Stratified Importance Sampling for Bernoulli Mixture Model of Portfolio Credit Risk

Sunggon Kim

University of Seoul

May 28th,2021

Overview

Introduction

Credit Portfolio

- We consider a credit portfolio consisting of m obligors whose defaults during a fixed time interval $[0, T]$ are dependent.
- The default probability of obligor j is $\bar{p}_j$, $j = 1, 2, \ldots, m$.
- The exposure at the default of obligor j is c_j, $j = 1, 2, \ldots, m$.
- Let Y_j, $j = 1, 2, \ldots, m$, be the default indicator of obligor j in time interval $[0, T]$, i.e.

$$Y_j = \begin{cases} 1, & \text{obligor } j \text{ defaults in } [0, T], \\ 0, & \text{obligor } j \text{ does not default in } [0, T]. \end{cases}$$

- We call $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$ the default vector.
- The sample space of $\boldsymbol{Y}$ is $\{0, 1\}^m$, which is denoted by $\mathcal{Y}$.

Loss of the credit porfolio and tail loss probability

- The loss of the portfolio in time interval $[0, T]$ is given by

$$L(\boldsymbol{Y}) = \sum_{j=1}^{m} c_j Y_j, \quad \boldsymbol{Y} \in \mathcal{Y}. \tag{1}$$

- For a large value of $x > 0$, the tail loss probability over the theshold x is defined as

$$\theta = \Pr\{L(\boldsymbol{Y}) > x\} \tag{2}$$

- We propose an importance sampling scheme to estimate θ under the Bernoulli mixture of model for the dependence of $Y_1, \ldots, Y_m$.

Bernoulli mixture model for dependent defaults of obligors

Bernoulli mixture model

- The dependency of default events is modeled by introducing a number of d common factors $\mathbf{\Psi} = (\Psi_1, \ldots, \Psi_d) \in \mathbb{R}^d$.
- When the value of $\mathbf{\Psi} = \boldsymbol{\psi}$ is given,
  - Default indicators $Y_1, \ldots, Y_m$ are conditionally independent.
  - The conditional default probability of obligor j is detetmined by $\boldsymbol{\psi}$.
- Let $p_j(\psi)$ be the conditional default probability of obligor j. Then, the conditional joint pmf of the default vector $\boldsymbol{Y}$ given $\mathbf{\Psi} = \boldsymbol{\psi}$ is as follows: for $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathcal{Y}$,

$$p(\boldsymbol{y}|\boldsymbol{\psi}) = \prod_{j=1}^{m} p_j(\boldsymbol{\psi})^{y_j} (1 - p_j(\boldsymbol{\psi}))^{1-y_j}. \tag{3}$$

Nominal probability measure

- Let $\mathcal{S} = \mathbb{R}^d \times \mathcal{Y}$ be the sample space of $(\boldsymbol{\Psi}, \boldsymbol{Y})$.
- Define $\mathbb{P}$ as the probability measure on $\mathcal{S}$ induced from the pdf $f(\boldsymbol{\psi}; \boldsymbol{\mu})$ of $\boldsymbol{\Psi}$ and the conditional pmf $p(\boldsymbol{y}|\boldsymbol{\psi})$ of $\boldsymbol{Y}$ given $\boldsymbol{\Psi} = \boldsymbol{\psi}$.
- We call $\mathbb{P}$ the nominal probability measure.
- For an event A depending on $(\boldsymbol{\Psi}, \boldsymbol{Y})$, we denote by $\mathbb{P}(A)$ the probability of the occurrence of the event A.
- Let $E_{\mathbb{P}}[h(\boldsymbol{\Psi}, \boldsymbol{Y})]$ be the expectation of $h(\boldsymbol{\Psi}, \boldsymbol{Y})$ for a real valued function $h(\boldsymbol{\psi}, \boldsymbol{y})$ under the probability measure $\mathbb{P}$.
- Let L be the abbreviation of $L(\boldsymbol{Y})$. Then,

$$\theta = \mathbb{P}(L > x).$$

The proposed scheme

Stratifying the sample space

- Without loss of generality, we assume that $c_1, \ldots, c_m$ are in their descending order.
- We stratify the sample space $\mathcal{S}$ according to the defaults of the first K obligors.
- $\mathcal{S}$ is stratified into the strata $\{\mathcal{S}_{\boldsymbol{h}} = \mathbb{R}^d \times \mathcal{Y}_{\boldsymbol{h}}, \boldsymbol{h} \in \{0,1\}^K\}$, where

$$\mathcal{Y}_{\boldsymbol{h}} = \boldsymbol{h} \times \{0,1\}^{m-K}, \quad \boldsymbol{h} \in \{0,1\}^K. \tag{4}$$

- We call $\mathcal{S}_{\boldsymbol{h}}$ the stratum $\boldsymbol{h}$, $\boldsymbol{h} \in \{0,1\}^K$.
- Let $\bar{p}_{\boldsymbol{h}} = \mathbb{P}((\boldsymbol{\Psi}, \boldsymbol{Y}) \in \mathcal{S}_{\boldsymbol{h}})(= \mathbb{P}(\boldsymbol{Y} \in \mathcal{Y}_{\boldsymbol{h}}))$, $\boldsymbol{h} \in \{0,1\}^K$.

The norminal probability measure on $\mathcal{S}_{\boldsymbol{h}}$

- We denote by $f_{\boldsymbol{h}}(\psi)$ the conditional pdf of $\boldsymbol{\Psi}$ given $\boldsymbol{Y} \in \mathcal{Y}_{\boldsymbol{h}}$.
- From the Bayes' theorem, we have that for $\boldsymbol{h} \in \{0,1\}^K$,

$$f_{\boldsymbol{h}}(\boldsymbol{\psi};\boldsymbol{\mu}) = \frac{f(\boldsymbol{\psi};\boldsymbol{\mu}) \prod_{j=1}^{K} p_j(\boldsymbol{\psi})^{h_j} (1 - p_j(\boldsymbol{\psi}))^{1-h_j}}{\bar{p}_{\boldsymbol{h}}}, \quad \boldsymbol{\psi} \in \mathbb{R}^d. \quad (5)$$

- Since $Y_1, Y_2, \ldots, Y_m$ are independent given $\boldsymbol{\Psi} = \boldsymbol{\psi}$, the conditional pmf of $\boldsymbol{Y}$ given $\boldsymbol{\Psi} = \psi$ and $\boldsymbol{Y} \in \mathcal{Y}_{\boldsymbol{h}}$ is given by

$$p_{\boldsymbol{h}}(\boldsymbol{y}|\psi) = \prod_{j=K+1}^{m} p_j(\boldsymbol{\psi})^{y_j} (1 - p_j(\boldsymbol{\psi}))^{1-y_j}, \quad \boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{h}}. \quad (6)$$

- Let $\mathbb{P}_{\boldsymbol{h}}$ be the probability measure on stratum $\boldsymbol{h}$ induced by the pdf $f_{\boldsymbol{h}}(\psi)$ of $\boldsymbol{\Psi}$ and the conditional pmf $p_{\boldsymbol{h}}(\boldsymbol{y}|\psi)$ of $\boldsymbol{Y}$ given $\boldsymbol{\Psi} = \psi$.
- Then, $\mathbb{P}_{\boldsymbol{h}}$ is the conditional probability measure of $\mathbb{P}$ given $(\boldsymbol{\Psi}, \boldsymbol{Y}) \in$ stratum $\boldsymbol{h}$.

A representation of θ in terms of $\mathbb{P}_{\boldsymbol{h}}$

- The tail loss probability over the threshold x is represented as

$$\begin{aligned} \theta &= \sum_{\boldsymbol{h} \in \{0,1\}^K} \bar{p}_{\boldsymbol{h}} \mathbb{P}_{\boldsymbol{h}}(L > x) \\ &= \sum_{\boldsymbol{h} \in \{0,1\}^K} \bar{p}_{\boldsymbol{h}} E_{\mathbb{P}_{\boldsymbol{h}}}[I(L > x)]. \end{aligned} \tag{7}$$

Importance distributions on each stratum

- Let $g_{\boldsymbol{h}}(\boldsymbol{\psi})$, $\boldsymbol{h} \in \{0,1\}^K$, be an importance pdf of $\boldsymbol{\Psi}$ in stratum $\boldsymbol{h}$.
- Let $q_{\boldsymbol{h},j}(\boldsymbol{\psi})$, $j = K+1, \ldots, m$, be an conditional importance default probability of obligor j given $\boldsymbol{\Psi} = \boldsymbol{\psi}$ in stratum $\boldsymbol{h}$.
- Then, the conditional importance pmf $q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi})$ of $\boldsymbol{Y}$ given $\boldsymbol{\Psi} = \boldsymbol{\psi}$ in stratum $\mathcal{S}_{\boldsymbol{h}}$ has the following form:

$$q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi}) = \prod_{j=K+1}^{m} q_{h,j}(\boldsymbol{\psi})^{y_j} \left(1 - q_{h,j}(\boldsymbol{\psi})\right)^{1-y_j}, \quad \boldsymbol{y} \in \mathcal{Y}_{\boldsymbol{h}}. \tag{8}$$

- We define $\mathbb{Q}_{\boldsymbol{h}}$, $\boldsymbol{h} \in \{0,1\}^K$, as the probability measure on stratum $\mathcal{S}_{\boldsymbol{h}}$ induced by the pdf $g_{\boldsymbol{h}}(\boldsymbol{\psi})$ of $\boldsymbol{\Psi}$ and the conditional pmf $q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi})$ of $\boldsymbol{Y}$ given $\boldsymbol{\Psi} = \boldsymbol{\psi}$.

A representation of θ in terms of $\mathbb{Q}_{\boldsymbol{h}}$

- The likelihood ratio of an event $(\boldsymbol{\psi}, \boldsymbol{y})$ on $\mathcal{S}_{\boldsymbol{h}}$ is

$$w_{\boldsymbol{h}}(\boldsymbol{\psi}, \boldsymbol{y}) = \frac{f_{\boldsymbol{h}}(\boldsymbol{\psi}; \boldsymbol{\mu}) p_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi})}{g_{\boldsymbol{h}}(\boldsymbol{\psi}) q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi})}, \quad (\boldsymbol{\psi}, \boldsymbol{y}) \in \mathcal{S}_{\boldsymbol{h}}. \tag{9}$$

- Then, it follows that

$$E_{\mathbb{P}_{\boldsymbol{h}}}[I(L > x)] = E_{\mathbb{Q}_{\boldsymbol{h}}}[w_{\boldsymbol{h}}(\boldsymbol{\Psi}, \boldsymbol{Y}) I(L > x)], \quad \boldsymbol{h} \in \{0,1\}^K, \tag{10}$$

- By applying Eq. (10) to Eq. (7), we have that

$$\begin{aligned} \theta &= \sum_{\boldsymbol{h} \in \{0,1\}^K} \bar{p}_{\boldsymbol{h}} E_{\mathbb{P}_{\boldsymbol{h}}}[I(L > x)] \\ &= \sum_{\boldsymbol{h} \in \{0,1\}^K} \bar{p}_{\boldsymbol{h}} E_{\mathbb{Q}_{\boldsymbol{h}}}[w_{\boldsymbol{h}}(\boldsymbol{\Psi}, \boldsymbol{Y}) I(L > x)]. \end{aligned} \tag{11}$$

Stratifed importance sampling estimation of the tail loss probability

- Suppose that we have a number of $n_{\boldsymbol{h}}$ random samples $\boldsymbol{\Psi}^{(\boldsymbol{h},1)}, \ldots, \boldsymbol{\Psi}^{(\boldsymbol{h},n_{\boldsymbol{h}})}$ from the pdf $g_{\boldsymbol{h}}(\boldsymbol{\psi})$ for each $\boldsymbol{h} \in \{0,1\}^K$.
- For each $\boldsymbol{\Psi}^{(\boldsymbol{h},i)}$, we also have a sample $\boldsymbol{Y}^{(\boldsymbol{h},i)} \in \mathcal{Y}_{\boldsymbol{h}}$ from the conditional pmf $q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\Psi}^{(\boldsymbol{h},i)})$.
- An unbiased estimator of θ is given by

$$\hat{\theta} = \sum_{\boldsymbol{h}\in\{0,1\}^K} \frac{\bar{p}_{\boldsymbol{h}}}{n_{\boldsymbol{h}}} \sum_{i=1}^{n_{\boldsymbol{h}}} w_{\boldsymbol{h}}(\boldsymbol{\Psi}^{(\boldsymbol{h},i)}, \boldsymbol{Y}^{(\boldsymbol{h},i)}) I(L^{(\boldsymbol{h},i)} > x), \tag{12}$$

where $L^{(\boldsymbol{h},i)} = L(\boldsymbol{Y}^{(\boldsymbol{h},i)})$.

Neymann allocation

- Let $\sigma_{\boldsymbol{h}}^2 = V_{\mathbb{Q}_{\boldsymbol{h}}}[w_{\boldsymbol{h}}(\boldsymbol{\Psi}, \boldsymbol{Y}) I(L > x)]$, $\boldsymbol{h} \in \{0,1\}^K$. Then, the variance of the estimator $\hat{\theta}$ is given by

$$V[\hat{\theta}] = \sum_{\boldsymbol{h} \in \{0,1\}^K} \frac{\bar{p}_{\boldsymbol{h}}^2 \sigma_{\boldsymbol{h}}^2}{n_{\boldsymbol{h}}}. \tag{13}$$

- The Neymann allocation minimizes $V[\hat{\theta}]$, i.e. when the sum of $n_{\boldsymbol{h}}$'s are given to be n, the optimal value of $n_{\boldsymbol{h}}$ is

$$n_{\boldsymbol{h}} = n \frac{\bar{p}_{\boldsymbol{h}} \sigma_{\boldsymbol{h}}}{\sum_{\boldsymbol{h} \in \{0,1\}^K} \bar{p}_{\boldsymbol{h}} \sigma_{\boldsymbol{h}}}, \quad \in \{0,1\}^K. \tag{14}$$

- In order to allocate the number of samples to each of the stata according to the Neymann allocation, we need to compute the value of $\sigma_{\boldsymbol{h}}$, which is not an easy task. Instead, we estimate it by a pilot simulation.

Exponential twisting to choose the optimal conditional pmf of $\boldsymbol{Y}$ given $\boldsymbol{\Psi}$

- For $j > K$ and $\boldsymbol{\psi} \in \mathbb{R}^d$, we let $q_j(t, \boldsymbol{\psi})$ be the exponentially twisted default probability of $p_j(\boldsymbol{y}|\boldsymbol{\psi})$, i.e.

$$q_j(t, \boldsymbol{\psi}) = \frac{p_j(\boldsymbol{\psi}) \exp(c_j t)}{1 - p_j(\boldsymbol{\psi}) + p_j(\boldsymbol{\psi}) \exp(c_j t)}, \quad t \geq 0, \tag{15}$$

  where t is the parameter of the twisting.

- An asymptotically optimal value of t is as follows (Glassermann and Li (2005)): if $x > E_{\mathbb{P}_{\boldsymbol{h}}}[L|\boldsymbol{\psi}]$,

$$t_{\boldsymbol{h}}(\boldsymbol{\psi}) = \text{the solution of } \boldsymbol{h} \cdot (c_1, \ldots, c_K) + \sum_{j=K+1}^{m} c_j q_j(t, \boldsymbol{\psi}) = x,$$

  otherwise, $t_{\boldsymbol{h}}(\boldsymbol{\psi}) = 0$.

Cross entropy method to choose the optimal pdf of $\boldsymbol{\Psi}$

- The zero variance pdf of $\boldsymbol{\Psi}$ on $\mathcal{S}_{\boldsymbol{h}}$ is proportional to

$$E_{\mathbb{P}_{\boldsymbol{h}}}[I(L > y)|\boldsymbol{\psi}] f_{\boldsymbol{h}}(\boldsymbol{\psi}; \boldsymbol{\mu}).$$

- We confine the parametric family of the importance pdf of $\boldsymbol{\Psi}$ to

$$\mathcal{F} = \{f(\boldsymbol{\psi}; \boldsymbol{\nu})\}.$$

- The optimal value of $\boldsymbol{\nu}$ in stratum $\boldsymbol{h}$ can be found by solving the following maximization problem (Chan and Kroese (2010)):

$$\boldsymbol{\nu}_{\boldsymbol{h}}^{*} = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \int E_{\mathbb{P}_{\boldsymbol{h}}}[I(L > y)|\boldsymbol{\psi}] \log f(\boldsymbol{\psi}; \boldsymbol{\nu})\, f_{\boldsymbol{h}}(\boldsymbol{\psi}; \boldsymbol{\mu})\, d\boldsymbol{\psi}, \tag{16}$$

or equivalently,

$$\boldsymbol{\nu}_{\boldsymbol{h}}^{*} = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \int E_{\mathbb{Q}_{\boldsymbol{h}}}\left[\frac{p_{\boldsymbol{h}}(\boldsymbol{Y}|\boldsymbol{\psi})}{q_{\boldsymbol{h}}(\boldsymbol{Y}|\boldsymbol{\psi})} I(L > y)|\boldsymbol{\psi}\right] \log f(\boldsymbol{\psi}; \boldsymbol{\nu})\, f_{\boldsymbol{h}}(\boldsymbol{\psi}; \boldsymbol{\mu})\, d\boldsymbol{\psi}, \tag{17}$$

- Suppose that we have samples $(\boldsymbol{\Psi}^{(1)}, \boldsymbol{Y}^{(1)}), \ldots, (\boldsymbol{\Psi}^{(M)}, \boldsymbol{Y}^{(M)})$ in stratum $\boldsymbol{h}$ from the distribution $f_{\boldsymbol{h}}(\boldsymbol{\psi}; \boldsymbol{\mu}) q_{\boldsymbol{h}}(\boldsymbol{y}|\boldsymbol{\psi})$. Then, the stochastic optimization problem corresponding to Eq. (17) is that

$$\boldsymbol{\nu}_{\boldsymbol{h}}^{*} = \underset{\boldsymbol{\nu}}{\operatorname{argmax}} \frac{1}{M} \sum_{i=1}^{M} \frac{p_{\boldsymbol{h}}(\boldsymbol{Y}^{(i)}|\boldsymbol{\Psi}^{(i)})}{q_{\boldsymbol{h}}(\boldsymbol{Y}^{(i)}|\boldsymbol{\Psi}^{(i)})} I(L^{(i)} > x) \log f(\boldsymbol{\Psi}^{(i)}; \boldsymbol{\nu}). \quad (18)$$

Numerical Results

Simlulation setting

- The number of obligors, m, are set to be 10^3.
- The marginal default probabilies $\bar{p}_1, \ldots, \bar{p}_m$ are generated m independently from the uniform distribution on $[0, 0.02]$.
- Two sets of exposures with different tail behavior are used.
  - Case 1 exposures was generated from the Pareto distribution with shape parameter 0.8.
  - Case 2 exposures was generated from the Pareto distribution with shape parameter 1.2.
- For simplicity, the factor variable Ψ follows the standard normal distribution $N(0, 1)$.
- For the mixing distribution, we consider the probit-normal distribution, i.e. given $\Psi = \psi$, the conditional default probability $p_j(\psi)$ of obligor j has the following form:

$$p_j(\psi) = \Phi(a_j\psi + b_j) \quad j = 1, \ldots, m, \tag{19}$$

where Φ is the c.d.f. of the standard normal.

- For the simulations with the case 1 exposures, we set

$$(a_1, a_2, \ldots, a_{m-1}, a_m) = (0.5, -3, 0.5, -3, \ldots, 0.5, -3), \tag{20}$$

and for the simulations with the case 2 exposures, we set

$$(a_1, a_2, \ldots, a_{m-1}, a_m) = (1, -1.2, 1, -1.2, \ldots, 1, -1.2). \tag{21}$$

- By setting $b_j = \sqrt{1 + a_j^2}\Phi^{-1}(\bar{p}_j)$, $j = 1, \ldots, m$, the default probability of the obligor j is equal to is $\bar{p}_j$
- We estimate the tail loss propoability using the following simulation scheme:
  - Crude Monte Carlo simulation (CMC)
  - Two-step importance sampling (CE-ET) : the same simulation scheme as the proposed scheme, but in this scheme the sample space is not stratified.
  - The proposed scheme (SCE-ET)

Estimated tail loss probability over various thresholds

Table: Estimated tail loss probabilties for various values of thresholds

	$\alpha = 0.8$		$\alpha = 1.2$	
k	threshold	$\hat{\theta}$	threshold	$\hat{\theta}$
1	11137.59	$1.16 \cdot 10^{-2}$	587.68	$1.31 \cdot 10^{-2}$
2	12681.42	$4.53 \cdot 10^{-3}$	881.52	$4.54 \cdot 10^{-3}$
3	14114.97	$1.43 \cdot 10^{-3}$	1175.37	$1.58 \cdot 10^{-3}$
4	14886.88	$5.66 \cdot 10^{-4}$	1469.21	$4.73 \cdot 10^{-4}$
5	15658.80	$1.61 \cdot 10^{-4}$	1763.05	$1.03 \cdot 10^{-4}$

- For each estimation of θ, we generated $n = 10^6$ number of $(\boldsymbol{\Psi}, \boldsymbol{Y})'s$.
- $\hat{\theta}$ is the one estimated by SCE-ET.

Standard errors of the estimations and their time-variance measure for the set of $\alpha = 0.8$

	s.e			time*variance		
k	CMC	CE-ET	SCE-ET	CMC	CE-ET	SCE-ET
1	$1.06 \cdot 10^{-4}$	$2.30 \cdot 10^{-5}$	$1.57 \cdot 10^{-5}$	$1.21 \cdot 10^{-6}$	$2.73 \cdot 10^{-7}$	$4.81 \cdot 10^{-8}$
2	$6.73 \cdot 10^{-5}$	$1.10 \cdot 10^{-5}$	$6.59 \cdot 10^{-6}$	$4.65 \cdot 10^{-7}$	$5.20 \cdot 10^{-8}$	$2.06 \cdot 10^{-8}$
3	$3.81 \cdot 10^{-5}$	$3.59 \cdot 10^{-6}$	$1.62 \cdot 10^{-6}$	$1.46 \cdot 10^{-7}$	$6.58 \cdot 10^{-9}$	$1.44 \cdot 10^{-9}$
4	$2.25 \cdot 10^{-5}$	$9.55 \cdot 10^{-7}$	$1.08 \cdot 10^{-6}$	$5.08 \cdot 10^{-8}$	$5.78 \cdot 10^{-10}$	$9.02 \cdot 10^{-10}$
5	$1.33 \cdot 10^{-5}$	$3.79 \cdot 10^{-7}$	$2.24 \cdot 10^{-7}$	$1.77 \cdot 10^{-8}$	$9.04 \cdot 10^{-11}$	$3.91 \cdot 10^{-11}$

- The time*variances of CE-ET and SCE-ET are about 10 to 10^3 times less than those of CMC.
- CE-ET and SCE-ET are about 10 to 10^3 times faster than CMC in terms of simulation time to obtain the same estimation error.
- SCE-ET is about 2 to 5 times faster than CE-ET except the case of threshold being equal to x_4.

Standard errors of the estimations and their time-variance measure for the set of $\alpha = 1.2$

k	s.e			time*variance		
	CMC	CE-ET	SCE-ET	CMC	CE-ET	SCE-ET
1	$1.12{\cdot}10^{-4}$	$3.17{\cdot}10^{-5}$	$2.08{\cdot}10^{-5}$	$1.27{\cdot}10^{-6}$	$3.58{\cdot}10^{-7}$	$2.13{\cdot}10^{-7}$
2	$6.72{\cdot}10^{-5}$	$1.26{\cdot}10^{-5}$	$6.68{\cdot}10^{-6}$	$4.52{\cdot}10^{-7}$	$5.95{\cdot}10^{-8}$	$2.18{\cdot}10^{-8}$
3	$4.00{\cdot}10^{-5}$	$4.94{\cdot}10^{-6}$	$2.33{\cdot}10^{-6}$	$1.60{\cdot}10^{-7}$	$9.36{\cdot}10^{-9}$	$2.60{\cdot}10^{-9}$
4	$2.19{\cdot}10^{-5}$	$1.75{\cdot}10^{-6}$	$5.31{\cdot}10^{-7}$	$4.83{\cdot}10^{-8}$	$1.14{\cdot}10^{-9}$	$1.20{\cdot}10^{-10}$
5	$1.05{\cdot}10^{-5}$	$4.15{\cdot}10^{-7}$	$1.12{\cdot}10^{-7}$	$1.11{\cdot}10^{-8}$	$6.14{\cdot}10^{-11}$	$5.49{\cdot}10^{-12}$

- The simulation with exposures of $\alpha = 1.2$ shows the similar behavior of the time*variance to the simulation with expsoures of $\alpha = 0.8$
- CE-ET and SCE-ET have about 10 to 10^4 times less time*variances compared to CMC.
- SCE-ET method shows the better performance than CE-EC for all thresholds.

Conclusion

Conclusion

- We proposed an importance sampling scheme to estimate the tail loss probability over a threshold.
- In the proposed scheme, the sample space of the factor variables and the default vectors is stratified according to the defaults of some obligors with heavy exposures.
- We proposed to find the optimal importance distribution of the factor variables and the conditional default probabilities of obligors on each stratum of the sample space.
- Numerical study shows that the proposed scheme is efficient compared to the existing methods.