

Estimating Portfolio Loss Probabilities with Optimal Risk Loading Coefficients and Fixed Dependency among Obligor

Xi He, Ioannis Akrotirianakis, Amit Chakraborty
Siemens Corporate Research

August 20, 2015

Abstract

We consider the problem of measuring risk of a portfolio comprising loans, bonds, and financial instruments, which is caused by possible default of its obligors. Specifically, we are interested in estimating probability that a portfolio incurs large loss over a fixed time horizon. One crucial concern of such problem is how to measure dependency among all obligors. Our idea is based on so-called threshold model and normal copula model, which represent dependency by multiple risk factors and their corresponding loading coefficients. In this paper, we develop several novel regression models to determine the risk loading coefficients instead of exploring other copula models to measure dependency among all obligors. Several efficient algorithms are also reviewed or developed aiming to solve the proposed regression models. Using these generated optimal risk loading coefficients, we derive a more robust large loss probability distribution. Numerical experiments show that we can efficiently estimate the large loss probabilities while make use of dependency knowledge among all obligors as much as possible.

1 Introduction

As one of the major concern of financial institutions, to measure and manage risk of a portfolio which caused by default of its obligors have led to signif-

icant interests over past decades. Such problem is generally referred to as credit risk problem. In particular, we are interested in performance measures such as the probability that the portfolio incurs large loss over a fixed time horizon. One attractive and powerful model to address it is so-called threshold model. In this model, each obligor in a portfolio is assigned a default probability and a default loss as two constants during the same fixed time horizon. By introducing a latent variable for each obligor, we claim that one obligor defaults when its latent variable exceeds a given threshold uniquely determined by its default probability. We model total loss of a portfolio as the summation of its all obligors' loss and we intend to know the probability that this total loss is greater than a global user-defined threshold. In real world, we have to consider dependency among all obligors. A very popular structure that has been widely used to measure the dependency is often imposed on the multivariate distributions. Among all of these, multivariate normal distribution and its related normal copula model are the most popular choice. The model also forms the basis of the J.P.Morgan's CreditMetrics and other management system (see [Mor97, BOW10]). Within normal copula model, dependency is introduced by a set of common risk factors and some factor loadings coefficients, which can simultaneously affect all obligors. Afterwards, we can express large loss probability with dependent obligors by combining the models introduced above (see [GL05]). However, there is no analytic approach to generate this large loss probability. Consequently, Monte Carlo (MC) method are used a lot to estimate this value.

One common observation is that the default probability of each obligor are always very small. Therefore, if we make use of classical MC method to simulate the large loss event, definitely, we need to generate a huge number of scenarios, which is inefficient and computationally expensive. Therefore, to change probability distribution of default and further take advantages of likelihood technique, i.e., so-called Importance Sampling (IS) (see [RK11]) turns out to be a proper approach to increase probabilities of these rare events and furthermore enhance efficiency of traditional MC. Glasserman and Li ([GL05]) developed a two-steo IS procedure to efficiently estimate the large loss probability and further derived asymptotic property for single-factor homogeneous portfolio.

In this paper, we develop several regression models to find optimal risk loading coefficients while explore existing dependency knowledge as much as possible. Also, some efficient algorithms are reviewed or developed to address those models. We highlight main contributions of this paper as fol-

lowing. First, our regression models provide a novel idea on representing or generating extreme dependency among obligors. Instead of exploring other structure models, e.g. Student t -copula model, which is derived from multivariate Student t distribution (see [CK10, BJZ08]), to illustrate the implications of extreme dependency among obligors. We take benefits of normal copula model and try to make use of risk loading coefficients to represent extreme dependency among obligors. Second, our model can be adopted to address the situation when there exists (partial) dependency knowledge among obligors, which may be provided from correlation table from market analysis [Ank10].

The rest of this article is organized as follows. In Section 2, we state normal copula model and introduce IS procedure for estimating portfolio credit risk. Difference between our work and previous work are also discussed in this section. We then develop two regression models to compute optimal risk factor loading coefficients in Section 3, where we consider two cases that with or without correlation knowledge, respectively. In the following Section 4, we introduce some efficient algorithms to address those regression models appeared in previous section. Performance of these models when applying on MC or IS procedure are shown via multiple numerical experiments in Section 5. At the end, in Section 6, we state our conclusion and discuss some further related topic.

2 Preliminaries

Throughout the whole article, we consider a portfolio \mathcal{P} composed by m obligors and each has index k from set $\{1, \dots, m\}$. We also assign p_k as default marginal probability of obligor k and let c_k be single loss resulting from default of obligor k . Y_k is an $\{0, 1\}$ indicator to show whether obligor k defaults or not. We say $Y_k = 1$ if obligor k defaults and vice verse. Now, we can then express total loss L of portfolio \mathcal{P} as

$$L = c_1 Y_1 + \dots + c_m Y_m. \quad (1)$$

Afterwards, if we choose a large user-defined threshold x , we are then interested in finding the probability

$$P(L > x) \quad (2)$$

or large loss probability distribution with respect to multiple loss level.

One may note that if all obligors in portfolio \mathcal{P} are independent with each other, we can split this large loss L on to m different obligors and estimate default loss of each obligor totally separated, which is trivial. Hence, we are interested in how to capture dependency among obligors. A common used structure model is so-called Copula Model, like normal copula model (see [GHS99, GL05, Gla05]) and/or Student t -copula model (see [CK10, BJZ08]). In this paper, we will mainly focus on normal copula model.

2.1 Normal Copula Model

To represent dependency among obligors, we first introduce a multivariate normal vector (X_1, \dots, X_m) as a latent variable. Then we express each indicator of obligor k by

$$Y_k = \mathbf{1}\{X_k > x_k\}, k = 1, \dots, m, \quad (3)$$

with x_k uniquely determined by its default marginal probability p_k .

Furthermore, we let each X_k have a standard normal distribution and set $x_k = \Phi^{-1}(1 - p_k) = -\Phi^{-1}(p_k)$, where $\Phi(\cdot)$ is cumulative normal distribution. Thus,

$$P(Y_k = 1) = P(X_k > x_k) = p_k. \quad (4)$$

Under this construction, dependency among obligors is determined by correlation among X_k s. Given normal copula model framework, we can express X_k through a factor model of the form

$$X_k = a_{k1}Z_1 + \dots + a_{kd}Z_d + b_k\epsilon_k. \quad (5)$$

Here d stands for the number of risk factors affecting all X_k s. Z_1, \dots, Z_d are called systematic risk factors and all have $\mathcal{N}(0, 1)$ standard norm distribution. We assume all these risk factors are independent with each other [GL05, BOW10]. One may note that those systematic risk factors are common among all obligors. Except this, for each obligor k , we denote ϵ_k as an idiosyncratic risk factor and also has $\mathcal{N}(0, 1)$ standard norm distribution. Respectively, $a_{k1}, \dots, a_{kd}, b_k$ with respect to each obligor k are called factor loading coefficients of obligor k . Within normal copula model, we expect X_k to have standard normal distribution as well, therefore, we force these loading coefficients of obligor k to satisfy

$$a_{k1}^2 + \dots + a_{kd}^2 + b_k^2 = 1, k = 1, \dots, m. \quad (6)$$

To make notation concisely, we let all a_{kj} , $k = 1, \dots, m$, $j = 1, \dots, d$ together consist of a matrix $A \in \mathbb{R}^{m \times d}$ and all b_k , $k = 1, \dots, m$ together consist of a vector $b \in \mathbb{R}^m$. We name A as loading matrix and b as loading vector of the portfolio \mathcal{P} , respectively.

Under this construction, we can derive that correlation between X_k and X_j is given by $a_k^T a_j$, where $k \neq j$. Note in previous work (see [GHS99, GL05, Gla05]), they assume that all loading coefficients are positive. This assumption guarantees that larger values of the factors Z_j , $j = 1, \dots, d$ lead to a larger number of default and correlation among all obligors are always positive. However, in our work, we don't assume this. In other words, we admit existence of negative correlation. It indicates that obligors maybe negatively correlated, which is realistic and important (e.g. see [EMS02, DMV09]).

Since we drop off the positive correlation assumption, which is different from previous work, therefore, given risk factor $Z = (Z_1, \dots, Z_d)^T$, the conditional default probability for the k th obligor becomes

$$\begin{aligned} p_k(Z) &= P(Y_k = 1|Z) \\ &= P(X_k > x_k|Z) = P(a_k^T Z + b_k \epsilon_k > x_k|Z) \\ &= \begin{cases} \Phi\left(\frac{a_k^T Z - x_k}{b_k}\right), & b_k > 0, \\ 1 - \Phi\left(\frac{a_k^T Z - x_k}{b_k}\right), & b_k < 0. \end{cases} \end{aligned}$$

Note that conditional on risk factors Z , we can specifically derive the default probabilities of each obligors. In other words, conditional on risk factors Z , the default event among all obligors are independent with each other. This is important to develop following efficient IS procedure to estimate large loss probability.

2.2 Importance Sampling for Portfolio Credit Risk

Important sampling (IS) is an efficient technique to estimate large loss probability $P(L > x)$, since it reduce variance a lot compared with naive MC approach, especially at large value x , see [GL05]. When all obligors are independent with each other, one common approach is to increase default probabilities with respect to each obligor. Replacing each default probability

p_k to a larger value q_k , we can derive the elementary IS formula

$$P(L > x) = \mathbf{E}_p[\mathbf{1}\{L > x\}] = \mathbf{E}_q[\mathbf{1}\{L > x\} \prod_{k=1}^m \left(\frac{p_k}{q_k}\right)^{Y_k} \left(\frac{1-p_k}{1-q_k}\right)^{1-Y_k}], \quad (7)$$

where the product term

$$\prod_{k=1}^m \left(\frac{p_k}{q_k}\right)^{Y_k} \left(\frac{1-p_k}{1-q_k}\right)^{1-Y_k}$$

is likelihood ratio relating the original distribution of (Y_1, \dots, Y_m) to the new one $(\hat{Y}_1, \dots, \hat{Y}_m)$, where $P(\hat{Y}_k = 1) = q_k \geq p_k$.

In practical, exponential twisting is used to generate q_k , where exponential twisting is introduced by

$$p_{k,\theta} = \frac{p_k e^{\theta c_k}}{p_k e^{\theta c_k} + 1 - p_k}, \quad (8)$$

in which θ is a properly determined parameter so that $q_k = p_{k,\theta}$. Note that under the exponential twisting, we have

$$\left(\frac{p_k}{p_{k,\theta}}\right)^{Y_k} \left(\frac{1-p_k}{1-p_{k,\theta}}\right)^{1-Y_k} = \exp(-\theta L + \phi(\theta)), \quad (9)$$

where

$$\phi(\theta) = \log \mathbf{E}[e^{\theta L}] = \sum_{k=1}^m \log(1 + p_k(e^{\theta c_k} - 1)). \quad (10)$$

Therefore, we derived a unbiased estimator

$$\hat{p}_x = \mathbf{1}\{L > x\} \exp(-\theta L + \phi(\theta)), \quad (11)$$

which means $P(L > x) = \mathbf{E}_\theta[\hat{p}_x]$.

To produce the best θ , we try to minimize variance, equivalently, the second moment of this estimator, where the second moment is given by

$$\begin{aligned} \mathbf{E}_\theta[\hat{p}_x^2] &= \mathbf{E}_\theta[\mathbf{1}\{L > x\} \exp(-2\theta L + 2\phi(\theta))] \\ &\leq \exp(-2\theta x + 2\phi(\theta)). \end{aligned} \quad (12)$$

Hence, we can find an upper bound of the second moment by minimizing the right hand side, where the minimum is attained at

$$\theta_x \begin{cases} \in \{\theta | \phi'(\theta) = x\}, & x > \phi'(0), \\ = 0, & x \leq \phi'(0). \end{cases} \quad (13)$$

Note that if we set $\theta_x = 0$, it means that we do not need to twist probability since the default event of obligor k is not rare.

So far, using this twist-probability approach, we can easily derive the large loss probability of independent obligors (we will further model this special case in Section 3.3). Moreover, as per dependent obligors, a two-step IS procedure is considered. The motivation is based on the following decomposition, called law of total variance, i.e., for any estimator \hat{p}_x of $P(L > x)$, we have

$$\text{Var}[\hat{p}_x] = \mathbf{E}[\text{Var}[\hat{p}_x|Z]] + \text{Var}[\mathbf{E}[\hat{p}_x|Z]]. \quad (14)$$

Actually, we want to reduce variance of the estimator \hat{p}_x , i.e., minimize $\text{Var}[\hat{p}_x]$. As we discussed in Section 2.1, conditional on Z , the obligors are independent, so $\text{Var}[\hat{p}_x|Z]$ is small if we choose θ_x according to (13). Therefore, we focus on the second term, where we have $\mathbf{E}[\hat{p}_x|Z] = P(L > x|Z)$. To minimize its variance, [GHS99] suggested to find the optimal solution of

$$\max_z P(L > x|Z = z) \exp(-z^T z/2), \quad (15)$$

because the solution can be used as the mean of approximating normal distribution for risk vectors $Z = (Z_1, \dots, Z_d)^T$.

Note that to exactly find the maximum is difficult since it's hard to derive $P(L > x|Z = z)$. Several well-defined approximations of $P(L > x|Z = z)$ are introduced (see [GL05] and its reference therein). Here in this paper, we adopt one approximation as following

$$\max_z F_x(z) - \frac{1}{2} z^T z, \quad (16)$$

where $F_x(z) = -\theta_x(z)x + \phi(\theta_x(z), z)$ and $\theta_x(z)$ is defined in (13).

We denote maximum of (16) by μ and then an efficient two-step IS procedure can be stated as

1. Sample Z from $\mathcal{N}(\mu, \mathbf{I})$.

2. Conditional on Z , generate the exponential twisting parameter $\theta_x(Z)$ from (13).
3. Return the estimator \hat{p}_x as

$$\mathbf{1}\{L > x\} \exp(-\theta_x(Z)L + \phi(\theta_x(Z), Z)) \cdot \exp(-\mu^T Z + \mu^T \mu/2). \quad (17)$$

The last term $\exp(-\mu^T Z + \mu^T \mu/2)$ is likelihood ratio relating density of $\mathcal{N}(0, \mathbf{I})$ distribution to $\mathcal{N}(\mu, \mathbf{I})$ distribution.

3 Optimal Factor Loading Coefficients

3.1 Motivation and Fundamental Setting

In previous work (see [GHS99, GL05, Gla05, CK10, BJZ08] etc.), systematic risk factor loading matrix $A \in \mathbb{R}^{m \times d}$ and idiosyncratic risk factor loading vector $b \in \mathbb{R}^m$ are either generated randomly or specified on purpose, which may not accurately reflect real situation in practical. In this section, we propose a regression model to (approximately) determine loading matrix A and loading vector b . Two benefits of this model is stated as follows. First, we intend to find proper A and b to build the probably strongest potential correlation among all obligors. Meanwhile, due to this loading A and b , we will increase probability of large loss and decrease the number of scenarios we need to generate for either MC method or IS procedure.

We assume that each obligor has a positive default probability. Based on this assumption and within normal copula model, for each obligor k , there will exist at least one factor loading $Z \in \mathbb{R}^d$ and $\epsilon_k \in R$, such that

$$X_k = a_k^T Z + \epsilon_k b_k \geq x_k. \quad (18)$$

Now, we assume that the loading coefficients A and b are known. Furthermore, if we generate N scenarios, i.e., N risk loading factors Z^1, \dots, Z^N s and $\epsilon_k^1, \dots, \epsilon_k^N$ for obligor k , where N is a sufficiently large number and all scenarios are generated from normal distribution. Thus, we're able to claim that there should be roughly $n_k = \lceil p_k N \rceil$ scenarios lead to default of obligor k . That's to say

$$a_k^T Z^i + b_k \epsilon_k^i \geq x_k, \quad i = 1, \dots, n_k. \quad (19)$$

We denote $\underline{N} = \lceil \min_k p_k N \rceil$, $\overline{N} = \lceil \max_k p_k N \rceil$ and thus we have

$$\underline{N} \leq n_k \leq \overline{N}, \quad k = 1, \dots, m. \quad (20)$$

Our idea is then to randomly generate n risk loading factor Z^1, \dots, Z^n so that $a_k^T Z^i + \epsilon_k^i b_k \geq x_k$ for all $k = 1, \dots, m$, where $i = 1, \dots, n$. Obviously, we should choose $n \in (\underline{N}, \overline{N})$. We intend to find proper risk loading coefficients A and b for all obligors, which make (19) true.

We randomly sampling risk loading factor $Z^i = [Z_1^i \ Z_2^i \ \dots \ Z_d^i]^T$, $i = 1, \dots, n$ from multi-normal distribution $\mathcal{N}(0, I)$ independently, and let

$$\mathbf{Z} = [Z^1 \ Z^2 \ \dots, Z^n]^T \in \mathbb{R}^{n \times d}. \quad (21)$$

Furthermore, with respect to obligor k , we randomly generate ϵ_k^i within standard normal distribution $\mathcal{N}(0, 1)$ and let

$$\epsilon^i = [\epsilon_1^i \ \epsilon_2^i \ \dots \ \epsilon_m^i]^T \in \mathbb{R}^m, \quad \epsilon_k = [\epsilon_k^1 \ \epsilon_k^2 \ \dots \ \epsilon_k^n]^T \in \mathbb{R}^n. \quad (22)$$

Note that for given i th sampling, i.e., Z^i and ϵ^i , the total loss of this portfolio \mathcal{P} is then determined.

If we force that all these risk factors will lead default of all m obligors, one straightforward model can be stated as follows,

$$\begin{cases} a_k^T Z^i + b_k \epsilon_k^i \geq x_k, & i = 1, \dots, n, \ k = 1, \dots, m, \\ a_k^T a_k + b_k^2 = 1, & k = 1, \dots, m, \end{cases} \quad (23)$$

where the equalities comes from (6).

It would be perfect if we can find a solution of (23), which means all obligors default with these n risk loading factor Z^i , ϵ^i , $i = 1, \dots, n$. Unfortunately, (23) are generally infeasible. For example, for specific i, k , by considering $a_k \leq \mathbf{e}$, where let $\mathbf{e} \in \mathbb{R}^d$ be all one vector, and $b_k \leq 1$, we have $a_k^T Z^i + b_k \epsilon_k^i \leq \|Z^i\|_1 + |\epsilon_k^i|$. Thus, if we have $x_k = \Phi^{-1}(1 - p_k) > \|Z^i\|_1 + |\epsilon_k^i|$, the system (23) is then insolvable.

To solve this conflict, we propose a least square regression model, whose purpose is to make $a_k^T Z^i + b_k \epsilon_k^i$ close to x_k as much as possible. Note that we may or may not have previous knowledge about correlation among all obligors. Hence, two models are introduced as follows. For the first model, we assume that there is no correlation information among obligors. For the second one, we assume that we know partial or even complete correlation information. Note that in second model, we include totally independent obligors as a special case.

3.2 Regression Model without Correlation Knowledge

In this section, we don't assume any correlation information with respect to all obligors (i.e, $a_k^T a_j$ can be any value). It is then possible to derive risk loading coefficients of each obligor separately. Actually, we are trying to find out an appropriate loading a_k and b_k from a regression problem as

$$\begin{aligned} \min_{a_k, b_k} \quad & \left\| \begin{bmatrix} \mathbf{Z} & \epsilon_k \end{bmatrix} \begin{bmatrix} a_k \\ b_k \end{bmatrix} - x_k \mathbf{e} \right\|_2^2 \\ \text{s.t.} \quad & \left\| \begin{bmatrix} a_k \\ b_k \end{bmatrix} \right\|_2^2 = 1, \end{aligned} \tag{24}$$

where $\mathbf{e} \in \mathbb{R}^n$ is an all one vector and \mathbf{Z}, ϵ_k are defined before. Note that equality constraints comes from (6). Different with (23), in this model, the problem is always feasible, thus we can try to find an optimal solution. The related algorithms to solve this model will be followed in later sections.

3.3 Regression Model with Correlation Knowledge

In this section, we assume that we have some (may not completely) correlation information among obligors. Let $\mathcal{C} \in \mathbb{S}^{m \times m}$ be the correlation matrix among obligors, i.e., $\mathcal{C}_{k,j} = \mathcal{C}_{j,k} = a_k^T a_j$, $k \neq j$. Here we assume that the specific correlation matrix \mathcal{C} is a possible correlation matrix, that is, \mathcal{C} has to be positive semidefinite. In the following content, we then can do Cholesky decomposition to matrix \mathcal{C} directly without bothering whether \mathcal{C} is positive semidefinite or not. Since if \mathcal{C} is not positive semidefinite, its Cholesky decomposition will fail.

Let $\hat{A} = [A \quad \text{diag}(b)]^T \in \mathbb{R}^{(d+m) \times m}$, $\hat{x} = [x_1 \mathbf{e} \quad \dots \quad x_m \mathbf{e}] \in \mathbb{R}^{n \times m}$ and denote $[\mathbf{Z} \quad \epsilon_1 \quad \dots \quad \epsilon_m]$ by $D \in \mathbb{R}^{n \times (d+m)}$. If we (partly) know the correlation information among obligors, we can develop following relaxed model

$$\begin{aligned} \min_{\hat{A} \in \mathbb{R}^{(d+m) \times m}} \quad & \|D\hat{A} - \hat{x}\|_F^2 \\ \text{s.t.} \quad & \text{diag}(\hat{A}^T \hat{A}) = \mathbf{I} \\ & (\hat{A}^T \hat{A})_{k,j} = \mathcal{C}_{k,j}, (k, j) \in \mathcal{I}, \end{aligned} \tag{25}$$

where $\|\cdot\|_F$ is Frobenius norm and \mathcal{I} is set of known correlation pairs. We state (25) as a relaxed model since we relax matrix variable \hat{A} to be arbitrary

matrix in $\mathbb{R}^{(d+m) \times m}$ while we originally require \hat{A} have the following form

$$\hat{A} = [A \quad \text{diag}(b)]^T \in \mathbb{R}^{(d+m) \times m}. \quad (26)$$

Note that (25) is in general hard to solve, we are going to discuss two trivial cases.

Case 1.

If all obligors are independent with each other, which means $a_k^T a_j = 0$ for all $k \neq j$, i.e., $\mathcal{C}_{k,j} = 0$, for all $k \neq j$. Therefore, In this case, the following regression model is proposed

$$\begin{aligned} \min_{\hat{A} \in \mathbb{R}^{(d+m) \times m}} \quad & \|D\hat{A} - \hat{x}\|_F^2 \\ \text{s.t.} \quad & \hat{A}^T \hat{A} = \mathbf{I}. \end{aligned} \quad (27)$$

Note that (27) is usually refereed as Quadratic Matrix Programming (QMP), which has been studied for decades (see [Bec07]). Specifically, it can be reformulated as a special case of Procrustes problem on the Stiefel manifold [EP99]. Note that (27) is an unbalanced problem since $\hat{A} \in \mathbb{R}^{(d+m) \times m}$ is not square, which is relatively hard to solve [EP99].

Case 2.

If we know complete correlation information among all obligors, that is, we know \mathcal{C} (Note \mathcal{C} should be a positive semidefinite matrix). We can then combine two constraints in (25) together by setting the main diagonal elements of \mathcal{C} all equal to one as $\hat{\mathcal{C}}$, see Figure 1 as following

PIN Portfolio	December 2001 - October 2009							
	Cash or Equivalent	Fund 1	Fund 2	Fund 3	Fund 4	Fund 5	Fund 6	Fund 7
Cash or Equivalent	1.00	-0.18	-0.08	-0.03	0.07	-0.15	-0.09	-0.09
Fund 1	-0.18	1.00	0.58	0.42	0.63	0.65	0.64	0.44
Fund 2	-0.08	0.58	1.00	0.19	0.41	0.58	0.54	0.34
Fund 3	-0.03	0.42	0.19	1.00	0.55	0.33	0.40	0.82
Fund 4	0.07	0.63	0.41	0.55	1.00	0.50	0.70	0.48
Fund 5	-0.15	0.65	0.58	0.33	0.50	1.00	0.52	0.47
Fund 6	-0.09	0.64	0.54	0.40	0.70	0.52	1.00	0.42
Fund 7	-0.09	0.44	0.34	0.82	0.48	0.47	0.42	1.00
AVERAGE CORRELATION = 0.5								

Figure 1: Correlation Matrix [Ank10]

In this case, we have the following regression model

$$\begin{aligned} \min_{\hat{A} \in \mathbb{R}^{(d+m) \times m}} \quad & \|D\hat{A} - \hat{x}\|_F^2 \\ \text{s.t.} \quad & \hat{A}^T \hat{A} = \hat{\mathcal{C}}. \end{aligned} \quad (28)$$

If furthermore, $\hat{\mathcal{C}}$ is positive definite, we can do Cholesky decomposition to $\hat{\mathcal{C}}$, i.e., $\hat{\mathcal{C}} = E^T E$, where E is inevitable lower triangular matrix. Let $\hat{A}_0 = \hat{A}E^{-1}$, we can reformulate the equality constraint in (28) as

$$\hat{A}_0^T \hat{A}_0 = (\hat{A}E^{-1})^T (\hat{A}E^{-1}) = \mathbf{I}. \quad (29)$$

Furthermore, we can reformulate (28) as following

$$\begin{aligned} \min_{\hat{A}_0 \in \mathbb{R}^{(d+m) \times m}} \quad & \|D\hat{A}_0 E - \hat{x}\|_F^2 \\ \text{s.t.} \quad & \hat{A}_0^T \hat{A}_0 = \mathbf{I}. \end{aligned} \quad (30)$$

This problem is called Weighted Orthogonal Procrustes Problem (WOPP) from [ZD06, FB12]. Since $\hat{A}_0 \in \mathbb{R}^{(d+m) \times m}$, this model is also treated as unbalanced. Note if $E = \mathbf{I}$, (30) will degenerate to (27). As far as we know, there is not closed-form solution to this unbalanced problem (30), even if $E = \mathbf{I}$, i.e. (27). Algorithms to solve such problems are discussed in [ZD06, FB12] and reference therein, however, we are not going to discuss these algorithms in this paper.

4 Optimization Approachs to Solve the Regression Model

Denote $\begin{bmatrix} \mathbf{Z} & \epsilon_k \end{bmatrix}$ by $F_k \in \mathbb{R}^{n \times (d+1)}$ and $\begin{bmatrix} a_k \\ b_k \end{bmatrix}$ by $l_k \in \mathbb{R}^{d+1}$, we can rewrite the problem as

$$\begin{aligned} \min_{l_k \in \mathbb{R}^{d+1}} \quad & \|F_k l_k - x_k \mathbf{e}\|_2^2 \\ \text{s.t.} \quad & \|l_k\|_2^2 = 1. \end{aligned} \quad (31)$$

One may note that this problem (31) is a QCQP (Quadratically Constrained Quadratic Programming) and it's nonconvex. As we known, in

general QCQP is NP-hard. Specifically, this kind of problem is often named as least squares problem with a quadratic equality constraint (LSQE)(see [ZH03]). And it can be treated as a special case in [EP99], by setting $m = 1$. Similar problem can be found in many important applications, (e.g. the Karush-Kuhn-Tucker condition for the quadratic inequality constrained least squares problem or using some regularization techniques to solve ill-posed least squares problem).

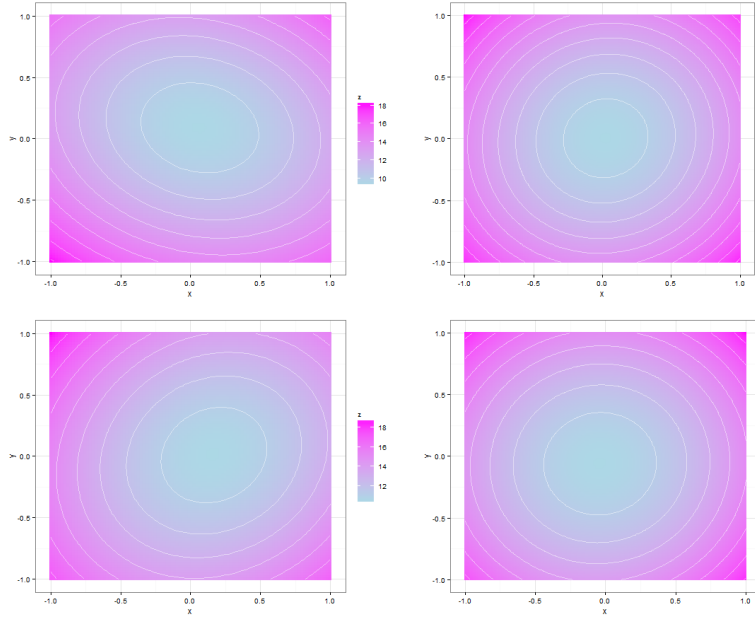


Figure 2: Objective Function Contour of (31)

Multiple algorithms are proposed to solve such kind of problems, we are going to discuss two kind of widely used algorithms as following.

4.1 SDP (Semi-definite Programming) Relaxation

One popular technique to solve (31) is to reformulate it as a SDP, which makes this QCQP problem tractable. we first derive its SDP by introducing a matrix variable $L_k = l_k l_k^T$, which implies that $(L_k)_{ij} = (l_k)_i (l_k)_j$. Moreover, we know that L_k is a symmetric positive semi-definite matrix and has rank

one. We then reformulate (31) as follows

$$\begin{aligned}
& \min_{L_k \in \mathbb{S}^{(d+1) \times (d+1)}, l_k \in \mathbb{R}^{d+1}} (F_k^T F_k) \bullet L_k - (2x_k \mathbf{e}^T) l_k + nx_k^2 \\
& \text{s.t.} \quad \mathbf{I} \bullet L_k = 1 \\
& \quad L_k - l_k l_k^T \succeq 0 \\
& \quad \text{rank}(L_k) = 1,
\end{aligned} \tag{32}$$

where $A \bullet B = \text{tr}(AB)$. It can be easily verify that (31) is equivalent to (32). To homogenate the objective function of (32), we can introduce new variables as

$$\hat{F}_k = \begin{bmatrix} nx_k^2 & x_k \mathbf{e}^T F_k \\ x_k F_k^T \mathbf{e} & F_k^T F_k \end{bmatrix}, \hat{I}_1 = \begin{bmatrix} 0 & 0^T \\ 0 & \mathbf{I} \end{bmatrix} \text{ and } \hat{I}_2 = \begin{bmatrix} 1 & 0^T \\ 0 & \mathbf{0} \end{bmatrix} \tag{33}$$

and set

$$\hat{L}_k = \begin{bmatrix} 1 & l_k^T \\ l_k & L_k \end{bmatrix}. \tag{34}$$

Consequently, we know $\hat{L}_k \succeq 0$ since we have $L_k - l_k l_k^T \succeq 0$ in (32). Note that so far we are keep reformulating our original problem (31) without reducing its difficult to be solved. By further dropping the only nonconvex rank-one constraint of (32), we can relax it as

$$\begin{aligned}
& \min_{L_k \in \mathbb{S}^{(d+1) \times (d+1)}} q(\hat{L}_k) := \hat{F}_k \bullet \hat{L}_k \\
& \text{s.t.} \quad \hat{I}_i \bullet \hat{L}_k = 1, i = 1, 2 \\
& \quad \text{diag}(\hat{L}_k) \leq 1 \\
& \quad \hat{L}_k \succeq 0,
\end{aligned} \tag{35}$$

where $\text{diag}(\hat{L}_k) \leq 1$ is a redundant constraint (since we have $l_k \in [-\mathbf{e}, \mathbf{e}]$ from (6)) and we add this dumb constraint to make it consistent with the one discussed in [Ye99]. Note that (35) is a convex SDP problem and is always strictly feasible, e.g.

$$\tilde{L}_k = \text{diag}(1, \frac{1}{\sqrt{d}}, \dots, \frac{1}{\sqrt{d}}) \succ 0 \tag{36}$$

is a strictly feasible solution.

As we know, SDP relaxation problems can be solved conveniently by standard convex optimization toolboxes, e.g. *SDPT3* [TTT99] and *Rcsdp* [Bra14]. We denote \hat{L}_k^* by optimal solution of (35). However, unfortunately, we can't generate optimal solution for original problem (31) from \hat{L}_k^* in general since \hat{L}_k^* may not be a rank one matrix. When $\text{rank}(\hat{L}_k^*) \neq 1$, we need adopt some approaches to recovery a feasible good solution \hat{l}_k^* from \hat{L}_k^* . There are multiple methods to achieve this goal. Here we use the following approach (see [Ye99]) to recovery approximated optimal solution \hat{l}_k^* of (31).

Note that \hat{L}_k^* is positive semi-definite from (35), we can find a factorization matrix $V \in \mathbb{R}^{(d+1) \times (d+1)}$ such that $\hat{L}_k^* = V^T V$ (Cholesky Decomposition). We denote $v_i, i = 1, \dots, d+1$ by each column of V , separately. Afterwards, we can generate a random vector u uniformly distributed on the $(d+1)$ dimensional unit ball and let

$$\hat{l}_k^* = D \text{sgn}(V^T u), \quad (37)$$

where

$$D = \text{diag}(\|v_1\|, \dots, \|v_{d+1}\|), \quad (38)$$

and $\text{sgn}(x)$ represents sign function of vector x . By this approach, optimality analysis follows in Theorem 1 derived from [Ye99].

Theorem 1. \hat{l}_k^* that derived from (37) is a feasible solution of (31) and the expected objective value $\mathbf{E}_u(q(\hat{L}_k^*))$ satisfies that

$$\frac{\bar{q} - \mathbf{E}_u(q(\hat{L}_k^*))}{\bar{q} - \underline{q}} \leq \frac{\pi}{2} - 1 \leq \frac{4}{7}, \quad (39)$$

where \bar{q} and \underline{q} stand for the maximal and minimal objective values, respectively.

4.2 Projection Method

The other efficient approach to solve (31) is to explore a projection method in [ZH03]. Instead of solving (31) directly or using SDP relaxation, we can solve its KKT conditions,

$$\begin{cases} 2(F_k^T F_k + \lambda \mathbf{I})l_k - 2x_k F_k^T e = 0, \\ \|l_k\|^2 - 1 = 0. \end{cases} \quad (40)$$

Thus, we can find optimal solution of (31) by solving equation system (40). Note that there is no restriction on λ since we have an equality constraint. Suppose $\lambda \neq -\sigma_j^2$, where σ_j is singularvalue of matrix F and thus, σ_j^2 is eigenvalue of matrix $F^T F$. Under this assumption, we then know $F^T F + \lambda \mathbf{I}$ is nonsingular and further we can derive l_k , which depends on λ , i.e.,

$$l_k(\lambda) = (F_k^T F_k + \lambda \mathbf{I})^{-1} F_k^T (x_k \mathbf{e}). \quad (41)$$

Let

$$f(\lambda) = \|l_k(\lambda)\|^2 = x_k^2 \|(F_k^T F_k + \lambda \mathbf{I})^{-1} F_k^T \mathbf{e}\|^2, \quad (42)$$

by considering the second equation of (40), we can try to find a proper λ such that

$$f(\lambda) = 1. \quad (43)$$

Note that $f''(\lambda) = 6\|y(\lambda)\|^2$, where $y(\lambda) = -l'_k(\lambda) = -(F_k^T F_k + \lambda \mathbf{I})^{-1} l_k(\lambda)$. It indicates that $f(\lambda)$ is convex in each continuous interval $(-\sigma_j^2, -\sigma_{j+1}^2)$ and strictly monotone decreasing in $(-\sigma_r^2, \infty)$, where σ_r is the smallest singular-value of matrix F , see Figure 3 from [ZH03].

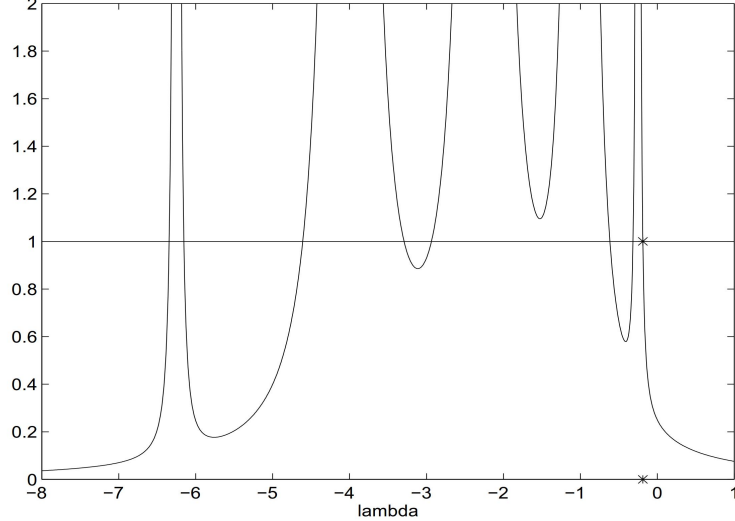


Figure 3: Function $f(\lambda)$

If we denote the largest solution of (43) by

$$\lambda^* = \max\{\lambda | (\lambda, l_k) \text{ that satisfies (40)}\}, \quad (44)$$

the following Theorem 2 is then given (see [EP99])

Theorem 2. *The Lagrange multiplier λ^* , corresponding to a minimum l_k^* of (31) satisfies*

$$\lambda^* \geq -\sigma_r^2. \quad (45)$$

If $\lambda^ > -\sigma_r^2$, then l_k^* can be uniquely determined by (41).*

Therefore, we can only focus on solving (43) in interval $[-\sigma_r^2, \infty)$. One widely used approach to achieve this is Newton method. However, it is hard to determine a good starting point. As we know, Newton method is sensitive of starting point. Without a good starting point, Newton method is not able to derive a reliable solution, even it may diverge. So, instead of using Newton method, a projection method is proposed.

Actually, in [ZH03], they find a nice property approximation $\phi^i(\lambda)$ for $f(\lambda)$ at point λ^i as

$$\phi^i(\lambda) = \frac{\|l_k^i\|^4}{\|l_k^i\|_2 + 2(y_k^i)^T l_k^i (\lambda - \lambda^i) + \|y_k^i\|^2 (\lambda - \lambda^i)^2}. \quad (46)$$

Afterwards, we are going to use the following method to derive λ^* ,

$$\lambda^{i+1} = \lambda^i + \begin{cases} -(l_k^i)^T y_k^i / (y_k^i)^T y_k^i, & \Delta^i \leq 0 \\ (\sqrt{\Delta^i} - (l_k^i)^T y_k^i) / (y_k^i)^T y_k^i, & \Delta^i > 0, \end{cases} \quad (47)$$

where

$$\Delta^i = ((l_k^i)^T y_k^i)^2 + (\|l_k^i\|^2 - 1) \|l_k^i\|^2 \|y_k^i\|^2. \quad (48)$$

Convergence result is also stated in [ZH03],

Theorem 3. *For any $\lambda_0 \in (\underline{\lambda}, -\sigma_r^2) \cup (-\sigma_r^2, \lambda^*)$, the iteration sequence $\{\lambda_k\}$ of (47) converges monotonically to λ^* , in which $\underline{\lambda}$ is the largest value satisfies*

$$f(\underline{\lambda}) \leq 1 \text{ and } f'(\underline{\lambda}) = 0. \quad (49)$$

Moreover, the sequence l_k^i corresponding to λ^i converges to l_k^ .*

5 Numerical Examples

In this section, we numerically show performance of our regression model (31) in Section 3.2. Note that our experiments are carried out using R 3.2.0 in x86_64 Windows platform.

5.1 Portfolio \mathcal{P}_1

The first example comes from the one in [GL05, Gla05] with slight modification. In this example, portfolio \mathcal{P}_1 is introduced, which consists of $m = 100$ obligors. As in [GL05], we set number of risk factors $d = 10$. Furthermore, the marginal default probability and single loss with respect to each obligor k have the following form,

$$p_k = 0.01 \cdot (1 + \sin(16\pi k/m)), k = 1, 2, \dots, m, \quad (50)$$

and

$$c_k = (\lceil 5k/m \rceil)^2, k = 1, 2, \dots, m. \quad (51)$$

Note that we choose number of samplings $n = 3(d + 1)$ and then use projection method to derive optimal loading coefficients, i.e., loading matrix $A \in \mathbb{R}^{100 \times 10}$ and loading vector $b \in \mathbb{R}^{100}$, from regression model (31) in Section 3.2. Afterwards, we randomly generate 500 scenarios for IS procedure and 5000 scenarios for MC to simulate the large loss probability distribution. Note we set threshold $x = 1\% \cdot \sum_{k=1}^m c_k = 110$.

We repeatedly generate risk factors Z and ϵ for four times and run simulation, respectively. Figure 4 then shows four large loss probability distribution corresponding to different optimal risk loading coefficients. The green and red solid curves show the 95% confidence interval of probability at each loss level by MC and IS, respectively.

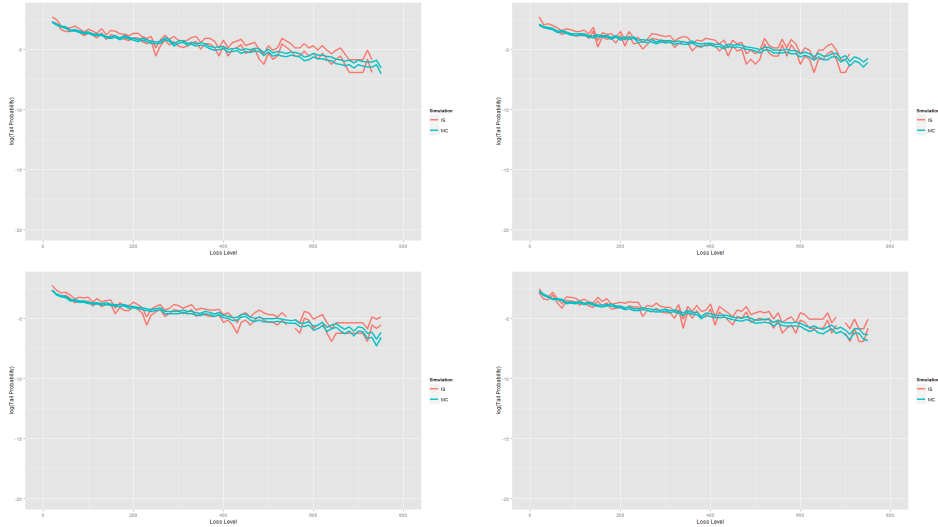


Figure 4: Four Optimal Loading Coefficients

It shows that by various samplings (\mathbf{Z}, ϵ) and their corresponding optimal risk loading coefficients, the large loss probability distribution is relatively stable. Actually, in all four cases, under 500 scenarios for IS and 5000 scenarios for MC, the largest total loss level is around 750, i.e., 68% of the largest total loss and its probability, $P(L > 750)$ is around $\exp(-7.5) \approx 5 \times 10^{-4}$. Besides, we can observe that under optimal risk loading coefficients, there is not much difference on efficiency by using IS or MC. One can compare these results with those in [GL05], where under the same setting as we have except for using optimal risk loading coefficients, IS is much more efficient than MC. The reason is that under strong correlation among obligors, even though default of each obligor is still rare, once one obligor default, other obligors in the portfolio with high correlation to the default obligor will more likely default as well. It makes the large loss event not that "much" rare, which reduce the efficiency of IS.

In Figure 5, we first generate risk factors (\mathbf{Z}, ϵ) once to calculate optimal loading coefficients. As per the rest of three cases, we produce entries a_{kj} of risk loading matrix A independently and uniformly from the interval $1/\sqrt{d} \times (-1, 1)$ where $d = 10$, and then derive elements b_k of risk loading vector b equals to $\xi_k \sqrt{1 - a_k^T a_k}$, where ξ_k is uniformly generated from $\{-1, 1\}$ and a_k is k th row of A . Besides, we randomly generated the risk loading coefficients as in [Gla05, GL05], i.e., to assume all loading coefficients are positive. Note that we set A and b in this way to guarantee latent variables X_k in Section 2 is in normal distribution. We repeatedly the procedure twice and after running simulation, we show large loss probability distribution with respect to different loss level under these six cases in Figure 5. The green and red solid curves show the 95% confidence interval of probability at each loss level by MC and IS, respectively.

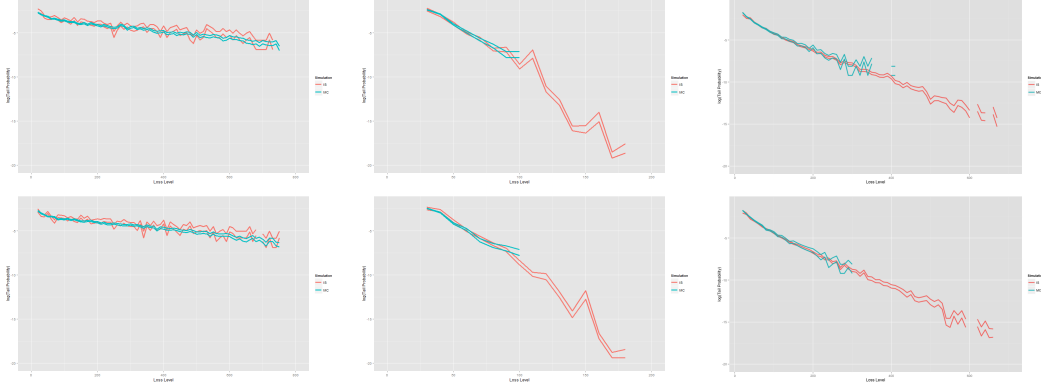


Figure 5: Two Optimal and Four randomly Generated Loading Coefficients

By looking into these six figures in Figure 5, one can observe that if we generate risk loading coefficients randomly, which indicates small correlation among obligors, the large loss event is very rare. Especially in the case that we do not assume all positive correlation among obligors. In second column, the largest loss level is around 175, i.e. around 15.9% and its loss probability, i.e., $P(L > 175)$, is around $\exp(-18) \approx 1.5 \times 10^{-8}$, which is extremely small and we think it underestimates the large loss probability. In third column, the largest loss level is around 650, i.e. around 59.1% and its loss probability, i.e., $P(L > 650)$, is around $\exp(-17) \approx 4.1 \times 10^{-7}$. As said in [BJZ08, CK10], in this setting, extreme dependency is not captured. However, this is different from our cases at first column, where we take concerns on dependency much. One other observation is that, as we discussed before, under small correlation among obligors, the large loss event is rare and it is the reason why IS is much more efficient than MC in randomly generated risk loading coefficients cases (the last two columns).

5.2 Portfolio \mathcal{P}_2

Our next example, \mathcal{P}_2 is a $d = 10$ factors model with $m = 15$ obligors. We set margin default probabilities p_k as (50) and corresponding loss c_k as (51) with respect to all obligors. In this example, we're concerning the correlation among all obligors. We choose $n = 3(d + 1)$ and randomly produced $Z \in \mathbb{R}^{33 \times 10}$ and $\epsilon \in \mathbb{R}^{33 \times 15}$. We then generated optimal risk loading coefficients from Section 3.2 and randomly generated it by the same approach as in Example 5.1. We repeat this procedure twice and observe there correlation

by vision correlation matrix $\mathcal{C} = AA^T$, i.e. each entries $\mathcal{C}_{kj} = a_k^T a_j$ is the correlation between obligor k and obligor j . The following six figures show the correlation (i.e., dependency) among all 15 obligors in two cases.

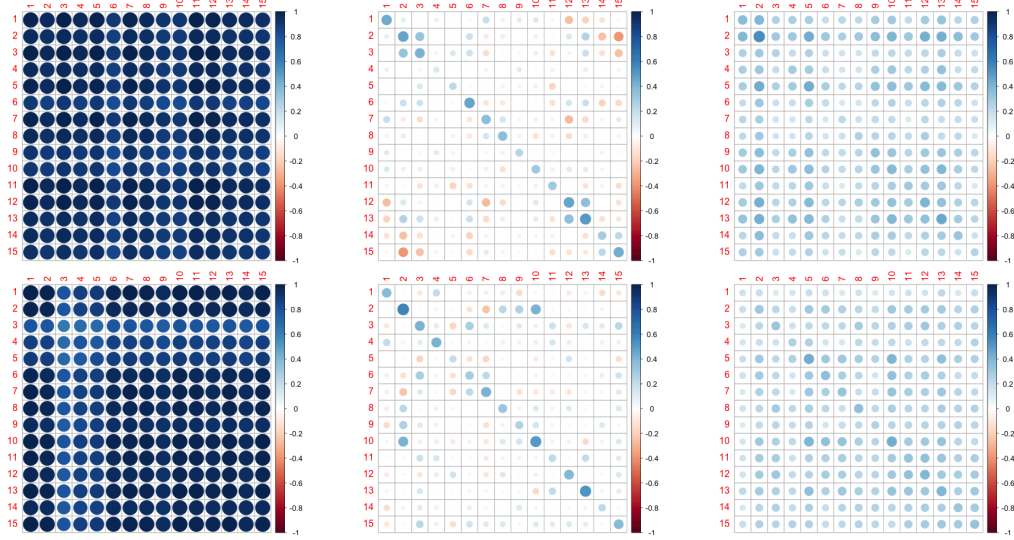


Figure 6: Correlation among all obligors by optimal and randomly generated coefficients

From Figure 6, we can find that the two correlation matrices in first column, which is generated from regression model (31), is much stronger than other four in second and third column. It then fits those simulation results in Example 5.1. One observation is that, even we do not assume all risk loading coefficients is positive, by our regression model in Section 3.2, we still show that correlation among all obligors are strictly positive, and furthermore are close to 1.

5.3 Comparison of SDP and PM in Section 4

In this comparison, we use the same setting as Example 5.2, where we have no idea about correlation among these m obligors, thus we use the model in section 3.2 to derive loading information A and b . We set various n to generate loading efficient a_k and b_k of last obligor $k = 100$ by using SDP method in Section 4.1 with solver *Rcsdp* (see [Bra14]) and Projection method in Section 4.2. We denote violation of the norm constraint and objective

function value at terminated point by *err.* and *obj.*, respectively. Results are stated in the following table.

$n/(d+1)$	1	2	3	5	10	20
<i>err.</i> (10^{-4})	1.13	0.22	1.44	1.34	1.65	1.19
<i>obj.</i>	5.012	8.507	11.835	14.906	23.549	33.902

Table 1: Solve (31) by PM

$n/(d+1)$	1	2	3	5	10	20
<i>err.</i> (10^{-4})	0	0	0	0	0	0.429
<i>obj.</i>	5.012	8.506	11.834	14.879	23.536	33.902

Table 2: Solve (31) by SDP

6 Conclusion

We introduced several regression models aiming to generate optimal risk loading coefficients and further review and propose some algorithms to solve these models. With respect to the first regression model (31) in Section 3, we do not have any correlation information among obligors. Thus, we can split the regression model so as to address each obligor separately. Corresponding algorithms in Section 4 and numerical examples in Section 5 show that under this setting, we can estimate large loss probability distribution stably and enhance correlations among obligors as much as possible. The rest of models targeted to find out optimal risk loading coefficients with (partly) known correlation information. However, it's still a problem on how to solve these models efficiently and accurately. There are also some further topics worthy to address, such as what if we do assume all risk loading coefficients are positive, where we need to modify our regression model (31) to have a positive cone constraint. Or we can also apply these regression models to *t*-copula model, which is also widely used in credit risk problem.

References

[Ank10] Lee Anke. Correlation—the basis of risk management, 2010.

- [Bec07] Amir Beck. Quadratic matrix programming. *SIAM Journal on Optimization*, 17(4):1224–1238, 2007.
- [BJZ08] Achal Bassamboo, Sandeep Juneja, and Assaf Zeevi. Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, 56(3):593–606, 2008.
- [BOW10] Christian Bluhm, Ludger Overbeck, and Christoph Wagner. *Introduction to credit risk modeling*. Crc Press, 2010.
- [Bra14] Hector Corrada Bravo. *Rcsdp: R interface to the CSDP semidefinite programming library*, 2014. R package version 0.1.53.
- [CK10] Joshua CC Chan and Dirk P Kroese. Efficient estimation of large portfolio loss probabilities in t-copula models. *European Journal of Operational Research*, 205(2):361–367, 2010.
- [DMV09] Joost Driessen, Pascal J Maenhout, and Grigory Vilkov. The price of correlation risk: Evidence from equity options. *The Journal of Finance*, 64(3):1377–1406, 2009.
- [EMS02] Paul Embrechts, Alexander McNeil, and Daniel Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pages 176–223, 2002.
- [EP99] Lars Eldén and Haesun Park. A procrustes problem on the stiefel manifold. *Numerische Mathematik*, 82(4):599–619, 1999.
- [FB12] Juliano B Francisco and Fermín S Viloche Bazán. Nonmonotone algorithm for minimization on closed sets with applications to minimization on stiefel manifolds. *Journal of Computational and Applied Mathematics*, 236(10):2717–2727, 2012.
- [GHS99] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical finance*, 9(2):117–152, 1999.
- [GL05] Paul Glasserman and Jingyi Li. Importance sampling for portfolio credit risk. *Management science*, 51(11):1643–1656, 2005.

- [Gla05] Paul Glasserman. Measuring marginal risk contributions in credit portfolios. *FDIC Center for Financial Research Working Paper*, (2005-01), 2005.
- [Mor97] JP Morgan. Creditmetrics-technical document. *JP Morgan, New York*, 1997.
- [RK11] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*, volume 707. John Wiley & Sons, 2011.
- [TTT99] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. Sdpt3a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581, 1999.
- [Ye99] Yinyu Ye. Approximating quadratic programming with bound and quadratic constraints. *Mathematical programming*, 84(2):219–226, 1999.
- [ZD06] Zhenyue Zhang and Keqin Du. Successive projection method for solving the unbalanced procrustes problem. *Science in China Series A*, 49(7):971–986, 2006.
- [ZH03] Zhenyue Zhang and Yushan Huang. A projection method for least squares problems with a quadratic equality constraint. *SIAM journal on matrix analysis and applications*, 25(1):188–212, 2003.