0329	0.0465	0.37	0.031	4
128	−68.98873	0.03048	0.04425	0.50	0.000	-69.0084	0.0107	0.0361	0.40	0.031	9	-69.0084	0.0106	0.0361	0.40	0.015	4
256	−69.03922	−0.02000	0.02555	0.40	0.000	-69.0645	-0.0453	0.0263	0.42	0.031	9	-69.0645	-0.0454	0.0262	0.42	0.046	4
512	−69.04422	−0.02500	0.01590	0.35	0.031	-69.0380	-0.0188	0.0156	0.35	0.078	7	-69.0380	-0.0191	0.0155	0.35	0.015	3
1024	−69.03775	−0.01853	0.01841	0.58	0.000	-69.0365	-0.0173	0.0180	0.57	0.078	7	-69.0365	-0.0178	0.0179	0.57	0.031	3
2048	−69.02686	−0.00764	0.00869	0.39	0.031	-69.0279	-0.0087	0.0086	0.39	0.140	7	-69.0279	-0.0091	0.0086	0.39	0.062	2
4096	−69.02766	−0.00844	0.00648	0.41	0.046	-69.0274	-0.0082	0.0065	0.41	0.281	7	-69.0274	-0.0086	0.0065	0.41	0.093	2
8192	−69.02342	−0.00420	0.00482	0.43	0.078	-69.0226	-0.0034	0.0049	0.44	0.531	7	-69.0226	-0.0047	0.0049	0.44	0.171	2
16384	−69.02336	−0.00414	0.00338	0.43	0.125	-69.0229	-0.0037	0.0034	0.43	1.125	7	-69.0229	-0.0045	0.0033	0.43	0.343	2
32768	−69.01656	0.00265	0.00257	0.46	0.234	-69.0168	0.0023	0.0025	0.46	3.046	8	-69.0168	0.0027	0.0025	0.46	0.765	2
65536	−69.02120	−0.00198	0.00180	0.46	0.484	-69.0206	-0.0013	0.0018	0.46	6.859	7	-69.0206	-0.0021	0.0018	0.46	1.968	2
131072	−69.01749	0.00172	0.00137	0.49	1.031	-69.0167	0.0024	0.0013	0.49	14.765	7	-69.0167	0.0017	0.0013	0.49	4.171	2
262144	−69.01886	0.00035	0.00109	0.56	2.140	-69.0180	0.0011	0.0011	0.56	29.953	7	-69.0180	0.0003	0.0010	0.56	8.796	2

4.0.9 Sequential Minimization

To solve the system (T), it is crucial to solve the minimization problem in (T), for which we have the following.

Lemma 4.0.3. *Let* $a = (a_1, \cdots, a_T, b_1, \cdots, b_T) \in A$ *and* $(X_{a,1}, \cdots, X_{a,T}) := (\Phi_1, \cdots, \Phi_T)$ *where* $\{\Phi_t\}_{t=1}^T$ *is given by (3.0.4) with* $(\eta_1, \cdots, \eta_T) \sim N(0, I)$.

(1) $X_{a,t} \sim N(\mu_t, \sigma_t^2)$ *where* (μ_t, σ_t^2) *are calculated iteratively by setting* $\mu_0 = 0, \sigma_0 = 0$ *and*

$$\begin{cases} \mu_t := a_t^2[\delta\mu_{t-1} - b_t], \\ \\ \sigma_t^2 := a_t^2 + \delta^2 a_t^4 \sigma_{t-1}^2, \end{cases} \qquad t = 1, \cdots, T.$$

(2) *If* $\xi \sim N(\mu, \sigma^2)$, *then* $\underset{(c_1,c_2)}{\operatorname{argmin}}\, V[c_1\xi^2 + c_2\xi - e^{-\xi}] = \Big(\frac{1}{2},\ -[1+\mu]\Big)e^{-\mu+\sigma^2/2}$.

(3) *For* ψ_t *given by (4.0.5),*

$$\underset{(\kappa,\ell)}{\operatorname{argmin}}\, \mathbf{V}\left[\psi_t(X_{at}, \kappa, \ell)\right] = (0, \nu h_t) + h_t e^{-\nu\mu_t + \nu^2\sigma_t^2/2}\Big(\frac{\nu^2}{2}, -\nu[1+\nu\mu_t]\Big)\ . \quad (4.0.7)$$

Proof. 1. Consider (3.0.4). Note that Φ_{t-1} depends only on $\eta_1, \cdots, \eta_{t-1}$, so it is independent of η_t. Since a linear combination of independent normal distributions is still normal, we see that each Φ_t is normal. Denote by μ_t and σ_t^2 the mean and variance of Φ_t. Taking the mean and variance of the equations in (3.0.4) we then obtain the induction formula for μ_t and σ_t.

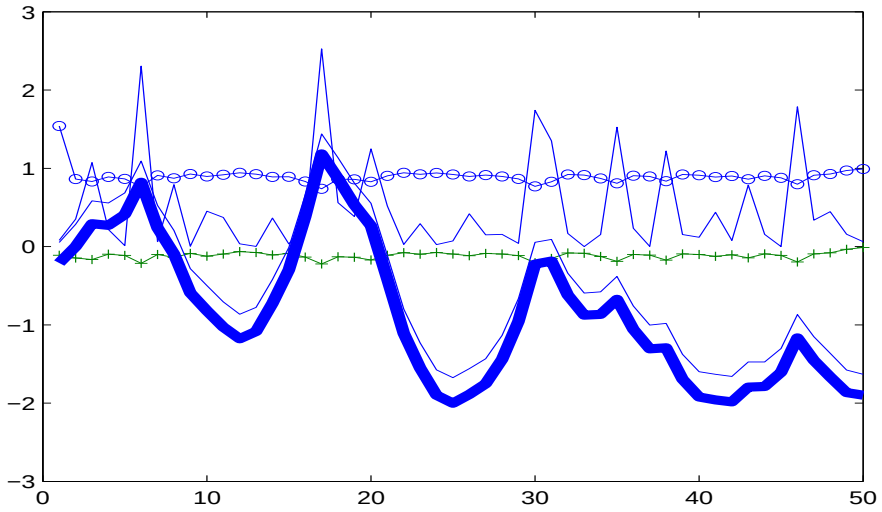
2. Let $v = \xi - \mu \sim N(0, \sigma^2)$. Then $E[v] = 0, E[v^2] = \sigma^2, E[v^3] = 0, E[v^4] = 3\sigma^2$ and

$$\mathbf{E}[ve^{-v}] = -\sigma^2 e^{\sigma^2/2}, \qquad \mathbf{E}[v^2 e^{-v}] = \sigma^2(1+\sigma^2)e^{\sigma^2/2}.$$

Set $u = e^{-v} - e^{\sigma^2/2}\{1 - v + \frac{1}{2}(v^2 - \sigma^2)\}$. We can check that $E[uv] = 0$ and $E[uv^2] = 0$. Hence,

$$\begin{aligned} \mathbf{V}[e^{-v} - c_1v^2 - c_2v] &= \mathbf{V}[u + e^{\sigma^2/2}\{1 - v + \tfrac{1}{2}(v^2 - \sigma^2)\} - c_1v^2 - c_2v] \\ &= \mathbf{V}[u] + \Big(c_1 - \tfrac{1}{2}e^{\sigma^2/2}\Big)^2\mathbf{V}[v^2] + \Big(c_2 + e^{\sigma^2/2}\Big)^2\mathbf{V}[v]. \end{aligned}$$

Figure 5: UIS Curves (a)



Dots on top are $\{a_t^*\}$ and in the middle are $\{b_t^*\}$. The thin curve is $\{h_t^*\}$, thick one is $\{x_t^*\}$ and dotted one is $\{x_t^* + \overline{x}_t\}$.

This implies that

$$\underset{(c_1,c_2)}{\operatorname{argmin}} \mathbf{V}[e^{-v} - c_1 v^2 - c_2 v] = (\tfrac{1}{2},\ -1)\, e^{\sigma^2/2}.$$

Consequently, using $\xi = v + \mu$ we have

$$\begin{aligned}
\underset{(c_1,c_2)}{\operatorname{argmin}} \mathbf{V}[e^{-\xi} - c_1\xi^2 - c_2\xi] &= \underset{(c_1,c_2)}{\operatorname{argmin}} \mathbf{V}[e^{\mu}(e^{-\xi} - c_1\xi^2 - c_2\xi)] \\
&= \underset{(c_1,c_2)}{\operatorname{argmin}} \mathbf{V}[e^{v} - c_1 e^{\mu} v^2 - (c_2 + 2\mu\, c_1) e^{\mu} v] \\
&= \Big(\tfrac{1}{2},\ -[1+\mu]\Big) e^{-\mu+\sigma^2/2}.
\end{aligned}$$

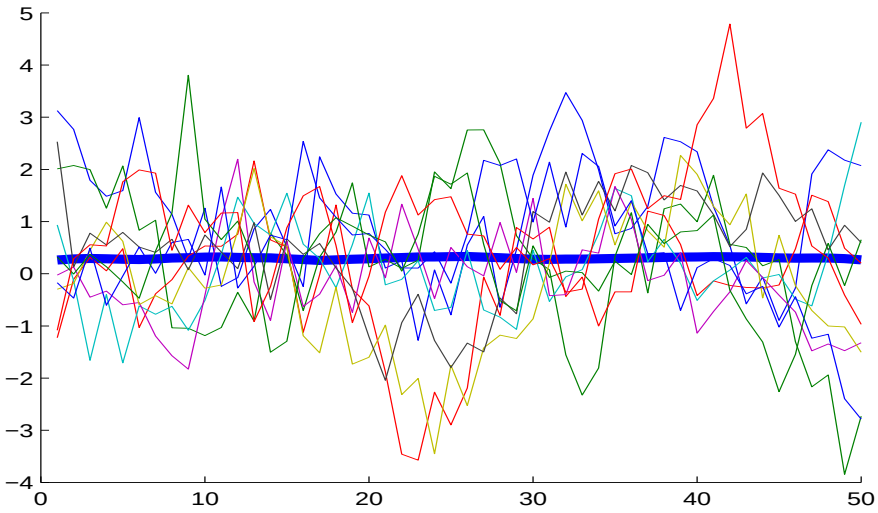
3. The third assertion is trivially true when $h_t = 0$. Hence assume $h_t > 0$. Set $\xi = \nu X_{a,t}$; then $\xi \sim N(\nu\mu_t, \nu^2\sigma_t^2)$. It follows from $\psi_t(z,\kappa,\ell) = \kappa z^2 + (\ell - \nu h_t)z - h_t(e^{-\nu z} - 1)$

that

$$\begin{aligned}\underset{(\kappa,\ell)}{\operatorname{argmin}}\ \mathbf{V}[\psi_t(X_{a,t},\kappa,\ell)] &= \underset{(\kappa,\ell)}{\operatorname{argmin}}\ \mathbf{V}\Big[\frac{\psi_t(X_{a,t},\kappa,\ell)}{h_t}\Big] \\ &= \underset{(\kappa,\ell)}{\operatorname{argmin}}\ \mathbf{V}\Big[\frac{\kappa}{\nu^2 h_t}\xi^2+\frac{\ell-\nu h_t}{\nu h_t}\xi-e^{-\xi}\Big].\end{aligned}$$

The third assertion of the Lemma that follows from the second assertion.

Figure 6: UIS Curves (b)



Ten sample paths (points) used in MCI. The middle one is expectation $\{\overline{x}_t\}$ of all sample paths, obtained by $\overline{x}_t = \{a_t^{*2}(\delta\overline{x}_{t-1}) - b_t^*\}$.

Now since (p_t, q_t) in (3.0.6) is equivalent to a in (3.0.7), we see that (T) can be solved as follows

Theorem 1. *A parameter* $a = (a_1, \cdots, a_T, b_1, \cdots, b_T) \in (0,\infty)^T \times R^T$ *is a solution*

of (T) if and only if there exist $(\mu_1, \cdots, \mu_T, \sigma_1, \cdots, \sigma_T)$ *such that*

$$\begin{cases} a_t &= \left(1 + \delta^2[1 - a_{t+1}^2 - \delta_{t1}] + \nu^2 h_t e^{-\nu\mu_t + \nu^2\sigma_t^2/2} \right)^{-\frac{1}{2}}, \\ b_t &= \delta a_{t+1}^2 b_{t+1} + \nu h_t[1 - (1 + \nu\mu_t)e^{-\nu\mu_t + \nu^2\sigma_t^2/2}], \\ \mu_t &= a_t^2[\delta\mu_{t-1} - b_t], \\ \sigma_t &= a_t\sqrt{1 + \delta^2 a_t^2 \sigma_{t-1}^2} \end{cases} \qquad \forall\, t = 1, \cdots, T. \tag{4.0.8}$$

Here $\sigma_0 := 0, \mu_0 := 0, a_{T+1} := 1, b_{T+1} := 0, \delta_{ts} := 1$ *if* $t = s$ *and* $\delta_{ts} := 0$ *if* $t \neq s$.

Remark 4.0.5. *(1) It is important to observe that the "optimal" parameters, being the solution of (4.0.8), do not depend on the CRNs. This is fundamentally different from the sequential "optimal" parameters obtained from the criterion (S) discussed in the earlier section. Since in using (S), most of the computing time is spent on the searching* $a^* = \lim_{k\to\infty} a^{(k)}$ *which depends on the CRN* $\{\{\eta_t^i\}_{t=1}^T\}_{i=1}^n$*, the advantage of using (4.0.8) is numerically efficient and theoretically sound.*

(2) At the optimal parameters, one finds that,

$$\mathbf{V}[\psi_t(X_{a,t}, p_t(a), q_t(a))] = h_t^2 e^{-2\nu\mu_t + \nu^2\sigma_t^2}\Big(e^{\nu^2\sigma_t^2} - 1 - (\nu\sigma_t)^2 - \tfrac{1}{2}(\nu\sigma_t)^4\Big).$$

Also, one can show that the optimal parameters satisfy

$$0 < a_t^2 \leqslant 1 \ (t = 2, \cdots, T), \qquad 0 < a_1^2 \leqslant \frac{1}{1-\delta^2}, \qquad 0 < \sigma_t \leqslant \frac{1}{1-\delta^2} \ \ (t = 1, \cdots, T).$$

Note that $r_t = \beta e^{v_t}\varepsilon_t$ *where* $v_t = \nu x_t$ *and* $V[v_t] = \nu^2/(1-\delta)^2 \geqslant (\nu\sigma_t)^2$*. We see that at least when the stochastic part* v_t *of the volatility is not too large, our SISMCI method will be very effective. We omit the details.*

4.0.10 Numerical Implementation

1. The flow of our new sequential importance sampling goes as follows.

Universal Sequential Importance Sampling Monte–Carolo Integration

1. Find x^* by solving $D\phi(x^*) = \mathbf{0}$.
Define $h = (h_1, \cdots, h_T)$ as in (4.0.7).
2. Solve $a^* = (a_1, \cdots, b_T, b_1, \cdot, b_T)$ from (4.0.8).
Define $p = (p_1, \cdots, p_T, q_1, \cdots, q_T)$ as in (3.0.6).
3. Generate independent $N(0,1)$ distributed random draws $\{\ \{\eta_t^i\}_{t=1}^T\ \}_{i=1}^n$.
Define $\{\{X_t^i\}_{t=1}^T\}_{i=1}^n$ by $X_t^i = a_t^2[\delta X_{t-1}^i - b_t] + a_t\eta_t^i$.
4. Produce the output as $\ell = \ell_n \pm \sigma_n/\sqrt{n}$, computed by

$$\psi^i := \sum_{t=1}^T \psi_t(X_t^i, p_t, q_t),\ \overline{e^\psi} := \sum_{i=1}^n \frac{e^{\psi^i}}{n},\ \ell_n := \psi_0(a) + \ln \overline{e^\psi},$$
$$\sigma_n^2 := \frac{1}{n-1}\sum_{i=1}^n \Big(\frac{e^{\psi^i}}{\overline{e^\psi}} - 1\Big)^2.$$

2. Numerical Algorithm Calculating x^*. Numerically, the point x^* in Lemma 4.0.2 can be solved from the equation $D\phi(x^*) = 0$ via Newton's iteration:

$$x^* = \lim_{k\to\infty} x^{(k)}, \qquad x^{(0)} = \mathbf{0}, \quad x^{(k+1)} = x^{(k)} - (D^2\phi(x^{(k)})^{-1} D\phi(x^{(k)}). \tag{4.0.9}$$

Since ϕ is convex, the Newton's iteration scheme (4.0.9) converges unconditionally and fast. The numerical evaluation of $(D^2\phi)^{-1}D\phi$ is also very easy. Note the following

$$\frac{\partial\phi}{\partial x_t} = \frac{\nu}{2} - \frac{\nu R_t^2 e^{-\nu x_t}}{2\beta^2} + [1+\delta^2]x_t - \delta[x_{t-1} + x_{t+1}] \quad \Big(x_0 \equiv \delta x_1,\ x_{T+1} \equiv \delta x_T\Big),$$
$$\frac{\partial^2\phi}{\partial x_t \partial x_t} = \frac{\nu^2 R_t^2}{2\beta^2} e^{-\nu x_t} + 1 + \delta^2[1 - \delta_{t1} - \delta_{tT}],$$
$$\frac{\partial^2\phi}{\partial x_t \partial x_{t-1}} = -\delta, \qquad \frac{\partial^2\phi}{\partial x_t \partial x_s} = 0 \text{ if } |t-s| > 1.$$

Hence the evaluation of $y = (D^2\phi(x^{(k)}))^{-1}D\phi(x^{(k)})$ can be put into the form of solving

$$\begin{pmatrix} a_1 & c_1 & 0 & \cdots & 0 \\ b_1 & a_2 & c_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{T-2} & a_{T-1} & c_{T-1} \\ 0 & \cdots & 0 & b_{T-1} & a_T \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-1} \\ y_T \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{T-1} \\ f_T \end{pmatrix}.$$

This can be solved by the following Gaussian Elimination procedure: For $t = 1$ to $(T-1)$ do $\{a_{t+1} = a_{t+1} - b_t c_t/a_t,\ f_{t+1} = f_{t+1} - f_t b_t/a_t\}$,

$y_T = f_T/a_T$, for t=(T-1) to 1 do $y_t = (f_t - c_t y_{t+1})/a_t$.

3. The Algebraic System (4.0.8). The universal optimal parameter is the solution of (4.0.8). We solve it by the following straightforward scheme: start from $a_t \equiv 1$ and $b_t \equiv 0$, update the unknowns by the right-hand sides of (4.0.8), in the order of $(\sigma_1, \mu_1), \cdots, (\sigma_T, \mu_T), (a_T, b_T), \cdots, (a_1, b_1)$.

4.0.11 A Numerical Simulation

We take the same data used in the previous section. The results are summarized in Figure 4 and Table 5.

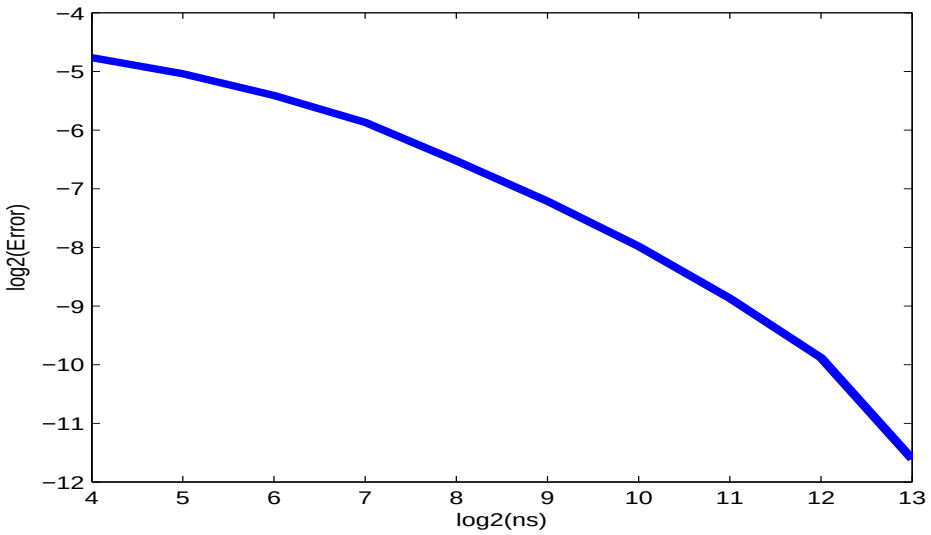
1. The Normalization. The Newton's iteration in finding x^* converges very fast; in about 5 iterations, we obtain x^* within 10^{-12} accuracy. Regarding $x^* = \{x_t^*\}_{t=1}^T$ as a path, we plot it as the thick curve in Figure 4 (a); the corresponding (positive) $\{h_t\}$ is plotted as the thin curve. In view of Figure 3 (b), one discovers that the path $\{x_t^*\}_{t=1}^T$ is close to all the mean paths of the results from the previous literature using sequential importance sample method. This is not a coincidence; it reflects the essence of importance sampling: important samples should be those that are near the point of maximum of the integrand, i.e., neat x^*.

Table 6: Numerical Comparison between UIS and SIS

n	L_n			Error (10^{-4})			$S_n/\sqrt{n}$ (10^{-4})			S_n			CPU Time		
	UIS	SIS (w. i.)	SIS	UIS	SIS (w. i.)	SIS	UIS	SIS (w. i.)	SIS	UIS	SIS (w. i.)	SIS	UIS	SIS (w. i.)	SIS
4	−68.9308	−69.0618	−68.9729	0.0885	−0.0426	0.0463	0.26605	0.01696	0.03241	0.53209	0.03391	0.06481	0.00	0.06	0.13
8	−68.9765	−68.8508	−69.0336	0.0427	0.1684	−0.0144	0.16309	0.06671	0.04901	0.46130	0.18869	0.13863	0.00	0.05	0.09
16	−68.8960	−69.0200	−69.1738	0.1232	−0.0008	−0.1546	0.09903	0.05598	0.04539	0.39613	0.22393	0.18157	0.00	0.05	0.11
32	−68.9600	−69.0799	−69.1139	0.0592	−0.0607	−0.0946	0.08075	0.05741	0.05943	0.45680	0.32478	0.33616	0.00	0.05	0.06
64	−69.0336	−69.0213	−69.0748	−0.0144	−0.0021	−0.0556	0.05626	0.05463	0.04975	0.45009	0.43702	0.39803	0.00	0.05	0.06
128	−69.0010	−69.0570	−69.0318	0.0182	−0.0377	−0.0125	0.04215	0.03299	0.03760	0.47692	0.37323	0.42536	0.00	0.08	0.06
256	−69.0099	−69.0581	−69.0234	0.0093	−0.0389	−0.0042	0.04187	0.02562	0.02831	0.66987	0.40997	0.45303	0.03	0.06	0.08
512	−69.0268	−69.0552	−69.0285	−0.0076	−0.0360	−0.0093	0.02619	0.01932	0.02148	0.59256	0.43721	0.48604	0.03	0.11	0.13
1024	−69.0151	−69.0316	−69.0153	0.0041	−0.0123	0.0039	0.01676	0.01317	0.01441	0.53623	0.42149	0.46124	0.03	0.17	0.19
2048	−69.0152	−69.0179	−69.0168	0.0040	0.0013	0.0024	0.01059	0.01002	0.00965	0.47937	0.45350	0.43691	0.05	0.25	0.34
4096	−69.0099	−69.0214	−69.0107	0.0094	−0.0022	0.0085	0.00735	0.00693	0.00678	0.47024	0.44334	0.43387	0.06	0.58	0.66
8192	−69.0192	−69.0198	−69.0195	0.0001	−0.0006	−0.0003	0.00525	0.00486	0.00498	0.47529	0.43994	0.45065	0.11	1.14	1.39
16384	−69.0189	−69.0188	−69.0188	0.0003	0.0005	0.0004	0.00356	0.00342	0.00350	0.45571	0.43838	0.44760	0.23	2.67	3.73
32768	−69.0193	−69.0179	−69.0194	0.0000	0.0013	−0.0002	0.00251	0.00238	0.00245	0.45382	0.43141	0.44266	0.52	6.63	7.69
65536	−69.0190	−69.0186	−69.0190	0.0002	0.0006	0.0002	0.00179	0.00171	0.00173	0.45928	0.43823	0.44357	1.34	17.34	20.95
131072	−69.0173	−69.0173	−69.0173	0.0019	0.0019	0.0019	0.00136	0.00121	0.00122	0.49283	0.43786	0.44198	2.78	103.06	110.44
262144	−69.0193	−69.0176	−69.0193	−0.0001	0.0017	−0.0001	0.00094	0.00086	0.00086	0.48368	0.43784	0.44159	5.70	450.22	485.73

2. The Universal Optimal Parameter. It takes about 25 iterations to obtain a fixed point of (4.0.8) with error smaller than 10^{-12} . The optimal parameter $\{a_t^*\}$, plotted as dots on top in Figure 4 (a), is basically the same as that from previous literature, which depend on the CRNs. Indeed our universal optimal parameter is the expectation over all possible CRNs of the local optimal parameter. When the sample size is large, the law of large number ensures that the local optimal parameters should be almost the same as the universal optimal parameter; see the last column in Table 3. Since the major portion of the drift for the important sample paths has been taken care of by the normalization, the universal parameter $\{b_t^*\}$ is quite small; see the dots in the middle in Figure 4 (a).

Figure 7: Sensitivity of Error Corresponding to Sample size

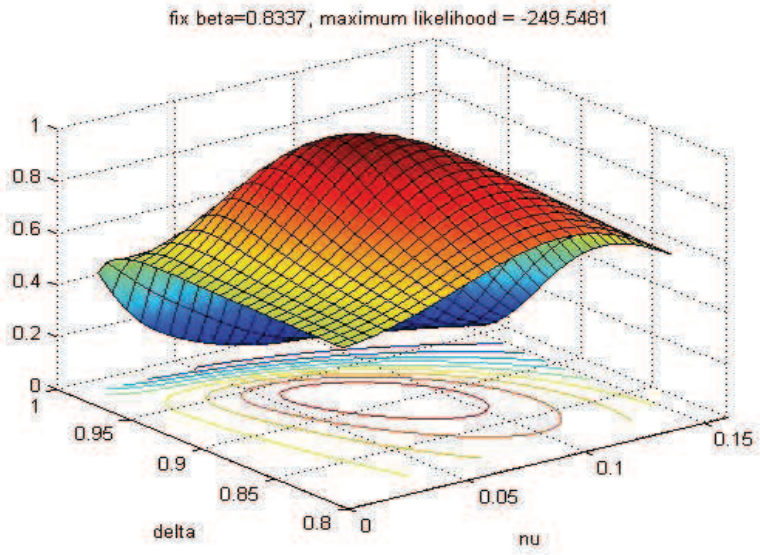


The curve represents $\log_2(Error)$ to $\log_2(n_s)$.

3. The Important Sample Paths. The mean of all important sample paths is $\{\bar{x}_t\}_{t=1}^T$ defined by $\bar{x}_t = a_t^2[\delta\bar{x}_{t-1} - b_t]$ with $\bar{x}_0 = 0$; see the thick almost horizontal cuve in the middle of Figure 4 (b). Denote $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_T)$. Then in the original coordinates

system, all the important sample points for the MCI is $x^* + \bar{x}$, instead of x^*, the point of maximum of integrand. This slight shift from x^* to $x^* + \bar{x}$ is indeed another fine tuning of the importance sampling method. After an "optimal" probability density function $\rho(\cdot, a^*)$ is chosen, major variation of the integrand $f(x)$ near its point x^* of maximum has been taken care of, so the center of sample points (= center of mass with density $\rho(\cdot, a^*)$) should be, or at least very close to, the point of maximum of the function $f(x)/\rho(x, a^*)$. The (coordinate-wise) positive shift from x^* to $x^* + \bar{x}$ reflects the fact that for every $h > 0$, the center of a mass with density $\exp(-h[e^{-\nu z} - 1 + \nu z])$ is positive.

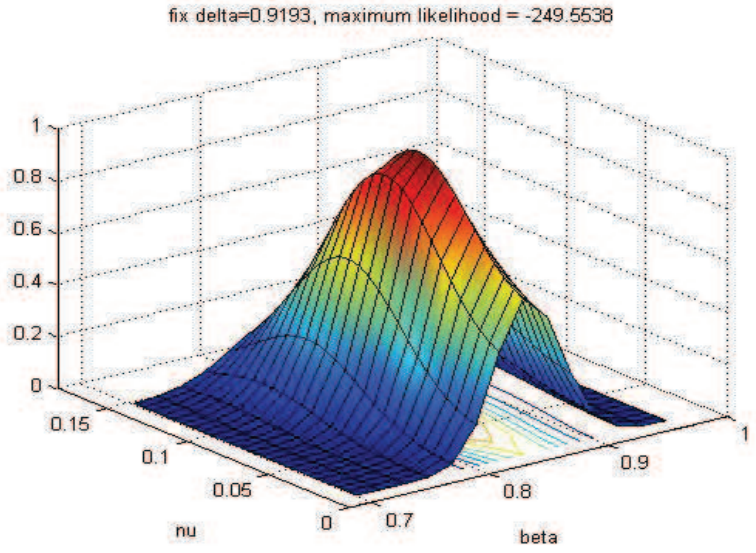
Figure 8: The Maximum Log Likelihood When $\beta = 0.8337$



It is not very difficult to see that if $\{X_t\}_{t=1}^T$ is a sample path used in our MCI, then the shifted path $\{X_t + x_t^*\}_{t=1}^T$ resembles the important sample paths used in [15]. Indeed, $x^* + \bar{x}$ should be the theoretical expectation over all CRNs of all the sample points.

4. Error Analysis. Since our parameter is universal, we can apply the central limit theorem to conclude that ℓ_n is asyptotically $N(\ell, \sigma^2/n)$ distributed so that $\ell = \ell_n \pm \sigma_n/\sqrt{n}$ where σ_n is the sample relative standard deviation. The result is displayed in Table 4. Basically, the sizes of the errors are comparable to that of previous literatures.

Figure 9: The Maximum Log Likelihood When $\delta = 0.9193$



4.0.12 Maximum Likelihood

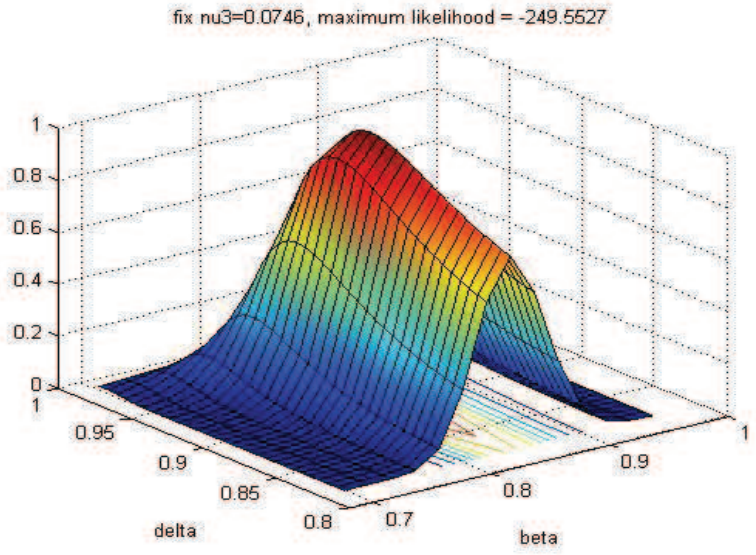
To test another sequence of data of size 200, with parameters $\beta = 0.8, \nu = 0.07, \delta = 0.9$, the maximum likelihood was applied to search the optimal parameters. The following graphs are used here to show what the likelihood function values look like around the optimal points.

4.0.13 Conclusion.

The new method proposed here has a few advantages over the one discussed in the previous section.

1. The use of universal parameter saves the computing time.
2. Theoretical analysis of the error for our new method is very simple since it involves only the classical analysis on the average of i.i.d random variables.

Figure 10: The Maximum Log Likelihood When $\nu = 0.0746$



3. Numerically the initial step of normalization can be totally omitted from numerical algorithm (with formula (4.0.8) revised of course); theoretically the normalization step allows us to see clearly how the method of SISMCI worked; it provides an important clue for further theoretical work, e.g. efficient sampling should have the property that the center of mass with density $\rho(\cdot, a^*)$ should be the point of maximum of the function $f(\cdot)/\rho(\cdot, a^*)$.

5.0 CONCLUSION

The sequential Importance Sampling depends on the particular common random draws to figure out the suitable fitted values for the suboptimal parameters. This is limited and inflexible. The estimators we derive in our another approach present a general formula which is flexible for financial applications. The new method proposed in this paper has quite a few advantages over the one discussed in the previous sections. The use of universal parameter saves the computing time and is more efficient. Theoretical analysis of the error for our new method is very simple since it involves only the classical analysis on the average of i.i.d random variables. Numerically the initial step of normalization can be totally omitted from numerical algorithm. Theoretically the normalization step allows us to see clearly how the method of SISMCI works. Finally it provides a valuable clue for further theoretical work.

BIBLIOGRAPHY

[1] Cartea, Alvaro and Figueroa, Marcelo G.(2004), "Pricing in electricity Markets: a mean reverting jump diffusion model with seasonality", *University of London*, bf Sep.

[2] Choi, J. E., (2004), "Estimation of Stochastic Volatility Models by Simulated Maximum Likelihood Method", *University of Waterloo*.

[3] DeJong, C. (2006), "The Nature of Power Spikes: A Regime-Switch Approach", *Studies in Nonlinear Dynamics&Econometrics*, **10**, No. 3, July.

[4] Durbin, J., and Koopmans, S.J. (1997), "Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models", *Biometrika*, **84**, 669–684.

[5] Fishman, G.S. (1996), "Monte Carlo Concepts, Algorithms, and Applications", *New York: Springer-Verlag*.

[6] Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, **57**, 1317-1339. - -(1996), "Monte Carlo Simulation and Numerical Integration," in *The Handbook of Computational Economics*, Vol. 1., eds. H. Amman, D. Kendrick, J. and Rust, Amsterdam: North Holland.

[7] Geweke, J., and Zhou, G., (1996), "Measuring the Pricing Error of the Arbitrage Price Theory," *Review of Financial Studies*, **9**, 557-587.

[8] Geweke, J., and Amisano, G., (2007), "Hierarchical Markov normal mixture models with applications to financial asset returns", *working paper.*

[9] Jaillet, P., Ronn, E. I., and Tompaidis, S. (2003), "Valuation of Commodity-Based Swing Options", *Management Science*, **Dec.**

[10] Melino, A., and Turnbull, S. M. (1990), "Pricing Foreign Currency Options with Stochastic Volatility", *Journal of Econometrics*, **45**, 239–265, North-Holland.

[11] Nelson, D. B. (1990), "ARCH Models as Diffusion Approximations", *Journal of Econometrics*, **45**, 7–38.

[12] Nelson, D. B. (1994), "Bayesian Analysis of Stochastic Volatility Models: Comment", *Journal of Business and Economic Statistics*, **Vol. 12**, 4,Oct., 403–406.

[13] Owen, A., and Zhou, Y. (2000), "Safe and Effective Importance Sampling," *Journal of the American Statistical Association*, **95**, 135–143.

[14] Richard, J.F., and Liesenfeld, R. (2004), "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models", *CAU Economic Working Paper*, 12–2004.

[15] Richard, J.F., and Zhang, W. (2007), "Efficient High-Dimensional Importance Sampling", *Journal of Econometrics*, **Vol 7**, Feburary.

[16] Stern, S. (1997), "Simulation-Based Estimators", *Journal of Economic Literature*, **35**, 2006-2039.

[17] Hull, J., and White, A., "The Pricing of Options on Assets with Stochastic Volatilities" *Journal of Finance*, **Vol. 42, No. 2**, pp. 281-300, June 1987.

[18] Shiryayev,A.N., *Probalility*, Springer-Verlag, New Yorm, 1979.

[19] Wilmott, P., *Derivatives: The Theory and Practice of Financial Engineering*, John wile & Sons, 1998.