AMSI VACATION RESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Efficient estimation of risk measures using Monte Carlo methods

Peter (Hoang Nhat Huy) Nguyen

Supervised by A/Prof Georgy Sofronov & Dr Houying Zhu

Department of Mathematics and Statistics

Macquarie University

February 23, 2021

Vacation Research Scholarships are funded jointly by the Department of Education, Skills and Employment and the Australian Mathematical Sciences Institute.

Contents

1 Abstract

Regular adverse market's movements emphasized the need for efficient risk estimation. In this statistical methodology driven project, we will explore different Monte Carlo methods to address the research question of how to achieve effective estimation for risk metric: Value at Risk and Expected Shortfall. To improve the efficiency of crude Monte Carlo methods, we will employ various variance reduction techniques such as importance sampling, control variate and stratification. The proposed methods then will be applied to share portfolio. Simulation and comparison will be carried out to verify the performance of the algorithms to acquire insights of the efficiency of the proposed methods.

2 Introduction

The coronavirus pandemic has brought about increasing market volatility and, at the same time, reminded us of the importance of risk management. Managing risk is a complex task for financial institutions and portfolio managers. There are two common questions needed to be addressed:

- What value of a given portfolio is at risk?
- If it does get bad, what is the expected loss?

Value at Risk (VaR) and Expected Shortfall (ES) are the risk measures used to answer these questions. Institutions are required to estimate the profit and loss distributions of portfolios, and thus compute risk measures that summarise these distributions. From the mathematical point of view, one needs to calculate the expected value of a random variable. More precisely, it results in integral and quantiles estimation, which is very often related to the estimation of rare event probabilities.

The effectiveness of the risk measures, however, depends on the accuracy of estimation. The Monte Carlo method is one of the most robust and straightforward ways to simulate and estimate the necessary risk measures. This report will focus on three well-known methods used in financial risk management: Control Variates, Importance Sampling and Stratification. Especially, we will see how they perform against normal risk factor and heavy tailed risk factor.

2.1 Statement of Authorship

A/Prof Sofronov suggested the project's main idea. Under the supervision of A/Prof Sofronov and Dr Zhu, Mr Nguyen introduced the theoretical algorithms, derived the mathematical model and implemented different variance reduction methods to estimate necessary quantities. Huy Nguyen performed numerical simulations and interpreted results with support from the two supervisors. AMSI funded the project.

3 Monte Carlo methods

In this section, we will discover different theoretical aspects of Monte Carlo methods to justify the proposed variance reduction methods, and to understand why these are superior to Crude Monte Carlo in terms of convergence speed. The quantity of interest here is $\mu = \mathbb{E}(f(X))$, with X follows probability density function (pdf) p_X.

3.1 Crude Monte Carlo

The Crude Monte Carlo method [5], expresses the estimated integral as the expected value of random variable $f(X)$: $\mu = \mathbb{E}(f(X))$. The n samples estimate of μ is given by

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

For example, suppose we estimate the integral $\int_0^\infty (x^2+2x+2)e^{-2x}\,dx$ using Crude Monte Carlo method. By transforming it to $\int_0^\infty (\frac{1}{2}x^2+x+1)2e^{-2x}\,dx$, we have $f(x) = \frac{1}{2}x^2+x+1$ and $p(x) = 2e^{-2x}$, which is an exponential distribution with scale parameter of 2. The Crude Monte Carlo method suggests us to generate random variable $X \sim \text{Exp}(\lambda = 2)$, and the integral is estimated by $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2}X_i^2 + X_i + 1)$. We know that the true value of the integral is 1.75, and the table below shows estimates with different number of samples.

n	$n = 100$	$n = 1000$	$n = 10{,}000$	$n \to \infty$ (True value)
Estimate	1.619	1.758	1.751	1.75

The variance of this estimate is

$$\text{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n}, \tag{1}$$

which means that the larger the number of simulations, the better the estimation. The convergence rate of Crude Monte Carlo is $O(1/n)$, which is quite slow as illustrated above. This opens opportunities for advanced methods to increase the rate of convergence.

3.2 Control Variate-Weighted Monte Carlo

3.2.1 Control Variate

Control Variate method is a variance reduction techniques used in Monte Carlo methods. It is useful when we knew the value of $\theta = \mathbb{E}(h(X))$, where $h(x) \approx f(x)$. A popular way to use the control variate is the regression control variate estimator below:

$$\hat{\mu}_\beta = \frac{1}{n}\sum_{i=1}^{n}\Big(f(X_i) - \beta h(X_i)\Big) + \beta\theta. \tag{2}$$

In order to justify its efficiency, we derive the expected value and variance of the regression estimator:

$$\mathbb{E}(\hat{\mu}_\beta) = \mu + \beta\theta - \beta\theta = \mu,$$

$$\text{Var}(\hat{\mu}_\beta) = \frac{1}{n}\left[\text{Var}(f(X) - 2\beta\text{Cov}(f(X), h(X)) + \beta^2\text{Var}(h(X))\right]. \tag{3}$$

The estimator is unbiased, and the variance suggests that we can choose efficient parameter β in order to minimize the theoretical standard error. Basic calculus gives the optimal β_{opt} of the following form:

$$\beta_{\text{opt}} = \frac{\text{Cov}(f(X), h(X))}{\text{Var}(h(X))}.$$

Substituting the optimal β into equation (3), with ρ as the correlation coefficient between $f(X)$ and $h(X)$, we have:

$$\text{Var}(\hat{\mu}_{\beta_{\text{opt}}}) = \frac{\sigma^2}{n}(1 - \rho^2). \tag{4}$$

Equation (4) proposes that we have effectively reduced the variance of the original Crude Monte Carlo as given in Equation (1). Additionally, we can see that the higher the correlation between $h(x)$ and $f(x)$, the faster our estimate converges to the true quantity.

3.2.2 Weighted Monte Carlo

Control Variate can be viewed as a weighted Monte Carlo method. This method assigns weights to different realizations of $f(X)$ from above. β is estimated by sample variance and covariance. Given m as sample mean of $f(X_i)$, p as sample mean of $h(X_i)$:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(f(X_i) - m)(h(X_i) - p)}{\sum_{i=1}^{n}(h(X_i) - p)^2}. \tag{5}$$

Substituting the estimated β in Equation (2), we have assigned different weights to $f(X_i)$. The estimator then is given by

$$\hat{\mu}_{\beta} = \sum_{i=1}^{n}\left[\frac{1}{n} + \frac{(p - h(X_i))(p - \mathbb{E}(h(X))}{\sum_{i=1}^{n}(h(X_i) - p)^2}\right] f(X_i). \tag{6}$$

3.3 Importance Sampling–Exponential Twisting

3.3.1 Importance Sampling

Importance Sampling is one of the most robust way for simulation. The idea is to sample more from interest region/distribution and assign appropriate weight to the realizations. Define domains D and Q corresponding to densities $p(x)$ and $q(x)$, respectively. Then

$$\mu = \mathbb{E}(f(X)) = \int_D f(x)p(x)\,dx = \int_Q \frac{f_X(x)p(x)}{q(x)}q(x)\,dx = \mathbb{E}_q\left(\frac{f(X)p(X)}{q(X)}\right),$$

$$\hat{\mu}_q = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)p(X_i)}{q(X_i)}, \quad X_i \sim q. \tag{7}$$

In order to justify its efficiency, we derive the expected value and variance of the Importance Sampling estimator and realize that the estimator is also unbiased.

$$\begin{aligned} \mathbb{E}_q(\hat{\mu}_q) &= \mu, \\ \mathrm{Var}_q(\hat{\mu}_q) &= \frac{\sigma_q^2}{n}, \\ \sigma_q^2 &= \int_Q \frac{(f(x)p(x))^2}{q(x)} dx - \mu^2. \end{aligned} \tag{8}$$

This suggests that a very low variance can be achieved when $q(x)$ has a heavier tail in the area of interest compare to $p(x)$.

3.3.2 Exponential Twisting

Exponential Twisting proposes one simple way to derive importance sampling density $q(x)$. Let $\psi(\theta) = \ln \int_{-\infty}^{\infty} e^{\theta x} d(F(x))$ be the cumulant generating function (cgf) of $p(x)$. Then, the cumulative distribution function (cdf) and subsequently, the pdf of the importance density could be derived, we have:

$$F_\theta(x) = \int_{-\infty}^{x} e^{\theta u - \psi(\theta)} dF(u), \qquad p_\theta(x) = e^{\theta x - \psi(\theta)} p(x). \tag{9}$$

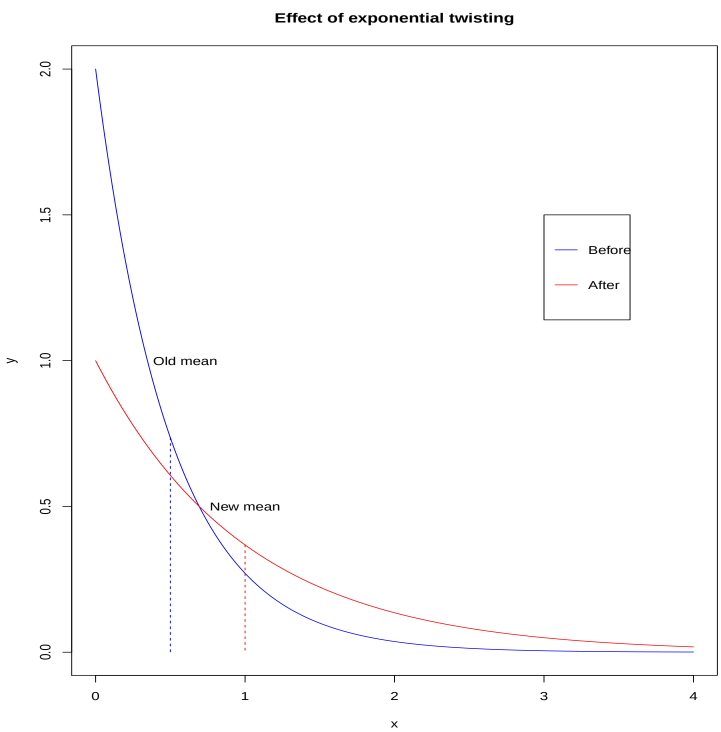

Figure 1: An example of exponential twisting applied to exponential distribution of rate parameter $\lambda = 2$ using $\theta = 1$. It results in a new distribution with a shifted mean equals $\frac{1}{\lambda - \theta}$.

Substituting in $w(x) = \frac{p(x)}{q(x)}$, the variance becomes:

$$\int_Q \frac{(f(x)p(x))^2}{q(x)} dx = \mathbb{E}_q(f(X)^2 w(X)^2). \tag{10}$$

Substituting the exponential twisted $w(X) = e^{\psi(\theta)-\theta X}$ in Equation (10), we have:

$$\begin{aligned}\int_Q \frac{(f(x)p(x))^2}{q(x)} dx &= \mathbb{E}_q(f(X)^2 w(X)^2) \\ &= \mathbb{E}_q(f(X)^2 e^{2\psi(\theta)-2\theta X}) \\ &= \mathbb{E}(f(X)^2 e^{\psi(\theta)-\theta X}).\end{aligned} \tag{11}$$

It is clear that choosing an effective parameter θ would be helpful in reducing the variance. We will derive its closed form formula for our problem of interest later on.

3.4 Stratification-Proportional Allocation

3.4.1 Stratified Sampling

The idea of stratified sampling could further improve the efficiency of importance sampling. The method splits the domain D into separate regions (strata) $D_j, j = 1, 2, \ldots, J$, then takes a certain sample points from each region. This would ensure the sufficient number of points needed from the region of interest. In order to use stratified sampling, one must know:

- Size of the regions, $w_j = P(X \in D_j)$, and
- How to sample from $X \sim p_j$, with $p_j(x) = w_j^{-1} p(x)\mathbb{I}\{x \in D_j\}$, where $\mathbb{I}\{x \in D_j\} = 1$ if $x \in D_j$ and $\mathbb{I}\{x \in D_j\} = 0$ otherwise.

Let $X_{ij} \sim p_j$ be i.i.d. samples, the stratification estimator is given by:

$$\hat{\mu}_{\text{strat}} = \sum_{j=1}^{J} \frac{w_j}{n_j} \sum_{i=1}^{n_j} f(X_{ij}). \tag{12}$$

We can see that the estimator is also unbiased since:

$$\begin{aligned}\mathbb{E}(\hat{\mu}_{\text{strat}}) &= \sum_{j=1}^{J} w_j \mathbb{E}\left(\frac{1}{n_j}\sum_{i=1}^{n_j} f(X_{ij})\right) \\ &= \sum_{j=1}^{J} \frac{p(x)}{p_j(x)} \sum_{D_j} p_j(x) f(x) dx \\ &= \int_D p(x) f(x) dx = \mu.\end{aligned} \tag{13}$$

The variance shows that the choice of strata sizes would affect the convergence speed:

$$\text{Var}(\hat{\mu}_{\text{strat}}) = \sum_{j=1}^{J} \frac{w_j^2}{n_j} \sigma_j^2. \tag{14}$$

3.4.2 Proportional Allocation

Proportional Allocation is a popular way to choose the stratum's sample size. It may not be the most efficient, but it is the simplest way to combine with Importance Sampling. This method allocates $n_j = nw_j$ points into stratum D_j. As a result, we have the following:

$$\hat{\mu}_{\text{prop}} = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j} f(X_{ij}),$$
$$\text{Var}(\hat{\mu}_{\text{prop}}) = \frac{1}{n}\sum_{j=1}^{J} w_j \sigma_j^2. \tag{15}$$

It can easily be proven that the estimator is also unbiased, and its variance is smaller than Crude Monte Carlo's, since weighted sum of each stratum's variance is small than the original variance.

4 Estimating risk measures using Monte Carlo methods

There are two measures of preference when financial institutions assess their portfolios' loss distribution. These are the Value at Risk and the Expected Shortfall. The two risk measures take into account the worst case scenarios or the extreme region of the loss distribution. In particular, let F_L be the cdf of portfolio's loss L:

- $(1-\alpha)\%$ VaR is the level of loss that institutions are $(1-\alpha)\%$ confident that portfolio will not lose more than $\text{VaR}_\alpha(L) = F_L^{-1}(1-\alpha)$.
- $(1-\alpha)\%$ ES is the expected level of loss, given that the loss has exceeded the confidence level, $\text{ES}_\alpha(L) = \frac{1}{\alpha}\int_\alpha^0 \text{VaR}_u du$.

Very often, the exact distribution of loss is hard to estimate and Monte Carlo methods are used to estimate these measures [4]. It is also noticeable of the importance of generating adequate amounts of extreme values in the tail area (α is around 1% to 2%), which Crude Monte Carlo might fail to deliver sufficient realizations of interest.

A common feature of financial models capturing the interaction between risk factor and loss distributions were the assumptions of normality [1]. The notorious Black-Scholes model assumes stock price follows Geometric Brownian Motion, which leads to normal expected return. In reality, financial asset returns' distribution might possess heavier tails than the normal distribution case. We will examine both cases with one of the most important market risk factors (stock price) and apply the three proposed Monte Carlo methods onto this portfolios.

Even though the methods are different in nature, there is one thing in common: all methods would produce realizations $L^{(i)}$ and assign weights $W^{(i)}$ to each one of them. If $L^{(1)}$ is the largest loss, $L^{(n)}$ is the smallest loss and after choosing our significance level α,

$$i_\alpha = \min\left[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\right]. \tag{16}$$

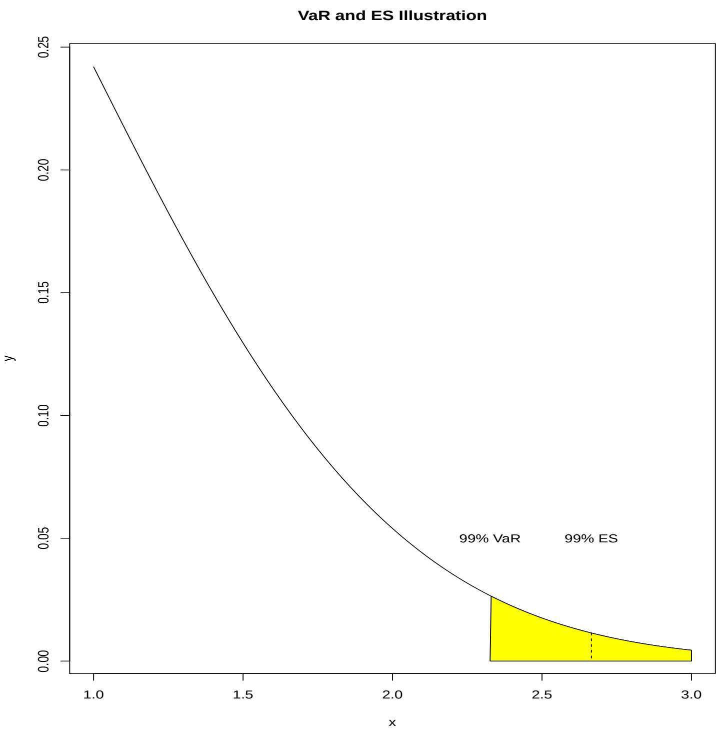

Figure 2: An example of VaR and ES for the standard normal distribution.

Then we have the following estimators for the VaR and the ES:

$$\widehat{\mathrm{VaR}}_\alpha = L^{(i_\alpha)},$$

$$\widehat{\mathrm{ES}}_\alpha = \frac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha-1} W^{(j)}L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha-1} W^{(j)}\right) L^{(i_\alpha)}\right].$$

4.1 Linear light tailed portfolio

One of the most fundamental model to value an asset price $S(t)$ is the Geometric Brownian Motion (GBM): $S(t)$ follows the GBM if

- $\frac{S(t_1)-S(t_0)}{S(t_0)}, \frac{S(t_2)-S(t_1)}{S(t_1)}, \ldots, \frac{S(t_k)-S(t_{k-1})}{S(t_{k-1})}$ are independent, and
- instantaneous changes are normally distributed, $\frac{dS(t)}{S(t)} = \mu dt + \sigma dW(t)$.

Because loss is calculated based on $\Delta S(t)$ rather than $\frac{\Delta S(t)}{S(t)}$, this can bring complications to our modelling. Such complications, however can be circumvented for our problem if we assume that for institutions, market risk factor such as stocks, derivatives can be liquidated quite quickly. This means that regulations usually require rather small horizon for evaluating the VaR or the ES. Because of that, we can say that $\Delta S(t)$ approximately follows normal distribution: $\Delta S(t) = S(t+\Delta t) - S(t) \approx S(t)\left(\frac{dS(t)}{S(t)}\right)$. Moreover, within such small horizon, it is safe to assume mean of $\Delta S(t)$ is 0. If the annual volatility of the stock is v, then the volatility of changes from [0,Δt] is $\sigma = v\sqrt{\Delta t}$. For a linear portfolio of shares, portfolio loss would be: $L = -\Delta S$, this means $L \sim N(0, \sigma^2 = v^2\Delta t)$. The sections below will address the specific algorithms of each Monte Carlo methods applied to this light tail case.

4.1.1 Control Variate-Weighted Monte Carlo

As realizations of $L_i = f(\Delta S_i) = -\Delta S_i$ can be generated from $N(0, \sigma_1^2 = v^2\Delta t)$, we could find a control variate that has high correlation with the original one. For this simple case, we could choose $h(\Delta S_i) = -2\Delta S_i$. In order to use this method, we need to know how to calculate the two quantities below:

$$\begin{aligned} L_i &= f(\Delta S_i), \Delta S_i \sim N(0, \sigma_1^2 = v^2\Delta t), \\ W_i &= \frac{1}{n} + \frac{(p - h(\Delta S_i))(p - E(h(\Delta S))}{\sum_{i=1}^{n}\Big(h(\Delta S_i) - p\Big)^2}. \end{aligned} \tag{17}$$

Algorithm 1 CV algorithm

1. Generate a vector S of N realizations of ΔS.
2. Create a $N \times 3$ empty matrix A.
3. Fill $A(i,1)$ with $f(S[i])$, $A(i,2)$ with $h(S[i])$, $A(i,3)$ with corresponding weights W_i calculated from equation (17).
4. Sort the matrix in decreasing order based on the first column.
5. Set $i_\alpha = \min\Big[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\Big]$.
6. Obtain $\widehat{\text{VaR}}_\alpha = L^{(i_\alpha)}$.
7. Obtain $\widehat{\text{ES}}_\alpha = \dfrac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha - 1} W^{(j)} L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha - 1} W^{(j)}\right) L^{(i_\alpha)}\right]$.

The algorithm shows one shortfall of Control Variate, it is that if α is small, then few simulations would fall into the area of interest [2]. Exponential Twisting method would prove to be superior in generating points in such extreme area.

4.1.2 Importance Sampling-Exponential Twisting

Exponential Twisting changes, or shifted our original density function with $p_\theta(x) = e^{\theta x - \psi(\theta)}p(x)$. In our specific case of $L \sim N(0, \sigma^2 = v^2\Delta t)$, our new loss distribution follows $L \sim N(\theta\sigma^2, \sigma^2 = v^2\Delta t)$. The only thing to do now is to find the efficient θ.

Now, let us focus only on the area of interest, $L > x$. Probability of this loss is $P(L > x)$, which can be connected to our integral estimation by $P(L > x) = E(\mathbb{I}\{L > x\})$. Using importance sampling principal, we acquire the estimator:

$$\hat{P}(L > x) = \frac{1}{n}\sum^{N} \mathbb{I}\{L > x\}\frac{p(L_i)}{q(L_i)}.$$

Exponential Twisting would give $\widehat{P}(L > x) = \frac{1}{n}\sum^{N} 1\{L > x\}e^{\psi(\theta)-\theta L_i}$. We should choose θ that could minimize the variance of this estimator. Let us examine the second moment of $\widehat{P}(L > x)$, which is $E_\theta(\mathbb{I}\{L > x\}e^{2\psi(\theta)-2\theta L_i})$. This is equivalent to equation (11). With the same modification, the second moment becomes $E(\mathbb{I}\{L > x\}e^{\psi(\theta)-\theta L_i})$, which is always less than or equal to $e^{\psi(\theta)-\theta x}$.

In order to minimize the standard error, the problem becomes choosing θ that minimizes $\psi(\theta) - \theta x$. This happens when $\psi'(\theta) = x$. Recall that loss distribution follows $N(0, \sigma_1^2 = v^2\Delta t)$ with the cgf $\psi(\theta) = \frac{\theta^2\sigma^2}{2}$. We derive the optimal $\theta = \frac{x}{\sigma^2}$. Note that even though x would be true quantile of interest, one can make a prediction close to the quantile.

In summary, to conduct the simulation, we need to know:

$$L_i = f(\Delta S_i), \Delta S_i \sim p_\theta,$$

$$W_i = \frac{1}{n}\frac{p(\Delta S_i)}{p_\theta(\Delta S_i)} = \frac{1}{n}e^{\psi(\theta)-\theta\Delta S_i}. \tag{18}$$

Algorithm 2 IS-ET algorithm

1. Generate a vector S of N realizations of $\Delta S \sim N(\theta\sigma^2, \sigma^2 = v^2\Delta t)$ and $L = f(\Delta S)$.
2. Create a $N \times 2$ empty matrix A, choosing x close to the true quantile and $\theta = \frac{x}{\sigma^2}$.
3. Fill $A(i, 1)$ with $L[i]$, $A(i, 2)$ with corresponding weights W_i calculated from equation (18).
4. Sort the matrix in decreasing order based on the first column.
5. Set $i_\alpha = \min\left[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\right]$.
6. Obtain $\widehat{\text{VaR}}_\alpha = L^{(i_\alpha)}$.
7. Obtain $\widehat{\text{ES}}_\alpha = \frac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha-1} W^{(j)}L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha-1} W^{(j)}\right)L^{(i_\alpha)}\right]$.

4.1.3 Stratified Sampling-Exponential Twisting

To even further reduce the variance, we apply stratification to sufficiently acquire the number of points of interest. The method will be used on the new exponential twisted distribution. When simulating N points, we might assign $N\alpha$ points to the tail area, thus creating two new strata. The problem here is trying to estimate the cutoff points between strata, which is simple to find out in our case.

Algorithm 3 SS-ET algorithm

1. Generate a vector S of N realizations of $\Delta S \sim N(\theta\sigma^2, \sigma^2 = v^2\Delta t)$ and $L = f(\Delta S)$. Acquire $N\alpha$ points above cutoff points and $(N - N\alpha)$ below cutoff points. Cutoff point $:= \sigma N^{-1}(1-\alpha)$.
2. Create a $N \times 2$ empty matrix A, choosing x close to true quantile and $\theta = \frac{x}{\sigma^2}$.
3. Fill $A(i,1)$ with $L[i]$, $A(i,2)$ with corresponding weights W_i calculated from equation (18).
4. Sort the matrix in decreasing order based on the first column.
5. Set $i_\alpha = \min\left[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\right]$.
6. Obtain $\widehat{\mathrm{VaR}}_\alpha = L^{(i_\alpha)}$.
7. Obtain $\widehat{\mathrm{ES}}_\alpha = \dfrac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha - 1} W^{(j)} L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha - 1} W^{(j)}\right) L^{(i_\alpha)}\right]$.

4.2 Linear heavy tailed portfolio

In reality it is witnessed that financial asset's return might follow heavier tail distribution with higher peak. This can be seen in stocks where there are small price changes in most of the times and few occurrences of drastic changes. Based on [2], one of the famous distributions to model heavy tailed price changes is Student's t-distribution with 3 to 7 degrees of freedom.

Suppose price changes X follow Student's t-distribution with degree of freedom ν. We have

$$\mathbb{E}(X) = 0, \qquad \mathrm{Var}(X) = \frac{\nu}{\nu - 2}.$$

We then continue to scale the t-distribution to achieve our necessary standard deviation. If our estimated volatility is σ, our scaled distribution should be $\sigma\sqrt{\frac{\nu-2}{\nu}}X$.

The sections below will address the specific algorithms of each Monte Carlo methods applied to this heavy tail case.

4.2.1 Control Variate

Just as the light tail case, we could choose $h(\Delta S_i) = -2\Delta S_i$ with ΔS_i follows scaled Student's t-distribution $\sigma\sqrt{\frac{\nu-2}{\nu}}\, t_\nu$. In order to use this method, we need to know how to calculate the two quantities below.

$$L_i = \sigma\sqrt{\frac{\nu-2}{\nu}}\, t_\nu,$$

$$W_i = \frac{1}{n} + \frac{(p - h(\Delta S_i))(p - \mathbb{E}(h(\Delta S))}{\sum_{i=1}^{n}\left(h(\Delta S_i) - p\right)^2}. \tag{19}$$

After that, we can simply apply Algorithm 1, the Control Variate algorithm.

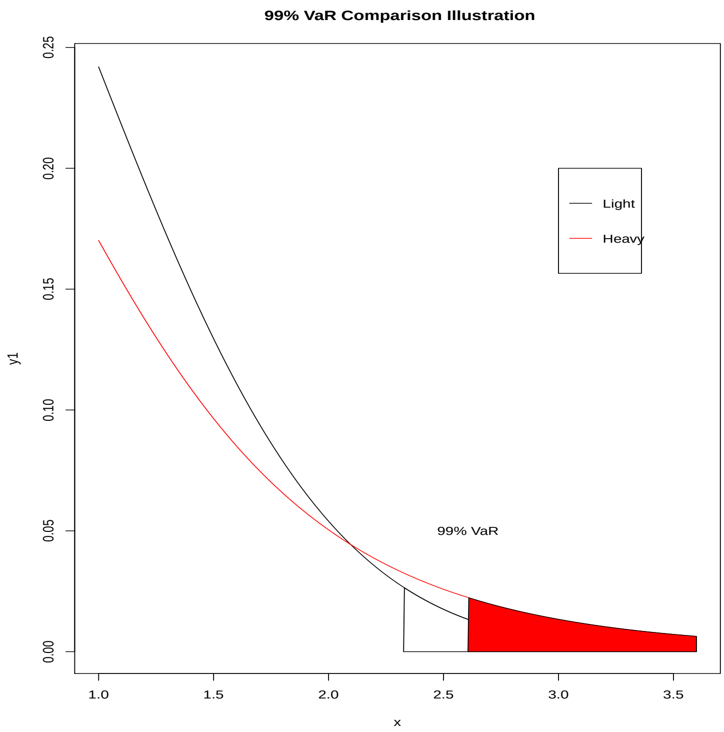

Figure 3: A comparison of the VaR for the standard normal and the scaled Student's t-distribution with 5 degrees of freedom.

4.2.2 Importance Sampling

Following the same logic from the light tailed case, exponential twisting requires a defined cgf and a good choice of parameter θ, which is dependent on the area of interest. For Student's t-distribution, the cgf does not exist. In [2], it was suggested a convenient but indirect method to approach this problem. We know that Student's t-distribution could be written as: $t_\nu = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$, with Z follows standard normal and Y follows Chi-squared with ν degree of freedom. This method takes advantage of the fact that conditional on variable Y, its moment generating function (mgf) takes the form of a scaled normal's mgf. To assist with choosing the most efficient θ, we further change our Student's t random variable $X = t_\nu = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$ into $X_x = \frac{Y}{\nu}(X - x) = Z\sqrt{\frac{Y}{\nu}} - x\frac{Y}{\nu}$. Let us construct our conditional moment generating function:

$$\mathbb{E}(e^{\theta X_x}|Y) = \exp\left\{-x\frac{Y}{\nu}\theta + \frac{1}{2}\frac{Y}{\nu}\theta^2\right\},$$

$$\mathbb{E}(\mathbb{E}(e^{\theta X_x}|Y)) = \mathbb{E}\left[\exp\left\{-x\frac{Y}{\nu}\theta + \frac{1}{2}\frac{Y}{\nu}\theta^2\right\}\right],$$

$$\mathbb{E}(e^{\theta X_x}) = \mathbb{E}\left[e^{(-\frac{x}{\nu}\theta + \frac{1}{2}\frac{\theta^2}{\nu})Y}\right] = \mathbb{E}\left[e^{a(\theta)Y}\right],$$

which is the chi-squared mgf with parameter $a(\theta)$.

We already know the moment generating function for chi-squared variable, which is $(1-2\theta)^{-\frac{\nu}{2}}$ with $\theta < \frac{1}{2}$.

Using the parameter $a(\theta)$ from above, we obtain the mgf, thus the cgf for the variable X_x: $\psi_x(\theta) = -\frac{\nu}{2}\ln\{1 - 2(-\frac{x}{\nu}\theta + \frac{1}{2}\frac{\theta^2}{\nu})\}$. Now we can exponentially twist the variable X_x instead of X. From section 4.1.2, our estimator would be $\widehat{P}(X > x) = \frac{1}{n}\sum^N \mathbb{I}\{X_i > x\}e^{\psi_x(\theta_x)-\theta_x X_{x,i}}$. The second moment is $\mathbb{E}_\theta(\mathbb{I}\{X > x\}e^{2\psi_x(\theta_x)-2\theta X_{x,i}})$, which is equivalent to $\mathbb{E}(\mathbb{I}\{X > x\}e^{\psi_x(\theta_x)-\theta X_{x,i}})$. The event $\{X > x\}$ happens when $\{X_x > 0\}$, which leads to $\mathbb{E}(\mathbb{I}\{X > x\}e^{\psi_x(\theta_x)-\theta X_x}) \leq e^{\psi_x(\theta_x)}$. The efficient choice of θ would minimize $e^{\psi_x(\theta_x)}$. From the cgf above, it is minimized when $\theta_x = x$.

The next step is to know how to sample $X = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$ from Z and Y after being exponentially twisted. In [3] it was derived their distributions (see Theorem 4.1). It shows that under p_θ:

$$Y \sim \text{Gamma}\left(\nu/2, 2/[1 - 2a(\theta)]\right), \qquad Z \sim N\left(\theta\sqrt{\frac{Y}{\nu}}, 1\right).$$

In summary, to conduct the simulation, we need to know:

$$\begin{aligned} L_i &= \frac{Z}{\sqrt{\frac{Y}{\nu}}}, \text{ under } p_\theta, \\ W_i &= \frac{1}{n}e^{\psi_x(\theta_x)-\theta_x X_x}, \\ \psi_x(\theta_x) &= \frac{-\nu}{2}\ln\left\{1 - 2(\frac{-x}{\nu}\theta + \frac{1}{2}\frac{\theta^2}{\nu})\right\}. \end{aligned} \tag{20}$$

Algorithm 4 IS-ET algorithm - Heavy tail

1. Choose x close to true quantile and $\theta_x = x$.
2. Create a $N \times 2$ empty matrix A, for each of N simulation:
   - Generate Y from $Y \sim \text{Gamma}\{\nu/2, 2/[1 - 2a(\theta)]\}$, generate Z from $Z \sim N\{\theta\sqrt{\frac{Y}{\nu}}, 1\}$.
   - Set $X = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$, $X_x = \frac{Y}{\nu}(X - x)$, $L = \sigma\sqrt{\frac{\nu-2}{\nu}}X$, $W = \frac{1}{n}e^{\psi_x(\theta_x)-\theta_x X_x}$ with $\psi_x(\theta_x)$ from equation (20).
3. Fill $A(i, 1)$ with $L[i]$, $A(i, 2)$ with corresponding weights W_i.
4. Sort the matrix in decreasing order based on the first column.
5. Set $i_\alpha = \min\left[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\right]$.
6. Obtain $\widehat{\text{VaR}}_\alpha = L^{(i_\alpha)}$.
7. Obtain $\widehat{\text{ES}}_\alpha = \frac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha-1} W^{(j)}L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha-1} W^{(j)}\right)L^{(i_\alpha)}\right]$.

4.2.3 Stratified Sampling

Just as illustrated in Section 4.1.3, stratification would help us acquire $N\alpha$ points for estimating the VaR and the ES inside the region of interest. Finding the cut off point separating the regions would be simple in our case, where we know the distribution already.

Algorithm 5 SS-ET algorithm - Heavy tail

1. Choose x close to true quantile and $\theta_x = x$.
2. Create a $N \times 2$ empty matrix A, acquire $N\alpha$ points above cutoff points and $(N - N\alpha)$ below cutoff points. Cutoff point $:= t^{-1}(1-\alpha, df = \nu)$. For each of N simulation:
   - Generate Y from $Y \sim \text{Gamma}\{\nu/2, 2/[1-2a(\theta)]\}$, Z from $Z \sim N\{\theta\sqrt{\frac{Y}{\nu}}, 1\}$.
   - Set $X = \frac{Z}{\sqrt{\frac{Y}{\nu}}}$, $X_x = \frac{Y}{\nu}(X - x)$, $L = \sigma\sqrt{\frac{\nu-2}{\nu}}X$ and $W = \frac{1}{n}e^{\psi_x(\theta_x)-\theta_x X_x}$ with $\psi_x(\theta_x)$ from equation (20).
3. Fill $A(i,1)$ with $L[i]$, $A(i,2)$ with corresponding weights W_i.
4. Sort the matrix in decreasing order based on the first column.
5. Set $i_\alpha = \min\left[j : \sum_{i=1}^{j} W^{(i)} \geq \alpha\right]$.
6. Obtain $\widehat{\text{VaR}}_\alpha = L^{(i_\alpha)}$.
7. Obtain $\widehat{\text{ES}}_\alpha = \frac{1}{\alpha}\left[\sum_{j=1}^{i_\alpha-1} W^{(j)}L^{(j)} + \left(\alpha - \sum_{j=1}^{i_\alpha-1} W^{(j)}\right) L^{(i_\alpha)}\right]$.

5 Convergence rate illustration

5.1 Light tail case

After knowing some of the theoretical algorithms above, we will apply those to our rare event simulations case. Figures 4, 6, 8, and 10 illustrate the convergence speed towards VaR and ES of Crude Monte Carlo, Control Variate, Importance Sampling and Importance Sampling plus Stratification, by slowly increasing the number of simulations.

It can be seen that as expected, Crude Monte Carlo's estimator displays much noise and fluctuations, even when after the fluctuations stabilise after 2000 simulations. Control Variate also seems to fail in this case. It might be because very few points are generated in the area of interest, which fails to correctly estimate the optimal coefficient β. Importance Sampling method is the most efficient, where Exponential Twisting has successfully shifted the distribution near the area of interest in order to generate more realizations. Stratification does not seem to add any value to Importance Sampling, since Importance Sampling has simulated enough points in that area already. Because of this, we would like to assess Stratification separately to the Heavy tail case.

5.2 Heavy tail case

Heavy tail rare event simulations require simulations in more extreme area, which means that we expect our four methods to perform worse than the Light tail case; see Figures 5, 7, 9, and 11. It can be seen that even Importance Sampling method is less stable before 500 simulations, but it is still the most powerful variance reduction method. After being separated for Importance Sampling, Stratification does not seem very attractive anymore. The method actually does not outperform Crude Monte Carlo and Control Variate when 1500 - 2000 simulations is reached, however, it holds its merits around 500 simulations.

6 Conclusion

We have reviewed four most popular Monte Carlo methods and compared them by estimating a rare event probability. The three variance reduction methods (Control Variate, Importance Sampling and Stratification) have interesting ideas in increasing the speed of convergence. Control Variate takes advantage of the correlation between functions to reduce standard error, Importance Sampling samples more points from the area of interest, while Stratification ensures “adequate” amount points in the area of interest. From the analysis, we have learned that while scarcity of realizations in the area of interest makes Control Variate inefficient, only getting adequate amount of realizations from Stratification is not optimal if we want small sample simulation. Therefore, the important aspect to remember when doing rare event probability estimation is to get as many points as possible in the extreme area of interest, which Importance Sampling is quite robust and powerful. Financial institutions, as a result might want to research more into this specific method as a way to reduce computational cost when estimating risk measures.

References

[1] Brendan O Bradley and Murad S Taqqu. Financial risk and heavy tails. In *Handbook of heavy tailed distributions in finance*, pages 35–103. Elsevier, 2003.

[2] Paul Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer Science & Business Media, 2013.

[3] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269, 2002.

[4] Philippe Muller. Computation of risk measures using importance sampling. Master’s thesis, ETH Zurich, 2010.

[5] Art B Owen. Monte carlo theory, methods and examples. *https://statweb.stanford.edu/ owen/mc/. Last access in Feb 2021.*

Appendix

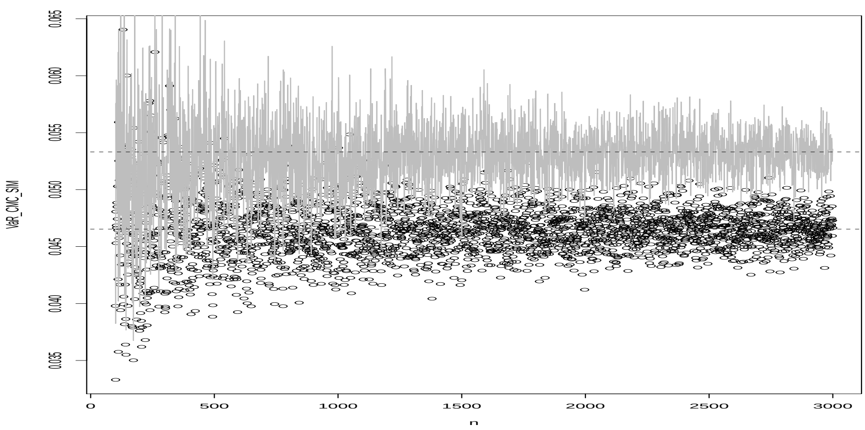

Figure 4: Graphical illustrations of Light tail convergence rate for CMC

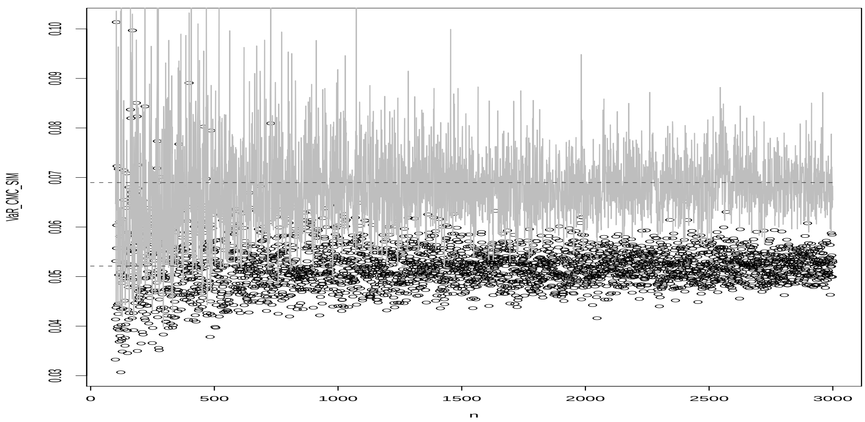

Figure 5: Graphical illustrations of Heavy tail convergence rate for CMC

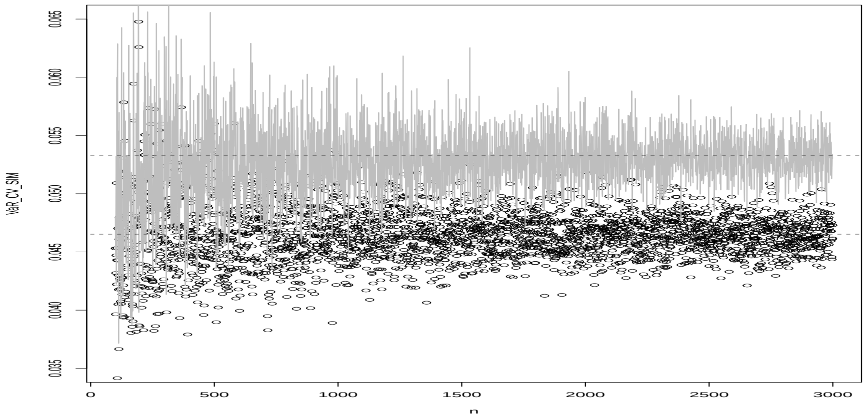

Figure 6: Graphical illustrations of Light tail convergence rate for CV

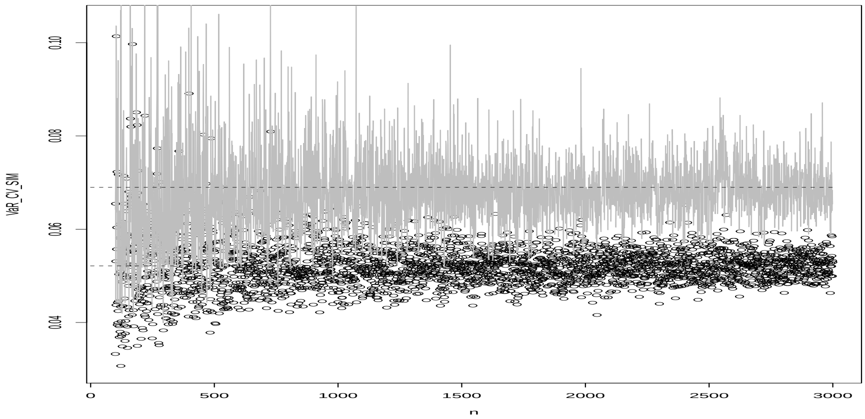

Figure 7: Graphical illustrations of Heavy tail convergence rate for CV

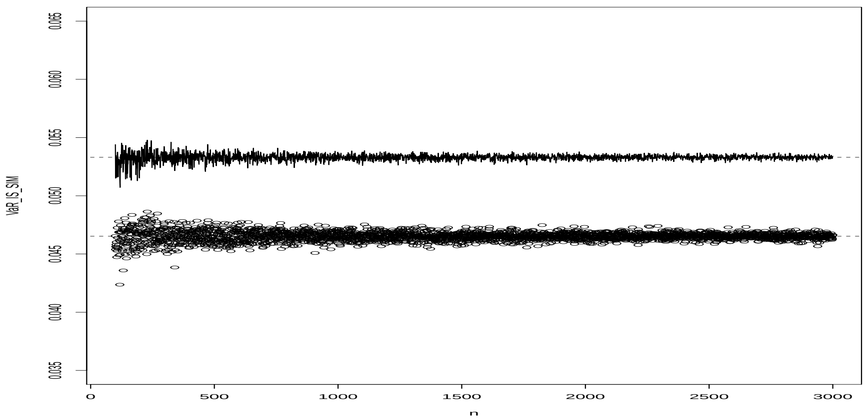

Figure 8: Graphical illustrations of Light tail convergence rate for IS

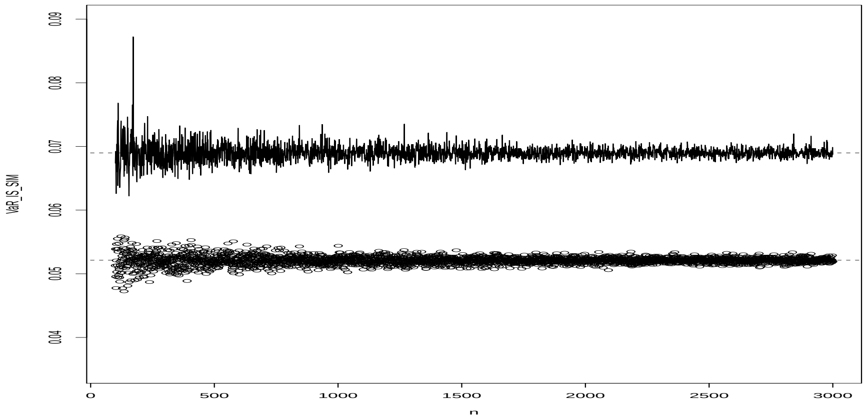

Figure 9: Graphical illustrations of Heavy tail convergence rate for IS

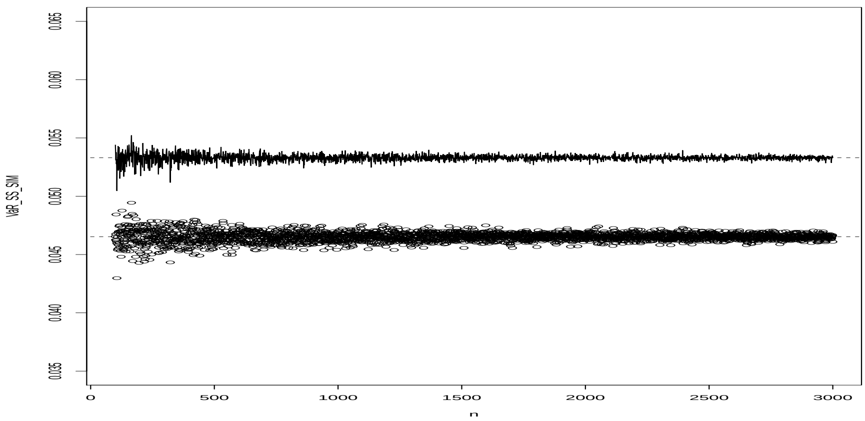

Figure 10: Graphical illustrations of Light tail convergence rate for SS+IS

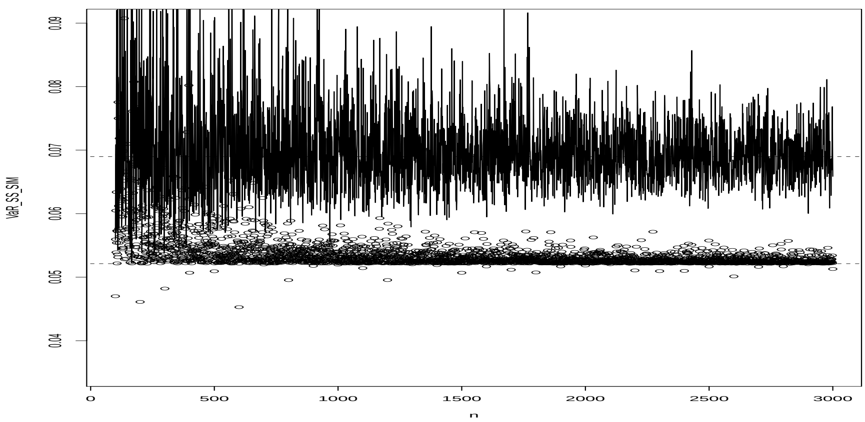

Figure 11: Graphical illustrations of Heavy tail convergence rate for SS