Simulating Risk Measures with Estimated Relative Errors

Wei Jiang
Department of Mathematics, National University of Singapore, A0120703@nus.edu.sg
Steven Kou
Risk Management Institute and Department of Mathematics, National University of Singapore, matsteve@nus.edu.sg

Risk measures, such as value-at-risk and expected shortfall, are widely used in stochastic models. With the necessary sample size being computed using asymptotic expansions of relative errors for a wide class of dependent samples, we propose a general framework to simulate these risk measures via a sorting algorithm. The asymptotic expansions appear to be new even for independent and identical samples. An extensive numerical study is conducted to compare the proposed algorithm against existing algorithms, showing that the new algorithm is easy to implement, fast and accurate, even at the 0.001 quantile level. Applications to the estimation of intra-horizon risk and to a comparison of the relative errors of value-at-risk and expected shortfall are also given.

Key words: relative errors, importance sampling, geometric α-mixing, order statistics, estimation time.
History:

1. Introduction

Risk measures, such as value-at-risk and expected shortfall, are widely used in risk management (see, e.g., Jorion 2007 and McNeil, Frey and Embrechts 2009) and inventory control (see, e.g. Chen et al. 2005, Tapiero 2005, Choi et al. 2011, Shapiro et al. 2014). Monte Carlo simulation is widely used to compute these risk measures, because the underlying stochastic models are too complicated to yield analytical solutions.

For example, the value-at-risk (VaR), defined as a quantile of a loss distribution, has been recommended by Basel II as a basis for the Bank of International Settlement's market-risk-based capital requirement, which has been implemented in over 60 member countries. In addition, the Bank of International Settlement is considering to also use the expected shortfall (ES), which is the expectation of the loss beyond certain VaR level. For a recent review of the regulations, see, e.g.,

the references cited in Kou and Peng (2014). Indeed, Basel Accords prescribe a standard procedure to calculate risk measures in two steps: In the first steps, model parameters are estimated based on samples from at least one year. In the second step, risk measures are computed via one of the 3 approaches: variance-covariance matrices, historical scenario simulation, and Monte Carlo simulation. As discussed in Hong and Liu (2009), Sun and Hong (2010) and Hong et al. (2014), the third approach, Monte Carlo simulation approach, is widely used due to its convenience to incorporate various risk models and in analyzing complex portfolios.

This paper attempts to propose a general framework to simulate these risk measures via an sorting algorithm, with the necessary sample size being computed using asymptotic expansions of relative errors for a wide class of dependent samples. To our best knowledge, the asymptotic expansion is new even for independent and identically distributed (i.i.d.) samples. An extensive numerical study shows that the algorithm appears to be accurate and fast, compared to the existing ones.

1.1. Three Model Settings and Our Contribution

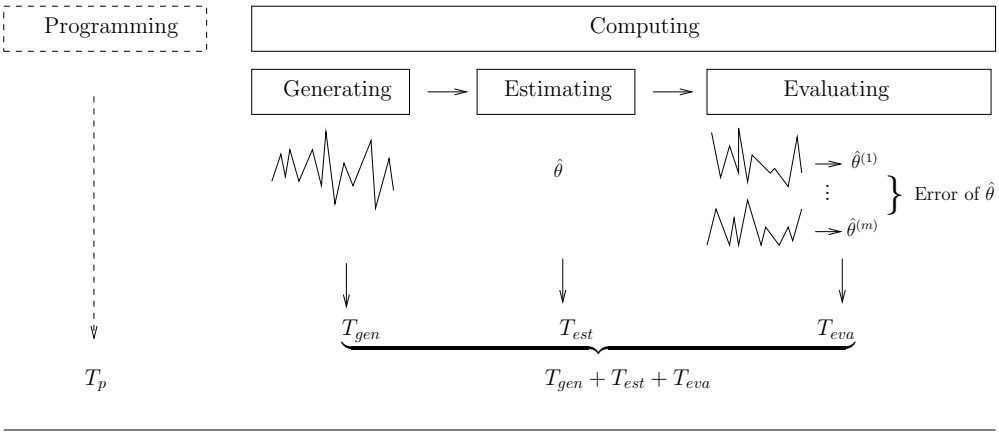
The first main setting is that we consider dependent samples, while many of the existing papers on simulating risk measures focus on i.i.d. samples. Many financial data (e.g. Campbell et al. 1997) and demand data in inventory control (see, e.g., Aviv 2003, Gilbert 2005) exhibit complicated time series dependence. Here we consider a wide class of dependent models, called geometric $\alpha-$mixing models, as introduced in a classical paper by Rosenblatt (1961). This class includes many commonly used time series models, such as ARMA (p,q), GARCH (p,q), Vasicek Model, some stochastic volatility models; see, e.g. Chen and Tang (2005) and Chen (2008).

Secondly, we take a standard approach to evaluate the accuracy of a simulation estimator, which is to use the relative error (RE), defined as the ratio of the standard deviation and the absolute mean value of the estimator, i.e. $RE(\hat{\theta}_n) = \sqrt{\text{var}(\hat{\theta}_n)}/|E(\hat{\theta}_n)|$, for an estimator $\hat{\theta}_n$ for θ, where n is the sample size; see Glasserman (2003).

Thirdly, one needs to clarify the meaning of simulation time. Consider a typical simulation procedure as shown in Figure 1. In addition to the programming time T_p, which is very important but not easy to quantify, there are three parts of the total execution time: the time for generating samples T_{gen}, the time for estimating risk measures T_{est}, and the time for evaluating the accuracy of the estimator T_{eva}. We propose to consider the total execution time, i.e. $T_{gen} + T_{est} + T_{eva}$, as a criterion to evaluate simulation time. The inclusion of T_{eva} appears to be a new suggestion in this paper. Note that the three parts of the total execution time all contribute differently to the total execution time. For example, a simulation algorithm may require very few simulation samples, hence small T_{gen}, but long estimation time for each simulation sample, hence large T_{est}. In addition,

complicated algorithms tend to make the evaluation step difficult, hence large T_{eva}. Later we shall compare the total execution time for various simulation algorithms with similar levels of relative errors.

Figure 1 Standard simulation procedures



It should be noted that our proposed algorithm is easy to program, as the samples are drawing directly from original loss processes and the estimators are based on order statistics which enable the use of "quick sort" algorithm. Thus the programming time T_p of our proposed algorithm is also reasonable.

The contribution of the current paper is fourfold: (1) We rigorously derive the expansions of relative errors for both VaR and ES under the $\alpha-$mixing model, which appear to be new even for i.i.d. samples; see Theorem 1 in Section 2. (2) With the necessary sample size being computed using the new asymptotic expansions, we propose a general framework to simulate these risk measures via a sorting algorithm; see Sections 3. (3) We conduct an extensive numerical study, comparing the proposed algorithm against existing algorithms by re-programming them using the same computer. Numerical results indicate that the new algorithm is easy to implement, fast and accurate, even at the 0.001 quantile level; see Tables 4, 5, 6, 7 in Section 4. Note that the standard risk management guidance (e.g. Basel Accords) in financial industry requires computing risk measures with moderate quantile, i.e. between 0.01 and 0.001, and our propose method can well satisfy that requirement. (4) We give two applications: The first one is the estimation of intra-horizon risk using S&P 500 index from 1999 to 2016 in Section 5, and the second one is a comparison of the relative errors of value-at-risk and expected shortfall in Section 6.

1.2. Literature Review

Many of the existing papers on Monte Carlo simulation of risk measures focus on i.i.d. samples; see Panel A in Table 1. In terms of dependent samples, a related paper is Heidelberger and Lewis (1984), in which a maximum transformation method is proposed to simulate VaR, with the main motivation being to save the computer storage cost (which was a significant factor then but perhaps no longer so now). More precisely, they suggested to regroup the original dependent samples into some subgroups, and then sort each subgroups to find the estimator of quantile for the original simulation data, yielding smaller storage cost; see Table 2. They also pointed out that the method may inflate the variance of the estimator and change the dependent structures after the regrouping.

Our method complements Heidelberger and Lewis (1984) in two ways. First, we sort the original samples directly, avoiding inflating the variance of estimators and changing the dependent structure of original samples (see Table 3). Secondly, the preservation of dependent structure, i.e. $\alpha-$mixing and stationary, allows us to use the relative error expansions, which ultimately enables us to perform simulation faster by estimating the necessary sample size required beforehand.

Table 1 Some literature on simulating VaR and ES

A: I.I.D. Samples	
Glasserman et.al (2000, 2002)	indirect importance sampling (IIS) and indirect importance sampling with stratification (IIS-Q), for VaR
Sun and Hong (2010)	direct importance sampling, for VaR & ES
Fuh et al. (2011)	indirect importance sampling on portfolio with heavy-tailed risk factors, for VaR
B: Dependent Samples	
Heidelberger and Lewis (1984)	maximum transformation, for VaR
This paper	sorted Monte Carlo with estimated RE, for VaR & ES

Table 2 Compared with Heidelberger and Lewis (1984)

This paper	Heidelberger and Lewis (1984)
$X_1, \cdots, X_n$	$X_1, \cdots, X_n$
$\Downarrow$ sort	$\Downarrow$ regroup, find minimal
$X_{(1)}, \cdots, X_{(n)}$	$Y_1, \cdots, Y_m,\ n = mv$
	$\Downarrow$ sort
	$Y_{(1)}, \cdots, Y_{(m)}$
$\Downarrow$	$\Downarrow$
$v_n = X_{([np])}$	$v_n = Y_{([mp^v+1])}$

Our expansions of relative errors for both VaR and ES under the geometric $\alpha-$mixing model appear to be new even in the special case of i.i.d. samples. As suggested in Hult and Svensson

Table 3 Technical comparison results with Heidelberger and Lewis (1984)

	This paper	Heidelberger and Lewis (1984)
Risk measures	VaR & ES	VaR
Stationary	available	not available
$\alpha-$mixing	available	not available
Storage cost	not reduced	reduced
Variance	not inflated	inflated
RE expansion	yes	no

(2009), one might calculate the relative errors heuristically by using the central limit theorem; more precisely, one can approximate the standard deviation and the absolute expectation of an estimator by their corresponding terms in the limiting normal distribution. Because the convergence in distribution does not imply the convergence in moments, there is a counterexample for this heuristics (see Example 1 in Appendix B of this paper) even in the case of i.i.d. samples. Some very useful almost sure and convergence-in-probability type expansions for both VaR and ES are given in Sun and Hong (2010) and Hong et al. (2014). Here the expansions for the relative errors require a different convergence: moment convergence (see Example 2 in Appendix B of this paper). A related literature is about asymptotic relative errors for estimators of tail probabilities; see, e.g., Dupuis et al. (2007), Bassamboo et al. (2007), Blanchet and Glynn (2008), Blanchet and Liu (2008) and Hult (2011). However, VaR and ES are related to quantile and are inverse problems of tail probabilities; how to analyze the relative errors from the inverse problem remains an open problem.

The key idea of getting the expansions for the relative errors is to get some useful expansions for the moments of VaR and ES, in Theorems 4 and 5, respectively. Here we get the expansions successfully mainly because we use 4 tools: (1) a Berry-Essen type bound for $\alpha-$mixing sequences from Tikhomirov (1980); (2) a Bernstein type inequality for $\alpha-$mixing sequences Merlevede, Peligrad and Rio (2009); (3) an improved argument inspired by Liu and Yang (2012). The above three tools lead to the moment expansion for VaR in Theorem 4 in the online supplement. (4) A similar connection between VaR and ES as in Sun and Hong (2010) yield a moment expansion for ES in Theorem 5 in the online supplement.

The rest of the paper is organized as follows: The asymptotic expansions for the RE's are given in Section 2. The algorithm is given in Section 3, along with an extension numerical comparisons of different estimators in Section 4. The algorithm is applied in Section 5 to estimate intra-horizon risk. Section 6 compares the RE's for Value-at-risk, median shortfall, and expected shortfall estimators.

Technical details, including the regular conditions and asset models, are given in the appendix, while the proofs of lemmas and theorems are presented in the E-Companion.

2. Analysis of Relative Errors

Consider a profit L, whose distribution is denoted by F; note that L is a loss when $L<0$. For a given level $p\in(0,1)$, VaR is defined as $v(p)\equiv v=\inf\{x:F(x)\geq p\}$, and ES is $c(p)\equiv c=E(L|L<v(p))$. Given time series (dependent) samples $\{L_i\}_{i=1}^n$, the empirical estimators for $v(p)$ and $c(p)$ are given by $v_n(p)=\inf\{x:F_n(x)\geq p\}, c_n(p)=v_n(p)-\frac{1}{np}\sum_{i=1}^n[v_n(p)-L_i]^+$, respectively, where $F_n(x)=\frac{1}{n}\sum_{i=1}^n I_{\{L_i\leq x\}}$ is the empirical distribution of $\{L_i\}_{i=1}^n$. In terms of the order statistics, they can also be written as

$$v_n(p)=\begin{cases} L_{([np])} & \text{if } np \text{ is integer,}\\ L_{([np]+1)} & \text{otherwise,}\end{cases} \tag{1}$$

$$c_n(p)=v_n(p)-\frac{1}{np}\sum_{i=1}^n(v_n(p)-L_i)1_{\{L_i\leq v_n(p)\}} \tag{2}$$

where $L_{(r)}$ is the $r-$th order statistic of $\{L_i\}_{i=1}^n$.

Our expansions for relative errors are for geometric $\alpha-$mixing models. The dependent series $\{L_i\}_{i=1}^n$ is said to be $\alpha-mixing$, as introduced in a classical paper by Rosenblatt (1961), if $\lim_{k\to\infty}\alpha(k)=0$, where

$$\alpha(k)=\sup_{A\in\mathscr{F}_1^i,B\in\mathscr{F}_{i+k}^n}|P(AB)-P(A)P(B)|,$$

and $\mathscr{F}_k^l$ be the $\sigma-$algebra of events generated by $\{L_i,k\leq i\leq l\}$ for $l\geq k$. The series is said to be geometric $\alpha-mixing$ if $\alpha(k)\leq c_\alpha\rho^k$ for some constants $c_\alpha>0$ and $\rho\in(0,1)$. Many popular econometric models belomg to the geometric $\alpha-$mixing class, including ARMA(p,q), GARCH(p,q), Diffusion Model (Vasicek Model), Stochastic Volatility Model.

2.1. Expansions for Relative Errors

Consider the following two general types of simulation models.

Type A: *Dependent samples from a stationary model.*

The losses $\{L_i\}_{i=1}^{m+1}$ form a stationary time series. Since many time series models are asymptotic stationary (e.g. ARCH model, Stochastic Volatility model), to get a stationary distributed loss sample and to meet this requirement approximately, we can drop the first m (i.e. $m\geq 10000$) data $\{L_i\}_{i=1}^m$, or in a short notation,

$$(L_1,\cdots,L_m),L_{m+1},\cdots,L_{m+N},$$

in which the simulation data $\{L_{m+1}, \cdots, L_N\}$ are approximately stationary. In contrast, a traditional way is to repeat simulation procedures N times to get N i.i.d loss samples $\{L_{m+1}^{(1)}, \cdots, L_{m+1}^{(N)}\}$, i.e.

$$\begin{array}{l} (L_1^{(1)}, \cdots, L_m^{(1)}), \quad L_{m+1}^{(1)} \\ (L_1^{(2)}, \cdots, L_m^{(2)}), \quad L_{m+1}^{(2)} \\ \cdots \\ (L_1^{(N)}, \cdots, L_m^{(N)}), \; L_{m+1}^{(N)}. \end{array}$$

However, in this way to get N samples mN data points are dropped. Therefore, it is worthwhile to consider Type A in simulation to use simulation data more efficiently. This type of samples has been considered in Heidelberger and Lewis (1984); see Table 4 in Section 4 for a numerical comparison.

Type B: *i.i.d samples from a (non-stationary) time series model.*

For non-stationary time series model, we generate N sample path to get N i.i.d loss samples at time t:

$$\begin{array}{l} (L_1^{(1)}, \cdots, L_{t-1}^{(1)}), \quad L_t^{(1)} \\ (L_1^{(2)}, \cdots, L_{t-1}^{(2)}), \quad L_t^{(2)} \\ \cdots \\ (L_1^{(N)}, \cdots, L_{t-1}^{(N)}), \; L_t^{(N)}. \end{array}$$

This type of samples has been considered in many exiting papers, e.g. Heidelberger and Lewis (1984), Sun and Hong (2010), Hong et al. (2014), Glasserman et al. (2000), Glasserman et al. (2002) and Fuh et al. (2011); see Tables 5, 6, 7 in Section 4 for numerical comparisons.

THEOREM 1. *(1)(Dependent samples) For the geometric α mixing series, we have the following expansions under Assumptions A, B, C,*

$$RE(v_n) = -\frac{\sigma_{n,v}}{vf(v)} n^{-1/2} + o(n^{-3/4}), \quad RE(c_n) = -\frac{\sigma_{n,c}}{cp} n^{-1/2} + o(n^{-3/4+\epsilon}) \tag{3}$$

where f is the density of L, $\sigma_{n,v}^2 = \{p(1-p) + 2\sum_{k=1}^{n-1} \gamma_1(k)\}$, $\gamma_1(k) = cov\{1_{\{L_1<v\}}, 1_{\{L_{k+1}<v\}}\}$, and $\sigma_{n,c}^2 = \{var[(v-L_1)^+] + 2\sum_{k=1}^{n-1} \gamma_2(k)\}$, $\gamma_2(k) = cov\{(v-L_1)^+, (v-L_{k+1})^+\}$.

(2)(IID samples) For i.i.d. samples, we have the following expansions under the conditions of Assumptions A1, B, C:

$$RE(v_n) = -\frac{\sigma_v}{vf(v)} n^{-1/2} + o(n^{-3/4}), \quad RE(c_n) = -\frac{\sigma_c}{cp} n^{-1/2} + o(n^{-3/4+\epsilon}) \tag{4}$$

where f is the density of L, where $\sigma_v^2 = var[I_{\{L\le v\}}] = p(1-p)$ and $\sigma_c^2 = var[(v-L)^+]$.

Proof. See E-Companion EC.2.. □

The key idea of proof is to get some useful expansions for the moments of VaR and ES first, in Lemmas 7 and 8, respectively. Here we get Theorems 4 and 5 successfully mainly because we use 4 tools: (1) a Berry-Essen type bound for $\alpha-$mixing sequences from Tikhomirov (1980); (2) a

Bernstein type inequality for $\alpha-$mixing sequences from Merlevede, Peligrad and Rio (2009); (3) an improved argument inspired by Liu and Yang (2012). The above three tools lead to the moment expansion for VaR VaR in Theorem 4 in the online supplement. (4) Then a similar connection between VaR and ES as in Sun and Hong (2010) yields a moment expansion for ES in Theorem 5 in the online supplement. Even in the case of i.i.d. cases, Theorems 1, 4 and 5 appear to be new; in the case of i.i.d., Hall and Martin (1988) and Reiss (1989) develop approximations for the second order moments and higher order of moments of VaR, respectively (i.e. similar to Theorem 4), all under stronger conditions.

Theorem 1 is not only useful in estimating necessary sample sizes need in a Monte Carlo simultion to achieve a given accuracy of relative error (to be discussed the next), but also to help to make comparison of REs between VaR and ES, which will be discussed in Section 5.

2.2. Numerical Performance of the Relative Error Expansions and Related Sample Size calculation

The expansions in Theorem 1 suggest the follow approximations for the relative errors of VaR v_n and ES c_n:

$$\hat{RE}(v_n) = -\frac{\hat{\sigma}_{n,v}}{v_n \hat{f}(v_n)} n^{-1/2}, \quad \hat{RE}(c_n) = -\frac{\hat{\sigma}_{n,c}}{c_n p} n^{-1/2}, \tag{5}$$

where $\hat{\sigma}_{n,v}, \hat{\sigma}_{n,c}$ are computed using the spectral method in Heidelberger and Welch (1981), and the density function $f(\cdot)$ can be estimated by the standard kernel method. To show the accuracy of the approximation, we also use the traditional resampling method to compute the relative errors, i.e. repeating the estimation many times to compute error. More precisely, if the estimate is $\hat{\theta}$, then we repeat the estimation m times and obtain $\hat{\theta}^{(1)}, \cdots, \hat{\theta}^{(m)}$. The resampled RE $\tilde{RE}(\hat{\theta})$ is computed by

$$\tilde{RE}(\hat{\theta}) = \frac{\sqrt{\sum_{i=1}^{m} (\hat{\theta}^{(i)} - \bar{\hat{\theta}})^2/(m-1)}}{\bar{\hat{\theta}}}, \tag{6}$$

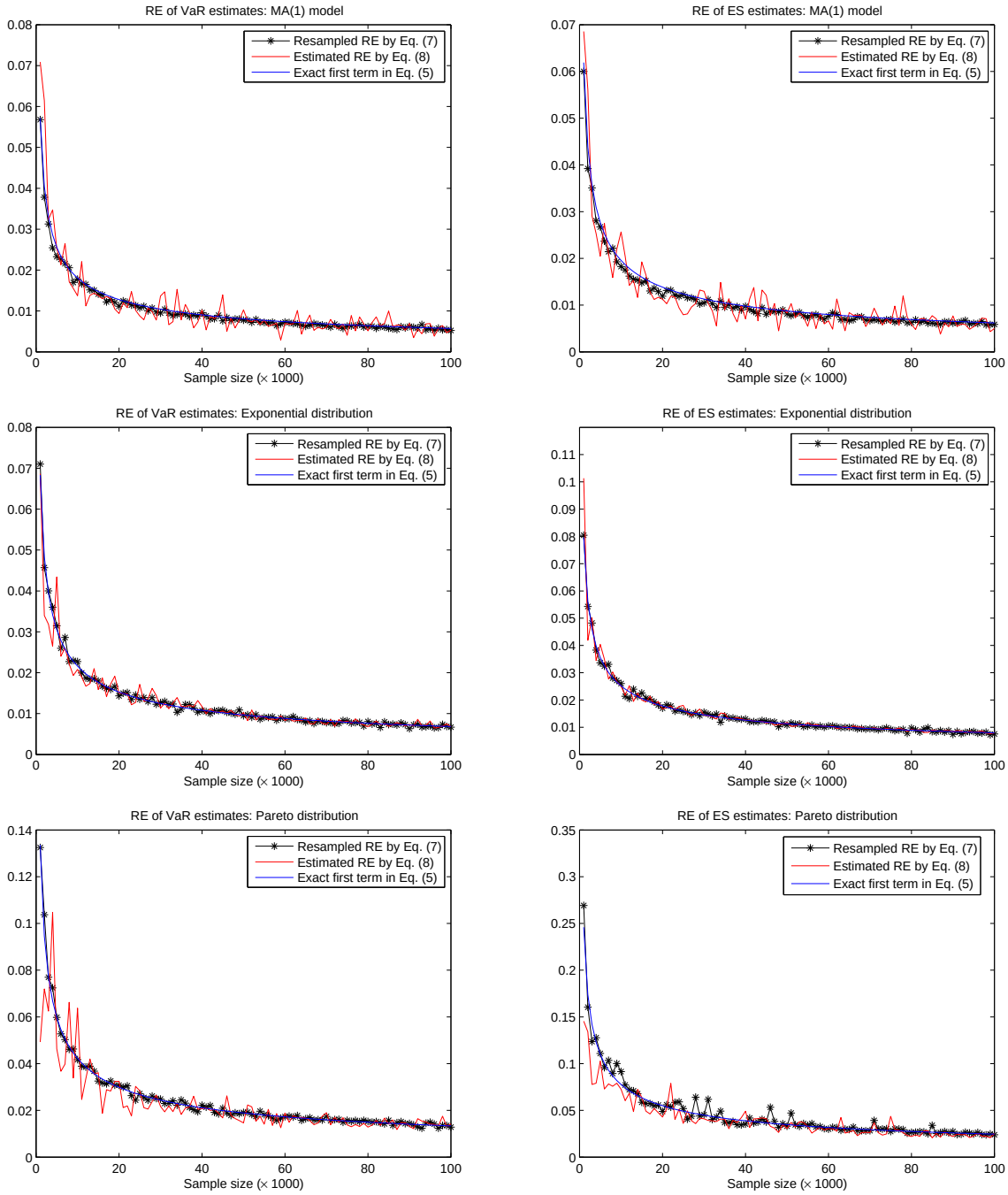
where $\bar{\hat{\theta}} = \sum_{i=1}^{m} \hat{\theta}^{(i)}/m$. Of course, the resampling method is a time consuming way while our approximations based on the expansion are much faster.

Figure 2 shows the performances of the estimated REs for both dependent samples and i.i.d samples. We use resampled REs obtained from 200 repeats as the true REs. The estimated REs match the true REs very well, especially for i.i.d samples with large sample size.

Approximations in Theorem 1 can help us to estimate the sample size needed to achieve a given level of relative error. More precisely, given a fixed level α for relative error, inverting (5) leads to n_α^v and n_α^c, which are the approximate sample size that makes relative errors of v_n and c_n to be around level α:

$$\hat{n}_\alpha^v = \frac{\hat{\sigma}_{n,v}^2}{v_n^2 \hat{f}(v_n)^2 \alpha^2}, \quad \hat{n}_\alpha^c = \frac{\hat{\sigma}_{n,c}^2}{c_n^2 p^2 \alpha^2}. \tag{7}$$

Figure 2 Estimated REs and Resampled REs for VaR and ES.



Note. The top two graphs report results for the MA(1) Model: $L_{t+1} = 0.5\epsilon_t + \epsilon_{t+1}$, $\epsilon_t \sim N(0,1)$. The other four graphs report results for i.i.d loss samples with exponential density $f(x) = e^x 1_{\{x<0\}}$, and Pareto density $f(x) = (1-x/3)^{-4} 1_{\{x<0\}}$. Here the level $p = 0.05$ and resampled relative errors are obtained from 200 repeated samplings.

One might think of using upper bounds to get some conservative estimation of the required sample sizes, instead of using the expansions. However, finding suitable bounds for the relative errors of VaR and ES may be difficult for two reasons: First, the bounds may not be available, especially

if the estimators involving importance sampling or more complicated models. For example, to our best knowledge, even in the case of i.i.d. samples, for direct importance sampling, no bounds for relative errors are available for importance sampling estimators (see, e.g., the discussions in Sun and Hong 2010 and Liu and Yang 2012); in addition, for indirect importance sampling method, even in the case of i.i.d. samples, the moments of tail probability estimators do not easily lead to a computable bound for the relative errors for quantile estimate (see, e.g., the discussions in Glasserman et al. 2000 and Glasserman et al. 2002), as the inversion of tail probabilities make the calculation more complicated. Second, although there are bounds available for direct estimators without importance sampling, the bounds may be too large, and may not get smaller with the increasing of sample size. For example, in the case of i.i.d. samples, Papadatos (1995) proposes the bound for variance of order statistic (without any importance sampling) $L_{(r)}(r=[np])$ However, for a fixed p, the upper bound is always bounded below by a constant, no matter how large the sample size n is.

3. Sorted Monte Carlo with Estimated Relative Errors

Based on the expansion in Theorem 1, we propose the following algorithm of sorted Monte Carlo with estimated relative errors as follows:

Algorithm of sorted Monte Carlo with estimated relative errors

I. Initialize: Set $i \to 0$, $N_0(e.g. \to 10000)$, $RE_0 \to \alpha_0(\alpha_0 \gg \alpha)$, the sample set $\mathcal{S}_0 \to \phi$, and its size $M_0 \to \sharp(\mathcal{S}_0)$.

II. S-step :

Sub-step 1: Generate N_i loss samples $\{L_1, \cdots, L_{N_i}\}$, let $M_{i+1} = M_i + N_i$, and form the sample set $\mathcal{S}_{i+1} = \mathcal{S}_i \bigcup \{L_1, \cdots, L_{N_i}\}$.

Sub-step 2: Sort the samples in $\mathcal{S}_{i+1}$ in ascending order, and determine $\hat{\theta}_{i+1} = v_{M_{i+1}}$ for VaR and $\hat{\theta}_{i+1} = c_{M_{i+1}}$ for ES by estimators (1) and (2).

Sub-step 3: Report $RE_{i+1} = \hat{RE}(v_{M_{i+1}})$ for VaR and $RE_{i+1} = \hat{RE}(c_{M_{i+1}})$ for ES according to (5).

Sub-step 4: If $RE_{i+1} < \alpha$, stop with output $\hat{\theta}_{i+1}$ and $\hat{RE} = RE_{i+1}$. Otherwise go to the R-step.

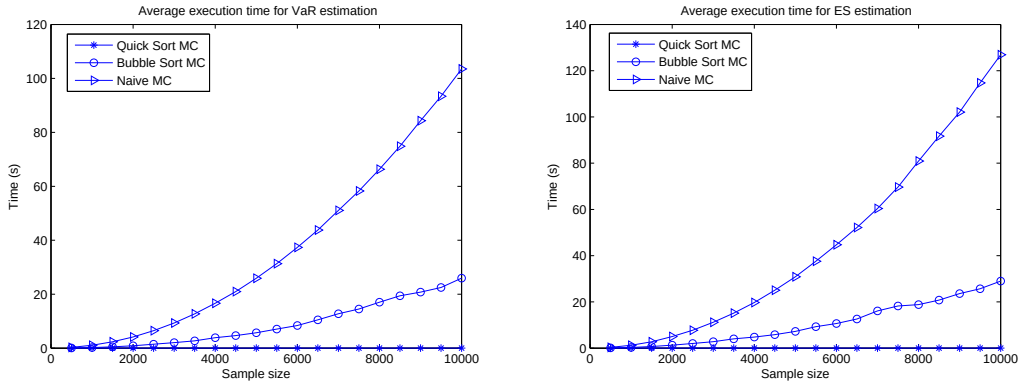
III. R-step: Compute the sample size $N_{i+1} = n_\alpha^v$ for VaR and $N_{i+1} = n_\alpha^c$ for ES by (7) and return to S-step by replacing i with $i+1$.

With fast computers and the availability of storage, several sorting methods can be applied to conduct Sub-step 2 in our simulation algorithm. Here we shall exam two popular sorting methds: Quick sort and Bubble sort. First, the following lemma reports the computation complexity of the two sorting algorithms and the naive Monto Carlo (without sorting).

LEMMA 1. *The complexity for naive Monte Carlo, Bubble sort Monte Carlo, and Quick sort Monte Carlo are* $O(N^2)$, $O(N^2)$, $O(N \log N)$, *respectively.*

Proof. See E-Companion EC.1. □

Figure 3 Execution Time for VaR and ES Estimations.



Note. Left one is for VaR estimation, right one is for ES estimation. Loss distributions here are with density function $f(x) = e^x$, $for\, x < 0$ and $f(x) = (1 - \frac{x}{3})^{-4}$, $for\, x < 0$ respectively. p is chosen to be 0.05. Average execution time are computed with 100 repeats.

Secondly, we report in Figure 3 the execution time for different simulation algorithms, which suggests that the Quick sort algorithm performs significantly faster than other two algorithms, consistent with the theoretical complexity results in Lemma 1. This numerical result is also consistent with the previous numerical comparions about sorting algorithms, e.g. in Puschner (1999), Knuth (1973) and Sedgewick (1989).

REMARKS: 1. There are other sorting methods in addition to the two sorting methods studied here. For example, Yaroslavskiy (2009) proposes a dual-pivot partitioning algorithm which outperforms the classic quicksort marginally under Java, and other multi-pivot quicksort methods are proposed and analyzed in Kushagra et al. (2014). However, dual-pivot and multi-pivot quicksort only outperform the classic one by a margin (Kushagra et al. 2014). Furthermore, the classic quicksort is easily implemented in the Matlab package.

2. The convenience of using sorting algorithm may disappear if we choose more complicated estimators of VaR and ES. For example, if we use importance sampling or nonparametric method, sorting techniques may not be applied directly and hence computation complexity increases.

4. Numerical Results

In this section we shall conduct an extensive numerical study, comparing the proposed algorithm with other algorithms by using the same computer and by re-programming the existing algorithms

Table 4 Time Series Factor Model (Type A)

Method	Estimate (SD)	Sample size	$T_{gen}+T_{est}$	T_{eva}	$T(=T_{gen}+$ $+T_{est}+T_{eva})$	Estimated RE (SD)	Resampled RE
Panel A: Portfolio with 10000 equally weighted stocks							
				VaR			
				$p=0.01$			
This paper (with estimated RE)	-23.599(0.1841)	263467	41.392	1.1277	**42.519**	0.0074(0.0013)	0.0078
This paper (no estimated RE)	-23.510(0.2398)	110000	33.227	3289.5	3322.7	N.A.	0.0102
Heidelberger & Lewis(1984)	-23.409(0.2271)	1750	16.221	1605.9	1622.1	N.A.	0.0097
				$p=0.001$			
This paper (with estimated RE)	-42.053(0.3448)	1435849	260.57	7.0381	**267.61**	0.0080(0.0016)	0.0082
This paper (no estimated RE)	-42.545(0.4552)	480000	144.67	14323	14467	N.A.	0.0107
Heidelberger & Lewis(1984)	-42.178(0.4471)	1600	168.79	16711	16879	N.A.	0.0106
				ES			
				$p=0.01$			
This paper (with estimated RE)	-30.986(0.2262)	722196	136.23	3.0413	**139.28**	0.0071(0.0020)	0.0073
This paper (no estimated RE)	-31.583(0.3411)	280000	89.783	8888.5	8978.3	N.A.	0.0108
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
				$p=0.001$			
This paper (with estimated RE)	-54.671(0.4538)	5603487	1112.7	22.618	**1135.3**	0.0067(0.0021)	0.0083
This paper (no estimated RE)	-54.517(0.6760)	2600000	697.97	69099	69797	N.A.	0.0124
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Panel B: Portfolio with 20000 equally weighted stocks							
				VaR			
				$p=0.01$			
This paper (with estimated RE)	-28.235(0.2231)	285385	78.295	1.2862	**79.581**	0.0074(0.0011)	0.0079
This paper (no estimated RE)	-28.448(0.2674)	120000	75.105	7435.3	7510.5	N.A.	0.0094
Heidelberger & Lewis(1984)	-28.336(0.2946)	1800	30.137	2983.6	3013.7	N.A.	0.0104
				$p=0.001$			
This paper (with estimated RE)	-47.344(0.3692)	1524346	482.36	8.2804	**490.64**	0.0073(0.0014)	0.0078
This paper (no estimated RE)	-47.131(0.4995)	500000	327.77	32449	32777	N.A.	0.0106
Heidelberger & Lewis(1984)	-47.020(0.4843)	1700	326.58	32331	32658	N.A.	0.0103
				ES			
				$p=0.01$			
This paper (with estimated RE)	-36.034(0.2414)	867756	248.09	3.7263	**251.81**	0.0063(0.0019)	0.0067
This paper (no estimated RE)	-36.313(0.4103)	290000	197.89	19591	19789	N.A.	0.0113
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
				$p=0.001$			
This paper (with estimated RE)	-59.597(0.4171)	6282067	2111.3	25.022	**2136.3**	0.0067(0.0018)	0.0070
This paper (no estimated RE)	-59.462(0.7670)	2700000	1415.9	140176	141592	N.A.	0.0129
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Note. VaR and ES of portfolio return $(1/n\sum_{i=1}^{n} r_{i,t})$ are computed. Portfolio includes n equally weighted stocks. Each stock return $r_{i,t}, i = 1,\cdots,n$ (normalized by multiply 100) is modeled by three factors $f_{1,t}, f_{2,t}, f_{3,t}$, i.e. $r_{i,t} = \alpha_i + \sum_{k=1}^{3} \beta_k^{(i)} f_{k,t} + \epsilon_{i,t}$, $\epsilon_{i,t} \sim N(0, 1/6)$. Three factors satisfy the AR(2) model, $f_{1,t} = 0.9 f_{1,t-1} - 0.2 f_{1,t-2} + \epsilon_t^{(1)}, \epsilon_t^{(1)} \sim N(0,1)$, ARCH(1) model, $f_{2,t} = 0.5 f_{2,t-1} + a_t, a_t = \sqrt{4 + 0.5 a_{t-1}^2}\epsilon_t^{(2)}, \epsilon_t^{(2)} \sim N(0,1)$, and SV (Stochastic Volatility) Model, $f_{3,t} = V_t \epsilon_t^{(3)}, \log V_t = 0.6 \log V_{t-1} + \eta_t, \epsilon_t^{(3)} \sim N(0,1), \eta_t \sim N(0,1), cov(\epsilon_t^{(3)}, \eta_t) = 0.5, cov(\epsilon_t, \eta_{t-j}) = 0$ for $j > 0$, respectively (see Chen and Tang, 2005). In Panel A, a portfolio with $n = 10000$ stocks is considered with $\alpha_i = -3 + 6i/10000$, $\beta_1^{(i)} = -4 + 8i/10000$, $\beta_2^{(i)} = 8i/10000$ and $\beta_3^{(i)} = -1 + 4i/10000$. In Panel B, another portfolio with $n = 20000$ stocks is considered with $\alpha_i = -8 + 6i/20000$, $\beta_1^{(i)} = -4 + 8i/20000$, $\beta_2^{(i)} = 8i/20000$ and $\beta_3^{(i)} = -1 + 4i/20000$. ES is only computed with our framework, because ES has not been considered in Heidelberger and Lewis (1984). The mean and SD of Estimated RE is based on 100 repeats, and resampled RE is computed with 100 replications. Each reported CPU time is in seconds.

in Heidelberger and Lewis (1984), Sun and Hong (2010) Glasserman et al. (2000), Glasserman et al. (2002) and Fuh et al. (2011) .

We shall conduct five numerical experiments to test the performances of our framework and other existing methods. The first two experiments consider Type A samples, and the rest three consider Type B samples. We compare simulation methods by simulation cost and accuracy, and

report the total execution time, i.e. $T_{gen}+T_{est}+T_{eva}$, where T_{gen} is the time for generating samples, T_{est} the time for estimating risk measures, and T_{eva} the time for evaluating the relative errors. More precisely, for each experiment we record the sample size, $T_{gen}+T_{est}$, T_{eva}, and $T=T_{gen}+T_{est}+T_{eva}$ with controlled RE to be around 1%.

Table 4 shows two experiments of portfolios with equally weighted stocks under factor models. Panel A considers a 10000-stock portfolio, while Panel B considers a 20000-stock portfolio. Risk measures of the portfolio returns and execution time for the maximum transformation method in Heidelberger and Lewis (1984) and for our algorithm are reported. Although our proposed algorithm has slightly larger $T_{gen}+T_{est}$ due to its R-step, the time in the evaluating part T_{eva} is much smaller thanks to our expansion in Theorem 1, resulting in much faster total execution time T. Hence, our framework has quite less T, comparing with the maximum transformation.

Table 5 reports numerical results of a portfolio which consists of call and put options. This portfolio is similar to one of the portfolios considered in Glasserman et al. (2002), except that all its underlying stocks satisfy Johnson NGARCH(1,1) model. We apply our method and the maximum transformation to simulate risk measures of portfolio value. Note the difficulty associated with Johnson time series innovation, if one attempts to do importance sampling. Similar to table 4, our framework outperforms the maximum transformation method as well.

Table 5 Value of a Portfolio of Call and Put Options with Non-stationary Time Series Returns (Type B)

Method	Estimate (SD)	Sample size	$T_{gen}+T_{est}$	T_{eva}	$T(=T_{gen}+$ $+T_{est}+T_{eva})$	Estimated RE (SD)	Resampled RE
	VaR						
	$p=0.01$						
This paper (with estimated RE)	-81.479(0.6599)	27643	5.5602	0.0094	**5.5696**	0.0082(0.0004)	0.0081
This paper (no estimated RE)	-81.881(0.8597)	15500	4.2679	422.52	426.79	N.A.	0.0105
Heidelberger & Lewis(1984)	-81.731(0.8664)	400	5.3006	524.76	530.06	N.A.	0.0106
	$p=0.001$						
This paper (with estimated RE)	-103.83(0.8618)	56479	13.627	0.0242	**13.651**	0.0084(0.0010)	0.0083
This paper (no estimated RE)	-104.06(1.0614)	36000	10.599	1049.3	1059.9	N.A.	0.0102
Heidelberger & Lewis(1984)	-104.19(1.0836)	90	12.177	1205.5	1217.7	N.A.	0.0104
	ES						
	$p=0.01$						
This paper (with estimated RE)	-90.946(0.7912)	30080	6.4069	0.0007	**6.4076**	0.0083(0.0005)	0.0087
This paper (no estimated RE)	-92.075(0.9575)	16000	4.5324	448.71	453.24	N.A.	0.0104
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	$p=0.001$						
This paper (with estimated RE)	-110.75(0.9414)	95170	21.254	0.0061	**21.261**	0.0080(0.0010)	0.0085
This paper (no estimated RE)	-110.66(1.0734)	52000	14.567	1442.1	1456.7	N.A.	0.0097
Heidelberger & Lewis(1984)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Note. VaR and ES of portfolio value $(V(t)-V(0))$ are computed. Portfolio includes shorting 10 ATM calls and 5 ATM puts on each of 10 uncorrelated stocks, while all options having a half-year maturity. We investigate losses over 10 days ($t=10$ days). All stock have an initial value of 100, and they satisfy Johnson NGARCH(1,1) process (model is estimated based on daily data, see Simonato and Sentoft (2015)): $\ln\frac{S_t}{S_{t-1}}=\alpha+\sigma_t\epsilon_t$, $\sigma_t^2=\beta_0+\beta_1\sigma_{t-1}^2+\beta_2\sigma_{t-1}^2(\epsilon-\theta)^2$, $\epsilon_t\sim J_{su}(a,b)$, where $\alpha=3.3\times10^{-4},\beta_0=1.1\times10^{-6},\beta_1=0.8664,\beta_2=0.0631,\theta=0.9937,a=0.3478,b=2.1610$. The mean and SD of Estimated RE is based on 100 repeats, and resampled RE is computed with 100 replications. Each reported CPU time is in seconds.

Table 6 Value of a portfolio of Call and Put Options with I.I.D. Gaussian Returns (Type B)

Method	Estimate (SD)	Sample size	$T_{gen}+T_{est}$	T_{eva}	$T(=T_{gen}+$ $+T_{est}+T_{eva})$	Estimated RE (SD)	Resampled RE
				VaR			
				$p=0.01$			
This paper (with estimated RE)	-160.33(1.3307)	23086	0.0731	0.0079	**0.0810**	0.0077(0.0003)	0.0083
This paper (no estimated RE)	-159.76(1.5656)	13000	0.0419	4.1487	4.1907	N.A.	0.0098
Sun and Hong (2010)	-159.38(1.6575)	300	0.0058	0.5743	0.5801	N.A.	0.0104
Glasserman et.al (2000)IIS	-158.93(1.5592)	135	0.1119	11.077	11.189	N.A.	0.0098
Glasserman et.al (2000)IIS-Q	-159.17(1.6596)	40	0.0995	9.8495	9.9490	N.A.	0.0108
				$p=0.001$			
This paper (with estimated RE)	-197.86(1.7687)	56316	0.1819	0.0211	**0.2030**	0.0083(0.0010)	0.0088
This paper (no estimated RE)	-197.46(2.0931)	34000	0.1105	10.947	11.058	N.A.	0.0106
Sun and Hong (2010)	-198.05(1.9104)	160	0.0031	0.3021	0.3052	N.A.	0.0097
Glasserman et.al (2000)IIS	-197.39(1.9937)	38	0.0439	4.3477	4.3916	N.A.	0.0101
Glasserman et.al (2000)IIS-Q	-196.96(2.0681)	12	0.0532	5.2666	5.3198	N.A.	0.0105
				ES			
				$p=0.01$			
This paper (with estimated RE)	-175.91(1.4248)	26551	0.0862	0.0006	**0.0868**	0.0078(0.0004)	0.0081
This paper (no estimated RE)	-176.92(1.7339)	15000	0.0483	4.7910	4.8394	N.A.	0.0098
Sun and Hong (2010)	-176.79(1.8033)	100	0.0018	0.1801	0.1819	N.A.	0.0102
Glasserman et.al (2000)IIS	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Glasserman et.al (2000)IIS-Q	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
				$p=0.001$			
This paper (with estimated RE)	-210.09(1.7437)	91474	0.2647	0.0083	**0.2730**	0.0081(0.0010)	0.0083
This paper (no estimated RE)	-209.87(2.1826)	57000	0.1864	18.456	18.643	N.A.	0.0104
Sun and Hong (2010)	-209.82(2.1401)	50	0.0011	0.1067	0.1078	N.A.	0.0102
Glasserman et.al (2000)IIS	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Glasserman et.al (2000)IIS-Q	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Note. VaR and ES of portfolio value $(V(t)-V(0))$ are computed. Portfolio is taken from Glasserman et al. (2000), consisting of shorting 10 ATM calls and 5 ATM puts on each of 10 uncorrelated stocks. All options have a half-year maturity, and all stocks have an initial value of 100 and an annual volatility of 0.3. Risk-free rate is 5%. We consider losses over 10 days ($t=0.04$ in year). For direct importance sampling in Sun and Hong (2010), we use the change of measure techniques from Glasserman et al. (2000). The mean and SD of Estimated RE is based on 100 repeats, and resampled RE is computed with 100 replications. Each reported CPU time is in seconds.

Table 6 and 7 report numerical results with i.i.d. samples taken from portfolios in the existing literature, with light and heavy tails, respectively. Note that the standard risk management guidance (e.g. Basel Accords) in financial industry requires computing risk measures with moderate quantile, i.e. between 0.01 and 0.001, and our propose method can well satisfy that requirement.

These numerical tables indicate that the new algorithm is easy to implement, fast and accurate, even at the 0.001 quantile level. There are two reasons for the good performance of our algorithm. First, when we only consider the execution time of S-step $T_{gen}+T_{est}$, it is usually small even when the sample size is extremely large. This is largely attributed to the quick sort algorithm. Secondly, our relative error expansions reduce the time of evaluating T_{eva} significantly.

5. Application 1: Intra-Horizon Risk

Our simulation algorithm can serve as an alternative to compute VaR with intra-horizon risk (VaR-I), which was introduced by Bakshi and Panayotov (2010) to incoporate path-dependent risk instead of just the risk at the end of a time horizon. More precisely, for a given stochastic process

Table 7 Value of a Portfolio of Call and Put Options with I.I.D. Heavy-tail Returns (Type B)

Method	Estimate (SD)	Sample size	$T_{gen}+T_{est}$	T_{eva}	$T(=T_{gen}+$ $+T_{est}+T_{eva})$	Estimated RE (SD)	Resampled RE
	VaR						
	$p=0.01$						
This paper (with estimated RE)	-59.089(0.4964)	312680	0.9519	0.2176	**1.1695**	0.0080(0.0010)	0.0084
This paper (no estimated RE)	-59.555(0.6015)	170000	0.6029	59.689	60.292	N.A.	0.0101
Glasserman et. al (2002)	-59.319(0.5932)	1000	1.3922	137.83	139.22	N.A.	0.0100
Fuh et. al (2011)	-59.191(0.6452)	850	13.481	1334.6	1348.1	N.A.	0.0109
	$p=0.001$						
This paper (with estimated RE)	-159.10(1.3046)	2266233	7.5765	1.1339	**8.7104**	0.0082(0.0011)	0.0084
This paper (no estimated RE)	-158.69(1.6980)	1400000	5.7016	564.46	570.16	N.A.	0.0107
Glasserman et. al (2002)	-158.73(1.5931)	930	1.2052	119.31	120.52	N.A.	0.0101
Fuh et. al (2011)	-159.43(1.6561)	800	9.7891	969.12	978.91	N.A.	0.0106
	$p=0.0005$						
This paper (with estimated RE)	-212.86(1.7029)	6251108	20.720	3.3899	**24.110**	0.0072(0.0016)	0.0080
This paper (no estimated RE)	-212.11(1.9514)	3800000	19.099	1890.8	1909.9	N.A.	0.0092
Glasserman et. al (2002)	-211.54(2.2423)	850	1.1334	112.21	113.34	N.A.	0.0100
Fuh et. al (2011)	-212.13(1.9516)	780	9.1624	907.08	916.24	N.A.	0.0092
	ES						
	$p=0.01$						
This paper (with estimated RE)	-103.35(0.8061)	1695129	4.9997	0.0718	**5.0715**	0.0078(0.0011)	0.0078
This paper (no estimated RE)	-103.55(1.0873)	910000	4.0777	403.69	407.77	N.A.	0.0105
Glasserman et. al (2002)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Fu et. al (2011)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
	$p=0.001$						
This paper (with estimated RE)	-268.68(2.2569)	12522977	34.615	0.3624	**34.977**	0.0082(0.0011)	0.0084
This paper (no estimated RE)	-269.38(2.8555)	6500000	32.216	3189.4	3221.6	N.A.	0.0106
Glasserman et. al (2002)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Fuh et. al (2011)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Note. VaR and ES of portfolio value $(V(t)-V(0))$ are computed. The portfolio is taken from Fuh et al. (2011), consisting of options with heavy-tailed underlying stock returns. More precisely, $L=(V(t)-V(0))=\sum_{i=1}^{m}(b_iX_i+\lambda_iX_i^2)$ with t-distributed risk factors $\{X_i\}_{i=1}^m$, where $X_i=\frac{Z_i}{\sqrt{Y/\nu}}$, $Z_i\sim N(0,1)$ and independent with $Y\sim\chi^2_\nu$. Parameters satisfy $\nu=5, b_j=0.1+j/100, \lambda_j=0.05\times j, j=1,\cdots,15$. The mean and SD of Estimated RE is based on 100 repeats, and resampled RE is computed with 100 replications. Each reported CPU time is in seconds.

$X_t, t\in[0,T], X_0=0$, with $X_T^{min}:=\min_{0<t<T}\{X_t\}$, the VaR-I is defined as the value of a quantile of the distribution for X_T^{min}; in particular, we consider the 10-day 1% VaR-I which satisfies

$$Prob(X_T^{min}\le -\text{VaR-I})=1\%, T=10 \text{ days}.$$

In some rare cases, e.g. the double exponential jump diffusion models in Kou (2002) and Kou and Wang (2003), an analytical solution for the VaR-I is available in terms of Laplace transforms. In general, previously there are two ways to compute the intra-horizon risk numerically, the partial integro-differential equation (PIDE) approach in Bakshi and Panayotov (2010) for Lévy jump models, and an approximation based on a displaced mixed-exponential model in Leippold and Vasiljevic (2016). Our simulation algorithm complement the existing numerical methods by offering a simple approach for general stochastic processes, with the total CPU time comparable (sometimes even shorter than) to the existing numerical methods, thanks to the relative error expansion and the quick sort. For example, it takes almost the same computational and programing effort for our simulation algorithm to compute VaR and VaR-I, while other methods may take significant longer time and may need more programming effort to compute VaR-I than VaR.

Here we consider three models (see Appendix C for details): Merton's jump diffusion model (MJD), double exponential model (DEM), and exponential dampened power law model (CGMY). MJD and DEM are finite activity Lévy processes, while the CGMY model is infinite activity Lévy processes. We include DEM, because the benchmark values are available from the analytical solutions in Kou and Wang (2003), and interestingly the largest VaR-I based on teh S&P 500 data from 1999 to 2016 was from the DEM.

We estimate the intra-horizon risk for S&P 500 index from Jan 1, 1999 to Nov 31, 2016. To estimate 10-day VaR-I, we follow the literature by using weekly data to estimate the parameters for the 3 models. Since the traditional MLE method may have multiple local maxima and may have difficulties of distinguishing high frequency small jumps from diffusions (Honore 1998), we consider both the maximum likelihood estimation (MLE) method and the threshold method in Dang and Forsyth (2015) and Forsyth and Vetzal (2016).

Table 8 reports the estimated parameters for three models with different estimation methods. For MJD and DEM models, the jump sizes from MLE method are quite larger than those from threshold method, indicating that MLE tends to over estimate the jumps due to it's drawback to distinguish small jumps from diffusions. See Dang and Forsyth (2015) and Forsyth and Vetzal (2016) for more discussion along this line. Since threshold method cannot be applied to infinite activity Lévy processes CGMY, we only report parameters from MLE method for this model.

Table 8 S&P 500: Model parameters estimates

Merton's jump-diffusion model (MJD)						
Method	λ	μ_J	σ_J	σ	Likelihood	
Threshold: $\alpha^{dwn}=-2.5, \alpha^{up}=2.5$	4.1756	-0.0124	0.0654	0.1257	N.A.	
Threshold: $\alpha^{dwn}=-3, \alpha^{up}=3$	1.4475	-0.0286	0.0865	0.1468	N.A.	
Threshold: $\alpha^{dwn}=-4, \alpha^{up}=4$	0.3897	-0.0319	0.1316	0.1631	N.A.	
MLE	22.3019	-0.0055	0.0296	0.1064	2196.1	
Double exponential model (DEM)						
Method	σ	λ	p	η_1	η_2	Likelihood
Threshold: $\alpha=2.5$	0.1257	4.1756	0.4133	0.0595	0.0630	N.A.
Threshold: $\alpha=3$	0.1468	1.4475	0.3462	0.0811	0.0866	N.A.
Threshold: $\alpha=4$	0.1631	0.3897	0.4286	0.1044	0.1341	N.A.
MLE	0.0953	46.5775	0.4139	0.0145	0.0164	2198.2
Exponential dampened power law model (CGMY)						
Method	C	G	M	Y	Likelihood	
MLE	5.3436	38.6766	54.0719	0.5	2198.6	

Note. The weekly data of S&P 500 in period Jan 1, 1999-Nov 31, 2016 is used to estimate 3 models: MJD, DEM, and CGMY. First, the return time-series is demeaned. Then, both the MLE method and the threshold method are implemented to estimate parameters according to Dang and Forsyth (2015) and Forsyth and Vetzal (2016) for the finite activity models, MJD and DEM. For infinite activity model CGMY, only the MLE is applicable. The threshold values $\alpha^{up}(\alpha^{dwn})$ for the MJD are used to indicate the occurrence of upward (downward) jumps, if the ratio of positive (negative) return to its volatility in this period is greater (smaller) than the threshold value $\alpha^{up}(-\alpha^{dwn})$. In DEM model, we use the same threshold value α for both upward and downward jumps. The chosen of these threshold values is based on Dang and Forsyth (2015) and Forsyth and Vetzal (2016).

After obtaining the estimated parameters, we report in Table 9 the VaR-I and VaR multiples at both $p = 0.01$ and $p = 0.001$ levels for different models. There are several interesting empirical observations. First, the VaR-I multiples are all greater than 1 for all models, indicating that the intro-horizon risk VaR-I is larger than the end of horizon risk VaR. Secondly, VaR-I multiples based on threshold method varies from 1.1056 to 1.6097 (1.8988 to 2.2947) for MJD model and from 1.0825 to 1.9397 (2.8385 to 3.0700) for DEM model at $p = 0.01(p = 0.001)$ level, indicating significant uncertainty in estimating risk. Thirdly, the largest VaR-I and VaR multiples seems to from DEM models, which might come from the fact that the DEM has the heaviest tail part among the 3 models.

Table 9 Multiples of Intra-horizon value-at-risk (VaR-I) and end-of-horizon value-at-risk (VaR)

	$p = 0.01$		$p = 0.001$	
Model(Method)	VaR-I	VaR	VaR-I	VaR
MJD (Threshold: $\alpha^{dwn} = -2.5, \alpha^{up} = 2.5$)	1.6097	1.4831	1.8988	1.7809
MJD (Threshold: $\alpha^{dwn} = -3, \alpha^{up} = 3$)	1.5342	1.3528	2.1630	2.0405
MJD (Threshold: $\alpha^{dwn} = -4, \alpha^{up} = 4$)	1.1056	1.0216	2.2947	2.1576
MJD (MLE)	1.2615	1.1750	1.3640	1.3055
MJD (MLE, Bakshi & Panayotov (2010), 1995-2005)	1.57	1.24	N.A.	N.A.
DEM (Threshold: $\alpha = 2.5$)	1.9397	1.8106	2.8350	2.7271
DEM (Threshold: $\alpha = 3$)	1.5854	1.3930	3.0700	2.9429
DEM (Threshold: $\alpha = 4$)	1.0825	1.0033	2.8460	2.6677
DEM (MLE)	1.3053	1.1848	1.4481	1.3613
DEM (MLE, Bakshi & Panayotov (2010), 1995-2005)	N.A.	N.A.	N.A.	N.A.
CGMY (MLE)	1.2590	1.1775	1.4225	1.3348
CGMY (MLE, Bakshi & Panayotov (2010), 1995-2005)	1.41	1.28	N.A.	N.A.

Note. The table reports multiples of the intra-horizon value-at-risk(VaR-I) and end-of-horizon value-at-risk(VaR) for S&P 500 index in period Jan 1, 1999-Nov 31, 2016 at $p = 0.01$ and $p = 0.001$ levels with a 2-week (10 days) horizon. The multiples are computed by $|\text{VaR-I}/(-2.32\hat{\sigma} + \hat{\mu})|$ and $|\text{VaR}/(-2.32\hat{\sigma} + \hat{\mu})|$ for $p = 0.01$ level and $|\text{VaR-I}/(-3.09\hat{\sigma} + \hat{\mu})|$ and $|\text{VaR}/(-3.09\hat{\sigma} + \hat{\mu})|$ for $p = 0.001$ level. As in Bakshi and Panayotov (2010), the denominators in these formulas are benchmark VaR, which are quantiles of Normal distribution $N(0, \hat{\sigma})$, where $\hat{\sigma}$ is the standard deviation of the return time series.

To compare the computational efficiency, we report the CPU time in table 10 for 3 methods, our simulation algorithm, the maximum transformation method, and the PIDE method, by re-programming all the method in the same computer. To give a benchmark comparison, we also report the results from the analytical inversion from Kou and Wang (2003). It is showed that when compute intra-horizon risk for MJD and CGMY models, our simulation method is faster than its two competitors. Our algorithm behaves well in this comparison. Note it takes almost the same computational and programing effort for our simulation algorithm to compute VaR and VaR-I, while the PIDE method may take significant longer time and may need more programming effort to compute VaR-I than VaR.

Table 10 Comparison of efficiency for intra-horizon risk estimation

Method	Risk Measure	Estimate (SD)	Sample size	$T_{gen}+T_{est}$	T_{eva}	$T(=T_{gen}+$ $+T_{est}+T_{eva})$	Estimated RE (SD)	Resampled RE
	Panel A: Merton's jump-diffusion model (MJD)							
				$p=0.01$				
This paper	VaR-I	1.2615(0.0119)	48231	66317	0.0183	**66.336**	0.0087(0.0009)	0.0095
	VaR	1.1750(0.0121)	51574	68.201	0.0247	68.226	0.0092(0.0010)	0.0103
Heidelberger & Lewis (1984)	VaR-I	1.2797(0.0143)	350	21.920	2170.1	2192.0	N.A.	0.0112
	VaR	1.1653(0.0133)	400	25.320	2506.7	2532.0	N.A.	0.0114
Bakshi & Panayotov (2010)	VaR-I	1.2724(N.A.)	N.A.	N.A.	N.A.	222.39	N.A.	N.A.
	VaR	1.1762(N.A.)	N.A.	N.A.	N.A.	14.208	N.A.	N.A.
				$p=0.001$				
This paper	VaR-I	1.3640(0.0126)	146790	162.51	0.0749	**162.59**	0.0086(0.0011)	0.0093
	VaR	1.3055(0.0133)	192570	261.85	0.1124	261.97	0.0090(0.0012)	0.0102
Heidelberger & Lewis (1984)	VaR-I	1.3722(0.0159)	120	79.663	7886.6	7966.3	N.A.	0.0116
	VaR	1.2991(0.0161)	140	98.823	9783.4	9882.3	N.A.	0.0124
Bakshi & Panayotov (2010)	VaR-I	1.3695(N.A.)	N.A.	N.A.	N.A.	218.55	N.A.	N.A.
	VaR	1.3101(N.A.)	N.A.	N.A.	N.A.	12.684	N.A.	N.A.
	Panel B: Double exponential model (DEM)							
				$p=0.01$				
This paper	VaR-I	1.3053(0.0116)	49605	60.658	0.0234	**60.682**	0.0080(0.0011)	0.0089
	VaR	1.1847(0.0115)	60332	98.019	0.0355	98.054	0.0091(0.0008)	0.0097
Heidelberger & Lewis (1984)	VaR-I	1.2967(0.0146)	420	32.782	3245.4	3278.3	N.A.	0.0113
	VaR	1.1859(0.0135)	500	36.325	3596.2	3632.5	N.A.	0.0114
Kou & Wang (2003)	VaR-I	1.3078(N.A.)	N.A.	N.A.	N.A.	2.0742	N.A.	N.A.
	VaR	1.1982(N.A.)	N.A.	N.A.	N.A.	26.032	N.A.	N.A.
				$p=0.001$				
This paper	VaR-I	1.4481(0.0131)	365081	410.31	0.1731	**410.48**	0.0084(0.0009)	0.0091
	VaR	1.3612(0.0133)	384120	477.11	0.1823	477.29	0.0087(0.0011)	0.0098
Heidelberger & Lewis (1984)	VaR-I	1.4417(0.0168)	200	156.41	15485	15641	N.A.	0.0117
	VaR	1.3667(0.0165)	250	203.73	20170	20373	N.A.	0.0121
Kou & Wang (2003)	VaR-I	1.4463(N.A.)	N.A.	N.A.	N.A.	2.7981	N.A.	N.A.
	VaR	1.3704(N.A.)	N.A.	N.A.	N.A.	21.371	N.A.	N.A.
	Panel C: Exponentially dampened power law model (CGMY)							
				$p=0.01$				
This paper	VaR-I	1.2590(0.0128)	45386	52.993	0.0156	**53.009**	0.0096(0.0012)	0.0102
	VaR	1.1775(0.0113)	71807	92.139	0.0394	92.178	0.0086(0.0014)	0.0096
Heidelberger & Lewis (1984)	VaR-I	1.2469(0.0137)	410	42.481	4205.6	4248.1	N.A.	0.0110
	VaR	1.1629(0.0221)	430	45.196	4474.5	4519.7	N.A.	0.0119
Bakshi & Panayotov (2010)	VaR-I	1.3224(N.A.)	N.A.	N.A.	N.A.	639.43	N.A.	N.A.
	VaR	1.1945(N.A.)	N.A.	N.A.	N.A.	17.272	N.A.	N.A.
				$p=0.001$				
This paper	VaR-I	1.4225(0.0147)	191260	230.16	0.1142	**230.27**	0.0094(0.0015)	0.0104
	VaR	1.3649(0.0117)	318193	445.98	0.1561	446.14	0.0075(0.0017)	0.0097
Heidelberger & Lewis (1984)	VaR-I	1.4408(0.0177)	130	128.41	12712	12840	N.A.	0.0123
	VaR	1.3658(0.0156)	200	185.06	18321	18506	N.A.	0.0114
Bakshi & Panayotov (2010)	VaR-I	1.4792(N.A.)	N.A.	N.A.	N.A.	859.95	N.A.	N.A.
	VaR	1.3905(N.A.)	N.A.	N.A.	N.A.	15.217	N.A.	N.A.

Note. A comparion of our simulation method with other methods in terms of computing time by controlling the relative error to be around 1%. Each simulation path has 4000 points. The PIDE methods for two models (MJD and CGMY) use 400×400 grids. Parameters for all underlying asset models are taken from the MLE method. The mean and SD of Estimated RE is based on 100 repeats, and resampled RE is computed with 100 replications. Each reported CPU time is in seconds.

6. Application 2: Comparison of Relative Errors for VaR, Median Shortfall, and Expected Shortfall

As risk measures, VaR and ES have their own advantages and disadvantages. Which one to choose depends on one's own internal and external objectives. The comparisons between VaR and ES have been done in terms of sub-additivity (Artzner et al. 1999, Rockafellar 2002), robustness (Cont et

al. 2010, Heyde, Kou and Peng 2013) and elicitablity (Kou and Peng 2014). With the expansions for the relative errors, we can offer a new theoretical comparison of VaR and ES in terms of relative errors for both dependent and i.i.d samples. Note that the larger relative errors, more difficult to simulate. Previously, Yamai and Yoshiba (2005) and Dowd and Cotter (2007) did numerical comparisons of relative errors for i.i.d samples.

THEOREM 2. **(VaR and ES)**

(1) (Dependent samples) Under the conditions of Assumptions A, B, C, we have

$$\lim_{n\to\infty}\frac{RE(v_n(p))}{RE(c_n(p))}=\frac{cp\sigma_{\infty,v}}{vf(v)\sigma_{\infty,c}}<\infty$$

where $\sigma^2_{\infty,v}=\{p(1-p)+2\sum_{k=1}^{\infty}\gamma_1(k)\}$, $\sigma^2_{\infty,c}=\{var[(v-L_1)^+]+2\sum_{k=1}^{\infty}\gamma_2(k)\}$.

(2) (I.I.D. samples) Under the conditions of Assumptions A1, B1, C, and assuming the existence of $\lim_{x\to-\infty}xh^{'}(x)$ *and* $\lim_{x\to-\infty}h(x)$*, where* $h(x)=\frac{F(x)}{xf(x)}$*, we have*

$$\lim_{p\to 0}\lim_{n\to\infty}\frac{RE(v_n(p))}{RE(c_n(p))}\le\frac{1}{\sqrt{2}}.$$

Proof. See E-Companion EC.3. □

Many distributions satisfy the regular conditions in Theorem 2 (2), i.e. Pareto distribution, exponential distribution, Normal distribution, Cauchy distribution, Weibull distribution, etc. In part (1) of Theorem 2, the limit of ratio of REs of VaR and ES is finite, implying that the convergence rate of $RE(c_n(p))$ is perhaps comparable to that of $RE(v_n(p))$. However, our numerical study of Figure 4 suggests that in some dependent cases, it is easier to simulate VaR than ES. The part (2) of Theorem 2 indicates that for i.i.d. losses, it is easier to simulate VaR than ES, which is also confirmed in our numerical study of Figure 4.

As in Kou and Peng (2014), median shortfall (MS) is proposed as an alternative risk measure to expected shortfall (ES) due to its elicitability and robustness. MS is defined as the median loss conditioning on the loss beyond certain VaR level. According to Kou and Peng (2014), MS at level p (i.e. $m(p)$) is exactly the VaR at level $\frac{p}{2}$ (i.e. $v(\frac{p}{2})$), i.e. $m(p)=v(\frac{p}{2})$. Naturally, median shortfall $m(p)$ can be estimated by estimator $v_n(\frac{p}{2})$. Under our framework, we can also compare the estimators of MS and ES with respect to their relative errors.

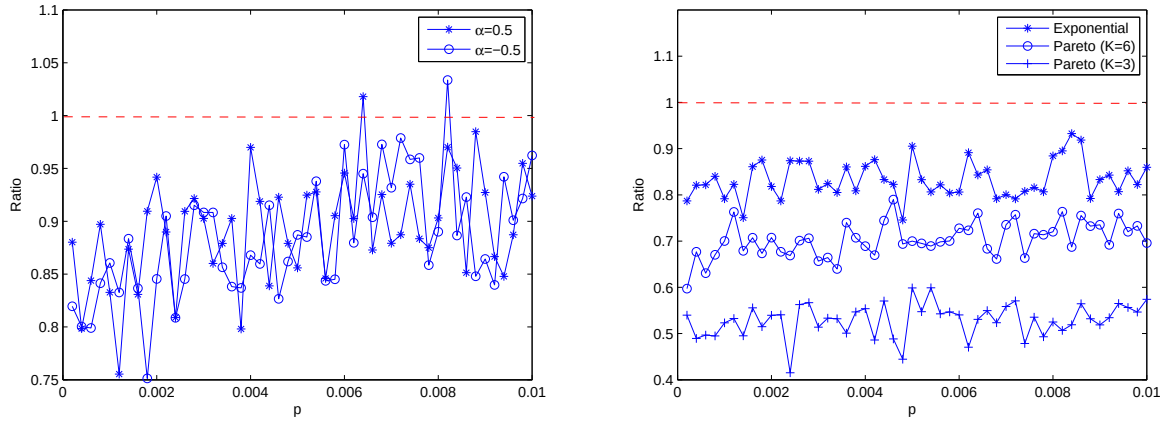
THEOREM 3. **(MS and ES)**

(1) (Dependent samples) Under the conditions of Assumptions A, B, C:

$$\lim_{n\to\infty}\frac{RE(m_n(p))}{RE(c_n(p))}=\frac{c(p)p\sigma_{\infty,m}}{m(p)f(m(p))\sigma_{\infty,c}}<\infty$$

where $\sigma^2_{\infty,m}=\{1/2p(1-p/2)+2\sum_{k=1}^{\infty}\gamma_3(k)\}$, $\gamma_3(k)=cov\{1_{\{L_1<v(p/2)\}},1_{\{L_{k+1}<v(p/2)\}}\}$, $\sigma^2_{\infty,c}=\{var[(v-L_1)^+]+2\sum_{k=1}^{\infty}\gamma_2(k)\}$.

Figure 4 Ratio of REs of VaR and ES as p varies.



Note. Left graph reports ratio for loss samples satisfy MA(1) Model: $L_{t+1} = 0.5\epsilon_t + \epsilon_{t+1}$, $\epsilon_t \sim N(0,1)$. Right graph reports the case of i.i.d. loss with p.d.f satisfies $f(x) = e^x 1_{\{x<0\}}$ and $f(x) = (1-\frac{x}{K})^{-K-1} 1_{\{x<0\}}$. Sample size is chosen to be 10^6, and p varies from 0.0002 to 0.01. Resampled REs are based on 200 times repetition. The ratio in most cases is less than 1, indicating that the relative error for VaR is smaller than that of ES for both models.

(2) (I.I.D. samples) Under the conditions of Assumptions A1, B1, C, denote $\frac{F(x)}{xf(x)} = h(x)$*, and assume the existence of* $\lim_{x\to-\infty} xh'(x)$ *and* $\lim_{x\to-\infty} h(x)$*. Moreover, we assume* $\exists\theta \le 1$ *such that* $\frac{v(p)f(v(p))}{p^\theta}$ *is a slowly varying function at* $p = 0$*. Then*

$$\lim_{p\to 0}\lim_{n\to\infty} \frac{RE(m_n(p))}{RE(c_n(p))} \le 1.$$

Proof. See E-Companion EC.3. □

Many commonly used distributions satisfy the conditions in Theorem 3 (2), including Pareto distribution, exponential distribution, Normal distribution, Cauchy distribution, Weibull distribution. Particularly, the slowly varying condition can also be satisfied by these distribution families. For instance, for exponential distribution, $\frac{v(p)f(v(p))}{p^\theta}$ is a log function for some θ and hence is a slowly varying function, and for Pareto distribution, $\frac{v(p)f(v(p))}{p^\theta}$ is constant for some θ and hence is also a slowly varying function.

7. Conclusion

This paper gives rigorous expansions of relative errors for both VaR and expected shortfall under the $\alpha-$mixing model, which appear to be new even for i.i.d. samples. With the necessary sample size being computed using the new asymptotic expansions, we propose a general framework to simulate these risk measures via a sorting algorithm. In our extensive numerical study, and by comparing with existing algorithms in Heidelberger and Lewis (1984), Sun and Hong (2010), Glasserman et al. (2000), Glasserman et al. (2002) and Fuh et al. (2011), we find that the new algorithm is fast and

accurate, even at the 0.001 quantile level. Two applications are also given, including the estimation of intra-horizon risk and a comparison of the relative errors of VaR and expected shortfall.

One should also be cautious about our numerical findings. In our numerical study the 0.001 quantile level is suitable for many applications, as the standard risk management guidance (e.g. Basel Accords) in financial industry requires computing risk measures with moderate quantile, i.e. between 0.01 and 0.001. However, if one wants to estimate VaR and expected shortfall at level below 0.001, then other simulation methods, such as the ones based on importance sampling may be more suitable, at least one should think of extending the method of importance sampling to the simulation of dependent samples.

Appendix A: Regular Conditions

Here we state some assumptions listed as regular conditions in our theorems and lemmas.

1. Assumption A: Sample $L_1, \cdots, L_n$ are strictly stationary and geometric $\alpha-$mixing. Each L_i is has a density f and a distribution function F.
2. Assumption A1: Samples $L_1, \cdots, L_n$ are i.i.d. Each L_i is has a density f and a distribution function F.
3. Assumption B: With $v = \inf\{x : F(x) \geq p\}$, the density f of F satisfies $vf(v) < 0$, and f is Hölder continuous with index $1/2 + \delta_0 (\delta_0 > 0)$ in a neighborhood of v, i.e. there exists a constant c_H such that
$$|f(y_1) - f(y_2)| \leq c_H |y_1 - y_2|^{1/2+\delta_0},$$
for all y_1 and y_2 in a neighborhood of v.
4. Assumption B1: With $v = \inf\{x : F(x) \geq p\}$, the density f of F satisties $xf(x) < 0$, for all $x \leq v$, and f is Hölder continuous with index $1/2 + \delta_0 (\delta_0 > 0)$ in a neighborhood of v, i.e. there exists a constant c_H such that
$$|f(y_1) - f(y_2)| \leq c_H |y_1 - y_2|^{1/2+\delta_0},$$
for all y_1 and y_2 in a neighborhood of v.
5. Assumption C: $E|L|^{2+\delta} < \infty$ for some $\delta > 0$.
6. Assumption C1: $E|L| < \infty$.

Many distributions satisfy Assumption B, i.e. Pareto distribution, exponential distribution, normal distribution, Cauchy distribution, Weibull distribution, etc. Assumption B1 is provided to ensure the existence of the limit for the ratio of REs. The distributions aforementioned also satisfy this assumption.

Appendix B: Two Counter Examples

Example 1. Let $Z_n = z + T_n + \frac{1}{\sqrt{n}}N(0,1)$, where z is a positive constant and $P(T_n = n) = 1/n, P(T_n = 0) = 1 - 1/n$. Suppose T_n and $N(0,1)$ are independent. Since $\sqrt{n}T_n \to 0$ in probability, we have $\sqrt{n}(Z_n - z) \Rightarrow N(0,1)$. Then the heuristic approximation suggests that the relative error of Z_n can be approximated by $1/(z\sqrt{n})$. However, since $var(Z_n) = n + 1/n$ and $EZ_n = z + 1$, we have $RE(Z_n) = \sqrt{var(Z_n)}/EZ_n = \sqrt{(n+1/n)}/(z+1) \to \infty$ as $n \to \infty$, which is different from the heuristic approximation of $1/(z\sqrt{n})$.

Example 2. (A) Consider $X_n(\omega) = \frac{1}{n}1_{\{\frac{1}{n^2} \le \omega \le 1\}} + (\frac{1}{n} + n^2)1_{\{0 \le \omega < \frac{1}{n^2}\}}$. Then $P(|n^{1-\epsilon}X_n| > \theta) = 1/n^2 \to 0$, and hence $X_n = o_p(n^{-1+\epsilon})$. We also have that $E(X_n) = 1 + O(\frac{1}{n})$. A heuristic approximation implies that relative error of X_n might be approximated by $o(n^{-1+\epsilon})$. However, since $EX_n^2 = O(n^2) \to \infty$ as $n \to \infty$, we have that $RE(X_n) = O(n^2) \to \infty$, which is different from the heuristic result based on convergence in probability. (B) Define random variables as $Y_n = 2^n 1_{U_n} + n^{-1}1_{V_n}$, where U_n, V_n are defined recursively: $U_1 = (0, 1/3) \cup (2/3, 1), V_1 = [1/3, 2/3]$, if $U_k = \cup_{i=1}^{2^k}(a_i, b_i), V_k = \cup_{i=1}^{2^k-1}[c_i, d_i]$, then $U_{k+1} = \cup_{i=1}^{2^k}(a_i, a_i + (b_i - a_i)/3) \cup (a_i + 2(b_i - a_i)/3, b_i), V_{k+1} = V_k \cup_{i=1}^{2^k} [a_i + (b_i - a_i)/3, a_i + 2(b_i - a_i)/3]$. Since $\sum_{k=1}^{\infty} P(|Y_n| > 1/n) = \sum_{k=1}^{\infty}(2/3)^k < \infty$, by Borel-Cantelli Lemma, we have $Y_n = O_{a.s.}(1/n)$. We also have that $EY_n = (4/3)^n + O(1/n) = O((4/3)^n)$. A heuristic approximation may suggest that the relative error of Y_n can be approximated by $O(n^{-1}(4/3)^{-n}) \to 0$. However, since $EY_n^2 = O((8/3)^n) \to \infty$, we have that $RE(Y_n) = O(2^n) \to \infty$, which is different from the heuristic result based on almost sure convergence.

Appendix C: Three Models in Estimating Intra-Horizon Risk

1. Merton's jump-diffusion model (MJD). Denote log return $X(t) = log(S(t)/S(0))$. Merton's jump diffusion model is given by $dX_t = \mu dt + \sigma dW_t + dJ_t$, where $\sigma > 0$ is the diffusion volatility, W_t is the standard Brownian motion, and $\{J_t, t \ge 0\}$ is a compound Poisson process having intensity $\lambda > 0$ and normally distributed jump sizes with mean μ and standard deviation σ_J. The Lévy measure is $k[x] = \frac{\lambda}{\sigma_J\sqrt{2\pi}} \exp\{-\frac{(x-\mu_J)^2}{2\sigma_J^2}\}, \mu_J \in \mathcal{R}, \sigma_J \in \mathcal{R}^+$. We assume the expected return is equal to zero, so the drift is $\mu = -(\sigma^2/2) - \lambda\Big(\exp(\mu_J + \sigma_J^2/2) - 1\Big)$. The characteristic function of log price X_t is $E[e^{iuX_t}] = \exp\Big(iu\mu t - \frac{u^2\sigma^2 t}{2} + \lambda t\Big(\exp(iu\mu_J - \frac{u^2\sigma_J^2}{2}) - 1\Big)\Big)$.

2. Double exponential model (DEM). Denote log return $X(t) = log(S(t)/S(0))$. The double exponential jump diffusion model is given by $dX(t) = \mu dt + \sigma dW(t) + dJ(t)$, where $J(t) = \sum_{i=1}^{N(t)} Y_i$ is the jump process with $N(t)$ a Poisson process with intensity λ. Y has double exponential density $f_Y(y) \sim p\eta_1 e^{-\eta_1 y}1_{\{y \ge 0\}} + q\eta_2 e^{\eta_2 y}1_{\{y<0\}}$, $p + q = 1$ and $\zeta := E(e^Y) - 1 = \frac{p\eta_1}{\eta_1 - 1} + \frac{q\eta_2}{\eta_2 + 1} - 1$. To ensure expected return to be zero, we have $\mu = r - \frac{1}{2}\sigma^2 - \lambda\zeta$. The characteristic function is $Ee^{iuX_t} = \exp\Big\{i\gamma ut - \frac{u^2\sigma^2 t}{2} + t\int_{-\infty}^{\infty}(e^{iuy} - 1 - iuy1_{\{|y| \le 1\}})\Pi(dy)\Big\}$, where $\Pi(dy) = \lambda \cdot f_Y(y)dy =$

$\lambda p\eta_1 e^{-\eta_1 y}1_{\{y\geq 0\}}dy + \lambda q\eta_2 e^{\eta_2 y}1_{\{y<0\}}dy$ and $\gamma = \mu + E[V1_{\{|V|\leq 1\}}] = r - \lambda\zeta + \lambda p(\frac{1-e^{-\eta_1}}{\eta_1} - e^{-\eta_1}) - \lambda q(\frac{1-e^{-\eta_2}}{\eta_2} - e^{-\eta_2})$.

3. Exponential dampened power law model (CGMY). Denote log return $X(t) = log(S(t)/S(0))$. The CGMY model is given by $dX_t = \mu dt + dJ_t$, where J_t is a pure-jump Lévy process with Lévy measure: $k[x] = C\frac{\exp\{-G|x|\}}{|x|^{1+Y}}1_{\{x<0\}} + C\frac{\exp\{-Mx\}}{x^{1+Y}}1_{\{x>0\}}$, where $G, M, C \in \mathcal{R}^+$. We assume the expected return is equal to zero, so the drift is $\mu = -C\Gamma(-Y)\{(M-1)^Y - M^Y + (G+1)^Y - G^Y\}$. The characteristic function is $E[e^{iuX_t}] = \exp\left(iu\mu t + C\Gamma(-Y)\left((M-iu)^Y - M^Y + (G+iu)^Y - G^Y\right)\right)$.

Acknowledgments

The research of Steven Kou is supported in part by MOE2014-T2-1-006 grant from MOE in Singapore, and NSFC grant 71571129 in China.

References

Artzner, P., F. Delbaen, J.-M. Eber, Health D. 1999. Coherent measures of risk. *Mathematical Finance* **9** 203-228.

Aviv, Y. 2003. A time-series framework for supply-chain inventory management. *Operations Research* **51** 210-227.

Bakshi, G. and Panayotov, G. 2010. First-passage probability, jump models, and intra-horizon risk. *Journal of Financial Economics* **95** 20-40.

Bassamboo, A., Juneja, S. and Zeevi, A. 2007, On the inefficiency loss of state-independent importance sampling in the presence of heavy-tails. *Operations Research letter* **35** 251-260.

Blanchet, J. and Glynn, P. 2008. Efficient rare event simulation for the maximum of heavy-tailed random walks. *The Annals of Applied Probability* **18** 1351-1378.

Blanchet, J. and Liu, J. 2008. State dependent importance sampling for regularly varying random walks. *Advances in Applied Probability* **40** 1104-1128.

Bosq, D. 1998. *Nonparametric Statistics for Stochastic Processes*. Springer Verlag.

Campbell, J., Lo., A.W., MacKinlay, A.C. 1997. *The Econometrics of Financial Markets*. Princeton University Press.

Chen, S. 2008. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* **6** 87-107.

Chen, S. and Tang, C. 2005. Nonparametric inference of Value-at-Risk for dependence financial returns. *Journal of Financial Econometrics* **3** 227-255.

Chen, Y., Xu, M., and Zhang, Z. (2009). A risk-averse newsvendor model with CVaR criterion. *Operations Research* **57** 1040-1044.

Choi, S., Ruszczinsky, A., and Zhao, Y. (2011). A multi-product risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research* **59** 346-364.

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* **10** 593-606.

Dang, D.M. and Forsyth, P.A. 2015. Better than pre-commitment mean-variance portfolio allocation strategies: a semi-self-financing Hamilton-Jacobi-Bellman equation. *European Journal of Operational Research* **250** 827-841.

David, H. A., Nagaraja, H. N. 2003. *Order Statistics*. Wiley.

Davydov, Yu. A. 1968. Convergence of distribution generated by stochastic processes. *Theory Probab. Appl.* **13** 730-737.

Dowd, K. and Cotter, J. 2007. Evaluating the precision of estimators of quantile-based risk measures. Working Paper.

Dupuis, P., Leder, K., and Wang, H. 2007. Importance sampling for sums of random variables with regularly varying tails. *ACM Transactions on Modeling and Computer Simulation* **17**.

Forsyth, P.A. and Vetzal, K.R. 2016. Dynamic mean variance asset allocation: tests for robustness. Working paper.

Fuh, C., Hu, I., Hsu, Y., and Wang, R. 2011. Efficient simulation of Value at Risk with heavy-tailed risk factors. *Operations Research* **59** 1395-1406.

Gilbert, K. 2005. An ARIMA Supply Chain Model. *Management Science* **51** 305-310.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. 2000. Variance reduction techniques for estimating Value-at-risk. *Management Science* **46** 1349-1364.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* **12** 239-269.

Glasserman, P. and Ruiz-Mata, J. 2006. Computing the credit loss distribution in the Gaussian copula model: a comparison of methods. *Journal of Credit Risk* **2** 33-66.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. Springer Verlag.

Hall, P. and Martin, M. A. 1988. Exact convergence rate of bootstrap quantile variance estimator. *Probability Theory and Related Fields* **80** 261-268.

Heidelberger, P. and Lewis, P. A. W. 1984. Quantile estimation in dependence sequences. *Operations Research* **32** 185-209.

Heidelberger, P., and Welch, P. 1981. A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM - Special issue on simulation modeling and statistical computing* **24** 233-245.

Hong, L. and Liu, G. 2011. Monte Carlo estimation of Value-at-Risk , conditional Value-at-Risk, and their sensitivities. *Proceedings of the 2011 Winter Simulation Conference* 95-107.

Hong, L., Hu, Z. and Liu, G. 2014. Monte Carlo methods for Value-at-Risk and conditional Value-at-Risk: a review. *ACM Transactions on Modeling and Computer Simulation* **24** 22:1-22:37.

Hong, L. and Liu, G. 2009. Simulating sensitivities of conditional value-at-risk. *Management Science* **55** 281-293.

Honore, P. 1998. Pitfalls in estimating jump diffusion models. *Working paper.*

Hult, H. and Svensson, J. 2009. Efficient calculation of risk measures by importance sampling-the heavy tailed case. Working paper.

Hult, H. and Svensson, J. 2011. On importance sampling with mixtures for random walks with heavy tails. *ACM Trans. Model. Comput. Simulation* **22**.

Jorion, P. 2007. *Value-at-Risk: the New Benchmark for Managing Financial Risk*, 3rd ed., McGraw-Hill, Boston.

Knuth, D. E. 1973. *The Art of Computer Programming.* Addison-Wesley.

Kou, S.G. 2002. A jump-diffusion model for option pricing. *Management Science* **48** 1086-1101.

Kou, S.G. and Peng, X. 2016. On the measurement of economic tail risk. *Operations Research* **64** 1056-1072.

Kou, S.G., Peng, X. H, and Heyde, C. C. 2013. External risk measures and Basel accords. *Mathematics of Operations Research* **38** 393-417.

Kou, S.G. and Wang, H. 2003. First passage times of a jump diffusion process. *Advances in Applied Probability* **35** 504-531.

Kushagra, S., Lopez-Oritz, A., Munro, J., and Qiao, A. 2014. Multi-pivot quicksort: theory and experiments. *Proceedings of the Meeting on Algorithm Engineering & Expermiments* 47-60.

Leippold, M. and Vasiljevic, N. 2016. Option-implied intra-horizon risk and first-passage disentanglement. Working paper.

Liu, J. and Yang, X. 2012. The convergence rate and asymptotic distribution of the bootstrap quantile variance estimator for importance sampling. *Adv. Appl. Prob* **44** 815-841.

McNeil, A., Frey R., Embrechts P. 2015. *Quantitative Risk Management: Concepts, Techiniques, Tools.* Princeton University Press, Princeton, NJ.

Merlevede, F., Peligrad, M., and Rio, E. 2009. Bernstein inequalities and moderate deviations under strong mixing conditions. *Institute of Mathematical Statistics Collections* **5** 273-292.

Papadatos, N. 1995. Maximum variance of order statistics. *Ann. Inst. Statist. Math.* **47** 185-193.

Petrov, V.V. 1975. *Sums of Independent Random Variables.* Springer Verlag.

Puschner, P. P. 1999. Real-Time performance of sorting algorithms. *Real-Time Systems* **16** 63-79.

Reiss, R.-D. 1989. *Approximate Distributions of Order Statistics.* Springer Verlag.

Rockafellar, R. T., and Uryasev, S. 2002. Conditional value-at-risk for general loss distribution. *Journal of Banking and Finance* **26** 1443-1471

Rosenblatt, M. 1961. Independence and dependence. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 431-443.

Sedgewick, R. 1989. *Algorithms*. Addison-Wesley.

Shapiro, A., Dentcheva, D. and A Ruszczyski. 2014. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM series on optimization.

Simonato, J. and Sentoft, L. 2015. Which pricing approach for options under GARCH with non-normal innovations? Working paper.

Sun, L. and Hong, L. 2010. Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters* **38** 246-251.

Tapiero, C. (2005). Value at risk and inventory control. *European Journal of Operational Research* **163** 769775.

Taylor, A. E. 1952. L'Hospital's Rule. *The American Mathematical Monthly* **59** 20-24.

Tikhomirov, A. N. 1980. On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory of Probability and its Application* **25** 790-809.

Viola, E. 2004. The complexity of constructing pseudo random generators from hard functions. *Computational Complexity* **13** 147-188.

Yaroslavskiy, V. 2009. Dual-pivot Quicksort algorithm. Working paper.

Yamai, Y. and Yoshiba, T. 2002. Comparative analyses of expected shortfall and Value-at-Risk: their estimation error, decomposition, and optimization. *Monetary and Economic Studies* **20**.

Yamai, Y. and Yoshiba, T. 2005. Value at risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance* **29** 997-1015.

Yokoyama, R. 1980. Moment bounds for stationary mixing sequences. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **52** 45-57.

E-Companion of "Simulating Risk Measures with Estimated Relative Error" by Wei Jiang and Steven Kou

EC.1. Proof of Lemma 1 and Related Algorithms

EC.1.1. Algorithms

Naive Monte Carlo: Estimation of VaR and ES without sorting method.

Step 1 Generate N loss samples $\{L_1, \cdots, L_N\}$.

Step 2 Define empirical function $F_N(x) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{L_i \le x\}}$. For $i = 1:N$, denote $T_i = F_N(L_i)$, if $T_i \ge p$, we have $Ch_i = 1$, else $Ch_i = -1$.

Step 3 Estimate of VaR and ES by $v_N = \min\{L_i * Ch_i, i = 1:N\}$ and $c_N = v_N - \frac{1}{Np}\sum_{i=1}^{N}(v_N - L_i)1_{\{L_i \le v_N\}}$.

Bubble/Quick sorted Monte Carlo The Bubble/Quick sorted Monte Carlo algorithm means only the S-step (from sub-step 1 to sub-step 3) in the algorithm of sorted Monte Carlo with estimated RE. Bubble (Quick) sorted Monte Carlo means the bubble (quick) sorting method is used in the sub-step 2.

EC.1.2. Proof of Lemma 1

Proof. Suppose the time required to do one simple action is bounded by t_1 and t_2 is the time to generate one loss sample. We try to find the relationship between t_2 and t_1 first. Suppose to simulate one portfolio loss sample we need M different random variables with length $n_1, \cdots, n_M$ and need K simple actions to put up them together. Then we have $t_2 = (\sum_{i=1}^{M} P(n_i)) \times t_1 + K \times t_1$, where $P(\cdot)$ is the polynomial complexity function in terms of the length of the generated random number (see Viola 2004). Given the above complexity of generating loss samples, we can clarify the complexity for each algorithm by working out the number of simple actions in each step of the Monte Carlo method.

More precisely, for naive Monte Carlo, given the number of loss samples N, the complexity of step 1 is $N \times t_2 = N \times (\sum_{i=1}^{M} P(n_i) + K) \times t_1$. Complexity of step 2 is $((2N+1) \times N \times t_1 + N \times t_1 = ((2N+1)N + N) \times t_1$. Complexity of step 3 is $(2N) \times t_1$ for VaR and $(2N) \times t_1 + (3N) \times t_1 + 3t_1 = (5N+3) \times t_1$ for ES. Then the summation of complexity for three steps is $O(N^2)$ and hence the complexity of naive Monte Carlo is $O(N^2)$. Similarly, for the Bubble (Quick) sorted Monte Carlo, complexity for sub-step 1 is $N \times t_2 = N \times (\sum_{i=1}^{M} P(n_i) + K) \times t_1$. Complexity of sub-step 2 is exactly the complexity of the bubble (quick) sorting algorithm. Complexity of sub-step 3 is $3t_1$ for VaR and $(3N+6) \times t_1$ for ES. Since complexity for Bubble Sort and Quick Sort are $O(N^2)$, $O(N \log N)$ respectively (see Knuth, 1973, Sedgewick, 1989), the computational complexity for Bubble sorted Monte Carlo and Quick sorted Monte Carlo are $O(N^2)$, $O(N \log N)$ respectively. □

EC.2. Proof of Theorem 1

Let's firstly state some useful lemmas. The following lemma is a simple extension of lemma 3 in Liu and Yang (2012).

LEMMA 2. *(1) Let $h(x)$ be a nonnegative function, such that there exists $\zeta_0 > 0$, $h(x) \le x^{\zeta_0}$ for all sufficiently large x. (2) $Q(v,n)$ is a function depending on v and n, such that $Q(v,n) \to 0$ as $n \to \infty$ or $v \to 0$. If for all $\zeta_1, \lambda > 0$, $n^{\lambda(\zeta_1-1)}Q(v,n) \to 0$ is satisfied, and there exists $\zeta_2 > 0$, such that $\lambda(\zeta_1+1) < \zeta_2$ and $n^{\zeta_2}Q(v,n) \to \infty$ then:*

$$\int_0^{n^\lambda} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n)))dx = \int_0^{n^\lambda} h(x)\Phi(-x)dx + o(Q(v,n)),$$

where $\Phi(\cdot)$ is the cumulative distribution function of normal distribution. In addition, we define $a(n,x) = o(x^{\zeta_1}Q(v,n))$ if $a(n,x)x^{-\zeta_1}Q(v,n)^{-1} \to 0$ as $n \to \infty$ uniformly for $x \in (\epsilon_1, n^\lambda)$ for any $\epsilon_1 > 0$.

Proof. We split the integral into two parts:

$$\begin{aligned}\int_0^{n^\lambda} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n))) = &\int_0^{(\log n)^2} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n))) \\ &+ \int_{(\log n)^2}^{n^\lambda} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n)))\end{aligned}$$

We firstly analyze the second term. Since $x \in ((\log n)^2, n^\lambda)$, then for large enough n, we have $-x(1 + o(x^{(\zeta_1-1)}Q(v,n))) < -\frac{(\log n)^2}{2}$ and hence $\Phi(-x + o(x^{\zeta_1}Q(v,n))) = \Phi(-x(1 + o(x^{(\zeta_1-1)}Q(v,n)))) \le \Phi(-\frac{(\log n)^2}{2})$. Then we obtain $\int_{(\log n)^2}^{n^\lambda} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n))) \le n^{\lambda(\zeta_0+1)}\Phi(-\frac{(\log n)^2}{2}) = o(Q(v,n))$.

Next, we consider the first term. Noticing that for all $\epsilon_1 \le x \le (\log n)^2$, we have

$$\begin{aligned}\frac{\Phi(-x + o(x^{\zeta_1}Q(v,n))) - \Phi(-x)}{\Phi(-x)x^{\zeta_1+1}Q(v,n)} &= \frac{\int_x^{x+o(x^{\zeta_1}Q(v,n))} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx}{\Phi(-x)x^{\zeta_1+1}Q(v,n)} \\ &= \frac{\inf\{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \frac{1}{\sqrt{2\pi}}e^{-(x+o(x^{\zeta_1}Q(v,n)))^2/2}\}o(x^{\zeta_1}Q(v,n))}{\Phi(-x)x \times x^{\zeta_1}Q(v,n)} \\ &\to 0\end{aligned}$$

where the last convergence part holds due to that $\frac{\inf\{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \frac{1}{\sqrt{2\pi}}e^{-(x+o(x^{\zeta_1}Q(v,n)))^2/2}\}}{\Phi(-x)x}$ is bounded and $\frac{o(x^{\zeta_1}Q(v,n))}{x^{\zeta_1}Q(v,n)} \to 0$. Hence there is $\Phi(-x + o(x^{\zeta_1}Q(v,n))) = (1 + o(x^{\zeta_1+1}Q(v,n)))\Phi(-x)$, for $x \in [\epsilon_1, (\log n)^2]$. Then we obtain

$$\begin{aligned}&\int_0^{(\log n)^2} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n))) \\ &= \int_{\epsilon_1}^{(\log n)^2} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n))) + \int_0^{\epsilon_1} h(x)\Phi(-x + o(x^{\zeta_1}Q(v,n)))\end{aligned}$$

$$= \int_{\epsilon}^{(\log n)^2} h(x)\Phi(-x)dx + \int_{\epsilon_1}^{(\log n)^2} h(x)\Phi(-x)o(x^{\zeta_1+1}Q(v,n))dx + \int_{0}^{\epsilon_1} h(x)[\Phi(-x)+o(x^{\zeta_1}Q(v,n))]dx$$
$$= \int_{0}^{(\log n)^2} h(x)\Phi(-x)dx + \int_{\epsilon_1}^{(\log n)^2} h(x)\Phi(-x)o(x^{\zeta_1+1}Q(v,n))dx + \int_{0}^{\epsilon_1} h(x)o(x^{\zeta_1}Q(v,n))dx$$
$$= \int_{0}^{(\log n)^2} h(x)\Phi(-x)dx + o(Q(v,n)).$$

The last equality holds due to that by the definition of $o(x^{\zeta_1+1}Q(v,n))$, for any ϵ_0, we can choose large enough n such that $[\int_0^{\epsilon_1} h(x)o(x^{\zeta_1}Q(v,n))dx]/Q(v,n) \le \epsilon_0/2$ and $[\int_{\epsilon_1}^{(\log n)^2} h(x)\Phi(-x)o(x^{\zeta_1+1}Q(v,n))dx]/Q(v,n) \le \epsilon_0/2$. The conclusion follows by combining the first term and the second term. □

The following lemma taken from Tikhomirov (1980) is the Berry-Essen type bound which evaluates the distance between an empirical distribution and the normal distribution for a sequence of dependent random variables.

LEMMA 3. *Let $\{X_j\}_{j\ge1}$ be a sequence of stationary $\alpha-$mixing random variables with zero mean and finite variances. There exist constants $c_\alpha > 0$ and $\rho \in (0,1)$ such that the $\alpha-$mixing coefficient $\alpha(n)$ satisfies $\alpha(n) \le c_\alpha \rho^n$ for all n. There also exist constants $\delta, 0 < \delta \le 1$ such that $\mathbf{E}|X_1|^{2+\delta} < \infty$. Then there is and A depending just on c_α, ρ and δ such that*

$$\sup_z |F_n(z) - \Phi(z)| \le A n^{\delta/2} \log^{1+\delta} n,$$

where $F_n(z) = P(\frac{1}{\sqrt{n}\sigma_n}\sum_{j=1}^n X_j < z), \sigma_n^2 = \frac{1}{n}\mathbf{E}(\sum_{j=1}^n X_j)^2$, and $\Phi(\cdot)$ is the normal distribution function.

The following lemma taken from Merlevede, Peligrad and Rio (2009) is the Bernstein type of inequality for dependent random variables. We can use it to estimate the tail probability of the sum of dependent random variables.

LEMMA 4. *Let $(X_j)_{j\ge1}$ be a sequence of centered real-valued random variables. Suppose that the sequence satisfies geometric $\alpha-$mixing with coefficient $\alpha(n) \le c_\alpha \rho^n$ and there exists a positive M such that $\sup_{i\ge1} ||X_i||_\infty \le M$. Then there are positive constants C_1 and C_2 depending only on ρ such that for all $n \ge 2$ and t satisfying $t < \frac{1}{C_1 M(\log n)^2}$, we have*

$$\log \mathbf{E}(\exp(tS_n)) \le \frac{C_2 t^2 (n\hat{\sigma}^2 + M^2)}{1 - C_1 t M (\log n)^2},$$

where $S_n = \sum_{i=1}^n X_i$ and the $\hat{\sigma}^2$ is defined by $\hat{\sigma}^2 = \sup_{i>0}(var(X_i) + 2\sum_{j>i}|cov(X_i, X_j)|)$.

In terms of probabilities, there is a constant C_3 depending only on ρ such that for all $n \ge 2$,

$$P(|S_n| \ge x) \le \exp\Big(-\frac{C_3 x^2}{\hat{\sigma}^2 n + M^2 + xM(\log n)^2}\Big).$$

The following lemma, which is a combination of Corollary 1.1 from Bosq (1998) and Lemma 2.1 from Davydov (1968), proves the boundedness of the covariance coefficient for two random variables with $\alpha-$mixing coefficient α.

LEMMA 5. *(1)Let X and Y be two real valued random variables with $\alpha-$mixing coefficient α such that $X \in L^q(P), Y \in L^r(P)$ where $q>1, r>1$ and $\frac{1}{q}+\frac{1}{r}=1-\frac{1}{p}$, then $|Cov(X,Y)| \leq 2p(2\alpha)^{1/p}||X||_q||Y||_r$ (Davydov's inequality) . In particular, if $X \in L^\infty(P), Y \in L^\infty(P)$ then $|Cov(X,Y)| \leq 4||X||_\infty||Y||_\infty\alpha$ (Billingsley's inequality) . (2) Let $Y_1, Y_2, \cdots$ be r.v.s centered at their expectation, bounded by M, say, and forming an $\alpha-$mixing sequence with mixing coefficients $\alpha(n)$ such that $\sum_{n=1}^{\infty}\alpha(n)<\infty$. Let ξ and η be r.v.'s such that ξ is $\mathcal{F}_1^k-$measurable, η is $\mathcal{F}_{k+n}^\infty-$measurable, $|\eta| \leq M_0$ and $\mathbf{E}|\xi|^q<\infty$ for some $q>1$. Then $|Cov(\xi,\eta)| \leq 6M_0\alpha(n)^{1-1/q}||\xi||_q$.*

The following lemma from Yokoyama (1980) gives moment bounds for $\alpha-$mixing sequences. We should notice that the condition in this lemma is satisfied for stationary geometric $\alpha-$mixing time series.

LEMMA 6. *Let $\{X_j\}$ be a strictly stationary $\alpha-$mixing sequence with $EX_1=0$ and $E|X|^{r+\delta}<\infty$ for some $r>2$ and $\delta>0$. If $\sum_{i=0}^{\infty}(i+1)^{r/2-1}[\alpha(i)]^{\delta/(r+\delta)}<\infty$, then there exists a constant K such that $E|S_n|^r \leq Kn^{r/2}$.*

With the help of lemmas 2, 3, 4, 5, and 6, we can prove the following Theorems 4 and 5 about the approximation of moments of VaR and ES, which is fundamental for our expansion of RE. Note that this theorem are for both dependent samples and i.i.d samples, while similar existing results of VaR are only applied to i.i.d samples under stronger conditions (see Hall and Martin (1988) and Reiss (1989)).

THEOREM 4. *Under the conditions of Assumptions A, B, C1, for integer $m>0$:*

$$E(v_n-v)^m=\frac{\sigma_{n,v}^m}{f(v)^m}EZ^m n^{-m/2}+o(n^{-m/2-1/4}). \tag{ec.1}$$

Where Z satisfies standard Normal distribution and $\sigma_{n,v}^2=\{p(1-p)+2\sum_{k=1}^{n-1}\gamma_1(k)\}$, $\gamma_1(k)=cov\{1_{\{L_1<v\}},1_{\{L_{k+1}<v\}}\}$.

Proof. Since $EX^m=\int_{x>0}mx^{m-1}P(X>x)dx-\int_{x<0}mx^{m-1}P(X<x)dx$, we only need to compute the following probability:

$$P(n^{1/2}(v_n-v)>x)=P(F_n(v+xn^{-1/2})<p)=P(\sum_{i=1}^{n}(I_{\{L_i\leq v+xn^{-1/2}\}}-p)<0).$$

For some $\lambda \in (0,\frac{1}{4m})$, we try to estimate this probability in the following three cases: $0<x\leq n^\lambda, n^\lambda \leq x \leq c\sqrt{n}$, and $x>c\sqrt{n}$. Here c is a very small constant comparing with v. The calculation of $P(n^{1/2}(v_n-v)<x)$ in the region $x<0$ is the same as that for $x>0$.

Case 1: $0 < x \leq n^{\lambda}$. Denote $W_{x,n,i} = I_{\{L_i \leq v+xn^{-1/2}\}} - F(v + xn^{-1/2})$ and let $\sigma_n(x) = \sqrt{\frac{1}{n} var(\sum_{i=1}^{n} W_{x,n,i})}$. We can see that $\{W_{x,n,i}\}$ is a sequence of stationary geometric $\alpha-$mixing random variables with zero mean and finite moments, then according to lemma 3, one obtains

$$\begin{aligned} P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0) &= P(\frac{\sum_{i=1}^{n} W_{x,n,i}}{\sqrt{n}\sigma_n(x)} < -\frac{F(v+xn^{-1/2}) - p}{\sigma_n(x)/\sqrt{n}}) \\ &= \Phi(-\frac{F(v+xn^{-1/2}) - p}{\sigma_n(x)/\sqrt{n}}) + D_3(x) \end{aligned}$$

where there exists a constant A such that $|D_3(x)| \leq An^{-1/2}\log^2 n$.

Case 2: $n^{\lambda} \leq x \leq c\sqrt{n}$. Since in this case $xn^{-1/2} \leq c$, where c is very small constant, hence $(F(v + xn^{-1/2}) - p) = \int_v^{v+xn^{-1/2}} f(x)dx \sim f(v)xn^{-1/2}$. We can easily verify that $\{W_{x,n,i}\}_{i\geq 1}$ is a sequence of centered real-valued random variables satisfying geometric $\alpha-$mixing and bounding by 1. Hence the conditions in lemma 4 are satisfied by $\{W_{x,n,i}\}_{i\geq 1}$ and we can use this Bernstein type of inequality for dependent case to obtain:

$$\begin{aligned} P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0) &= P(\sum_{i=1}^{n} W_{x,n,i} < -n(F(v+xn^{-1/2}) - p)) \\ &\leq \exp\Big(-\frac{C_3(n(F(v+xn^{-1/2}) - p))^2}{\hat{\sigma}_1^2 n + 1 + (n(F(v+xn^{-1/2}) - p))(\log n)^2}\Big), \end{aligned}$$

where $\hat{\sigma}_1^2 = \sup_{i>0}(var(W_{x,n,i}) + 2\sum_{j>i}|cov(W_{x,n,i}, W_{x,n,j})|)$. Since $\{W_{x,n,i}\}_{i>0}$ is a stationary sequence, we have $\hat{\sigma}^2 = var(W_{x,n,1}) + 2\sum_{j>1}|cov(W_{x,n,1}, W_{x,n,j})|$. For small c, $var(W_{x,n,1}) \sim var(W_{0,n,1}) = F(v)(1 - F(v))$. Furthermore, based on Billingsley's inequality in lemma 5 (1), $cov(W_{x,n,i}, W_{x,n,j})$ is bounded by geometric form of $\alpha-$mixing coefficient $\alpha(|j-i|) \leq c_\alpha \rho^{|j-i|}$, and hence $\sum_{j>i}|cov(W_{x,n,i}, W_{x,n,j})|$ is bounded. Therefore $\hat{\sigma}_1^2$ is bounded. Then we have

$$\begin{aligned} \exp\Big(-\frac{C_3(n(F(v+xn^{-1/2}) - p))^2}{\hat{\sigma}^2 n + 1 + (n(F(v+xn^{-1/2}) - p))(\log n)^2}\Big) &\sim \exp\Big(-\frac{C_3 f(v)x^2 n}{\hat{\sigma}_1^2 n + 1 + f(v)xn^{-1/2}(\log n)^2}\Big) \\ &\leq \exp\Big(-\kappa_1 x\Big) \\ &\leq x^{-s}, \end{aligned}$$

where s can take a large enough positive integer and κ_1 is some constant depending on $C_3, f(v)$ and $\hat{\sigma}_1$. Hence we prove that for $n^{\lambda} \leq x \leq c\sqrt{n}$, $P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0) \leq x^{-s}$.

Case 3: $x > c\sqrt{n}$. Let $\Delta = \frac{F(v+xn^{-1/2}) - p}{\sigma_n(x)/\sqrt{n}}$. Since $P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0)$ is a non-increasing function of x, hence $P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0 | x > c\sqrt{n}) \leq P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0 | x = c\sqrt{n}) \leq (c\sqrt{n})^{-s}$. Then there are two subcases: (1)for $c\sqrt{n} \leq x \leq n$, one sees: $P(\sum_{i=1}^{n}(I_{\{L_i \leq v+xn^{-1/2}\}} - p) < 0) \leq (c\sqrt{n})^{-s} = O(x^{-s/2})$. (2)For $x > n$, we still use the same way as in case 2. As $x > n$, $\frac{x}{\sqrt{n}} > \sqrt{n} \gg 2|v|$, then $1 \geq F(v + xn^{-1/2}) > F(\frac{1}{2}xn^{-1/2}) > \frac{1}{2}$ and $1 - F(v + xn^{-1/2}) = EI_{\{L \geq v+xn^{-1/2}\}} \leq \frac{E|L|}{\frac{1}{2}xn^{-1/2}}$. Along with $\sigma_n(x) = \sqrt{\frac{1}{n} var(\sum_{i=1}^{n} W_{x,n,i})} \leq \sqrt{\sum_{i=1}^{n} EW_{x,n,i}^2} \leq$

$\sqrt{n(1-F(v+xn^{-1/2}))F(v+xn^{-1/2})} \le \sqrt{n\frac{E|L|}{\frac{1}{2}xn^{-1/2}}}$, we have $\Delta \ge \frac{1/2-p}{\sqrt{\frac{E|L|}{\frac{1}{2}xn^{-1/2}}}} \ge \kappa_2 \frac{\sqrt{x}n^{-1/4}}{\sqrt{E|L|}}$, where κ_2 is some constant depending on p. Similar to case 2, we apply lemma 4 and obtain

$$\begin{aligned} P(\sum_{i=1}^{n}(I_{\{L_i\le v+xn^{-1/2}\}}-p)<0) &= P(\sum_{i=1}^{n}W_{x,n,i} < -n(F(v+xn^{-1/2})-p)) \\ &= P(\frac{\sum_{i=1}^{n}W_{x,n,i}}{\sqrt{n}\sigma_n(x)} < \frac{F(v+xn^{-1/2})-p}{\sigma_n(x)/\sqrt{n}}) \\ &\le \exp\Big(-\frac{C_3\Delta^2}{\hat{\sigma}_2^2 n+1+\Delta(\log n)^2}\Big), \end{aligned}$$

where $\hat{\sigma}_2^2 = \sup_{i>0}(var(W_{x,n,i}/(\sqrt{n}\sigma_n(x))) + 2\sum_{j>i}|cov(W_{x,n,i}/(\sqrt{n}\sigma_n(x)), W_{x,n,j}/(\sqrt{n}\sigma_n(x)))|)$. We should notice that based on Billingsley's inequality in lemma 5 (1), both $\hat{\sigma}_2^2 \times (n\sigma_n(x)^2)$ and $\sigma_n(x)^2$ are bounded. Hence $n\hat{\sigma}_2^2$ is bounded, and we have

$$\begin{aligned} \exp\Big(-\frac{C_3\Delta^2}{\hat{\sigma}_2^2 n+1+\Delta(\log n)^2}\Big) &\le \exp\Big(-\frac{C_3(\kappa_2\sqrt{x}n^{-1/4}/\sqrt{E|L|})^2}{\hat{\sigma}_2^2 n+1+\kappa_2\sqrt{x}n^{-1/4}/\sqrt{E|L|}(\log n)^2}\Big) \\ &\le \exp\left(-\kappa_3 x^{1/4}\right) \\ &\le x^{-s}, \end{aligned}$$

where κ_3 is some constant depending on C_3, κ_2, and $E|L|$. Therefore, we prove that for $x > c\sqrt{n}$, tail probability $P(\sum_{i=1}^{n}(I_{\{L_i\le v+xn^{-1/2}\}}-p)<0)$ is also bounded by x^{-s}.

Summarizing case 2 and case 3, one obtains

$$\int_{n^\lambda}^{\infty} mx^{m-1}P(v_n > v+\frac{x}{\sqrt{n}})dx \le \int_{n^\lambda}^{\infty} mx^{m-1}\cdot x^{-s/2}dx = O(n^{-t})$$

where t can be a large enough integer. Then we obtain

$$\begin{aligned} \int_0^{\infty} mx^{m-1}P(v_n > v+\frac{x}{\sqrt{n}})dx &= \int_0^{n^\lambda} mx^{m-1}P(v_n > v+\frac{x}{\sqrt{n}})dx + O(n^{-t}) \\ &= \int_0^{n^\lambda} mx^{m-1}[\Phi(-\Delta)+O(n^{-1/2}log^2 n)]dx + O(n^{-t}) \\ &= \int_0^{n^\lambda} mx^{m-1}\Phi(-\Delta)dx + O(n^{m\lambda-1/2}log^2 n) + O(n^{-t}) \qquad \text{(ec.2)} \end{aligned}$$

To compute the term $\int_0^{n^\lambda} mx^{m-1}\Phi(-\Delta)dx$ in (ec.2), we need to estimate $\Delta = \frac{F(v+xn^{-1/2})-p}{\sigma_n(x)/\sqrt{n}}$. Firstly, we consider the denominator of Δ and prove that $\sigma_n(x) = (1+O(xn^{-1/2}))\sigma_n(0)$. By definition,

$$\sigma_n(x)^2 = \frac{1}{n}var(\sum_{i=1}^{n}W_{x,n,i}) = var(W_{x,n,1}) + 2\sum_{i=1}^{n-1}(1-\frac{i}{n})cov(W_{x,n,1}, W_{x,n,i+1}).$$

Here $var(W_{x,n,1}) = F(v+xn^{-1/2})(1-F(v+xn^{-1/2})) = (1+O(xn^{-1/2}))F(v)(1-F(v)) = (1+O(xn^{-1/2}))var(W_{0,n,1})$. For the covariance part, since $W_{x,n,i}$ is centered real-value r.v, we have

$|cov(W_{x,n,1}, W_{x,n,i+1}) - cov(W_{0,n,1}, W_{0,n,i+1})| = |EW_{x,n,1}W_{x,n,i+1} - EW_{0,n,1}W_{0,n,i+1}| \le |E(W_{x,n,1} - W_{0,n,1})W_{x,n,i+1}| + |E(W_{x,n,i+1} - W_{0,n,i+1})W_{0,n,1}|$. Since $|W_{x,n,1}| \le 1$ and $E|(W_{x,n,i+1} - W_{0,n,i+1}]|^q < \infty$ for any $q \ge 1$, according to lemma 5 (2), if we take $1/q = 1 - \epsilon_2$ for some small $\epsilon_2 > 0$, then

$$\begin{aligned}
|E(W_{x,n,i+1} - W_{0,n,i+1})W_{x,n,1}| &\le 6\alpha(i)^{1-1/q}[E|W_{x,n,i+1} - W_{0,n,i+1}|^q]^{1/q} \\
&\le 6(c_\alpha \rho^i)^{\epsilon_2}[E|W_{x,n,i+1} - W_{0,n,i+1}|]^{1-\epsilon_2} \\
&\le 6c_\alpha^{\epsilon_2}\rho^{\epsilon_2 i}[E|1_{\{v \le L_{i+1} \le v + xn^{-1/2}\}} - (F(v + xn^{-1/2}) - F(v))|]^{1-\epsilon_2} \\
&\le 6c_\alpha^{\epsilon_2}\rho^{\epsilon_2 i}[2(1 - (F(v + xn^{-1/2}) - F(v)))(F(v + xn^{-1/2}) - F(v))]^{1-\epsilon_2} \\
&\le 6c_\alpha^{\epsilon_2}\rho^{\epsilon_2 i}O(xn^{-1/2})^{1-\epsilon_2}.
\end{aligned}$$

Similarly, we have $|E(W_{x,n,1} - W_{0,n,1})W_{x,n,i+1}| = 6c_\alpha^{\epsilon_2}\rho^{\epsilon_2 i}O(xn^{-1/2})^{1-\epsilon_2}$. Hence we have

$$\begin{aligned}
\sigma_n(x)^2 &= (1 + O(xn^{-1/2}))var(W_{0,n,1}) + 2\sum_{i=1}^{n-1}(1 - \frac{i}{n})[cov(W_{0,n,1}, W_{0,n,i+1}) + 12c_\alpha^{\epsilon_2}\rho^{\epsilon_2 i}O(xn^{-1/2})^{1-\epsilon_2}] \\
&= var(W_{0,n,1}) + 2\sum_{i=1}^{n-1}(1 - \frac{i}{n})cov(W_{0,n,1}, W_{0,n,i+1}) + O(xn^{-1/2})^{1-\epsilon_2} \\
&= \sigma_n(0)^2(1 + O(xn^{-1/2})^{1-\epsilon_2}),
\end{aligned}$$

where the last equality holds because $\sigma_n(0)$ is bounded.

Next, we show that $\Delta = \frac{F(v+xn^{-1/2}) - F(v)}{\sigma_n(x)/\sqrt{n}} = \frac{xf(v)}{\sigma_n(0)} + \inf\{O(x^{3/2+\delta_0}n^{-1/4-\delta_0/2}), O(x^{2-\epsilon_2}n^{-1/2+1/2\epsilon_2})\}$. With Assumption B about the Hölder continuous of the density function $f(\cdot)$, we have

$$\begin{aligned}
F(v + xn^{-1/2}) - F(v) &= \int_v^{v+xn^{-1/2}} f(u)du = f(v)xn^{-1/2} + \int_v^{v+xn^{-1/2}} (f(u) - f(v))du \\
&\le f(v)xn^{-1/2} + \int_v^{v+xn^{-1/2}} c_H(u - v)^{1/2+\delta_0}du \\
&= f(v)xn^{-1/2} + O(x^{3/2+\delta_0}n^{-3/4-1/2\delta_0}).
\end{aligned}$$

Using $\sigma_n(x)^2 = (1 + O(xn^{-1/2}))\sigma_n(0)^2$, we have

$$\begin{aligned}
\Delta &= \frac{F(v + xn^{-1/2}) - F(v)}{\sigma_n(x)/\sqrt{n}} = \frac{f(v)xn^{-1/2} + O(x^{3/2+\delta_0}n^{-3/4-1/2\delta_0})}{\sigma_n(0)\sqrt{(1 + O(xn^{-1/2})^{1-\epsilon_2})}/\sqrt{n}} \\
&= \frac{1}{\sigma_n(0)}(xf(v) + O(x^{3/2+\delta_0}n^{-1/4-1/2\delta_0}))(1 + O(xn^{-1/2})^{1-\epsilon_2}) \\
&= \frac{xf(v)}{\sigma_n(0)} + \inf\{O(x^{3/2+\delta_0}n^{-1/4-\delta_0/2}), O(x^{2-\epsilon_2}n^{-1/2+1/2\epsilon_2})\},
\end{aligned}$$

where $\sigma_n(0)^2 = p(1-p) + 2\sum_{i=1}^{n-1}(1 - i/n)\gamma_1(i), \gamma_1(i) = cov(1_{\{L_1 < v\}}, 1_{\{L_{i+1} < v\}})$. Then according to lemma 2, one obtains

$$\int_0^{n^\lambda} mx^{m-1}\Phi(-\Delta)dx = \int_0^{n^\lambda} mx^{m-1}\Phi(-\frac{xf(v)}{\sigma_n(0)})dx + o(n^{-1/4}).$$

Since $\lambda \in (0, \frac{1}{4m})$, we have $m\lambda - 1/2 < -1/4$. Using (ec.2) we have

$$\begin{aligned}
\int_0^\infty mx^{m-1}P(v_n - v > x)dx &= \frac{1}{n^{m/2}}\int_0^\infty mx^{m-1}P(v_n > v + \frac{x}{\sqrt{n}})dx \\
&= \frac{1}{n^{m/2}}\Big(\int_0^{n^\lambda} mx^{m-1}\Phi(-\Delta)dx + O(n^{m\lambda-1/2}log^2 n) + O(n^{-t})\Big) \\
&= \frac{1}{n^{m/2}}\int_0^{n^\lambda} mx^{m-1}\Phi(-\frac{xf(v)}{\sigma_n(0)})dx + o(n^{-1/4-m/2}) \\
&= \frac{\sigma_n(0)^m}{n^{m/2}f(v)^m}\int_0^{n^\lambda} mx^{m-1}\Phi(-x)dx + o(n^{-1/4-m/2}). \\
&= \frac{\sigma_n(0)^m}{n^{m/2}f(v)^m}\Big(\int_0^\infty mx^{m-1}\Phi(-x)dx - \int_{n^\lambda}^\infty mx^{m-1}\Phi(-x)dx\Big) + o(n^{-1/4-m/2}) \\
&= \frac{\sigma_n(0)^m}{n^{m/2}f(v)^m}\int_0^\infty mx^{m-1}\Phi(-x)dx + o(n^{-1/4-m/2})
\end{aligned}$$

where the last equality holds due to that $\int_{n^\lambda}^\infty mx^{m-1}\Phi(-x)dx \le \sqrt{\Phi(-n^\lambda)}\int_{n^\lambda}^\infty mx^{m-1}\sqrt{\Phi(-x)}dx = O(\sqrt{\Phi(-n^\lambda)}) = o(n^{-1/4-m/2})$.

Approximations of $P(v_n - v < x)$ for $x < 0$ are the same. Therefore,

$$E(v_n - v)^m = n^{-m/2}\frac{\sigma_n(0)^m}{f(v)^m}EZ^m + o(n^{-1/4-m/2}).$$

where Z is standard normally distributed r.v. Furthermore, since $E|1_{\{L<v\}}|^q < \infty$ for any $q > 0$, lemma 5 implies that $|\gamma_1(i)|$ is bounded by some geometric series and we have $\sum_{i=1}^{n-1} i/n\gamma_1(i) = O(n^{-1})$. Hence

$$\sigma_n(0)^2 = p(1-p) + 2\sum_{i=1}^{n-1}(1 - i/n)\gamma_1(i) = p(1-p) + 2\sum_{i=1}^{n-1}\gamma_1(i) + O(n^{-1}).$$

Then we can replace $\sigma_n(0)^2$ by $\sigma_{n,v}^2 := p(1-p) + 2\sum_{i=1}^{n-1}\gamma_1(i)$ and obtain

$$E(v_n - v)^m = \frac{\sigma_{n,v}^m}{f(v)^m}EZ^m n^{-m/2} + o(n^{-1/4-m/2}),$$

from which the result is proved. □

THEOREM 5. *Under the conditions of Assumptions A, B, C, for $m = 1, 2$, we have:*

$$E(c_n - c)^m = \frac{\sigma_{n,c}^m}{p^m}EZ^m n^{-m/2} + o(n^{-m/2-1/4+\epsilon}). \tag{ec.3}$$

where Z satisfies standard Normal distribution and $\sigma_{n,c}^2 = \{var[(v - L_1)^+] + 2\sum_{k=1}^n \gamma_2(k)\}$, $\gamma_2(k) = cov\{(v - L_1)^+, (v - L_{k+1})^+\}$.

Proof. There are three steps to prove this result.

Step 1. We prove that ES estimator c_n satisfies equation $c_n = v - \frac{1}{np}\sum_{i=1}^n (v - L_i)^+ + B_n$, where $|B_n| \le \frac{1}{p}|v_n - v|(|F_n(v_n) - F_n(v)| + |F_n(v_n) - F(v)|) \le \frac{1}{p}|v_n - v|(2|F_n(v_n) - F_n(v)| + |F_n(v) - F(v)|)$.

This result is based on derivations in Sun and Hong (2010). By definition, one obtains $c_n = v_n - \frac{1}{np}\sum_{i=1}^{n}(v_n - L_i)^+ = v - \frac{1}{np}\sum_{i=1}^{n}(v - L_i)^+ + (v_n - v) - \frac{1}{np}\sum_{i=1}^{n}((v_n - L_i)^+ - (v - L_i)^+)$. Hence, $B_n = (v_n - v) - \frac{1}{np}\sum_{i=1}^{n}((v_n - L_i)^+ - (v - L_i)^+) = (v_n - v) - \frac{1}{np}\sum_{i=1}^{n}((v_n - v)1_{\{L_i \le v_n\}}) - \frac{1}{np}\sum_{i=1}^{n}(v - L_i)(1_{\{L_i \le v_n\}} - 1_{\{L_i \le v\}})$. Since $(v_n - v) - \frac{1}{np}\sum_{i=1}^{n}((v_n - v)1_{\{L_i \le v_n\}}) = \frac{1}{p}(v_n - v)(F(v) - F_n(v_n))$ and $|\frac{1}{np}\sum_{i=1}^{n}(v - L_i)(1_{\{L_i \le v_n\}} - 1_{\{L_i \le v\}})| \le \frac{1}{p}|v_n - v||F_n(v_n) - F_n(v)|$, thus $|B_n| \le \frac{1}{p}|v_n - v|(|F_n(v_n) - F_n(v)| + |F_n(v_n) - F(v)|) \le \frac{1}{p}|v_n - v|(2|F_n(v_n) - F_n(v)| + |F_n(v) - F(v)|)$.

Step 2. We prove $E|B_n|^k = o(n^{-k/2-1/2+\epsilon})$ for $k > 0$. According to step 1, $|B_n|^k$ can be bounded by $\frac{1}{p^k}|v_n - v|^k|F_n(v_n) - F_n(v)|^k$ and $\frac{1}{p^k}|v_n - v|^k|F_n(v) - F(v)|^k$. We firstly consider the bound of $\frac{1}{p^k}|v_n - v|^k|F_n(v_n) - F_n(v)|^k$,

$$\begin{aligned} & E\frac{1}{p^k}|v_n - v|^k|F_n(v_n) - F_n(v)|^k \\ \le\ & \frac{1}{p^k}E(|\frac{1}{n}\sum_{i=1}^{n}|v_n - v|1_{\{v < L_i \le v_n\}}| + |\frac{1}{n}\sum_{i=1}^{n}|v_n - v|1_{\{v_n < L_i \le v\}}|)^k \end{aligned}$$

By Hölder's inequality, $E(|\frac{1}{n}\sum_{i=1}^{n}|v_n - v|1_{\{v<L_i\le v_n\}}|)^k \le E|v_n - v|^k 1_{\{v<L_1\le v_n\}} \le (E|v_n - v|^{k/(1-1/q)})^{1-1/q}(E1_{\{v\le L_1\le v_n\}})^{1/q}$. According to Theorem 4,we obtain $E|v_n - v|^t = O(n^{-t/2})$. Let $\theta = n^{-\frac{t}{2(t+1)}}$,

$$\begin{aligned} E1_{\{v\le L_1\le v_n\}} &= E1_{\{v<L_1\le v_n\}}1_{\{|v_n - v|\ge\theta\}} + E1_{\{v<L_1\le v_n\}}1_{\{|v_n-v|<\theta\}} \\ &\le E1_{\{|v_n - v|\ge\theta\}} + E1_{\{v<L_1\le v+\theta\}} \\ &\le E\frac{(v_n - v)^t}{\theta^t} + 2f(v)\theta = O(n^{-t/(2(t+1))}). \end{aligned}$$

Then one obtains $E(|\frac{1}{n}\sum_{i=1}^{n}|v_n - v|1_{\{v<L_i\le v_n\}}|)^k = (O(n^{-k/(2(1-1/q))})^{1-1/q}(O(n^{-t/(2(t+1))}))^{1/q} = O(n^{-k/2-1/2+\epsilon})$, as $q \to 1$. Result about $E(|\frac{1}{n}\sum_{i=1}^{n}|v_n - v|1_{\{v_n<L_i\le v\}}|)^k$ is exactly the same and hence: $E\frac{1}{p^k}|v_n - v|^k|F_n(v_n) - F_n(v)|^k = O(n^{-k/2-1/2+\epsilon})$.

Next, we consider the bound of $\frac{1}{p^k}|v_n - v|^k|F_n(v) - F(v)|^k$. Hölder's inequality implies that $E|v_n - v|^k|F_n(v) - F(v)|^k \le \sqrt{E|v_n - v|^{2k}E|F_n(v) - F(v)|^{2k}}$. We already show that $E|v_n - v|^{2k} = O(n^{-k})$. Let $X_i = 1_{\{L_i\le v\}} - F(v), i = 1, \cdots, n$, then $F_n(v) - F(v) = \frac{1}{n}\sum_{i=1}^{n}X_i$. Since $E|X_i|^t < \infty, \forall t \in [1, \infty)$, by lemma 6, we have $E|F_n(v) - F(v)|^{2k} = O(n^{-k})$. Hence one obtain $E\frac{1}{p^k}|v_n - v|^k|F_n(v) - F(v)|^k = O(n^{-k})$. Combining with the previous results, we obtian $E|B_n|^k = o(n^{-k/2-1/2+\epsilon})$.

Step 3. With the results in Step 1 and Step 2, we can finish the proof. For $m = 1$, one sees

$$\begin{aligned} Ec_n &= E(v - \frac{1}{np}\sum_{i=1}^{n}(v - L_i)^+ + B_n) \\ &= v - \frac{1}{p}E(v - L_1)^+ + EB_n = c + o(n^{-1/2+\epsilon}). \end{aligned}$$

For $m = 2$, denoting $A_n = \frac{1}{np}\sum_{i=1}^{n}(E(v - L_i)^+ - (v - L_i)^+)$, then $E(c_n - c)^2 = E(A_n + B_n)^2 = EA_n^2 + EB_n^2 + 2EA_nB_n$. Under Assumptions A, B, C, we directly compute $EA_n^2 = \frac{1}{n^2p^2}\Big(\sum_{i=1}^{n} var((v -$

$L_i)^+) + 2\sum_{i=1}^{n-1}(1-i/n)cov((v-L_1)^+,(v-L_{i+1})^+)\Big) = \frac{1}{p^2}\sigma_{n,c}^2 n^{-1} + O(n^{-2})$, where the last equality holds due to that $\sum_{i=1}^{n-1}\frac{i}{n}cov((v-L_1)^+,(v-L_{i+1})^+) = O(n^{-1})$ based on lemma 5. We also have $E(B_n)^2 = o(n^{-3/2+\epsilon})$ base on results in step 2. Noticing that $EA_nB_n \le \sqrt{EA_n^2EB_n^2} = o(n^{-1-1/4+\epsilon})$, then we have $E(c_n - c)^2 = \frac{\sigma_{n,c}^2}{p^2}n^{-1} + o(n^{-1-1/4+\epsilon})$, from which the conclusion follows. □

Proof of Theorem 1. By Theorem 4, we have $var(v_n) = \frac{\sigma_{n,v}^2}{nf(v)^2} + o(n^{-5/4})$ and $E(v_n) = v + o(n^{-3/4})$. According to the definition of relative error, one obtains

$$RE(v_n) = \sqrt{\frac{var(v_n)}{(Ev_n)^2}} = \sqrt{\frac{\frac{\sigma_{n,v}^2}{nf(v)^2} + o(n^{-5/4})}{(v + o(n^{-3/4}))^2}} = -\frac{\sigma_{n,v}}{vf(v)}n^{-1/2} + o(n^{-3/4}).$$

Similarly, by Theorem 5, we obtain $var(c_n) = \frac{\sigma_{n,c}^2}{np^2} + o(n^{-5/4+\epsilon})$ and $E(c_n) = c + o(n^{-3/4+\epsilon})$. Then, we have

$$RE(c_n) = \sqrt{\frac{var(c_n)}{(Ec_n)^2}} = \sqrt{\frac{\frac{\sigma_{n,c}^2}{np^2} + o(n^{-5/4+\epsilon})}{(c + o(n^{-3/4+\epsilon}))^2}} = -\frac{\sigma_{n,c}}{cp}n^{-1/2} + o(n^{-3/4+\epsilon}).$$

The proof for i.i.d samples is quite similar. □

EC.3. Proof of Theorem 2 and Theorem 3

Proof of Theorem 2. For part (i), according to Theorem 1, we have

$$\lim_{n\to\infty}\frac{RE(v_n(p))}{RE(c_n(p))} = \lim_{n\to\infty}\frac{cp\sigma_{n,v}}{vf(v)\sigma_{n,c}} = \lim_{n\to\infty}\frac{cp\sigma_{\infty,v}}{vf(v)\sigma_{\infty,c}} < \infty.$$

Due to lemma 5, we know that both $|\gamma_1(k)|$ and $|\gamma_2(k)|$ are bounded by some geometric series. Hence $\sigma_{n,v}$ and $\sigma_{n,c}$ converge to $\sigma_{\infty,v}$ and $\sigma_{\infty,c}$, respectively, where $\sigma_{\infty,v}^2 = \{p(1-p) + 2\sum_{k=1}^{\infty}\gamma_1(k)\}$, $\sigma_{\infty,c}^2 = \{var[(v-L_1)^+] + 2\sum_{k=1}^{\infty}\gamma_2(k)\}$.

Before the proof of part (ii), we state a useful result coming from Taylor (1952) in which L'Hospital's rule is extended to a general case which includes both $+\infty$ and $-\infty$.

LEMMA 7. *Let c and A be extended real numbers (i.e. real numbers, $+\infty$, and $-\infty$). Functions f and g are assumed to have first order derivative on an open interval $\mathcal{I}$ with an endpoint c. It is also assumed that $\lim_{x\to c}\frac{f'(x)}{g'(x)} = A$. If either $\lim_{x\to c}f(x) = \lim_{x\to c}g(x) = 0$ or $\lim_{x\to c}|g(x)| = \infty$, then: $\lim_{x\to c}\frac{f(x)}{g(x)} = A$.*

Then we come to the proof of part (ii). Denote $\rho := \lim_{x\to-\infty}h(x)$, where $\rho \in [-\infty, \infty]$. We first prove that $-1/2 \le \rho \le 0$. With the help of lemma 7, we have

$$\begin{aligned}\lim_{v\to-\infty}\frac{F(v)}{vf(v)} &= \lim_{v\to-\infty}\frac{vF(v)}{v^2f(v)}\\ &= \lim_{v\to-\infty}\frac{\int_{-\infty}^{v}xF(x)dx}{\int_{-\infty}^{v}x^2f(x)dx}\\ &= \lim_{v\to-\infty}\frac{1/2v^2F(v) - 1/2\int_{-\infty}^{v}x^2f(x)dx}{\int_{-\infty}^{v}x^2f(x)dx} \ge -\frac{1}{2}.\end{aligned}$$

Since $h(x) \leq 0$ as x is negative, we have $\lim_{v\to-\infty} h(v) \in [-1/2, 0]$. Furthermore, due to the existence of $\lim_{x\to-\infty} xh^{'}(x)$, with the above lemma for L'Hospital's rule, we can obtain $\lim_{v\to-\infty} vh^{'}(v) = \lim_{v\to-\infty} \frac{h(v)}{-log(-v)} = 0$.

According to Theorem 1, we only need to compute the ratio $\frac{c^2p^2\sigma_v^2}{v^2f(v)^2\sigma_c^2}$. Let K_1 denote $EL1_{\{L\leq v\}}$ and K_2 denote $EL^21_{\{L\leq v\}}$. Denote $\int_{-\infty}^{v} F(x)dx$ and $\int_{-\infty}^{v} xF(x)dx$ as T_1 and T_2. Then we have $K_1 = vp - T_1$ and $K_2 = v^2p - 2T_2$. The limit of relative errors' ratio is

$$\begin{aligned}
\lim_{n\to\infty}[\frac{RE(v_n(p))}{RE(c_n(p))}]^2 &= \frac{c^2p^2\sigma_v^2}{v^2f(v)^2\sigma_c^2} \\
&= \frac{p(1-p)K_1^2}{v^2f(v)^2[v^2p(1-p) - 2vK_1(1-p) + K_2 - K_1^2]} \\
&= \frac{K_1^2h(v)^2(1-p)/p}{K_1^2 - K_2 - v^2p(1-p) + 2v(1-p)K_1} \\
&= \frac{K_1^2h(v)^2(1-p)}{-(1-p)(K_1 - vp)^2 + (K_1^2 - pK_2)} \\
&= \frac{K_1^2h(v)^2(1-p)}{-(1-p)T_1^2 - [2p(vT_1 - T_2) - T_1^2]} \\
&= \frac{K_1^2h(v)^2(1-p)}{-pT_1^2 - 2p(vT_1 - T_2)}. \qquad \text{(ec.4)}
\end{aligned}$$

Next we apply lemma 7, the general L'Hospital's rule, to find asymptotic representations for T_1 and $vT_1 - T_2$. Note that

$$\begin{aligned}
\lim_{v\to-\infty} \frac{T_1}{vph(v)} &= \lim_{v\to-\infty} \frac{p}{ph(v) + vfh(v) + vh'(v)p} \\
&= \lim_{v\to-\infty} \frac{1}{h(v) - 1 + vh'(v)} = \frac{1}{\rho - 1}.
\end{aligned}$$

Thus, as $v \to -\infty$, there is an asymptotic representation $T_1 = \frac{1}{\rho-1}vph(v) + o(vph(v))$. Moreover, for $vT_1 - T_2$,

$$\begin{aligned}
\lim_{v\to-\infty} \frac{vT_1 - T_2}{v^2ph(v)^2} &= \lim_{v\to-\infty} \frac{T_1 + vp - vp}{2vph(v)^2 + v^2fh(v)^2 + 2v^2ph(v)h'(v)} \\
&= \lim_{v\to-\infty} \frac{\frac{1}{\rho-1}vph(v) + o(vph(v))}{2vph(v)\cdot h(v) + vph(v) + 2vph(v)\cdot vh'(v)} \\
&= \lim_{v\to-\infty} \frac{\frac{1}{\rho-1} + o(1)}{2h(v) + 1 + 2vh'(v)} = \frac{1}{(1+2\rho)(\rho-1)}.
\end{aligned}$$

Therefore $vT_1 - T_2 = \frac{1}{(1+2\rho)(\rho-1)}v^2ph(v)^2 + o(v^2ph(v)^2)$, as $v \to -\infty$.

With these two asymptotic results, it is possible to compute the limit of (ec.4) as follows

$$\begin{aligned}
&\lim_{v\to-\infty} \frac{K_1^2h(v)^2(1-p)}{-pT_1^2 + 2p(vT_1 - T_2)} \\
&= \lim_{v\to-\infty} \frac{(vp - T_1)^2h(v)^2(1-p)}{-pT_1^2 + 2p(vT_1 - T_2)}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{v\to-\infty} \frac{v^2p^2h(v)^2(1+\frac{1}{1-\rho}h(v)+o(h(v)))^2(1-p)}{-p(vph(v)+o(vph(v)))^2-2p(\frac{1}{(1+2\rho)(\rho-1)}v^2ph(v)^2+o(v^2ph(v)^2))}\\
&= \lim_{v\to-\infty} \frac{(1+\frac{1}{1-\rho}h(v)+o(h(v)))^2(1-p)}{-p(1+o(1))^2-2(\frac{1}{(1+2\rho)(\rho-1)}+o(1))}\\
&= \frac{1}{2}\cdot\frac{1+2\rho}{1-\rho}\le\frac{1}{2},
\end{aligned}
$$

where the last inequality holds due to the condition $\rho\le 0$. Thus there exists an v_0, such that for $\forall v<v_0$, $\frac{K_1^2h(v)^2(1-p)}{-pT_1^2+2p(vT_1-T_2)}\le\frac{1}{2}$ is satisfied. Furthermore, let $p_0=F(v_0)$, then for $\forall p<p_0$, we obtain $\lim_{n\to\infty}\frac{Re(v_n(p))}{Re(c_n(p))}\le\frac{1}{\sqrt{2}}$. □

Proof of Theorem 3. Notice the facts $m_n(p)=v_n(p/2)$ and $m(p)=v(p/2)$. Proof of part (1) is similar to the proof of part (i) in Theorem 2. For part (ii),

$$
\begin{aligned}
\lim_{n\to\infty}[\frac{Re(m_n(p))}{Re(c_n(p))}]^2 &= \lim_{n\to\infty}[\frac{RE(v_n(p/2))}{RE(c_n(p))}]^2=\frac{c^2p^2[\sigma_v(p/2)]^2}{v(p/2)^2f(v(p/2))^2\sigma_c^2}\\
&= \frac{2K_1^2h(v(p/2))^2(1-p/2)}{-pT_1^2-2p(vT_1-T_2)}.
\end{aligned}
$$

Using asymptotic representation $T_1=\frac{1}{\rho-1}vph(v)+o(vph(v))$ and $vT_1-T_2=\frac{1}{(1+2\rho)(\rho-1)}v^2ph(v)^2+o(v^2ph(v)^2)$ from Theorem 2, one gets

$$
\begin{aligned}
&\lim_{v\to-\infty}\frac{2K_1^2h(v(p/2))^2(1-p/2)}{-pT_1^2+2p(vT_1-T_2)}\\
&= \lim_{v\to-\infty}2\frac{(vp-T_1)^2h(v(p/2))^2(1-p/2)}{-pT_1^2+2p(vT_1-T_2)}\\
&= \lim_{v\to-\infty}2\frac{v^2p^2h(v(p/2))^2(1+\frac{1}{1-\rho}h(v)+o(h(v)))^2(1-p/2)}{-p(vph(v)+o(vph(v)))^2-2p(\frac{1}{(1+2\rho)(\rho-1)}v^2ph(v)^2+o(v^2ph(v)^2))}\\
&= \lim_{v\to-\infty}2\frac{(1+\frac{1}{1-\rho}h(v)+o(h(v)))^2(1-p/2)}{-p(1+o(1))^2-2(\frac{1}{(1+2\rho)(\rho-1)}+o(1))}\times\frac{h(v(p/2))^2}{h(v)^2}\\
&= \frac{1+2\rho}{1-\rho}\times\frac{(p/2)^2v^2f(v)^2}{p^2v(p/2)^2f(v(p/2))^2}\\
&= \frac{1+2\rho}{1-\rho}\times\frac{1}{4}\frac{\mathscr{L}(p)^2p^{2\theta}}{\mathscr{L}(p/2)^2(p/2)^{2\theta}}\le 1,
\end{aligned}
$$

where the last inequality holds due to the facts about slowly varying function $\lim_{p\to0}\mathscr{L}(p)/\mathscr{L}(p/2)=1$. Moreover, when $\rho\neq0$, this is an strict inequality. □