VaR/CVaR Estimation under Stochastic Volatility Models

Chuan – Hsiang Han[1]
Wei – Han Liu[2,*]
Tzu – Ying Chen[3]

This version: 11 January 2013

.

1. Department of Quantitative Finance, National Tsing-Hua University, Hsinchu, Taiwan 30013, R.O.C., E-mail: chhan@mx.nthu.edu.tw. Work supported by NSC 97-2115-M-007-002-MY2, Taiwan.
2. Department of Finance, La Trobe University, Melbourne, Australia. E-mail: weihanliu2002@yahoo.com.
3. Department of Finance, National Taiwan University, Taipei, Taiwan. E-mail: cmqf727@hotmail.com
* Corresponding author

Title: VaR/CVaR Estimation under Stochastic Volatility Models

Abstract: This paper proposes an improved procedure for stochastic volatility model estimation with an application to Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) estimation. This improved procedure is composed of the following instrumental components: Fourier transform method for volatility estimation, and importance sampling for extreme event probability estimation. The empirical analysis is based on several foreign exchange series and the S&P 500 index data. In comparison with empirical results by RiskMetrics, historical simulation, and the GARCH(1,1) model, our improved procedure outperforms on average.

Keywords: stochastic volatility, Fourier transform method, importance sampling, (conditional) Value-at-Risk, backtesting.

JEL classification: C13; C14; C63

Section 1: Introduction

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are two of the popular risk measures for determining the regulatory risk capital amount. There are two major directions for Value-at-Risk (VaR) and conditional Value-at-Risk (CVaR) estimation, two of the most popular risk measures: modeling the return distribution and capturing the volatility process (Jorion 2007). For the former direction, various techniques are employed for modeling the whole return distribution or just the tail areas, including known parametric distribution, kernel density approximation, and extreme value theory, etc (Tsay 2010). The latter direction mostly relies on discrete-time volatility models such as EWMA (Exponentially Weighted Moving Average model) and GARCH to capture the volatility process. See Jondeau et al. (2007) for further details.

One of the innovative alternatives for risk measurement is to apply stochastic volatility models because they are more appropriate to represent the behavior of financial returns, from a theoretical point of view (Pederzoli 2006). The stochastic volatility models under the continuous-time framework are known for capturing some stylized features of financial data. Those models are intensively applied to option pricing and hedging issues. Some innovative procedures are proposed and representative procedures are refined. For example, Fouque et al. (2000) derive option pricing and hedging approximation formula under stochastic volatility models by means of a singular perturbation technique. Under an equivalent martingale (risk-neutral) probability measure, Lehar et al. (2002) calibrate a stochastic volatility model. Those two studies incorporate an option pricing approximation for model calibration and use simulated option price change to approximate VaR estimate, which is based on the empirical percentile of the simulated data. While these approximations are instrumental in reducing computational cost, there is still room of improvement for estimation accuracy because those efforts are exclusively based on approximations.

Distinct from previous studies, we choose to value exact computation and incorporate the help of refined efficient simulation method for more accurate estimation. Accordingly, there is no need to incorporate approximation results and the approximation errors can be avoided. However, there are two major hurdles in this alternative direction, including (1) unstable parameter estimation subject to practical data constraints, and (2) lack of efficient computational technique for accurately estimating risk measures. We propose an improved procedure for VaR and CVaR estimation with stochastic volatility models under the historical (or physical) probability measure by proposing improvements in the following two aspects: (1) improved stochastic volatility model estimation scheme by refining the Fourier transform method (Malliavin and Mancino (2002, 2009)), and (2) enhanced importance sampling for estimating extreme event probability.

For the first aspect of the improvements, we propose a refined procedure for stochastic volatility parameter estimation. There are various estimation procedures for stochastic volatility model parameter estimation. The major procedures include method of moments, generalized method of moments, maximum likelihood estimators, quasi maximum likelihood, etc. See Broto and Ruiz (2004) for details. Among them, Yu (2010) emphasizes simulation-based estimation methods for its superior performance. Simulation-based estimation methods include at least simulated maximum likelihood, simulated generalized method of moments, efficient method of moments, indirect inference and Markov chain Monte Carlo, etc. While acknowledged for the help form simulation-based estimation methods, all these aforementioned estimation methods depend on the assumed parametric probability distribution of the underlying asset return.

Among the non-parametric approaches, several recent literatures exploit the quadratic variation formula to estimate the integrated volatility. See Zhang et al. (2005) and references therein. Meddahi (2002) concludes that quadratic variation plays a central role in the option pricing literature. In particular, when there are no jumps, quadratic variation equals the integrated volatility highlighted by Hull and White (1987). In diffusion models, the volatility refers to either the instantaneous diffusion coefficient or the quadratic variation over a given time period. The latter is often called the integrated volatility. Using integrated volatility to approximate instantaneous or spot volatility is possibly infeasible because its differentiation procedures may be numerically unstable and its modeling performance varies with data frequency, as cautioned by Malliavin and Mancino (2009). The two authors propose a non-parametric Fourier transform method to estimate spot volatility under continuous semi-martingale models. Notably, this method makes the estimation feasible by relying on the integration of the time series, rather than on its differentiation as in previous literature. The estimation process is based on the computation of Fourier coefficients of the variance process, rather than on quadratic variation which demands additional assumptions for estimation. The authors conclude that this approach is particularly suitable for the analysis of high frequency time series and for the computation of cross volatilities. We thus adopt this Fourier transform method by Malliavin and Mancino (2009) as the framework for stochastic volatility model estimation.

For the second aspect of improvement, refined brute force is instrumental for risk measurement (Gregoriou 2009) and it is also essential for VaR and CVaR estimation under stochastic volatility models. In general, there is no closed-form solution for these risk measures of interest under stochastic volatility models and we need to seek help from simulation. Among the major simulation methods, importance sampling is especially helpful for the estimation issues of the tail areas (McNeil, Frey, and Embrechts 2005). Importance sampling can effectively improve convergence of sample means particularly in rare event simulation while direct Monte Carlo simulation suffers pitfalls like variance augmentation and slow convergence (Glasserman 2003, Lemieux 2009). We further propose an enhanced version of importance sampling for estimating extreme event probability. The theoretical background of our proposed importance sampling combines the large deviation theory (Bucklew 2004) which has the averaging effect on realized variance (Fouque et al. 2000) and provides a sharp estimate for the decay rate of small probabilities. This methodology is useful in handling the heavy (or fatter) tail distributions induced by stochastic volatility models.

Empirical analyses confirm the outperformance of VaR estimation by our proposed improved procedure which integrates the Fourier transform method and the refined importance sampling technique.

Two datasets are used for empirical examination: the first one contains three foreign exchange series (January 5, 1998 to July 24, 2009) and the second one contains S&P 500 index and its VIX[1] (the measure of the implied volatility of S&P 500 index options, January 3, 2005 to July 24, 2009). Both data periods cover both tranquil and turbulent times. Three popular types of backtesting are conducted for model evaluation and performance comparison in VaR estimation. Our proposed procedure significantly outperforms especially at 99% VaR estimates, as compared with RiskMetrics, historical simulation, and GARCH(1,1) model. This outperformance matches the demands from the Basel II Accord implementation (Jorion 2007) for determining regulatory risk capital level, i.e. risk measurement at 99% confidence level.

The organization of this paper is as follows. Section 2 introduces the general one-factor stochastic volatility model, the extreme event probability estimation, and their relationship with VaR and CVaR estimation. Section 3 reviews the Fourier transform method, one of the nonparametric approaches to estimate volatility in time series. Section 4 discusses the construction of the efficient importance sampling estimators for extreme event probabilities, then solve for VaR and CVaR estimation. Section 5 investigates backtesting results of VaR estimation over three foreign exchange rate series and S&P 500 index with its VIX, and compare these results with some well known methods such as RiskMetrics, historical simulation, and GARCH(1,1). Section 6 concludes.

Section 2: VaR and CVaR Estimation under Stochastic Volatility Models

[1] Chicago Board Options Exchange Volatility Index, a popular measure of the implied volatility of S&P 500 index options, http://www.cboe.com/

The Black-Scholes model is fundamental in option pricing theory under no-arbitrage condition (Hull 2008), which simply assumes that log returns of risky asset prices are normally distributed. A stochastic volatility model is an extension of the Black-Scholes model which relaxes the assumption of constant volatility and allows volatility to be driven by other processes. Under a probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{0\le t\le T}\right)$, a general form of one-factor stochastic volatility model is defined by

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma_t S_t W_t^{(0)}, \\ \sigma_t &= f(Y_t), \\ dY_t &= c(Y_t)dt + g(Y_t)\left(\rho dW_t^{(0)} + \sqrt{1-\rho^2}\, dW_t^{(1)}\right), \end{aligned}$$

(2.1)

where S_t denotes the underlying asset price process with a constant growth rate μ and a random volatility σ_t. The vector $\left(W_t^{(0)}, W_t^{(1)}\right)$ consists of two independent standard Brownian motions and ρ denotes the instantaneous correlation coefficient satisfying $|\rho| \le 1$. We further assume that the volatility function f and coefficient functions of Y_t, namely $(c(y), g(y))$, satisfy classical assumptions such that the whole dynamic system (2.1) fulfills the existence and uniqueness conditions for the strong solution of stochastic differential equations (Oksendal 1998). The stochastic volatility model (2.1) is one-factor because its random volatility σ_t is driven by a single factor process Y_t, also known as the driving volatility process. This process is often assumed mean-reverting. Typical examples include Ornstein-Uhlenbeck process and square-root process like Heston model (Heston 1993)[2].

---

[2] For these two processes, their coefficient functions $(c(y), g(y))$ are $(\alpha(m-y), \beta)$ and $(\alpha(m-y), \beta\sqrt{y})$, respectively, where y denotes the variable of driving volatility. Other model parameters (α, m, β) denote the rate of mean reversion, long-run mean, and the volatility of volatility, respectively. The volatility function $f(y)$ is customarily chosen as $\exp(y/2)$ and $\sqrt{y}$, respectively.

For financial applications such as option pricing, hedging, and risk management, one often needs to compute the following conditional expectation under model (2.1) given a Markovian assumption:

$$P(t,x,y) = E\{e^{-r(\tau-t)}H(S_\tau) \mid S_t = x,\ Y_t = y\}, \tag{2.2}$$

where the value function P may represent an option price or a hedging ratio given state variables of asset price x and its driving volatility y at time $t \geq 0$. Other parameters or variables in (2.2) include the discounting rate r, the payoff function H, and the exercise time $\tau\left(\geq t\right)$, that can be either a fixed maturity, say $\tau = T$, or a stopping time. Equation (2.2) is not necessarily defined under the real world P-measure unless a derivative pricing problem is considered. In this case, Equation (2.2) has to be evaluated under an equivalent martingale measure so that the no-arbitrage condition holds. Above all, as an example in risk management, Equation (2.2) can be used to depict the tail areas, i.e. specific probability levels for extreme events under the historical or physical probability measure.

For our purpose to estimate VaR and CVaR, we shall first consider the estimation of an extreme event probability under the general formulation in (2.2). We do this by choosing the discounting rate $r = 0$, the payoff function as a rare-event indicator $H(x) = \mathbf{I}(x \leq D)$, where D denotes a threshold or cutoff point, and the exercise time $\tau = T > 0$ a fixed date. Hence, the time-T probability of an extreme event for logarithmic returns conditional at time 0 is defined by

$$P(0,x,y;D) = E\{\mathbf{I}(\ln(S_T / S_0) \leq D \mid S_0 = x,\ Y_0 = y\}. \tag{2.3}$$

There are two major cases where the probability P is rather small: (1) large negative threshold D, and (2) small expiration time T. The statistics of Value-at-Risk, denoted by VaR_α, is the $(1\text{-}\alpha)\times 100$ percentile of logarithmic returns, where significance level $0 \le \alpha \le 1$. Hence, the exact calculation of VaR_α ends up solving a nonlinear equation

$$1-\alpha = P(0,x,y;VaR_\alpha)\,. \tag{2.4}$$

CVaR is simply a conditional expectation given that losses are greater than or equal to the VaR_α. Due to the complexity of stochastic volatility models, there is no closed-form solution in general for either the pricing or hedging value defined in (2.2) or the extreme event probability P defined in (2.3). Thus, computational challenges arise when one needs to obtain VaR_α by solving Equation (2.4).

Section 3: Volatility Estimation: Fourier Transform Method

The Fourier transform method (Malliavin and Mancino 2002, 2009) is a nonparametric method to estimate a multivariate volatility process. Its main idea is to reconstruct volatility as the time series in terms of sine and cosine functions under the following continuous semi-martingale assumption: Let u_t be the log-price of an underlying asset S at time t, i.e. $u_t = \ln S_t$, and follow a diffusion process

$$du_t = \mu_t dt + \sigma_t dW_t, \tag{3.1}$$

where μ_t is the instantaneous growth rate and W_t is a one-dimensional standard Brownian motion. The time series volatility σ_t can be estimated by following the steps below.

Step 1: Compute the Fourier coefficients of the underlying u_t as follows:

$$a_0(du) = \frac{1}{2\pi}\int_0^{2\pi} du_t\ , \tag{3.2}$$

$$a_k(du) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)du_t \,, \tag{3.3}$$

$$b_k(du) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)du_t \,, \tag{3.4}$$

for any $k \geq 1$, so that $u(t) = a_0 + \sum_{k=1}^{\infty}\left[-b_k(du)\cos(kt) + a_k(du)\sin(kt)\right]$. Note that the original time interval $[0,T]$ can always be rescaled to $[0,2\pi]$ as shown in above integrals.

Step 2: Compute the Fourier coefficients of variance σ_t^2 as follows:

$$a_k(\sigma^2) = \lim_{N\to\infty}\frac{\pi}{2N+1}\sum_{s=-N}^{N-k}\left[a_s^*(du)a_{s+k}^*(du) + b_s^*(du)b_{s+k}^*(du)\right], \tag{3.5}$$

$$b_k(\sigma^2) = \lim_{N\to\infty}\frac{\pi}{2N+1}\sum_{s=-N}^{N-k}\left[a_s^*(du)b_{s+k}^*(du) - b_s^*(du)a_{s+k}^*(du)\right], \tag{3.6}$$

for $k \geq 0$, in which $a_s^*(du)$ and $b_s^*(du)$ are defined by

$$a_s^*(du) = \begin{cases} a_s(du), & if \ \mathrm{s} > 0 \\ 0, & if \ \mathrm{s} = 0 \\ a_{-s}(du), & if \ \mathrm{s} < 0 \end{cases} \quad and \quad b_s^*(du) = \begin{cases} b_s(du), & if \ \mathrm{s} > 0 \\ 0, & if \ \mathrm{s} = 0 \\ -b_{-s}(du), & if \ \mathrm{s} < 0. \end{cases}$$

Step 3: Reconstruct the time series of variance σ_t^2 by

$$\sigma_t^2 = \lim_{N\to\infty}\sum_{k=0}^{N}\varphi(\delta k)[a_k(\sigma^2)\cos(kt) + b_k(\sigma^2)\sin(kt)] \,, \tag{3.7}$$

where $\varphi(x) = \frac{\sin^2(x)}{x^2}$ is a smooth function with the initial condition $\varphi(0) = 1$ and δ is a smoothing parameter typically specified as $\delta = \frac{1}{50}$ (Reno 2008).

From Equations (3.2)-(3.4), it is observed that the integration error of Fourier coefficients is adversely proportional to data frequency. This Fourier transform method is easy to implement because, as shown in (3.5) and (3.6), the Fourier coefficients of the variance can be approximated by a finite sum of multiplications of a^* and b^*. This integration method can accordingly avoid drawbacks inherited from those traditional methods based on the differentiation of quadratic variation.

3.1 Stochastic Volatility Model Estimation

Given that the volatility time series is estimated by Fourier method by Malliavin and Mancino (2002, 2009), we proceed to estimate stochastic volatility model parameters. Assuming that the volatility process $\sigma_t = \exp(Y_t/2)$ and the driving volatility process Y_t is governed by the Ornstein-Uhlenbeck process, i.e.

$$dY_t = \alpha\left(m - Y_t\right)dt + \beta dW_t, \tag{3.8}$$

Based on the estimated variance by Fourier transform method specified by (3.7), we further estimate model parameters (α, β, m) of $Y_t = \ln \sigma_t^2$ in (3.8) by means of the maximum likelihood method. For a given set of observations $Y_1, \ldots, Y_N$, the likelihood function is given as

$$L(\alpha,\beta,m) = \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\beta^2\Delta_t}} \exp\left\{-\frac{1}{2\beta^2\Delta_t}\left[Y_{t+1} - \left(\alpha m\Delta_t + \left(1-\alpha\Delta_t\right)Y_t\right)\right]^2\right\}, \tag{3.9}$$

where Δ_t and N denote the length of discretized time interval and sample size, respectively. This likelihood function is obtained by discretizing the stochastic differential equation (3.8). Taking the natural logarithm and ignoring the constant term, the log-likelihood becomes

$$\ln L(\alpha,\beta,m) \propto N\ln\beta + \frac{1}{2\beta^2\Delta_t}\sum_{t=1}^{N-1}\left[Y_{t+1} - \left(\alpha m\Delta_t + \left(1-\alpha\Delta_t\right)Y_t\right)\right]^2.$$

By maximizing the right hand side over the parameters (α,β,m), we obtain the following maximum likelihood estimators

$$\hat{\alpha} = \frac{1}{\Delta_t}\left[1-\frac{\left(\sum_{t=2}^{N}Y_t\right)\left(\sum_{t=1}^{N-1}Y_t\right)-(N-1)\left(\sum_{t=1}^{N-1}Y_tY_{t+1}\right)}{\left(\sum_{t=1}^{N-1}Y_t\right)^2-(N-1)\left(\sum_{t=1}^{N-1}Y_t^2\right)}\right], \tag{3.11}$$

$$\hat{\beta}=\sqrt{\frac{1}{N\Delta_t}\sum_{t=1}^{N-1}\left[Y_{t+1}-\left(\alpha m\Delta_t+\left(1-\alpha\Delta_t\right)Y_t\right)\right]^2}, \tag{3.12}$$

$$\hat{\mathrm{m}}=\frac{-1}{\hat{\alpha}\Delta_t}\left[\frac{\left(\sum_{t=2}^{N}Y_t\right)\left(\sum_{t=1}^{N-1}Y_t^2\right)-\left(\sum_{t=1}^{N-1}Y_t\right)\left(\sum_{t=1}^{N-1}Y_tY_{t+1}\right)}{\left(\sum_{t=1}^{N-1}Y_t\right)^2-(N-1)\left(\sum_{t=1}^{N-1}Y_t^2\right)}\right]. \tag{3.13}$$

These are estimators of mean-reverting rate, volatility of volatility, and long-run mean, respectively. The parameters in the stochastic volatility model specified by (3.8) are thus estimated and obtained. This estimation framework is flexible enough to be extended to local volatility based Heston model or the hybrid model (a combination of both models). When a correlation between the Brownian motions defined by (3.1) and (3.8) is modeled, one can apply the multivariate Fourier estimation method proposed by Barucci and Mancino (2010). Detailed discussions are summarized in Han (2013.

Section 4: Importance Sampling: Variance Reduction

When dealing with sparse observations in the tails, the basic Monte Carlo simulation is handicapped for its undesirable properties, e.g. large relative error and data clustering around the center, etc. Importance sampling is one of the major methods of variance reduction to improve the convergence of basic Monte Carlo method. The fundamental idea behind importance sampling is to relocate the original density function to the area of interest with properly assigned weights. The relocated density typically incurs more occurrences of rare events so that a more accurate estimate for a small probability can be achieved. This technique is extremely helpful in rare event simulation. See Bucklew (2004) for discussions on importance sampling and extreme event probability estimation.

There are two major categories of the studies to investigate methods of importance sampling and their efficiency (Lemieux 2009). The first category aims to reduce the variance of an importance sampling estimator as much as possible. This approach often ends up solving a fully nonlinear optimization problem, possibly in high dimension, or solving a simplified optimization problem derived from some approximation techniques. The second category emphasizes on minimizing the variance rate of an importance sampling estimator. The notion of variance rate is defined as the difference between the decay rate of the second moment and the decay rate of the square of the first moment. It is treated as a measure of efficiency for importance sampling estimators. When zero variance rate (note: not variance itself) is achieved, the corresponding importance sampling estimator is known as asymptotically optimal or efficient. The second category is extensively applied to problems of rare event simulation.

Our proposed importance sampling estimation algorithm emerges from the second category. In contrast with the first category, the second category offers noticeable advantages, such as easy implementation, reduced computational cost, and analytical tractability. The first two advantages are manifest in itself, as opposed to the solving high-dimensional nonlinear optimization problems in the first category. The third advantage helps link our proposed algorithm for extreme event simulations with large deviation theory (Bucklew 2004) which provides sharp estimates for the decay rate of small probabilities.

We proceed to introduce the proposed algorithm of enhanced importance sampling with respect to the Black- Scholes model and stochastic volatility models. According to the ergodic property of the averaged variance process, the constant volatility of the Black-Scholes model can be viewed as a limiting case of some stochastic volatility model. See Fouque et al. (2000) for details. Namely, one can treat a stochastic volatility model as a perturbation around the Black-Scholes model. It is natural to first study the Black-Scholes model so as to investigate an importance sampling estimator for stochastic volatility model. Based on the large deviation principle of normal random variables, an efficient importance sampling algorithm and its variance analysis based on the Black-Scholes model are established in Section 4.1. Under a stochastic volatility environment, we first carry out its limiting volatility, or called effective volatility, then apply the aforementioned importance sampling estimator to the Black-Scholes model, and then to stochastic volatility models. This is detailed in Section 4.2.

4.1 Black-Scholes Model

Since the Black-Scholes model assumes that the risky asset price follows a geometric Brownian motion $dS_t = \mu S_t dt + \sigma S_t dW_t$, its logarithmic return $\ln\left(S_T/S_0\right)$ is normally distributed for $T > 0$. As a result, the extreme event probability with the threshold D, denoted by $P\left(0, S_0\right)$, admits the closed-form solution

$$\begin{aligned} P\left(0,S_0\right) &= E\left\{\mathbf{I}\left(\ln\left(S_T/S_0\right) \le D\right) \middle| S_0\right\} \\ &= N\left(\frac{D-\left(\mu-\sigma^2/2\right)T}{\sigma\sqrt{T}}\right), \end{aligned} \tag{4.1}$$

where $N(\cdot)$ denotes the cumulative normal integral function. We remark that in the case of VaR estimation, D is equal to VaR_α so that $E\left\{\mathbf{I}\left(S_T \le S_0 \exp\left(D\right)\right)\right\} = \left(1\text{-}\alpha\right)\times 100\%$ [3].

A basic Monte Carlo method provides an unbiased estimator for the extreme event probability $P\left(0, S_0\right)$ defined in (4.1) by the sample mean of extreme event indicators

$$P\left(0, S_0\right) \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{I}\left(\ln\left(S_T^{(i)}/S_0\right) \le D\right) \tag{4.2}$$

where N is the total number of i.i.d. simulated random samples and $S_T^{(i)}$ denotes the *i*-th simulated asset price at time T.

Next we investigate an efficient importance sampling estimator to estimate $P\left(0, S_0\right)$. By Girsanov theorem (Oksendal 1998), one can construct an equivalent probability measure $\tilde{P}$ defined by the Radon-Nikodym derivative $\frac{dP}{d\tilde{P}} = Q_T\left(h\right) = \exp\left(\int_0^T h\left(s, S_s\right) d\tilde{W}_s - \frac{1}{2}\int_0^T \left\|h\left(s, S_s\right)\right\|^2 ds\right)$, where $\tilde{W}_t = W_t + \int_0^t h(s, S_s) ds$ is a Brownian motion under $\tilde{P}$ provided that the process $h(s, S_s)$ satisfies Novikov's condition to ensure certain integrability of the function h such that Q_t is a martingale for $0 \le t \le T$.

The proposed importance sampling scheme is determined by a constant drift change h in order to satisfy the intuition of "the expected asset value S_T under the new probability measure is equal to its threshold $S_0 \exp\left(D\right)$," i.e.

$$\tilde{E}\left\{S_T \middle| F_0\right\} = S_0 \exp\left(D\right). \tag{4.3}$$

This intuition can be rigorously verified based on the construction of exponential change of measure

[3] Expressions (2.3) and (2.4) give another way of representation.

(Glasserman 2003; Han 2010). Hence, the extreme event $\{\ln(S_T/S_0) \le D\}$, when D is negatively large and/or T is small, is no longer rare under the new probability measure and the accuracy of Monte Carlo simulation can be improved significantly. Using the log-normal density of S_T, the criterion (4.3) results in a unique drift change

$$h = \frac{\mu}{\sigma} - \frac{D}{\sigma T}. \tag{4.4}$$

Therefore, under the new probability measure $\tilde{P}$ defined by the Radon-Nikodym derivative

$$Q_T(h) = \exp\left(h\tilde{W}_T - \frac{h^2 T}{2}\right), \tag{4.5}$$

the extreme probability defined in (4.1) can be re-expressed as

$$P(0, S_0) = \tilde{E}\left\{\mathbf{I}\left(\ln\left(S_T/S_0\right) \le D\right) Q_T(h) \middle| \, \mathrm{S}_0\right\}, \tag{4.6}$$

where the underlying risky-asset process is governed by $dS_t = (\mu - \sigma h)S_t dt + \sigma S_t d\tilde{W}_t$ due to this change of measure. The unbiased importance sampling estimator of $P(0, S_0)$ is

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{I}\left(\ln\left(S_T^{(i)}/S_0\right) \le D\right) Q_T^{(i)}(h). \tag{4.7}$$

In short, the direct Monte Carlo simulation is applied to (4.6) and the simulated samples are relocated to cluster around the quantity of interest, i.e. VaR and CVaR estimates at extreme probability levels, under the new probability measure. We can accordingly enhance the accuracy of estimates while avoiding the undesirable properties of basic Monte Carlo simulation.

The following theorem confirms that our proposed importance sampling estimator (4.7) is asymptotically optimal or efficient. We show that its variance rate approaches zero when extreme events occur. That is, the decay rate of the second moment of $\mathbf{I}\left(\ln\left(S_T/S_0\right) \le D\right) Q_T(h)$ is twice of the decay rate of its first moment under some scaling scenarios. Thus, we can present the following theorem for our importance sampling method.

Theorem 1 *Under the Black-Scholes model, the variance rate of the proposed importance sampling scheme defined in (4.6) approaches zero in cases of (1) very short maturity, (2) default threshold is negatively large. That is, the proposed importance sampling estimator (4.7) is asymptotically optimal or efficient under some scaling scenarios in time and space.*

The complete proof set is exhibited in the Appendix A, involving decay rate estimation of the first and second moments of importance sampling estimators under a spatial scale and/or a time scale. Our discussion on importance sampling not only serves as an alternative computation for the closed-form solution (4.1), but also lays a foundation to treat similar problems under stochastic volatility models.

4.2 Stochastic Volatility Model

In general, there is no closed-form solution for the evaluation problem defined in (2.2) under the stochastic volatility model specified by (2.1). Monte Carlo simulation is a generic approach to solve for this problem. In the last two decades, most previous literatures focus on variance reduction techniques under stochastic volatility models. Willard (1996) develops a conditional Monte Carlo scheme by conditioning on the driving volatility process. Fournie et al. (1997) and Fouque and Han (2004) apply regular and/or singular perturbation techniques to develop methods of importance sampling, control variate, or estimators combined with these two methods. Heath and Platen (2002) use an option price approximation with deterministic volatility to construct a control variate method. Fouque and Han (2007) generalize this approach to option price approximation with random volatility under multi-factor stochastic volatility models, and provide a variance analysis given two well-separated time scales. Han and Lai (2010) develop generalized control variate methods combined with Quasi Monte Carlo for enhanced reduction of variance. However, these control variate methods and importance sampling techniques are criticized as computationally intensive because of a required priori approximation to the evaluation problem (2.2).

Alternatively, we propose an enhanced importance sampling estimator to estimate the extreme event probability under stochastic volatility models. One of the advantages is that no prior knowledge about the unknown quantity defined in (2.2) is required, and the computational cost is significantly reduced accordingly. The downside is that the sample variance of this new estimator may be greater than those obtained from perturbation methods. However, our simulation and empirical studies indicate that the proposed estimator can indeed produce both efficient and unbiased estimate of the extreme event probability.

Our proposed importance sampling estimators under stochastic volatility models is based on the following ergodic property of the average of the variance process,

$$\frac{1}{T}\int_0^T f(Y_t^\varepsilon)^2 dt \xrightarrow{a.s.} \bar{\sigma}^2, \text{for } \varepsilon \to 0 \tag{4.8}$$

where ε denotes a small time scale and the driving volatility process in (3.8), Y_t^ε denotes a fast mean-reverting process. The *effective volatility* $\bar{\sigma}$, the averaged estimate, is a constant defined by the square root of the expectation of the variance function $f(\cdot)^2$ with respect to the limiting distribution of Y_t^ε; namely, $\bar{\sigma}^2 = \int f^2(y)d\Phi(y)$, where $\Phi(y)$ denotes the invariant distribution of the fast varying process Y_t^ε. Ornstein-Uhlenbeck process is a typical example of the stochastic volatility model. Under the fast mean-reverting assumption, coefficient functions of Y_t defined in (2.1) are chosen as $c(y) = \frac{1}{\varepsilon}(m-y)$ and $g(y) = \sqrt{\frac{2\nu}{\varepsilon}} = \beta$, so that the invariant distribution Φ is simply a Gaussian with mean m and variance ν. These results are thoroughly discussed in Fouque et al. (2000).

The limiting result (4.8) suggests a change of probability measure as follows. By substituting $\bar{\sigma}$ into σ shown in (4.4), a Radon-Nykodym derivative is defined as $Q_T\left(h(\bar{\sigma})\right) = \exp\left(h(\bar{\sigma})\tilde{W}_T^{(0)} - \frac{h^2(\bar{\sigma})T}{2}\right)$, so that $\tilde{W}_t^{(0)} = W_t^{(0)} + h(\bar{\sigma})t$ is a Brownian motion under the new probability measure denoted by $\tilde{P}$. Therefore, the extreme event probability defined in (2.3) can be re-expressed as

$$P(0,S_0,Y_0) = \tilde{E}\{\mathbf{I}(\ln(S_T/S_0) \le D)\; Q_T(h(\bar{\sigma})) \mid S_0, Y_0\}, \tag{4.9}$$

where the underlying risky-asset process is governed by $dS_t = \left(\mu - \sigma_t h(\bar{\sigma})\right)S_t dt + \sigma_t S_t d\tilde{W}_t^{(0)}$ and the dynamics of Y_t is changed accordingly. The unbiased importance sampling estimator for $P(0,S_0,Y_0)$ becomes

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left(\ln\left(S_T^{(i)}/S_0\right) \le D\right) Q_T^{(i)}\left(h(\bar{\sigma})\right) \tag{4.10}$$

We can apply this enhanced importance sampling to VaR estimate at extreme levels.

Notice that our proposed Radon-Nykodyn derivative $Q_T\left(h(\bar{\sigma})\right)$ considers only the averaging effect

$\bar{\sigma}$ and does not take the correlation coefficient into consideration. It is because $\bar{\sigma}$ corresponds to the first order effect, while ρ corresponds to the second order effect. According to the perturbation analysis of Fouque et al. (2011) and Fouque and Han (2004), the variance reduced from the second order effect is negligible so that our importance sampling estimator is an unbiased one with lower variance level.

4.3 Conditional VaR

CVaR, also known as expected shortfall, is qualified as a coherent risk measure but no VaR[4]. CVaR is defined as a conditional expectation $E\{X|X < c\}$,where X variable represents the loss observations and $c = VaR_\alpha$ satisfies $E\left\{\mathbf{I}\left(X \le VaR_\alpha\right)\right\} = \left(1-\alpha\right)\times 100\%$. The basic Monte Carlo algorithm to calculate CVaR, i.e. $E\{X|X < c\}$ is as follows:

$$n_c = \sum_{i=1}^{N} \mathbf{I}(X^{(i)} < c), \text{ where } N \text{ is the total number of simulations.}$$

$$E\{X|X < c\} \approx \frac{1}{n_c}\sum_{i=1}^{n_c} X^{(i)}, \text{ for each } X^{(i)} < c. \tag{4.11}$$

By selecting the likelihood function $Q_T = \frac{dP}{d\tilde{P}}$, a new probability measure $\tilde{P}$ is defined and we can derive the following importance sampling estimator for the conditional expectation:

$$\begin{aligned} E\left\{X|X < c\right\} &= \frac{E\left\{X\mathbf{I}\left(X < c\right)\right\}}{E\left\{\mathbf{I}\left(X < c\right)\right\}} \\ &= \frac{\tilde{E}\left\{X\mathbf{I}\left(X < c\right)Q_T\right\}}{\tilde{E}\left\{\mathbf{I}\left(X < c\right)Q_T\right\}} \\ &\approx \frac{\frac{1}{N}\sum_{i=1}^{N} X^{(i)}\mathbf{I}\left(X^{(i)} < c\right)Q_T\left(X^{(i)}\right)}{\frac{1}{N}\sum_{i=1}^{N}\mathbf{I}\left(X^{(i)} < c\right)Q_T\left(X^{(i)}\right)} \end{aligned}$$

4 Artzner et al. (1999) provide some criteria for qualifying a coherent risk measure.

$$
\begin{aligned}
&= \frac{\sum_{i=1}^{n_c} X^{(i)} Q_T\left(X^{(i)}\right)}{\sum_{i=1}^{n_c} Q_T\left(X^{(i)}\right)} \\
&= \sum_{i=1}^{n_c} X^{(i)} q_{iT},\ \textit{for each } X^{(i)} < c,
\end{aligned}
\tag{4.12}
$$

where $q_{iT} = Q_T(X^{(i)}) / \sum_{i=1}^{n_c} Q_T(X^{(i)})$. In (4.12), we have expressed the definition of a conditional expectation (first line), the same change of probability measure (second line), the same Monte Carlo simulation (third line), and straightforward calculation. Under this importance sampling algorithm, $E\left\{X \middle| X < c\right\}$ is approximated by the sum of a collection of random samples $X^{(i)}$, $i = 1, \cdots, n_c$, multiplied by its corresponding possibly non-equal weight q_i[5]. The approximation of the standard error (s.e.) in this non-equally weighted case is $s.e. \approx \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (X^{(i)} - \hat{m})^2 q_i}$, for each $X^{(i)} < c$, where $\hat{m}$ denotes the sample mean.

When $c = VaR_\alpha$, this conditional expectation becomes $CVaR_\alpha = E\left\{X \middle| X < VaR_\alpha\right\}$. Its computation can be reduced to:

$$
\begin{aligned}
E\left[X \mid X < VaR_\alpha\right] &= \frac{E\left[X\mathbf{I}\left(X < VaR_\alpha\right)\right]}{\left(1-\alpha\right) \times 100\%} \\
&= \frac{1}{1-\alpha} E\left[\mathbf{XI}\left(X < VaR_\alpha\right) Q_T\right] \\
&\approx \frac{1}{1-\alpha} \frac{1}{N} \sum_{i=1}^{N} X^{(i)} \mathbf{I}\left(X^{(i)} < VaR_\alpha\right) Q_T^{(i)}
\end{aligned}
\tag{4.13}
$$

Under the assumption that VaR_α is estimated, the estimator (4.13) is unbiased for estimating CVaR. Next we show one asymptotic property of CVaR under the standard normal assumption.

---

[5] This is different from the basic Monte Carlo estimator defined in (28), in which the weight associated with each random sample is equal to $1/n_c$ uniformly.

Lemma 1 *When X is a standard normal,* $\lim_{c\to-\infty}\frac{E\{X|X<c\}}{c}=1$. *This implies that when significance level* α *approaches one, i.e.* $c = VaR_\alpha$ *approaches negative infinity,* $CVaR = E\{X|X<c\}$ *is asymptotically equal to its* VaR_α.

The proof can be obtained directly by using the exact calculation

$$E\{X|X<c\} = \frac{-e^{-c^2/2}}{\sqrt{2\pi}N(c)} \tag{4.14}$$

and the approximation $N(c) \approx \frac{-1}{\sqrt{2\pi}c}e^{-c^2/2}$ for negatively large c. This lemma alerts that normal approximation can possibly underestimate CVaR, particularly when fat- (or fatter-) tailed distribution is considered. Table 2 in the following Section 4.4 summarizes the numerical comparisons about the discrepancy to estimate CVaR by using a normal approximation and the proposed importance sampling under a stochastic volatility model.

4.4 Numerical Examples

Two sets of numerical experiments are conducted to demonstrate the improved efficiency of the proposed importance sampling algorithm. The first set takes the Black-Scholes model as a benchmark. Closed-form solutions for extreme event probabilities (associated with $VaR_{95\%}$ and $VaR_{99\%}$) and their corresponding CVaR estimates are used to compare numerics estimated from the basic Monte Carlo and the enhanced importance sampling algorithm. The second set concentrates on CVaR estimation under a stochastic volatility model with various values of correlation. CVaRs are calculated from (1) a normal approximation and (2) importance sampling algorithm.

Table 1 summarizes the numerical results of the extreme event probability and CVaR estimates obtained from the exact solution, the basic Monte Carlo, and the enhanced importance sampling algorithm. Under the Black-Scholes model assumption as specified by (2.1), the model parameters are set as $\mu = 0$, $\sigma = 0.3$, and $T = 1/252$ (one trading day), and the number of simulations is set as 1,000,000. In the first column, two loss thresholds (D) are exactly $VaR_{95\%}$ and $VaR_{99\%}$, specified as the empirical VaR values, so that extreme event probabilities are 0.05 and 0.01, respectively. Standard errors obtained from importance sampling are all significantly smaller than those from the basic Monte Carlo method. Variance reduction ratios for extreme event probability and CVaR estimation are ranged from 4 to 12 in the case of $D = -0.0313$, and from 36 to 60 in the case of $D = -0.0441$. The variance reduction is more significant as it moves toward more extreme losses. The enhanced importance sampling is confirmed to give significant performance especially when the extreme event probability is small, say 0.01, in this simulation study. Note that each estimated CVaR is close to the loss threshold D. This numerical result coincides with the approximation predicted by Lemma 1. That is, CVaR asymptotically coincides VaR at extreme significance levels.

We further check the estimation performance among the Brownian motions specified in (2.1) with different correlation scenarios. Table 2 summarizes the CVaR estimates by the two implemented methods: normal approximation and the enhanced importance sampling. The parameters of the stochastic volatility model defined in (2.1) and the Ornstein-Uhlenbeck volatility process in (3.9) are chosen as μ=0, $m=-5$, $\alpha=5$, $\beta=1$, $S_0=50$, $Y_0=-3$, T=$1/252$ (one day), and the confidence level of VaR is specified as 99%. Different correlation values among the Brownian motions and corresponding $VaR_{99\%}$ estimated by our proposed importance sampling method specified by (4.10) are listed in Column 1 and Column 2, respectively. The number of simulations is uniformly set as 10,000 for basic Monte Carlo defined in (4.11) and the importance sampling defined in (4.13). CVaR estimates, defined as $E\left[\ln\left(S_T/S_0\right)\middle|\ln\left(S_T/S_0\right)<VaR_{99\%}\right]$ are calculated by two methods: a normal approximation and the enhanced importance sampling. The normal approximation, denoted by N. Approx. in Column 3, is derived based on the Black-Scholes model specified as (4.1) with the spot volatility $\sigma=\exp(Y_0/2)$, where Y_0 is the initial value of driving volatility process. That is, we approximate CVaR by assuming that $-\ln\left(S_T/S_0\right)$ is normally distributed with mean $\left(\exp\left(Y_0\right)/2-\mu\right)T$ and variance $\exp\left(Y_0\right)T$. Column 4 reports CVaR estimated by the enhanced importance sampling with standard errors in parenthesis. These standard errors are all significantly small relative to their corresponding CVaR estimates. Those CVaR estimates obtained by normal approximation in column 3 fall outside of 99% confidence interval of CVaR estimates obtained by the enhanced importance sampling in Column 4. This indicates that normal approximation for CVaR estimation may incur significant errors. Our proposed importance sampling method outperforms traditional normal approximation. It is worth noting that those relative errors between the normal approximation (Column 3) and importance sampling (Column 4) range from 10.03% to 12.62%. These significant discrepancies are consequences of the fat tail brought by the stochastic volatility model.

In summary, Table 1 confirms the outperformance of our enhanced importance sampling method for the risk measurement at extreme significance levels under the Black-Scholes model framework. Table 2 exhibits more accurate CVaR estimates than normal approximation under stochastic volatility models. It can be concluded that our proposed importance sampling method indeed contribute to provide more accurate estimates of risk measures at extreme significance levels. In addition, our importance sampling algorithm provides even the significant advantage of saving computing efforts and execution time, as compared with the basic Monte Carlo method under stochastic volatility model[6].

Section 5: Backtesting for VaR Estimation

Backtesting helps avoid model misspecification and differentiate the model performance from a faulty model under special conditions. In effect, backtesting can balance Type I against Type II statistical errors in VaR estimation. There are two major criteria for backtesting: unconditional rate of exceedances (UC) and independence of the exceedances (IND). The significance level for backtesting represents the maximum probability of observations exceeding VaR estimates if the model is correctly calibrated.

Unconditional rate of exceedances is used to check if the number of exceedances (the case in which the

[6] Under the configuration of our PC (Intel CPU Core 2 Duo 2.4GHz), it is rather time-consuming to solve for VaR under stochastic volatility models without the help of the proposed importance sampling algorithm. Under Matlab environment, we use the basic Monte Carlo estimator to approximate the extreme event probability defined in (2.3), then use a nonlinear Matlab solver, say fzero.m file, to solve for $VaR_{99\%}$ defined in (2.4). Even if the number of simulations increases to 250,000 and the execution time exceeds three minutes, we are still unable to solve for a single $VaR_{99\%}$ estimate. This indicates that VaR estimation is a challenging task for the Monte Carlo simulation under stochastic volatility models. In contrast, our proposed importance sampling algorithm takes only several seconds to get a $VaR_{99\%}$ estimate. Further, without a variance reduction, using the basic Monte Carlo method to estimate VaR is expected to consume tremendous computing resources.

actual loss is larger than VaR estimate) exceeds the level as specified by the significance level. Under the null hypothesis that the significance level is the true probability of exceedances occurring, the test statistics are a log-likelihood ratio specified as:

$$LR_{UC} = -2\ln\left[(1-p)^{T-N}\,p^{N}\right] + 2\ln\left\{\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right\} \sim \chi^2(1), \tag{5.1}$$

where T is total number of days and N is the number of exceedances. This asymptotically follows a Chi-square distribution with one degree of freedom (Kupiec 1995).

For the independence test of the exceedances, the first job is to set up a series which indicates if the daily VaR estimate is exceeded or not. If the VaR estimate is not exceeded by the actual loss, the exceedance indicator is set at 0, or 1 otherwise. The next job is to observe the switches of exceedances. Table 3 shows the construction of a table of conditional exceptions. The log-likelihood test statistics are specified as:

$$LR_{IND} = -2\ln\left[(1+\pi)^{(T_{00}+T_{10})}\,\pi^{(T_{01}+T_{11})}\right] + 2\ln\left[(1-\pi_0)^{T_{00}}\,\pi_0^{\,T_{01}}\,(1-\pi_1)^{T_{10}}\,\pi_1^{\,T_{11}}\right] \sim \chi^2(1), \tag{5.2}$$

where T_{ij} denotes the number of days in which state j occurred in one day while it was state i the previous day. Moreover, π_i represents the probability of observing an exceedance conditional on state i the previous day. It asymptotically follows a Chi-square distribution with one degree of freedom. The first term is specified under the hypothesis that the exceedances are independent across the sample, or $T_1 = T_0 = T_{11} = \dfrac{(T_{00}+T_{11})}{T}$. The second term is the maximized likelihood for the observed data. This test helps confirm if the exceedances are serially correlated, i.e. to examine whether the model makes systematic errors in the VaR estimates.

The conditional coverage (CC) test is designed to simultaneously test if the VaR violations are independent and the average number of exceedances is correct. The test statistics for conditional coverage are actually the sum of the test statistics of unconditional coverage and independence, i.e.

$$LR_{CC} = LR_{UC} + LR_{IND}.$$

These three types of backtesting- unconditional coverage, independence, and conditional coverage- are regarded as the minimum set of required tests to help validate VaR estimation performance (Christoffersen 1998).

5.1 Empirical Analysis

VaR and CVaR, two of the most widely used risk measures, are used as the criteria for performance comparison. The competing methods include historical simulation, RiskMetrics[7], and GARCH(1,1)[8]. The first one is acknowledged as model-free and easy to implement. The latter two are known for being robust in capturing volatility process. All of the three methods are commonly accepted as benchmark models for VaR estimation. Two datasets are used for empirical examination. The first dataset contains three exchange rate series against the US Dollar: Japanese Yen (JPY), Singapore Dollar (SGD), and Canadian Dollar (CAD).

[7] RiskMetrics is also called exponentially weighted moving average method and this method is designed to represent the finite-memory property. This method is specified to model the volatility process as: $\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2$, where r_{t-1} denotes return rate at time (t-1). The decaying factor (λ) in RiskMetrics model is set as 0.94 throughout this paper.

[8] For return series $r_t | F_{t-1} = \mu_t + a_t = \mu_t + \sigma_t\varepsilon_t,\ \varepsilon_t \sim N(0,1)$, where F_{t-1} : information set, μ_t : conditional mean, and σ_t^2 : conditional variance. GARCH(1,1) model is specified as $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta\sigma_{t-1}^2$. VaR estimation under GARCH(1,1) can be found in Hull (2008).

The data covers the period from January 5, 1998 to July 24, 2009, with 2890 daily observations. The dataset is collected from the official website of the Central Bank of the Republic of China (http://www.cbc.gov.tw/). The daily observations are taken as the natural logarithmic returns in two consecutive trading days and are denoted as $X_t : X_t = \ln\left(\frac{r_t}{r_{t-1}}\right)$, where r_t is the daily exchange rate at date t. Descriptive statistics and time series plot of the six series (3 original series and 3 corresponding return series) are summarized in Table 4 and Figure 1, respectively. Figure 2 shows the estimates of the three major Ornstein-Uhlenbeck volatility model parameters in the respective return series by the Fourier transform scheme: α (right), β (bottom left), and m (top-left) which represents mean-reverting rate, volatility of volatility, and long-run mean, respectively in the volatility process. Moving window of fixed 500 observations is set in estimation, in order to demonstrate the time-varying properties of the parameters in the volatility process. Noteworthy, those estimates of the three parameters in each series show significant spikes near the data period end. They reflect the dynamic variance along the major crisis since late 2007. The estimates of VaR and CVaR by the proposed importance sampling algorithm at 95% and 99% significance levels are plotted in Figure 3. In general, those VaR estimates are close and below the actual loss levels. This indicates that those VaR estimates can serve well as a safety net for risk management purpose because of their appropriate forecasts of actual loss levels.

Three methods of backtesting (unconditional coverage, independence, and conditional coverage) are employed for performance examination and the outcomes are summarized in Table 5. The significance level for rejecting backtesting is set as 10%. For VaR estimation, the proposed improved procedure (denoted as

SV) dominates at 99% significance level. The overall evidence shows that this proposed procedure overwhelmingly outperforms the other competitive models, with 2 exceptions at 95% significance level. The outperformance of this procedure is manifest at 99% extreme significance level. GARCH is evaluated as a competent competitor but its satisfactory performance is constrained at 95% significance level. The underperformance of historical simulation and RiskMetrics can be attributed to their rigid structure of adjustment to the volatility process. Accordingly, their responding adjustment is not fast enough to capture the vibrant market dynamics.

The second dataset, downloaded from Yahoo! Finance website (http://finance.yahoo.com/), is composed of two series: daily observations of S&P 500 and its VIX data. The data coverage expands from January 3, 2005 to July 24, 2009, consisting of 1138 daily observations. VIX is an annualized volatility index of S&P 500, which is used as the major measure to predict the market volatility level of the following 30 calendar days. The second dataset is selected to stress test the proposed procedure because the data series covers to the recent global financial crisis since 2007. The first 500 observations, which is the time period before the major financial crisis, are used as a warm-up period for estimation. Both data series are treated in the same manner as that of the first dataset to obtain their return series. Descriptive statistics and time series plot of both series are summarized in Table 6 and Figure 4, respectively.

There are noticeable correlation between S&P 500 and VIX (-0.67839 for original series and -0.71487 for log return series). VIX is usually noticed as a major volatility index for the S&P 500 stock index (Duan and Yeh 2010). Due to the significant correlation between VIX and S&P 500 index, it is expected that these properties are expected to be auxiliary in VaR estimation of S&P 500 series. Hence, we use VIX data for the correlation estimate for our proposed procedure[9].

We first use the S&P 500 and VIX data in the corresponding moving windows (fixed size of 500 daily observations) to calculate the correlation (ρ , Rho) between them (Figure 5) and introduce the estimated ρ into the proposed correction scheme[10]. The parameter estimates (m,α,ρ,β)(starting from top left panel, in clockwise order) of this scheme for S&P 500 are demonstrated in Figure 6. Again, the rugged and spiky curves demonstrate the intense volatility along the estimation process, especially since the included crisis period. The estimates of VaR and CVaR are plotted in Figure 7. The backtesting outcomes also give favorable conclusion on our proposed improved procedures performance (Table 7). Noteworthy, the outperformance is especially significant at extreme 1% significance level for VaR estimation, which is stipulated by the Basel II Accord. The turbulent dynamics during the financial crisis are satisfactorily captured by our proposed method. Overall, our proposed procedure outperforms in VaR and CVaR estimation under stochastic volatility models as compared to traditional benchmark methods, i.e. historical simulation, RiskMetrics, and GARCH(1,1).

The empirical analysis of the second dataset indicates that auxiliary series (VIX) showing significant correlation with the target series (S&P 500) can be employed to measure the volatility process and applied

[9] We do not treat the S&P 500 and its VIX data series to form a hypothetical bivariate portfolio.
[10] ρ is one of the major parameters in our proposed procedure. In contrast, we assume ρ =0, for the respective series in the first foreign exchange rate dataset which is designed for univariate VaR estimation.

to our proposed procedure. Even when the correlation estimated by the Fourier transform method exhibits significant estimation error, the proxy series VIX with our proposed procedure can also deliver superior performance.

Section 6: Conclusion

We investigate VaR and CVaR estimation under stochastic volatility models by proposing an improved procedure and comparing its estimation performance with major traditional methods. Two major hurdles in the estimation process are tackled: (1) unstable parameter estimation subject to practical data constraint, and (2) lack of efficient computational technique for accurately estimating extreme event probabilities. The first hurdle is overcome by the Fourier transform method to estimate volatility. The second hurdle is overcome by a variance reduction procedure for VaR and CVaR estimation via the enhanced importance sampling.

Two datasets are selected for empirical examination; the first one contains three exchange rate series and the second includes S&P500 index and its VIX. The long data period covers recent financial turmoil since 2007 so as to stress test the VaR and CVaR estimation performance and examine their performance in capturing the dynamic stochastic volatility. Three essential types of backtesting are preceded for performance evaluation: unconditional coverage, independence, and conditional coverage. Backtesting outcomes show that our improved procedure under stochastic volatility models outperforms in VaR estimation at the 99% significance level over classical benchmark methods: RiskMetrics, historical simulation, and the GARCH(1,1) model. The proposed procedure is confirmed to contribute to give more accurate VaR and CVaR estimate at extreme significance levels so as to satisfy the Basel II Accord requirement.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D., 1999. Coherent measures of risk. Mathematical Finance, 9, 203-28.

[2] Barucci, Emilio, and Mancino, Maria Elvira, 2010. Computation of volatility in stochastic volatility models with high frequency data. International Journal of Theoretical and Applied Finance, 13(5), 767-787.

[3] Broto, Carmen, and Ruiz, E., 2004. Estimation methods for stochastic volatility models: A survey. Journal of Economic Surveys, 18(5), 613-649.

[4] Bucklew, James, A., 2004. Introduction to Rare Event Simulation: Springer.

[5] Christoffersen, P., 1998. Evaluating interval forecast. International Economic Review, 3(4), 841-862.

[6] Duan, Jin-Chuan, and Chung-Ying Yeh. 2010. Jump and volatility risk premiums implied by VIX. Journal of Economic Dynamics and Control, 34(11), 2232-2244.

[7] Finner, H., Dickhaus, T. and Roters, M. 2008. Asymptotic tail properties of Student's t-distribution. Communications in Statistics-Theory and Methods, 37(2), 175-179.

[8] Fournie, E., Lebuchouix, J. and Touzi, N. 1997. Small noise expansion and importance sampling, Asymptotic Analysis, 14, 361-376.

[9] Fouque, J.-P. and Han, C.-H., 2004. Variance reduction for Monte Carlo methods to evaluate option prices under multi-factor stochastic volatility models, Quantitative Finance, 4(5), 597- 606.

[10] Fouque, J.-P. and Han, C.-H., 2007. A martingale control variate method for option pricing with stochastic volatility. ESAIM: Probability and Statistics, 11, 40-54.

[11] Fouque, J.-P., Papanicolaou, G. and Sircar, R., 2000. Derivatives in Financial Markets with Stochastic Volatility: Cambridge University Press.

[12] Fouque, J.-P., Papanicolaou, G., Sircar, R., and Sølna, K. 2011. Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives: Cambridge University Press.

[13] Glasserman, Paul, 2003. Monte Carlo Methods for Financial Engineering: Springer.

[14] Gregoriou, Greg N. 2009. The VaR Implementation Handbook: McGraw-Hill.

[15] Han, C.-H., 2011. Efficient Importance Sampling Estimation for Joint Default Probability: the First Passage Time Problem in Stochastic Analysis with Financial Applications. (Proceedings of 2009 Workshop on Stochastic Analysis & Finance), Editors: A. Kohatsu-Higa, N. Privault, and S.-J. Sheu. Progress in Probability, Vol. 65: Birkhauser.

[16] Han, C.-H., Lai, Y., 2010. Generalized control variance methods for pricing Asian options, Journal of Computational Finance, 14(2), 87-118.

[17] Han, C.-H., 2013. Instantaneous Volatility Estimation by Fourier Transform Methods in Handbook of Financial Econometrics and Statistics, Editor: C.-F. Lee: Springer.

[18] Heath, D. and Platen, E., 2002. A variance reduction technique based on integral representations, Quantitative Finance, 2(5), 362-369.

[19] Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options, Review of Financial Studies, 6(2), 327-343.

[20] Hull, John, and A. White. 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance, 42(2), 281-300.

[21] Hull, J., 2008. Options, Futures, and Other Derivatives. 7th edition: Prentice Hall.

[22] Jorion, P., 2007. Value-at-Risk: The New Benchmark for Managing Risk. 3nd edition: McGraw-Hill.
[23] Jondeau, E., Ser-Huang Poon and Rockinger, M., 2007. Financial Modeling under Non-Gaussian Distributions: Springer.
[24] Lehar, A., Scheicher, M. and Schittenkopf, C., 2002. GARCH vs. stochastic volatility: Option pricing and risk management. Journal of Banking & Finance, 26(2), 323-345.
[25] Lemieux, Christiane, 2009. Monte Carlo and Quasi-Monte Carlo Sampling: Springer.
[26] Malliavin, P., and Mancino, M. E., 2002. Fourier series method for measurement of multivariate volatilities. Finance and Stochastics, 6(1), 49-61.
[27] Malliavin, P. and Mancino, M. E., 2009. A Fourier transform method for nonparametric estimation of multivariate volatility. Annals of Statistics, 37(4), 1983-2010.
[28] McNeil, Alexander J., Rudiger Frey, and Paul Embrechts. 2005. Quantitative Risk Management: Concepts, Techniques and Tools: Princeton University Press.
[29] Meddahi, Nour. 2002. "A theoretical comparison between integrated and realized volatility." Journal of Applied Econometrics, 17(5), 479-508.
[30] Oksendal, Bernt, 1998. Stochastic Differential Equations. 5th Edition.Springer.
[31] Pederzoli, Chiara. 2006. "Stochastic volatility and GARCH: A comparison based on UK stock data." European Journal of Finance, 12(1), 41-59.
[32] Reno, R., 2008. Nonparametric estimation of the diffusion coefficient of stochastic volatility models. Econometric Theory, 24(5), 1174-1206.
[33] Tsay, Ruey S. 2010. Analysis of Financial Time Series. 3rd ed: John Wiley & Sons.
[34] Willard, G., 1996. Calculating prices and sensitivities for path-dependent derivatives securities in multifactor models, PhD thesis, Washington University in St. Louis.
[35] Yu, Jun., 2010. Simulation-based Estimation Methods for Financial Time Series Models. In Handbook of Computational Finance, edited by J.-C. Duan, J. E. Gentle and W. Hardle: Springer.
[36] Zhang, L., Mykland, P., 2005. A tale of two time scales: Determining integrated volatility with noise high frequency data. Journal of American Statistics, 100(472), 1394-1411.

Appendix A: Proof of Efficient Importance Sampling Estimator

Proof.

Firstly, we derive closed-form solutions for the first and second moments, denoted by P_1 and P_2 respectively, of any importance sampling scheme induced by a constant drift change h. Since $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$ is log-normally distributed, we obtain the following closed-form solution for P_1:

$$P_1 = E\left\{\mathbf{I}(\ln\frac{S_T}{S_0} \le D)\right\}$$
$$= N\left(\frac{D - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\right).$$

From Equation (4.6), the second moment P_2 is defined by

$$P_2 = \tilde{E}\left\{\mathbf{I}(\ln\frac{S_T}{S_0} \le D)Q_T^2(h)\right\},$$

where the measure change is given by $Q_T(h) = \exp\left(h\tilde{W}_T - \frac{h^2T}{2}\right)$, so that the asset price dynamics becomes $dS_t = (\mu - \sigma h)S_t dt + \sigma S_t d\tilde{W}_t$ under the new probability measure $\tilde{\mathrm{P}}$. Rewrite P_2 as

$$P_2 = e^{h^2T}\tilde{E}\left\{\mathbf{I}(\ln\frac{S_T}{S_0} \le D)e^{2h\tilde{W}_T - \frac{(2h)^2T}{2}}\right\}$$
$$= e^{h^2T}\hat{E}\left\{\mathbf{I}(\ln\frac{S_T}{S_0} \le D)\right\},$$

where $\exp\left(2h\tilde{W}_T - \frac{(2h)^2T}{2}\right)$ is the Radon-Nykodym derivative to further change probability measure from $\tilde{\mathrm{P}}$ to $\hat{\mathrm{P}}$ such that $\hat{W}_t := \tilde{W}_t - 2ht$ is a standard Brownian motion. Hence under $\hat{\mathrm{P}}$, the dynamics of S_t becomes $dS_t = (\mu + \sigma h)S_t dt + \sigma S_t d\hat{W}_t$ so that we get the closed form for P_2:

$$P_2 = e^{h^2T}N\left(\frac{D - (\mu + \sigma h - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}\right).$$

In order to obtain zero variance rate, the key step is to choose a peculiar drift-change parameter h. According to (4.4), $h = \frac{\mu T - D}{\sigma T}$ so that the associated probability measure can incur more extreme events.

Secondly, we estimate decay rates of P_1 and P_2 under three scaling scenarios in time and space. When time scale T is small, we set $T = \varepsilon << 1$. It is easy to see that $\frac{D - (\mu - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \approx \frac{D}{\sigma\sqrt{\varepsilon}}$ so that $P_1 \approx \frac{\sigma\sqrt{\varepsilon}}{\sqrt{2\pi}D} e^{-\frac{D^2}{2\sigma^2\varepsilon}}$ by using the normal approximation $N(-x) \approx \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}}$, where x is positively large. Because $h \approx -\frac{D}{\sigma\varepsilon}$ and $\frac{D - (\mu + \sigma h - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} \approx \frac{2D}{\sigma\sqrt{\varepsilon}}$, the approximation $P_2 \approx e^{\frac{D^2}{\sigma^2\varepsilon}} N\left(\frac{2D}{\sigma\sqrt{\varepsilon}}\right) \approx e^{\frac{D^2}{\sigma^2\varepsilon}} \frac{1}{\sqrt{2\pi}\frac{2D}{\sigma\sqrt{\varepsilon}}} e^{-\frac{2D^2}{\sigma^2\varepsilon}} = \frac{\sigma\sqrt{\varepsilon}}{\sqrt{2\pi}2D} e^{-\frac{D^2}{\sigma^2\varepsilon}}$ is obtained. Therefore we get the following decay rates for the first two moments under a small time scale T:

$$\lim_{\varepsilon \to 0} \varepsilon \log P_1 = -\frac{D^2}{2\sigma^2},$$

$$\lim_{\varepsilon \to 0} \varepsilon \log P_2 = -\frac{D^2}{\sigma^2}.$$

These results show that the decay rate of the second moment is twice of the decay rate of the first moment, which implies $P_1^2 \approx P_2$ as ε goes to zero, so that an asymptotic zero variance rate for the importance sampling (4.7) is justified.

Similar results can be obtained under a small spatial scale, i.e. $D = -1/\sqrt{\varepsilon}$ for $\varepsilon << 1$. It is easy to check that

$$P_1 \approx \frac{1}{\sqrt{2\pi}} \sigma\sqrt{T\varepsilon}\, e^{-\frac{1}{2\sigma^2 T\varepsilon}}$$

and

$$\begin{aligned} P_2 &\approx e^{-\frac{1}{\sigma^2 T\varepsilon}} N\left(\frac{-2}{\sigma\sqrt{T\varepsilon}}\right) \\ &\approx e^{-\frac{1}{\sigma^2 T\varepsilon}} \frac{1}{\sqrt{2\pi}} \frac{\sigma\sqrt{\mathrm{T}\varepsilon}}{2} e^{\frac{-2}{\sigma^2 T\varepsilon}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{\sigma\sqrt{\mathrm{T}\varepsilon}}{2} e^{\frac{-1}{\sigma^2 T\varepsilon}}. \end{aligned}$$

By inspection, $\lim_{\varepsilon\to 0}\varepsilon \log P_2 = 2\lim_{\varepsilon\to 0}\varepsilon \log P_1 = \frac{-1}{\sigma^2 T}$ is obtained so that an asymptotic zero variance rate is confirmed.

When maturity is short and default threshold is large, one can expect the increase of decay speed of these moments. Let $D = -1/\sqrt{\varepsilon}$ and $T = \varepsilon$ for $\varepsilon << 1$, then one can obtain the following decay rate estimates $\lim_{\varepsilon\to 0}\varepsilon^2 \log P_2 = 2\lim_{\varepsilon\to 0}\varepsilon^2 \log P_1 = \frac{-1}{\sigma^2}$. Note that the scaling order is ε^2 in this scenario which is faster than ε in previous two scenarios.

Table 1: Estimates of Extreme Event Probability and Its CVaR with Two Different Loss Thresholds.

Loss Threshold (D)	Extreme Event Probability				CVaR			
	Exact	BMC	IS	VR	Exact	BMC	IS	VR
-0.0313	0.05	0.0510 (0.0002)	0.0500 (0.0001)	4	-0.0392	-0.0391 (0.0017)	-0.0392 (0.0005)	12
-0.0441	0.01	0.0099 (9.88E-0.5)	0.0100 (1.64E-05)	36	-0.0505	-0.0506 (0.0031)	-0.0505 (0.0004)	60

Remark:

1. Extreme event probability is defined as $P(0,S_0)=E\left[\mathbf{I}\left(\ln\left(S_T/S_0\right)\le D\right)\right]$ and its CVaR is defined as $E\left[\ln\left(S_T/S_0\right)\middle|\ln\left(S_T/S_0\right)\le D\right]$.
2. Exact, BMC, IS, and VR denote the closed-form solution, the Basic Monte Carlo method, the enhanced importance sampling, and the variance reduction ratio, respectively. VR is defined as $VR=\left(\frac{\text{standard } error \text{ of } BMC}{\text{standard } error \text{ of } IS}\right)^2$.
3. Sample means and standard errors shown in parenthesis are reported in columns of BMC and IS.

Table 2: CVaR Approximation under Various Values of Correlation.

1	2	3	4
Correlation	$VaR_{99\%}$	CVaR$_{99\%}$	
		N. Approx.	IS
0.8	-0.0339	-0.0347	-0.0386 (7.5073E-05)
0.4	-0.0335	-0.0343	-0.0378 (7.3833E-05)
0	-0.0323	-0.0331	-0.0367 (7.2498E-05)
-0.4	-0.0317	-0.0325	-0.0366 (7.2739E-05)
-0.8	-0.0310	-0.0319	-0.0351 (6.9643E-05)

Remark: Given five scenarios of correlation (ρ) listed in Column 1 and estimated $VaR_{99\%}$ reported in Column 2. Two sets of CVaRs estimates are reported in Columns 3 and 4. The former is based on the approximation obtained from the closed-form solution under a normality assumption. The latter is estimated via the enhanced importance sampling in (4.13). N. Approx. and IS denote the normal approximation and the enhanced importance sampling, respectively. Relative error denotes the ratio of the discrepancy between the estimates by IS and N. Approx., divided by corresponding IS estimates.

Table 3: Construction of Conditional Exceptions

	conditional		
	Day before		
Current day	**No exception**	**exception**	**unconditional**
no exception	$T_{00} = T_0\left(1-\pi_0\right)$	$T_{10} = T_1\left(1-\pi_1\right)$	$T\left(1-\pi\right)$
exception	$T_{01} = T_0\left(\pi_0\right)$	$T_{11} = T_1\left(\pi_1\right)$	$T\left(\pi\right)$
Total	T_0	T_1	$T = T_0 + T_1$

Remark: T_{ij} denotes the number of days in which state j occurred in one day while it was state i the previous day. π_i represents the probability of observing an exceedances conditional on state i the previous day.

Table 4: Descriptive Statistics of the three Foreign Exchange Rate Data

Panel 1: Original Daily Data

	JPY	SGD	CAD
Minimum	87.915	1.34665	0.9218
1st Quantile	107.525	1.575763	1.1683
Mean	114.5035	1.653406	1.334586
Median	115.265	1.6875	1.357725
3rd Quantile	120.35	1.737975	1.506788
Maximum	147.41	1.85325	1.6147
Standard Deviation	10.00708	0.119753	0.188313
Skewness	0.254427	-0.7232320	0.295178
Kurtosis	0.435041	-0.2974507	1.273715

Panel 2: Daily Return Data

	JPY	SGD	CAD
Minimum	-0.04565	-0.03523	-0.03737
1st Quantile	-0.00384	-0.00184	-0.00301
Mean	-0.00012	-5.9E-05	-9.3E-05
Median	0	-7.2E-05	-5E-05
3rd Quantile	0.004003	0.001681	0.002774
Maximum	0.049073	0.031479	0.033091
Standard Deviation	0.007231	0.003634	0.005528
Skewness	-0.43005	-0.20578	0.043318
Kurtosis	3.67931	9.871163	4.425934