Submitted to iINFORMS Journal on Computing
manuscript (Please, provide the manuscript number!)

Online Risk Monitoring Using Offline Simulation

Guangxin Jiang
School of Management, Shanghai University, Shanghai, China, gxjiang@shu.edu.cn

L. Jeff Hong
Department of Management Science, School of Management, Fudan University, Shanghai, China, hong_liu@fudan.edu.cn

Barry L. Nelson
Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208, nelsonb@northwestern.edu

Estimating portfolio risk measures and classifying portfolio risk levels in real time are important yet challenging tasks. In this paper we propose to build a logistic regression model using data generated in past simulation experiments and to use the model to predict portfolio risk measures and classify risk levels at any time. We further explore regularization techniques, simulation model structure, and additional simulation budget, to enhance the estimators of the logistic regression model to make its predictions more precise. Our numerical results show that the proposed methods work well. Our work may be viewed as an example of the recently proposed idea of simulation analytics, which treats a simulation model as a data generator and proposes to apply data analytics tools to the simulation outputs to uncover conditional statements. Our work shows that the simulation analytics idea is viable and promising in the field of financial risk management.

Key words: simulation analytics; logistic regression; lasso; classification; Monte Carlo simulation; variance reduction
History: Received

1. Introduction [1]

In portfolio risk management, simulation studies are often used to estimate portfolio risk measures, e.g., exceedance probabilities, values-at-risk or conditional values-at-risk. These studies often produce accurate estimates of risk measures, as long as the stochastic processes of the underlying risk factors may be simulated. However, they can also be very time consuming, especially when portfolios consist of multiple derivative products whose prices also need to be determined by additional simulation efforts. This type of problem is known as nested estimation and it has been studied extensively in the areas of simulation and financial engineering (see, for instance, Gordy and Juneja 2010, Liu and Staum 2010, Broadie et al. 2011,

[1] A preliminary version of this work appeared in the Proceedings of the Winter Simulation Conference (Jiang et al. 2016), which only outlines the basic logistic regression approach without detailed analysis. Parts of Sections 2 and 3 originally appeared in the conference paper. In this paper, we add the full technical analysis of the basic logistic regression approach. More importantly, we propose three performance enhancing techniques with full technical analysis, and provide abundant numerical examples to illustrate all the methods.

Sun et al. 2011). Nearly all estimation methods proposed in the simulation literature consider the estimation problem only once, i.e., one is interested only in estimating a risk measure at the current time point given the current values of all underlying risk factors, and researchers often argue that these methods may be implemented over night or over the weekend so that the long computational time needed to run them is not a huge barrier. In practice, however, portfolio risk measures are often needed in real time as the current values of the underlying risk factors change. For instance, the famous "4:15 report" of J. P. Morgan requires the company to consolidate the risks of all trading desks based on the closing values of the underlying factors, available within 15 minutes after the market closes everyday, and the company uses it to decide whether the risk is under control (Jorion 2006). We call this the *online risk monitoring problem*. The estimation methods available in the simulation literature cannot solve this problem directly, because a portfolio may contain thousands of financial instruments based on many risk factors, and the risk factors may change in seconds. As we need to estimate the risk measures in real time based on these risk factors, we cannot guarantee that the desirable simulation results can be obtained in time, even though with parallel computing resources.

We take a different view. We ask whether we can use the sample paths and derivative prices from a database of retained past simulation results to estimate online risk measures based on the current values of the underlying risk factors, without running additional simulation experiments. From a mathematical point of view, the traditional approaches estimate an unconditional risk measure, which is a fixed value; while what we need is an approach that estimates a conditional risk measure, which is a function of the values of the underlying risk factors at a future time. This motivates us to consider the *simulation analytics* approach recently proposed by Nelson (2016).

The main idea of the simulation analytics approach is to treat the simulation model as a generator of multiple (often a large number of) replications of system dynamics over time and to apply data mining and data analytics tools to mine the data and to estimate conditional statements. To make an analogy, we consider a recommendation problem, which is a classical example of data mining/data analytics (see, for instance, Hastie et al. 2011). In that problem, the available data include the personal attributes and purchasing records of many individuals, and the goal is to estimate the probability that an incoming individual with certain attributes will make a purchase of a certain product (known as an estimation problem) and to decide whether to recommend a certain product to this individual (known as a classification problem). In the online risk monitoring problem, suppose we have many sample paths (i.e., the data) generated from the simulation model, and in each sample path (i.e., individual), we have the future evolution of risk factors (i.e., attributes) and its loss at a relevant day (i.e., purchasing record). Our goal is to estimate the risk measure given the risk factors at a future time and to classify whether the risk is too high. By this analogy, we conclude that the online risk monitoring problem is indeed analogous to a recommendation problem and the simulation analytics approach can be used to solve it.

In this article we consider a specific online risk monitoring problem. Suppose that an ordinary Monte Carlo simulation study was conducted at an initial time point (time 0) to evaluate the (unconditional)

exceedance probability (i.e., the probability that the loss is greater than a given threshold) at an important future date (time T). To estimate the exceedance probability, the simulation study generated many sample paths from the initial state of the underlying risk factors at time 0 to time T and evaluated the portfolio loss (including possibly pricing of derivatives) for each sample path at time T. Our goal is to use the same sample paths to estimate the same exceedance probability at time T conditional on the state of the underlying risk factors at any time $t \in [0, T]$ and to classify the portfolio risk at time t into either "safe" or "dangerous" based on whether the exceedance probability is below or above a certain threshold. The problem was motivated by our conversations with investment practitioners in the insurance industry, who complained that their company does not set risk limits dynamically based on portfolio status and, instead, uses static limits. In this paper, we analyze the problem in the context of financial simulation, which typically assume that the simulation model is the true model, and do not consider the model misspecification issue.

Regression is commonly used in estimating the conditional relationship. For instance, Longstaff and Schwartz (2001) proposed to use linear regression to approximate continuation values in the American option pricing problem. To estimate the conditional probability and to classify the risk category, logistic regression is used, and the maximum likelihood method is applied to estimate the parameters of the model. According to generalized linear model theory (Fahrmeir and Kaufmann 1985), we will show that the risk estimators are strongly consistent and asymptotically normal. In addition, the classification error goes to zero exponentially fast as the sample size $n \to \infty$. *This result shows that risk classifications, which are often the main purpose of online risk monitoring, are significantly easier than risk estimations.*

The logistic regression approach is treated as a baseline method in this paper. We then propose three performance enhancing techniques, by taking advantage of either existing data analytics tools or the knowledge of the simulation models, to further improve the baseline method. Firstly, considering that the number of risk factors is often large in practical situations, we propose to use L_1-regularization (often known as lasso) to conduct variable selection and to improve prediction accuracy. The basic idea of lasso is to tradeoff a small increase of bias to obtain a large decrease of variance, so that the estimation and classification can be done more precisely. It is a standard tool in data analytics to handle high-dimension data, and we show that it also works well in the context of online risk monitoring. We also prove that the risk estimators using lasso are strongly consistent and asymptotical normally, and the classification error goes to zero exponentially fast as the sample size $n \to \infty$.

Secondly, noticing that our data are generated from simulation, we may use our knowledge of the simulation model (i.e., the data generating process) to develop more efficient tools. This is a unique feature of simulation analytics. In typical data analytics problems, data are observed and the data generating processes are unknown. To improve the baseline method, we utilize gradient estimation in simulation to develop a method that perturbs all simulated sample paths so that they are more spread out. Through both theoretical analysis and numerical studies, we show that the perturbation method can significantly improve the quality

of both risk estimators and risk classifiers, especially when the time is close to the beginning of the planning horizon where sample paths are more clustered together.

Thirdly, in some practical situations we may have time to conduct a small number of additional simulation experiments. Therefore, we also consider how to incorporate the additional simulation data into the analysis to improve the quality of the risk estimators and risk classifiers. In particular, we propose two approaches, one is to combine two estimators that are derived from the original data and the new data, respectively, and the other is to combine the original data and the new data together to compute a new estimator. Through both theoretical analysis and numerical studies, we show that using additional simulation data can improve the performance in both risk estimations and risk classifications.

Literature Review

Our work is related to three lines of literature. The first is on portfolio risk measurement. Glasserman et al. (2000, 2002) were among the first to study how to estimate portfolio values-at-risk. In particular, Glasserman et al. (2000) considered the case where the risk factors follow light-tailed distributions and Glasserman et al. (2002) considered the case of heavy-tailed distributions. Glasserman and Li (2005), Bassamboo et al. (2008) and Glasserman et al. (2008) studied portfolio credit risks. For large portfolios, large pool approximation can be used to study the risk measures, see Iscoe and Kreinin (2010) and Sirignano and Giesecke (2018). All of these papers study situations where the loss of the portfolio may be calculated easily through closed-form expressions or delta-gamma approximations, and consider how to apply variance reduction techniques so that the risk measures may be estimated more accurately.

Sometimes, estimation of portfolio losses need additional simulation experiments. For instance, a portfolio may contain derivatives whose values need to be priced through simulation. Nested simulation approaches are often used in these situations. Lee and Glynn (2003) studied the general formulation of nested simulation and considered how to balance the simulation effort in the inner and outer levels, while Gordy and Juneja (2010) applied it to portfolio risk measurement. Liu and Staum (2010) used stochastic kriging to improve the estimation efficiency, and Broadie et al. (2011) designed an adaptive method to allocate the simulation effort in inner and outer levels. To reduce the large amount of simulation effort needed in the inner level, Broadie et al. (2015) proposed a regression method and Hong et al. (2017) proposed a kernel method to avoid nested simulations. Interested readers may see Hong et al. (2014) for a recent comprehensive review on simulation methods in estimating risk measures.

The second related line of literature is regression. Regression is a standard technique used in data analytics and statistical learning to construct functional relations and to classify instances (see, for instance, Hastie et al. (2011) for a thorough introduction of regression techniques in statistical learning). In this article our goal is to estimate conditional probabilities and use them for classification. Therefore, it is natural for us to consider logistic regression. For background on logistic regression and how to use maximum likelihood to estimate model parameters, readers are referred to the monograph of Hosmer and Lemeshow (2004). In statistical learning, lasso is often used with regression to select variables and to improve prediction preci-

sion, especially when the number of variables is large. It was first introduced by Tibshirani (1996) and it can be applied to not only linear regression models but also generalized regression models that include logistic regression. Regression has also been used at the interface of stochastic simulation and financial engineering. For instance, as we mentioned earlier, Broadie et al. (2015) proposed a regression method to estimate portfolio risk measures. Sirignano and Giesecke (2018) used logistic regression to approximate the transition function in loan-level models. Longstaff and Schwartz (2001) used a regression method to price American style options. Generally, pricing American style options is equivalent to solving dynamic programming problems, where regression is often used to approximate the value functions for backward induction. In our method, regression is used repeatedly to approximate the functions of conditional exceedance probabilities, which are used directly for online risk monitoring.

The third related line of literature is on re-use of simulation experiments. Liu et al. (2010) proposed to run simulation experiments to construct a good "database" and use it for future estimations. They call the approach "simulation on demand". Rosenbaum and Staum (2015) further developed this idea into database Monte Carlo simulation that uses the database to construct control variates to reduce the variances of the estimators. Similar to our idea of re-using simulation data, Feng and Staum (2017) proposed the concept of green simulation that uses retained simulation data as a complementary resource to new simulation data by employing a change of probability measures. Unlike the approaches mentioned above that either pre-run simulation experiments (as in simulation on demand) or convert the old simulation data to the new data (as in green simulation), our approach learns (or mines) the simulation data to uncover conditional relationships available in the data.

The rest of this paper is organized as follows. We formulate the online risk monitoring problem in Section 2 and introduce the logistic regression based methods in Section 3. Three performance enhancing techniques, lasso, perturbation, and additional simulation, are discussed in detail in Sections 4, 5, and 6, respectively. Numerical results are presented in Section 7, followed by conclusions and discussions in Section 8.

2. Problem Statement

Suppose that $\mathbf{S}(t) = (S_1(t), S_2(t), \ldots, S_m(t))^\top$ is a vector of the underlying risk factors, which may include prices of stocks and bonds, stochastic interest rates, etc, and that $\mathbf{S}(t)$ follows a Markov process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a natural filtration $\mathcal{F}_t$ that governs the evolution of the process. Consider a portfolio with k financial products, e.g., stocks, bonds and derivatives, whose values at time t are denoted by $V_i(t), i = 1, \ldots, k$, which depend on the realizations of the underlying risk factors $\mathbf{S}(t)$. For convenience of the notation, we let $\mathbf{V}(t) = (V_1(t), \ldots, V_k(t))^\top$. Furthermore, suppose that the positions on the financial products are $\mathbf{w} = (w_1, \ldots, w_k)^\top$. Then, the value of the portfolio at time t is

$$\Phi(t) = \sum_{i=1}^{k} w_i V_i(t) = \mathbf{w}^\top \mathbf{V}(t). \tag{1}$$

Let $L(t) = \Phi(0) - \Phi(t)$. Then, $L(t)$ is the loss of the portfolio at time t.

Suppose that there is a fixed future time T at which portfolio loss $L(T)$ needs to be evaluated and further actions need to be taken based on $L(T)$. For instance, T may be the end of the fiscal year at which portfolio loss needs to be reported to shareholders or T is the end of the investment cycle at which bonuses are distributed. Then, the managers of the portfolio may be interested in estimating the loss distribution and the portfolio risk measures at time T. In this paper, we consider the estimation of the exceedance probability, i.e., $\Pr\{L(T) > y\}$ for some important threshold value y. Notice that, if $\Pr\{L(T) > y\}$ may be estimated for any y, we may use it to obtain other risk measures, such as value-at-risk and conditional value-at-risk (Glasserman et al. 2000). In the simulation literature, much has been done in estimating the unconditional probability, i.e., $\Pr\{L(T) > y\}$, but we are interested in estimating the conditional probability $\Pr\{L(T) > y|\mathcal{F}_u\}$, which denotes the exceedance probability given the information up to time $u \in (0, T)$. By the Markov property of the underlying risk factors $\mathbf{S}(t)$, we have $\Pr\{L(T) > y|\mathcal{F}_u\} = \Pr\{L(T) > y|\mathbf{S}(u)\}$. The conditional probability is useful because, at any time $u \in (0, T)$, given the realization of $\mathbf{S}(u)$ observed in the market, we can tell the probability that the portfolio loss exceeds the threshold y at the important future time T and we may use the probability as a risk monitoring tool to determine whether the portfolio risk is under control.

2.1. The Data

When the portfolio is formed at time 0, a thorough simulation study is typically conducted to analyze the risk profile of the portfolio and report the risk measures to relevant stakeholders. In this article we suppose that the unconditional exceedance probability $\Pr\{L(T) > y\}$ is estimated through a nested Monte Carlo simulation study. For this simulation study, one often has more time available to run simulation experiments, and the risk measures are often estimated accurately. Then, after the simulation study, we have n simulated sample paths of the underlying risk factors, denoted by $\mathbf{S}_1(t), \mathbf{S}_2(t), \ldots, \mathbf{S}_n(t)$ for $0 \le t \le T$. These sample paths are often simulated under the real probability measure and they are the output of the outer level simulation in a nested simulation study. Moreover, we also have the values of the financial products at time T evaluated based on each simulated realization of the underlying risk factors. We denote them as $\mathbf{V}_1(T), \mathbf{V}_2(T), \ldots, \mathbf{V}_n(T)$. Notice that these financial products may include complicated financial derivatives whose values do not have closed-form expressions. Then, an inner level simulation study under the risk neutral probability measure may need to be used to estimate the $\mathbf{V}_i(T)$ values. In this article, we assume that these values can be estimated so accurately that they are effectively without estimation error. Then, given the weights $\mathbf{w}$ of the portfolio, we can easily calculate the portfolio loss at time T based on the simulated realizations of the underlyings, and we denote them as $L_1(T), L_2(T), \ldots, L_n(T)$.

2.2. Online Risk Monitoring

Once the portfolio is constructed, the portfolio managers need to constantly monitor the risk of the portfolio. For instance, they may need to estimate the exceedance probability at any real time (instead of the simulated

time) given the market conditions, i.e., the real realization of the underlyings $\mathbf{S}(u)$, and decide whether the portfolio is safe or not at time u. We call the first the "online risk estimation problem" and the second the "online risk classification problem". For both problems, we want to use the simulated data in Section 2.1 to avoid simulating more data, so that both problems may be solved quickly to meet the practical requirements.

In the online risk estimation problem, our goal is to estimate $p_u(\mathbf{S}(u)) = \Pr\{L(T) \geq y|\mathbf{S}(u)\}$ for any $u \in (0,T)$ in real time. Notice that $p_u(\mathbf{S}(u))$ is a function of $\mathbf{S}(u)$. Therefore, our goal is to estimate a function, which is often called a regression problem in the field of statistical learning (Hastie et al. 2011), and the regression is on $\mathbf{S}(u)$ for fixed u. If the function is estimated, we may then plug in $\mathbf{S}(u)$ observed at real time u to estimate the exceedance probability $p_u(\mathbf{S}(u))$. Notice that this may be done very quickly if the function has been estimated before hand.

In the online risk classification problem, our goal is to classify the portfolio risk into two categories, "safe" and "dangerous", based on the exceedance probability $p_u(\mathbf{S}(u))$ in real time. For instance, we may set $\alpha \in (0,1)$ as a threshold and classify the portfolio risk is dangerous if $p_u(\mathbf{S}(u)) > \alpha$ and safe otherwise. In practice, we may set $\alpha = 0.05$ or 0.1. Risk classification allows risk managers to know immediately whether actions need to be taken to control the portfolio risk. One may further extend the number of categories from two to a higher number in the risk classification problem. This may lead to a risk ratings that resemble to the credit ratings, e.g., the AAA to D levels, used by credit rating agencies such as Standard & Poor's and Moody's. Notice that, once the function $p_u(\mathbf{S}(u))$ is estimated, the classification can also be solved immediately given the values of $\mathbf{S}(u)$ at the time $u \in (0,T)$.

2.3. From Online Risk Monitoring to Online Risk Control

Notice that online risk monitoring is only the first step to online risk control, which is often the goal of risk management. For instance, if the risk classification indicates that the portfolio risk level is dangerous, risk managers may decide to change some positions and add some additional products. To apply the simulation analytics approach to estimate the risk of this new portfolio, however, there may be some difficulties. The major difficulty is that the new products may depend on new risk factors whose sample paths are not in our data. Moreover, even when the new products do not involve new risk factors, how their values at time T depend on existing sample paths of underlying risk factors may still be unknown and require additional simulation effort to evaluate. In these situations, we have to run additional simulation experiments, because the information cannot be mined if it is not in the data.

However, if the online risk control involves only adjusting the positions of existing financial products, i.e., changing the weights $\mathbf{w}$, without adding new products, we can still apply the simulation analytics approach to estimate the risk of the new portfolio. To do that, suppose the weights at time u are now $\mathbf{w}(u)$. Based on the new weights, we may reconstruct the portfolio value $\Phi(T)$ based on Equation (1) and calculate the portfolio loss $L(T)$ for the new portfolio, assuming the portfolio will be held until time T. Once we update the data, we are back to the situation that is considered in Section 2.2. Notice that the reconstruction of the

data does not need additional simulation experiments. Therefore, it can be done efficiently and it will not cause much delay in providing the risk estimates and the risk classifications.

3. Logistic Regression

As stated in Section 2, our goal is to estimate $p_u(\mathbf{S}(u)) = \Pr\{L(T) \geq y|\mathbf{S}(u)\}$ and use it for classification for any $u \in (0,T)$. To do that, notice that we have the simulated sample paths $\{\mathbf{S}_i(t), 0 \leq t \leq T\}$, $i = 1,2,\ldots,n$, and the corresponding portfolio loss $L_i(T)$ for each sample path i. Let

$$Y = \begin{cases} 1 & \text{if } L(T) \geq y, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We have $p_u(\mathbf{S}(u)) = \mathrm{E}[Y|\mathbf{S}(u)]$. Based on the simulated sample paths, we have n observations of Y, denoted by $Y_1, Y_2, \ldots, Y_n$. Then, to estimate the regression function $p_u(\mathbf{S}(u))$, we have n observations of the input-output pair for given u, and denoted by $\{(\mathbf{S}_1(u), Y_1), (\mathbf{S}_2(u), Y_2), \ldots, (\mathbf{S}_n(u), Y_n)\}$, where the inputs are also called predictors, input variables or features, and the outputs are also called dependent variables or responses or classes or labels in the areas of statistical learning.

Notice that both the regression problem and the classification problem are classical supervised learning problems. There are parametric and nonparametric approaches that can be used to solve the problems. For our problems, because the data were collected from a separate simulation study, the number of observations is typically not very large. Moreover, as the portfolio may include many assets, the dimension of the features, i.e., the underlying risk factors $\mathbf{S}(u)$, may be quite large. Therefore, the observations may be scattered sparsely in the feature space. These characteristics typically make nonparametric methods, such as k-nearest neighbors and kernel methods, less effective due to a high level of variance of the predictions (Hastie et al. 2011). Therefore, in this paper, we consider parametric methods that generally trade off variance for bias but may be effective if the parametric model is chosen appropriately.

3.1. Logistic Regression Model and Maximum Likelihood Estimation

Because the response Y is a Bernoulli random variable, to model $p_u(\mathbf{S}(u))$, a natural choice is a logistic regression model (Hosmer and Lemeshow 2004). Let $\mathbf{X}(\cdot) : \Re^m \to \Re^d$ denote a set of basis functions computed from $\mathbf{S}(u)$, we then propose the following logistic regression model

$$\log\left(\frac{p_u(\mathbf{S}(u))}{1 - p_u(\mathbf{S}(u))}\right) = \boldsymbol{\beta}(u)^\top \mathbf{X}(\mathbf{S}(u)), \tag{3}$$

where $\boldsymbol{\beta}(u) = (\beta_1(u), \ldots, \beta_d(u))^\top$ is the vector of coefficients. It is worthwhile noting that both $\mathbf{X}(\cdot)$ and $\boldsymbol{\beta}(u)$ depend on the time u. Therefore, for different time points $u \in (0,T)$, we may use different basis functions and obtain different coefficients.

To use the logistic regression model in Equation (3) we need to specify the basis functions $\mathbf{X}(\cdot)$. One way is to specify based on the properties of the loss function, and another is to use polynomials of $\mathbf{S}(u)$ due to the well-known Weierstrass approximation theorem (Rudin 1991). In this paper we suggest the second way

and we propose to use linear and individual quadratic functions, i.e., $\mathbf{X}(\mathbf{S}(u)) = (1, \mathbf{S}(u), \mathbf{S}^2(u))^\top$, where $\mathbf{S}(u)^2 = (S_1^2(u), \ldots, S_m^2(u))^\top$. Notice that this is the simplest form of polynomials of $\mathbf{S}(u)$ that one may use and it should also be the first form of polynomials for one to consider. Fortunately, our numerical results show that this approach works well for both the regression problem and the classification problem. Notice that the first component of $\mathbf{X}$ is typically 1, so β_1 denotes the intercept. More details on the appropriateness of the logistic regression model are provided in the online supplement.

The parameters of the logistic regression model of Equation (3) are typically estimated by the maximum likelihood (ML) method. For the convenience of notation, we omit the time index u when it does not cause confusion. That is, in Equation (3), we let $\mathbf{X}$ denote $\mathbf{X}(\mathbf{S}(u))$ and $\boldsymbol{\beta}$ denote $\boldsymbol{\beta}(u)$, and

$$g(\mathbf{X}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}^\top \mathbf{X})/(1 + \exp(\boldsymbol{\beta}^\top \mathbf{X})).$$

Then, the log likelihood function is

$$\log \ell(\boldsymbol{\beta}|\mathbf{X}, Y) = Y \log\left(g(\mathbf{X}, \boldsymbol{\beta})\right) + (1 - Y) \log\left(1 - g(\mathbf{X}, \boldsymbol{\beta})\right). \tag{4}$$

For any $u \in (0, T)$, we have the training data $\{(\mathbf{X}_i, Y_i), i = 1, \ldots, n\}$, where $\mathbf{X}_i = \mathbf{X}(\mathbf{S}_i)$,

$$\begin{aligned} L_n(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i \log\left(g(\mathbf{X}_i, \boldsymbol{\beta})\right) + (1 - Y_i) \log\left(1 - g(\mathbf{X}_i, \boldsymbol{\beta})\right) \right\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i \boldsymbol{\beta}^\top \mathbf{X}_i - \log\left(1 + \exp(\boldsymbol{\beta}^\top \mathbf{X}_i)\right) \right\}, \end{aligned} \tag{5}$$

and the ML estimator $\hat{\boldsymbol{\beta}}_n$ is given by

$$\hat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta} \in \Re^d} L_n(\boldsymbol{\beta}), \tag{6}$$

and the maximization problem may be solved numerically and efficiently by the coordinate descent algorithm (Hastie et al. 2011).

3.2. Online Risk Estimation

In the risk estimation problem our goal is to estimate $p_u(\mathbf{S}_r(u)) = \Pr\{L(T) \geq y | \mathbf{S}_r(u)\}$ given that we have observed $\mathbf{S}_r(u)$ at time u. We use the notation $\mathbf{S}_r(u)$ to denote it is the real-world observation of $\mathbf{S}(u)$, instead of a simulated observation. Nevertheless, we assume that it has the same distribution as the simulated observations. When using the logistic regression model, we estimate $p_u(\mathbf{S}_r(u))$ by

$$\hat{p}_u(\mathbf{S}_r(u)) = g(\mathbf{X}_r, \hat{\boldsymbol{\beta}}_n) = \frac{\exp\left(\hat{\boldsymbol{\beta}}_n^\top \mathbf{X}_r\right)}{1 + \exp\left(\hat{\boldsymbol{\beta}}_n^\top \mathbf{X}_r\right)}, \tag{7}$$

where $\mathbf{X}_r = \mathbf{X}(\mathbf{S}_r(u))$ is known at time u and $\hat{\boldsymbol{\beta}}_n$ is the ML estimator of the unknown parameter $\boldsymbol{\beta}$ calculated using the training data, i.e., the simulated observations of $\{(\mathbf{X}_i, Y_i), i = 1, 2, \ldots, n\}$ at time u. Based

on the generalized linear model theory in Fahrmeir and Kaufmann (1985), we can assess the quality (i.e., consisitency and asymptotic normality) of the risk estimator $\hat{p}_u(\mathbf{S}_r(u))$ easily. To do that, we make the following assumptions.

ASSUMPTION 1. *The observations* $\{(\mathbf{X}_i, Y_i), i = 1, 2, \ldots, n\}$ *are independent observations of* $(\mathbf{X}, Y)$ *and, given* $\mathbf{X}$*,* Y *is a Bernoulli random variable with* $\Pr(Y = 1|\mathbf{X}) = g(\mathbf{X}, \boldsymbol{\beta}_0)$.

Notice that the independence condition is easily satisfied because $\{(\mathbf{X}_i, Y_i), i = 1, 2, \ldots, n\}$ are calculated based on sample paths that are simulated independently. Therefore, Assumption 1 basically assumes that the logistic regression model of Equation (3) is the true model and the true parameter is $\boldsymbol{\beta}_0$. This is a typical assumption that is made in parametric statistical estimations, and we can only build the properties of the estimators based on this assumption. Nevertheless, we have to keep in mind that models are just approximations and they may introduce bias of which we are not aware.

Let $L(\boldsymbol{\beta})$ denote the expectation of the log-likelihood function, i.e.,

$$L(\boldsymbol{\beta}) = \mathrm{E}\left[\log \ell(\boldsymbol{\beta}|\mathbf{X}, Y)\right] = \mathrm{E}\left[Y\boldsymbol{\beta}^\top \mathbf{X} - \log\left(1 + \exp(\boldsymbol{\beta}^\top \mathbf{X})\right)\right]. \tag{8}$$

Assumption 1 also implies that

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta} \in \Re^d} L(\boldsymbol{\beta}). \tag{9}$$

Let $H(\boldsymbol{\beta})$ denote the Hessian matrix of the log-likelihood function, i.e., $H(\boldsymbol{\beta}) = \nabla^2_{\boldsymbol{\beta}} \log \ell(\boldsymbol{\beta}|\mathbf{X}, Y)$. One can easily verify that

$$H(\boldsymbol{\beta}) = -\frac{e^{\boldsymbol{\beta}^\top \mathbf{X}}}{\left(1 + e^{\boldsymbol{\beta}^\top \mathbf{X}}\right)^2} \mathbf{X}\mathbf{X}^\top.$$

We make the following assumptions on $H(\boldsymbol{\beta})$, where the expectations $\mathrm{E}(\cdot)$ are taken with respect to the distribution of $(\mathbf{X}, Y)$.

ASSUMPTION 2. *The Fisher information matrix* $J = -\mathrm{E}\left[H(\boldsymbol{\beta}_0)\right]$ *exists and is positive definite.*

ASSUMPTION 3. *There exists a neighborhood of* $\boldsymbol{\beta}_0$*, denoted as* $\mathcal{N}(\boldsymbol{\beta}_0)$*, such that* $\mathrm{E}\left[\sup_{\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0)} \|H(\boldsymbol{\beta})\|\right] < \infty$.

Assumptions 2 and 3 are typical assumptions used to analyze the asymptotic properties of the ML estimators (see, for instance, Fahrmeir and Kaufmann 1985 and Newey and McFadden 1994), and they can be satisfied easily in our framework, see the online supplement. Then, we can obtain the asymptotic properties of the online risk estimator $\hat{p}_u(\mathbf{S}_r(u))$ defined in Equation (7). The following theorem can be found in Jiang et al. (2016) without proof, and we provide the detailed proof in the online supplement. Again, we want to emphasize that the result is conditional on $\mathbf{S}_r(u)$ (and also $\mathbf{X}_r = \mathbf{X}(\mathbf{S}_r(u))$).

THEOREM 1. *Suppose that Assumptions 1, 2 and 3 are satisfied. Then,* $\hat{p}_u(\mathbf{S}_r(u)) \to p_u(\mathbf{S}_r(u))$ *a.s. and*

$$\sqrt{n}\left[\hat{p}_u(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u))\right] \xrightarrow{d} N(0, D),$$

as $n \to \infty$*, where* $D = c\mathbf{X}_r J^{-1} \mathbf{X}_r^\top$*, and* $c = \exp(2{\boldsymbol{\beta}_0}^\top \mathbf{X}_r)/(1 + \exp({\boldsymbol{\beta}_0}^\top \mathbf{X}_r))^4$.

Theorem 1 states that, under the assumption that the logistic regression model is the true model, the estimated conditional probability is a consistent estimator of the true conditional probability and it has an asymptotic normal distribution. As it has been demonstrated empirically in online supplement that the logistic regression model is a good model for the conditional probability, Theorem 1 also shows that our proposed approach can be used to solve the online risk estimation problem. Furthermore, as the logistic regression may be done very quickly given the sample paths or can even be done before $\mathbf{S}_r(u)$ is observed for any given u, the proposed approach can be used for risk estimation in real time.

3.3. Online Risk Classification

Sometimes we are concerned about the classification problems, i.e., whether the portfolio is safe or not at a given time. Of course, an accurately estimated logistic regression model may lead to good classifications. However, a coarsely estimated logistic regression may also give acceptable classifications. For example in Figure 1, the true boundary is the solid line, and classifies the points into two categories. With a less accurate boundary, the dash line, the classification accuracy is almost as good. This motivates us to consider the accuracy of classification instead of the accuracy of prediction as in Section 3.2, especially in high-dimensional cases.

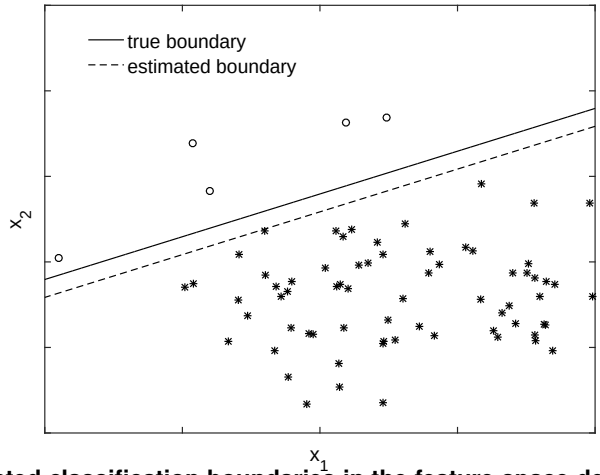

Figure 1 True and estimated classification boundaries in the feature space defined by (x_1, x_2), where $*$ and $\circ$ define observations of different categories

Suppose that α is the threshold probability of safe/dangerous zone. If the exceedance probability $p_u(\mathbf{S}_r(u))$ is less than or equal to α, we say the portfolio is in the safe zone. Otherwise, the portfolio is in the dangerous zone. In practice, $p_u(\mathbf{S}_r(u))$ is unknown and we estimate it by $\hat{p}_u(\mathbf{S}_r(u))$. Let I and $\hat{I}$ denote the safe/dangerous indictors under the true and estimated probabilities, respectively, i.e.,

$$I = \begin{cases} 1 & \text{if } p_u(\mathbf{S}_r(u)) \le \alpha \\ 0 & \text{if } p_u(\mathbf{S}_r(u)) > \alpha \end{cases}, \qquad \hat{I} = \begin{cases} 1 & \text{if } \hat{p}_u(\mathbf{S}_r(u)) \le \alpha \\ 0 & \text{if } \hat{p}_u(\mathbf{S}_r(u)) > \alpha \end{cases}. \tag{10}$$

Then, $\hat{I} \neq I$ denotes a misclassification, and we are also interested in understanding how the misclassification probability, i.e., $\Pr\{\hat{I} \neq I\}$, converges to zero as the sample size $n \to \infty$. To analyze $\Pr\{\hat{I} \neq I\}$,

we first establish a large-deviation result on $\Pr\{\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| > \delta\}$ and use it as a bridge. In the rest of this section, we drop the subscript u for convenience, and we use $\mathbf{X}$ to denote $\mathbf{X}(\mathbf{S}_r(u))$ and X_j to denote the jth component of $\mathbf{X}$, $j = 1, 2, \ldots, d$. This is not to be confused with $\mathbf{X}_i$ which is the ith observation of $\mathbf{X}$, and the boldface denotes that it is a vector. We make the following assumption on X_j.

ASSUMPTION 4. *For each $j = 1, 2, \ldots, d$, there exists a constant $\omega_j > 0$ such that* $\mathrm{E}\left(e^{\omega_j |X_j|}\right) < \infty$.

In this paper X_j is typically a polynomial function of the risk factors $\mathbf{S}(u)$. Assumption 4 basically requires that all risk factors are light-tailed. However, if some of risk factor, say $S_j(u)$, is heavy tailed, we may redefine it so that it is light-tailed. For instance, if $S_j(u)$ follows a lognormal distribution, we may redefine it by taking its logarithm. Based on Assumption 4, we establish an exponential rate of convergence of the ML estimator $\hat{\boldsymbol{\beta}}_n$ to the true parameter $\boldsymbol{\beta}_0$, as summarized in the following theorem, which is a standard large deviation result in ML estimator (see Fu et al. 1993), and has also been stated in Jiang et al. (2016) but without proof. We provide our own proof of this theorem (see the online supplement), because similar proof techniques are also used in the proof of Theorem 5.

THEOREM 2. *Suppose that Assumptions 1, 2, and 4 hold. Then, for any $\delta > 0$, there exists a positive constant $\bar{c}(\delta)$ such that*

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| > \delta\right\} \geq \bar{c}(\delta).$$

Based on Theorem 2, we have the following theorem showing that both the probability of a large deviation of the estimated exceedance probability from its true value, i.e., $\Pr\{|\hat{p}_u(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u))| > \delta\}$ for any $\delta > 0$, and the probability of misclassification, i.e., $\Pr\{\hat{I} \neq I\}$, converge exponentially fast as $n \to \infty$. The proof is included in the online supplement.

THEOREM 3. *Suppose that Assumptions 1, 2, and 4 hold. If $\|\mathbf{X}_r\| \neq 0$, then for any $\delta > 0$, there exists $\tilde{c}(\delta) > 0$ such that*

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{|\hat{p}_u(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u))| > \delta\right\} \geq \tilde{c}(\delta). \tag{11}$$

Furthermore, if $\|\mathbf{X}_r\| \neq 0$ and $p_u(\mathbf{S}_r(u)) \neq \alpha$, then there exists a constant $c_0 > 0$ such that

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{I \neq \hat{I}\right\} \geq c_0. \tag{12}$$

Theorem 3 indicates that the misclassification probability converges to zero with an exponential rate. This shows that the risk classification problem is in general easier than the risk estimation problem considered in Section 3.2, as we showed intuitively in Figure 1. In online risk monitoring, we are often interested more in risk classification than in risk estimation; Theorem 3 shows that we can often expect the proposed method to give a good risk classification even when the sample size is not large.

The logistic regression model considered in this section serves as a baseline model in this paper. It is quite straight-forward once the online risk monitoring problems are defined as we do. In Sections 4 to 6,

we consider various techniques to enhance the performance of the baseline model, based on either more advanced data analytics tools or the knowledge and the flexibility of the simulation model itself. We show that these performance enhancing techniques can indeed improve the qualities of risk estimators and risk classifiers, compared to the baseline model, and they may be used together to achieve further improvements.

4. Regularization Through Lasso

A large financial portfolio may include many financial products whose prices are affected by multiple underlying risk factors. Therefore, in a simulation study, we often simulate the dynamics of a large number of risk factors. This imposes a challenge to the online risk monitoring problems that we consider in this paper. In particular, when the number of risk factors is large, the number of predictors (i.e., the basis functions of the risk factors) used in the logistic regression model is often larger. Then, the ML estimate of the model parameters may have a large amount of variability, known as overfitting, and thus reduce the precision of risk estimation and risk classification. Furthermore, due to limited positions in relevant financial instruments, hedging or other reasons, the risk exposure to some risk factors may be quite small or even negligible at some time. This motivates us to consider regularization techniques, commonly used in statistical learning (e.g., Chapter 6 of James et al. 2013), to reduce the number of predictors in the model or to shrink the estimated coefficients. The basic idea is to tradeoff a small increase of bias to obtain a large decrease of the variance, thus improving the overall prediction accuracy of the model. A model with a smaller number of predictors also brings better model interpretability. In online risk monitoring, this is particularly important because it shows clearly to managers the risk factors that they need to monitor closely. In this section we propose to use L_1 regularization, also known as lasso, to improve the logistic regression model. Compared to other regularization techniques, such as L_2 regularization, lasso not only improves prediction accuracy, but also reduces the number of non-zero predictors, thus improving the model interpretability. More explanation on the variable-selection property of lasso is included in the online supplement.

Under the logistic regression model, i.e., Equation (3), the lasso estimator is

$$\hat{\boldsymbol{\beta}}_n^{\lambda_n} = \arg\max_{\boldsymbol{\beta}\in\Re^d} \left\{ L_n(\boldsymbol{\beta}) - \lambda_n \|\boldsymbol{\beta}\|_1 \right\}, \tag{13}$$

where $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^d |\beta_j|$ is the L_1-norm of the vector $\boldsymbol{\beta}$. Compared to the ML estimator in Equation (6), the lasso estimator adds a shrinkage penalty term $\lambda_n\|\boldsymbol{\beta}\|_1$ in the maximization problem, where $\lambda_n > 0$ is known as the tuning parameter. It is clear that the penalty term shrinks the lasso estimator towards zero when compared to the ML estimator $\hat{\boldsymbol{\beta}}_n$. In particular, $\hat{\boldsymbol{\beta}}_n^{\lambda_n} = \hat{\boldsymbol{\beta}}_n$ if $\lambda_n = 0$, and $\hat{\boldsymbol{\beta}}_n^{\lambda_n} = \mathbf{0}$ if $\lambda_n = +\infty$.

In the implementation of lasso, the tuning parameter λ_n is critical to the prediction accuracy of the model. To select a good λ_n, we suggest using k-fold cross validation with $k = 5$ or $k = 10$ to find the λ_n value that minimizes the average classification errors of the k test sets (see Chapter 5.1 of James et al. (2013) for an introduction). Once the tuning parameter is chosen, one can use all the training data to refit the lasso model, i.e., Equation (13), to find the lasso estimate $\hat{\boldsymbol{\beta}}_n^{\lambda_n}$. For a given tuning parameter, we suggest using the coordinate descent algorithm of Friedman et al. (2010) to fit the lasso model.

Similar to Section 3, in this section, we also study the consistency, asymptotic normality and the large deviation of the lasso estimators of the exceedance probability as well as the misclassification probability. We show that the lasso estimators preserve all the asymptotic properties that the ML estimators have. To study asymptotic properties of the lasso estimators, we need a stronger assumption on the Hessian matrix $H(\boldsymbol{\beta})$ than Assumption 2, which is commonly used in analyzing lasso estimators (see, for instance, Fan and Li 2001). In the assumption, we let $H_{jk}(\boldsymbol{\beta})$ and β_l denote the (j,k)th element of $H(\boldsymbol{\beta})$ and lth element of $\boldsymbol{\beta}$, respectively, for any $1 \le j,k,l \le d$.

ASSUMPTION 5. *There exists a neighborhood of $\boldsymbol{\beta}_0$, denoted as $\mathcal{N}(\boldsymbol{\beta}_0)$, such that for any $\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}_0)$ and any $1 \le j,k,l \le d$, $\partial_{\beta_l} H_{jk}(\boldsymbol{\beta})$ exists and there exists a function M_{jkl} such that $|\partial_{\beta_l} H_{jk}(\boldsymbol{\beta})| \le M_{jkl}(\mathbf{X},Y)$ and $\mathrm{E}[M_{jkl}(\mathbf{X},Y)] < \infty$.*

We have the following lemma on the consistency and asymptotic normality of the lasso estimator $\hat{\boldsymbol{\beta}}_n^{\lambda_n}$.

LEMMA 1 **(Fan and Li (2001))**. *Suppose that Assumptions 1, 2 and 5 holds. If $\lambda_n \to 0$ as $n \to \infty$, then $\hat{\boldsymbol{\beta}}_n^{\lambda_n} \to \boldsymbol{\beta}_0$ in probability as $n \to \infty$. Furthermore, if $\sqrt{n}\lambda_n \to 0$ as $n \to \infty$, then*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^{\lambda_n} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \mathbb{N}\left(\mathbf{0}, J^{-1}\right)$$

as $n \to \infty$.

REMARK 1. Even though we attribute Lemma 1 to Fan and Li (2001), we want to point out the lemma is an adaptation of Theorem 1 of Fan and Li (2001) to our context. Therefore, for completeness, we also provide the proof of the lemma in the online supplement.

REMARK 2. Furthermore, we want to emphasize that these asymptotic results are meaningful, even though the tuning parameter λ_n is chosen by cross validation in the actual implementation of lasso. We do expect λ_n chosen by cross validation to get smaller as n increases because we have enough data to tell which predictors matter without regularization.

Let $\hat{p}_u^{\lambda_n}(\mathbf{S}_r(u)) = g(\mathbf{X}_r, \hat{\boldsymbol{\beta}}_n^{\lambda_n}) = \exp(\mathbf{X}_r^\top \hat{\boldsymbol{\beta}}_n^{\lambda_n})/(1+\exp(\mathbf{X}_r^\top \hat{\boldsymbol{\beta}}_n^{\lambda_n}))$ be the lasso estimator of the exceedance probability at any time $u \in (0,T)$ after observing $\mathbf{S}_r(u)$ as well as $\mathbf{X}_r = \mathbf{X}(\mathbf{S}_r(u))$. Similar to Theorem 1, based on the consistency and asymptotic normality of the lasso estimator (i.e., Lemma 1), by the continuous mapping theorem and the delta method (Van der Vaart 2000), we can prove the following Theorem on the consistency and asymptotic normality of $\hat{p}_u^{\lambda}(\mathbf{S}_r(u))$. We omit the proof here.

THEOREM 4. *Suppose that Assumptions 1, 2, and 5 holds. If $\lambda_n \to 0$ as $n \to \infty$, then $\hat{p}_u^{\lambda_n}(\mathbf{S}_r(u)) \to p_u(\mathbf{S}_r(u))$ in probability as $n \to \infty$. Furthermore, if $\sqrt{n}\lambda_n \to 0$ as $n \to \infty$, then*

$$\sqrt{n}\left[\hat{p}_u^{\lambda_n}(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u))\right] \xrightarrow{d} N(0,D)$$

as $n \to \infty$, where $D = c\mathbf{X}_r J^{-1} \mathbf{X}_r^\top$ and $c = \exp(2\boldsymbol{\beta}_0{}^\top \mathbf{X}_r)/(1+\exp(\boldsymbol{\beta}_0{}^\top \mathbf{X}_r))^4$.

In the following theorem, we establish the large deviation result of the lasso estimator $\hat{\boldsymbol{\beta}}_n^{\lambda_n}$. Indeed, for the theorem to hold, we need the Fisher information matrix to be positive definite. Therefore, Assumptions 2 is used. The proof of the theorem is included in the online supplement.

THEOREM 5. *Suppose that Assumptions 1, 2, and 4 hold and $\lambda_n \to 0$ as $n \to \infty$. Then, for any $\epsilon > 0$, there exists a positive constant $\bar{c}^{\lambda}(\epsilon)$ such that*

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{ \left\| \hat{\boldsymbol{\beta}}_n^{\lambda_n} - \boldsymbol{\beta}_0 \right\| > \epsilon \right\} \geq \bar{c}^{\lambda}(\epsilon).$$

REMARK 3. Notice that, by Equations (9), (6) and (13), $\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}\in\Re^d} \; L(\boldsymbol{\beta})$, $\hat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}\in\Re^d} \; L_n(\boldsymbol{\beta})$, and $\hat{\boldsymbol{\beta}}_n^{\lambda_n} = \arg\max_{\boldsymbol{\beta}\in\Re^d} \; L_n^{\lambda_n}(\boldsymbol{\beta})$, where $L_n^{\lambda_n}(\boldsymbol{\beta}) = L_n(\boldsymbol{\beta}) - \lambda_n \|\boldsymbol{\beta}\|_1$. Therefore, $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_n$ and $L_n^{\lambda_n}(\boldsymbol{\beta})$ are all solutions to corresponding optimization problems. The proofs of both Theorems 2 and 5 depend critically on Proposition 4.32 in Bonnans and Shapiro (2000), which bounds the differences between the optimal solutions, i.e., $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$ and $\|\hat{\boldsymbol{\beta}}_n^{\lambda_n} - \boldsymbol{\beta}_0\|$, by the differences between the objective functions, i.e., $\|L_n(\boldsymbol{\beta}) - L(\boldsymbol{\beta})\|$ and $\|L_n^{\lambda_n}(\boldsymbol{\beta}) - L(\boldsymbol{\beta})\|$, respectively. For more details, one may refer to the online supplement.

Similar to the proof of Theorem 3, we can prove the following theorem on the probability of a large deviation of the estimated exceedance probability from its true value, i.e., $\Pr\left\{ \left| \hat{p}_u^{\lambda_n}(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u)) \right| > \delta \right\}$ for any $\delta > 0$, and the probability of misclassification, i.e., $\Pr\{\hat{I}^{\lambda_n} \neq I\}$. We omit the proof here.

THEOREM 6. *Suppose that Assumptions 1, 2, and 4 hold. If $\|\mathbf{X}_r\| \neq 0$, then for any $\epsilon > 0$, there exists $\tilde{c}^{\lambda}(\epsilon) > 0$ such that*

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{ \left| \hat{p}_u^{\lambda_n}(\mathbf{S}_r(u)) - p_u(\mathbf{S}_r(u)) \right| > \delta \right\} \geq \tilde{c}^{\lambda}(\delta).$$

Furthermore, if $\|\mathbf{X}_r\| \neq 0$ and $p_u(\mathbf{S}_r(u)) \neq \alpha$, then there exists a constant $c_0^{\lambda} > 0$ such that

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left\{ I \neq \hat{I}^{\lambda_n} \right\} \geq c_0^{\lambda}. \tag{14}$$

REMARK 4. We prove Theorems 5 and 6 in our context. However, these results are of interest beyond the context of online risk monitoring, because logistic regression and lasso are widely used in other contexts as well.

5. Variance Reduction Through Perturbation

So far in this paper we have taken a data analytics approach to analyzing the simulation output through logistic regression for risk estimation and risk classification. However, there is a fundamental difference between typical data analytics problems and simulation analytics problems. In simulation analytics problems, the analysts know the underlying data generating processes (i.e., the simulation model) and therefore have much more information than their peers in solving other types of analytics problems. In this section we ask how we might be able to use the information to develop more efficient tools for risk estimation and risk classification. In particular, we propose a method that perturbs the sample paths generated by the simulation

study and then uses a first-order Taylor expansion to approximate the portfolio value under the perturbed sample paths. We show that the ML estimators of the logistic regression parameters using the perturbed sample paths have smaller variances than those using the original sample paths, at least when the time is close to 0.

5.1. Basic Idea of the Perturbation Method

Notice that, for the logistic regression model to work, we only need to ensure that the sample paths deliver the correct default probabilities when conditioned on $\mathbf{S}(u)$ (see Assumption 1). Therefore, we only need to make sure that $\mathbf{S}(t)$ follows the correct dynamics when $t \in (u, T]$. This motivates us to perturb the initial values $\mathbf{S}(0)$ of the sample paths so that the correct dynamics are used for all $t \in (0, T]$ but the sample paths are more spread out. We can then use the new sample paths to fit the logistic regression model for any $u \in (0, T)$.

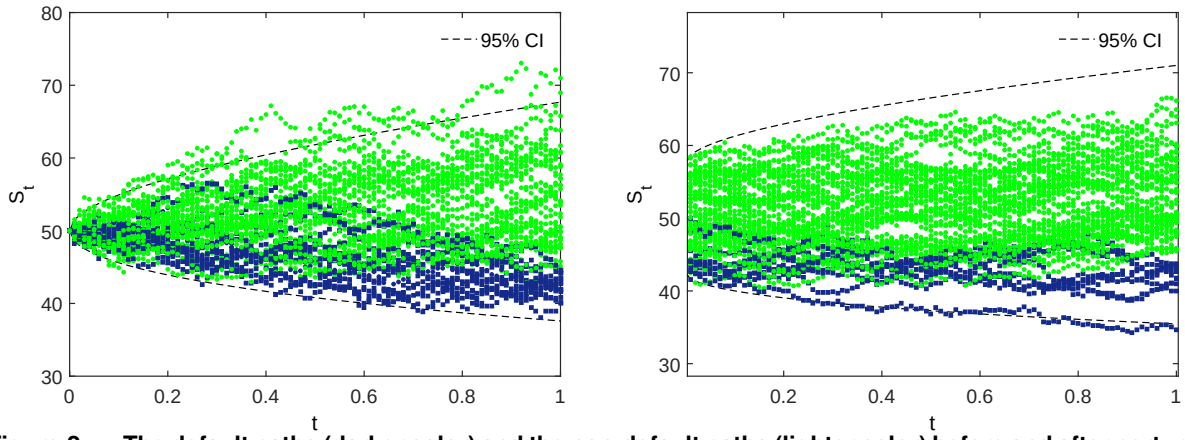

Figure 2 The default paths (darker color) and the non-default paths (lighter color) before and after perturbing the initial value

Consider a simple example where there is only one risk factor that is modeled as a geometric Brownian motion. In the left panel of Figure 2, we plot the generated sample paths. We color a path darker if it leads to default and lighter otherwise, and we also plot the 95% range of the data. From the plot we see that the paths with different colors are easier to differentiate when the time u is large because they are more separated, but more difficult to differentiate when u is small because they are more clustered together. Suppose that, instead of simulating all sample paths of risk factor from the initial value $S(0)$, we perturb it by a small amount, say $S(0)(1 + \delta U)$, where $\delta > 0$ is the perturbation size and $U \in (-1, 1)$ can be chosen deterministically or randomly, and simulate sample paths from the new initial values. We can then evaluate the portfolio risk at T and color the sample paths, as in the right panel of Figure 2. From the plot we see that the sample paths with different colors are more separated, especially when u is small. Therefore, intuitively, we would expect the perturbed sample paths lead to better risk estimation and risk classification. In the left and right panels of Figure 3, we plot the estimated risk boundaries for this one dimensional example using the original and perturbed sample paths, respectively. From the plots we see clearly that the perturbed sample paths lead to better risk estimation and risk classification.

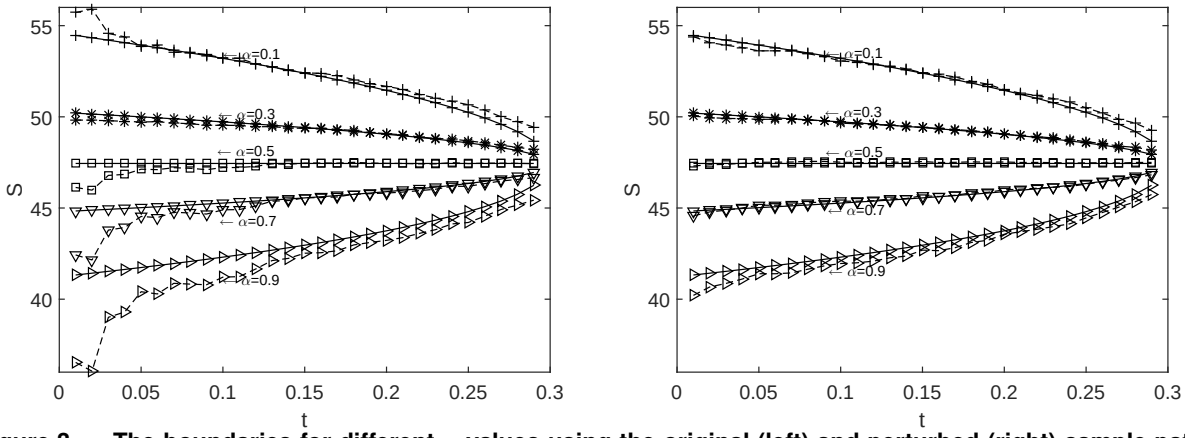

Figure 3 **The boundaries for different α values using the original (left) and perturbed (right) sample paths. The solid lines are the true boundaries, and the dashed lines are the estimated boundaries via logistic regression**

The perturbation method can be extended easily to multi-dimensional risk factors by perturbing $S_i(0)$ to $S_i(0)(1+\delta_i U_i)$ for any risk factor $i = 1, 2, \ldots, m$, where $U_1, \ldots, U_m \in (-1, 1)$. In Section 5.2 we show that, under the assumption that the sample paths are simulated from the perturbed initial values, the perturbation method works well when u is small. In particular, when $u \to 0$, the variance reduction ratio, i.e., the variance of the original estimator over that of the new estimator, goes to infinity. In Section 5.3 we show how to implement the perturbation method using the existing sample paths.

5.2. Analysis of the Perturbation Method

Let $\mathbf{S}'(u) = (S'_1(u), \ldots, S'_m(u))^\top$ denote the perturbed sample paths for any $u \in [0, T]$, where $S'_i(u)$ is simulated from a randomly perturbed initial value $S_i(0)(1 + \delta_i U_i)$ with $\delta_i > 0$ and $U_i \in (-1, 1)$ chosen determinsitcally or randomly for all $i = 1, 2, \ldots, m$. Let $\mathbf{X}'(u) = \mathbf{X}(\mathbf{S}'(u))$, let $\hat{\boldsymbol{\beta}}'_n$ denote the ML estimator of $\boldsymbol{\beta}_0$ of the logistic regression model (3) using the perturbed sample paths, and $J'(u)$ denote the Fisher information matrix when using the perturbed data. By Lemma C.1 in the online supplement, we have $\sqrt{n}[\hat{\boldsymbol{\beta}}'_n - \boldsymbol{\beta}_0] \xrightarrow{d} \mathbb{N}(\mathbf{0}, J'^{-1}(u))$ as $n \to \infty$ for any $u \in (0, T)$.

Notice that, for the original sample paths, all observations have the same $\mathbf{S}(0)$ value thus same $\mathbf{X}(\mathbf{S}(0))$ value at time 0. Then, intuitively, the logistic regression model cannot be fitted if $\mathbf{X}$ has more than one dimension, i.e., $d \geq 2$. For the perturbed data, however, all observations have different $\mathbf{S}'(0)$ values thus different $\mathbf{X}'(0)$ values. Therefore, the logistic regression model may be fitted if there are enough observations. This intuitively explains why the perturbation method works well at least when u is close to 0. In the rest of this subsection, we prove this intuition in a rigorous way.

To simplify the notation, we let $\mathbf{X}(u) = \mathbf{X}(\mathbf{S}(u))$ and let $\mathbf{X}(0) = \mathbf{x}$ to denote it is a deterministic vector. Furthermore, let $J_{ii}^{-1}(u)$ and $J'^{-1}_{ii}(u)$ denote the ith diagonal elements of $J^{-1}(u)$ and $J'^{-1}(u)$, respectively, for any $u \in (0, T)$. Notice that they are the asymptotic variances of $\hat{\beta}_i$ and $\hat{\beta}'_i$, respectively, for all $i = 1, 2, \ldots, d$. Let $r_i(u) = J_{ii}^{-1}(u)/J'^{-1}_{ii}(u)$ denote the asymptotic variance reduction ratio of the perturbation method for estimating β_{0i} for any $u \in (0, T)$. The following theorem shows that at least $d - 1$ terms of

$r_1(u), \ldots, r_d(u)$ go to infinity if u goes down to zero. Therefore, when u is close to zero, the perturbation method leads to a significant variance reduction effect compared to using the original sample paths. The proof is given in the online supplement.

THEOREM 7. *Suppose that Assumptions 1 and 2 are satisfied for any $u \in (0,T)$. Furthermore, suppose that $\mathbf{X}(u)$ is right continuous a.s. at $u=0$, $\sup_{u\in(0,T)} \mathbb{E}[\|\mathbf{X}(u)\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$, and $J'(0)$ is of full rank. Then, $r_i(u) \to \infty$ as $u \to 0^+$ for at least $d-1$ values of $i \in \{1,2,\ldots,d\}$, where $u \to 0^+$ means u goes to zero from the positive side.*

5.3. Implementation of the Perturbation Method

To implement the perturbation method, we need to construct the sample paths from randomly perturbed initial values using the original sample paths and also evaluate the portfolio loss for each newly constructed sample path. In some cases, these tasks can be done exactly. But in general cases, these can only be done approximately. First, we consider how to construct sample paths. In many models of risk factors, including the geometric Brownian motion model, Heston's stochastic volatility model, Merton's jump diffusion model, and others (see Hull 2014), $S_j(u)/S_j(0)$ is not a function of $S_j(0)$. For these models, one can construct the new sample path of risk factor j exactly by setting

$$S_j'(u) = \frac{S_j(u)}{S_j(0)} S_j'(0) = S_j(u)[1+\delta U_j] \tag{15}$$

for any $u \in (0,T]$. For more general models of risk factors, we assume that the derivative process $dS_j(u)/dS_j(0)$ is available along with the sample path $S_j(u)$. Notice that $dS_j(u)/dS_j(0)$ is known as the sample path derivative. It is often very easy to simulate and requires very little additional effort, when it is simulated along with $S(u)$ (Broadie and Glasserman 1996). Then, we may apply Taylor's first-order approximation and approximate $S'(u)$ by

$$S_j'(u) \approx S_j(u) + \frac{dS_j(u)}{dS_j(0)} \left[S_j'(0) - S_j(0)\right] = S_j(u) + \frac{dS_j(u)}{dS_j(0)} S_j(0)\delta U_j \tag{16}$$

for any $u \in (0,T]$.

Second, we consider how to reconstruct the portfolio loss. By Equation (1), the portfolio loss satisfies $L(T) = \Phi(0) - \Phi(T) = \Phi(0) - \mathbf{w}\mathbf{V}(T) = \Phi(0) - \sum_{i=1}^k w_i V_i(T)$. Then to reconstruct the portfolio loss for the new sample paths, we need to compute $\mathbf{V}'(T)$, which includes the values of the financial products at time T. Notice that $\mathbf{V}'(T)$ depends on $\mathbf{S}'(T)$. If there exists closed-form expressions, e.g., Black-Scholes formula for European call and put options, then $\mathbf{V}'(T)$ may be computed analytically. Otherwise, we assume that the price sensitivities of $V_i(T)$ with respect to $S_j(T)$, i.e., $\partial V_i(T)/\partial S_j(T)$, are available for all $i = 1,2,\ldots,k$ and $j = 1,2,\ldots,m$. Indeed, $dV_i(T)/dS_j(T)$ are known as the Greeks in financial risk management and they typically calculated when evaluating $V_i(T)$ (see, for instance, Broadie and Glasserman (1996) and

Liu and Hong (2011) on simulating the Greeks). Then, we may apply Taylor's first-order approximation to approximate $V_i'(T)$ by

$$V_i'(T) \approx V_i(T) + \sum_{j=1}^{m} \frac{\partial V_i(T)}{dS_j(T)} \left[S_j'(T) - S_j(T)\right],$$

where $S_j'(T) - S_j(T)$ may be computed using either Equation (15) or (16) by setting $u = T$.

Lastly, we discuss how to generate the initial values $\mathbf{S}'(0)$. Based on our current construction, $S_j'(0) = S_j(0)[1 + \delta U_j]$, where $\mathbf{U} = (U_1, \ldots, U_m)^\top$ is uniformly distributed in the hyperbox $(-1, 1)^m$. However, the randomness in $\mathbf{U}$ also introduces additional variance in the estimation of $\hat{\boldsymbol{\beta}}_n'$. This motivates us to use low dispersion sequences to spread out the observations of $\mathbf{U}$ more evenly in the hypercube to reduce the variance of the ML estimator. In our implementations, we use the Sobol sequence (Niederreiter 1988). In the online supplement, we compare the uniformness between two-dimensional uniform random vector and the Sobol sequence.

6. Additional Simulation

In previous sections, we assumed that all sample paths are generated at time 0. In some practical situations, one may be able to simulate more sample paths at some time point between 0 and T. For instance, when there is a weekend, one may decide to add more sample paths simulated from the current values of the risk factors. In this section, we consider how to incorporate the additional data into the analysis to improve the quality of the risk estimators and risk classifiers. In particular, we consider two methods. In the first method, we use the two sets of data to generate two estimators separately and then combine the two estimators; and in the second method, we combine the two sets of data together and use logistic regression on the combined data set to obtain a new estimator. We prove that both methods lead to estimators that have smaller variances than the one without using the additional data. The numerical experiments conducted in Section 7.2 show that both methods have comparable performance in both risk estimations and risk classifications. However, compared to the first method, the second one is easier to extend to the situations where there are multiple sets of additional data (simulated at different time points) and is easier to implement. Therefore, we prefer and recommend the second method.

6.1. Combining Estimators

Let $\{\tilde{\mathbf{S}}(t), u' \le t \le T\}$ be the additional sample paths generated from the time point $u' \in (0, T)$, and $\tilde{Y}$ be the corresponding exceendance indicator, defined similarly as in Equation (2). For any $u \in [u', T]$, one may obtain two estimators, denoted as $\hat{p}_u^0$ and $\hat{p}_u^1$, based on the original data and the additional data, respectively. Then, we may use the following combined estimator

$$\tilde{p}_u = \nu_0 \hat{p}_u^0 + \nu_1 \hat{p}_u^1,$$

where $\nu_0 > 0$ and $\nu_1 > 0$ are the corresponding weights with $\nu_0 + \nu_1 = 1$.

Notice that the original sample paths and the additional sample paths are independent. Then,

$$\mathrm{Var}\left[\tilde{p}_u\right] = \mathrm{Var}\left[\nu_0 \hat{p}_u^0 + \nu_1 \hat{p}_u^1\right] = \nu_0^2 \mathrm{Var}\left[\hat{p}_u^0\right] + \nu_1^2 \mathrm{Var}\left[\hat{p}_u^1\right].$$

Therefore, we propose to find the optimal weights ν_0^* and ν_1^* to minimize the variance of the combined estimator, i.e., to solve the following optimization problem:

$$\begin{aligned} \min_{\nu_0,\nu_1} \quad & \nu_0^2 \mathrm{Var}\left[\hat{p}_u^0\right] + \nu_1^2 \mathrm{Var}\left[\hat{p}_u^1\right], \\ \text{s.t.} \quad & \nu_0 + \nu_1 = 1, \\ & 0 \le \nu_0, \nu_1 \le 1. \end{aligned}$$

It is easy to see that the optimal weights are

$$w_0^* = \frac{\mathrm{Var}\left[\hat{p}_u^1\right]}{\mathrm{Var}\left[\hat{p}_u^0\right] + \mathrm{Var}\left[\hat{p}_u^1\right]} \text{ and } w_1^* = \frac{\mathrm{Var}\left[\hat{p}_u^0\right]}{\mathrm{Var}\left[\hat{p}_u^0\right] + \mathrm{Var}\left[\hat{p}_u^1\right]}. \tag{17}$$

In addition, it is clear that the combined estimator with the optimal weights has a smaller variance to both $\hat{p}_u^0$ and $\hat{p}_u^1$ have. Therefore, by adding additional sample paths, one can improve the quality of the original estimator $\hat{p}_u^0$. To use the combined estimator in practice, one first estimates the variances of $\hat{p}_u^0$ and $\hat{p}_u^1$ using the corresponding sample paths, and then uses these variance estimates to approximate the optimal weights and to calculate the combined estimator.

6.2. Combining Data

Notice that the original sample paths $\{\mathbf{S}(t), 0 \le t \le T\}$ and the additional sample paths $\{\tilde{\mathbf{S}}(t), u' \le t \le T\}$ follow the same simulation model from u' to T, conditioned on $\mathbf{S}(u')$ and $\tilde{\mathbf{S}}(u')$ respectively. Then, for any $u \in [u', T]$, given $\mathbf{S}(u) = \mathbf{s}$ and $\tilde{\mathbf{S}}(u) = \mathbf{s}$, the probability mass functions of Y and $\tilde{Y}$ are the same. Based on the logistic regression model, we have

$$\Pr\{Y = 1 | \mathbf{S}(u) = \mathbf{s}\} = \Pr\{\tilde{Y} = 1 | \tilde{\mathbf{S}}(u) = \mathbf{s}\} = \frac{e^{\boldsymbol{\beta}_0^\top \mathbf{X}(\mathbf{s})}}{1 + e^{\boldsymbol{\beta}_0^\top \mathbf{X}(\mathbf{s})}}.$$

Therefore, the likelihood functions of these two data sets are the same, and we may conduct logistic regression on the combined dataset. In the rest of this subsection, we show that, by combing the two datasets together, we may estimate the exceedance probabilities with smaller asymptotic variances, compared to the original estimator without adding the additional sample paths.

Recall that $\mathbf{X}$ denotes $\mathbf{X}(\mathbf{S}(u))$. For the convenience of notation, we also use $\tilde{\mathbf{X}}$ to denote $\mathbf{X}(\tilde{\mathbf{S}}(u))$ when $u \in [u', T]$ is fixed. By Theorem 1, we know that the asymptotic variance of the estimator $\sqrt{n}\,\hat{p}_u^0$ is $D = c\mathbf{X}_r^\top J^{-1} \mathbf{X}_r$, where n is the original sample size, $c = \exp(2{\boldsymbol{\beta}_0}^\top \mathbf{X}_r)/(1 + \exp({\boldsymbol{\beta}_0}^\top \mathbf{X}_r))^4$ with $\mathbf{X}_r = \mathbf{X}(\mathbf{S}_r)$, and $J = \mathbb{E}[\exp(\boldsymbol{\beta}_0^\top \mathbf{X})/(1 + \exp(\boldsymbol{\beta}_0^\top \mathbf{X}))^2 \mathbf{X}\mathbf{X}^\top]$. So we can use D/n to approximate the variance of the estimator $\hat{p}_u^0$. Similarly, let $\hat{p}_u^1$ denote the estimator derived from the additional simulation data, then its variance is approximately $\tilde{D}/\tilde{n}$, where $\tilde{n}$ is the sample size of the additional data and $\tilde{D} = c\mathbf{X}_r^\top \tilde{J}^{-1} \mathbf{X}_r$ with $\tilde{J} = \mathbb{E}[\exp(\boldsymbol{\beta}_0^\top \tilde{\mathbf{X}})/(1 + \exp(\boldsymbol{\beta}_0^\top \tilde{\mathbf{X}}))^2 \tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top]$. Let $\tilde{p}_u$ denote the estimator derived from the combined dataset.

Consider the likelihood function on the combined dataset

$$l_{n+\tilde{n}}(\boldsymbol{\beta}) = \prod_{i=1}^{n} g(\mathbf{X}_i, \boldsymbol{\beta})^{Y_i} \left(1 - g(\mathbf{X}_i, \boldsymbol{\beta})\right)^{1-Y_i} \prod_{j=1}^{\tilde{n}} g(\tilde{\mathbf{X}}_j, \boldsymbol{\beta})^{\tilde{Y}_j} \left(1 - g(\tilde{\mathbf{X}}_j, \boldsymbol{\beta})\right)^{1-\tilde{Y}_j},$$

and recall that $g(\mathbf{X},\boldsymbol{\beta})=\exp(\boldsymbol{\beta}^\top\mathbf{X})/(1+\exp(\boldsymbol{\beta}^\top\mathbf{X}))$, so the averaged log likelihood function is

$$L_{n+\tilde{n}}(\boldsymbol{\beta}) = \frac{1}{n+\tilde{n}}\Bigg\{\sum_{i=1}^{n}\left[Y_i\boldsymbol{\beta}^\top\mathbf{X}_i-\log\left(1+\exp(\boldsymbol{\beta}^\top\mathbf{X}_i)\right)\right] + \sum_{j=1}^{\tilde{n}}\left[\tilde{Y}_j\boldsymbol{\beta}^\top\tilde{\mathbf{X}}_j-\log\left(1+\exp(\boldsymbol{\beta}^\top\tilde{\mathbf{X}}_j)\right)\right]\Bigg\}.$$

Suppose that there is a sampling regime that makes the additional simulation data size $\tilde{n}=\tilde{n}(n)$, which is related to the original simulation data size n, and $\lim_{n\to\infty}\tilde{n}(n)/n\to r$. Under this regime, we only need to focus on the asymptotic analysis with $n\to\infty$.

Notice that $(\mathbf{X}_i,Y_i), i=1,2,\ldots,n$ are i.i.d., and $(\tilde{\mathbf{X}}_j,\tilde{Y}_j), j=1,2,\ldots,\tilde{n}$ are i.i.d., then according to Theorems 1 and 2 in Hoadley (1971), the Fisher information matrix is given by

$$\bar{J}=\frac{1}{1+r}J+\frac{r}{1+r}\tilde{J}. \tag{18}$$

Then, the variance of $\tilde{p}_u$ is approximated by $\bar{D}/(n+\tilde{n})$, where $\bar{D}=c\mathbf{X}_r^\top\bar{J}^{-1}\mathbf{X}_r$.

To compare the the asymptotic variance of $\tilde{p}_u$ with those of $\hat{p}_u^0$ and $\hat{p}_u^1$, we need the following lemma, whose proof is provided in the online supplement.

LEMMA 2. *Suppose that Σ_1 and Σ_2 are both $d\times d$ positive definite matrices, and $\Sigma=\Sigma_1+\Sigma_2$. For every non-zero column vector $\mathbf{X}_r\in\Re^d$,*

$$\mathbf{X}_r^\top\Sigma^{-1}\mathbf{X}_r<\mathbf{X}_r^\top\Sigma_i^{-1}\mathbf{X}_r, i=1,2. \tag{19}$$

Based on Lemma 2, we have the following proposition that shows $\bar{D}/(n+\tilde{n})$ is smaller than those of D/n and $\tilde{D}/\tilde{n}$, where D/n and $\tilde{D}/\tilde{n}$ approximate the variances of $\hat{p}_u^0$ and $\hat{p}_u^1$, respectively. The proof is given in the online supplement.

PROPOSITION 1. *Let $r=\tilde{n}/n$. If J and $\tilde{J}$ are positive definite, then $\bar{D}/(n+\tilde{n})<\min\{D/n,\tilde{D}/\tilde{n}\}$.*

7. Numerical Examples

In this section we conduct a sequence of numerical experiments to study the behavior of our proposed methods in online risk estimation and classification, and to compare their performance. Notice that, in this paper, we focus on designing simulation methods. Therefore, we assume that simulation models are valid models without misspecification errors. For calibrating simulation models and discussions on model misspecification, readers are referred to Alexander (2001), Belomestny and Reib (2006), and White (1982).

In Sections 7.1 to 7.3, we use only the linear basis functions in logistic regression, and in Section 7.4, we use both linear and individual quadratic basis functions. In all these examples, let $\lambda_n=\lambda/n$, and we use 5-fold cross validation to determine the tuning parameter λ when implementing lasso, and use the perturbation size $\delta=5\%$ when implementing the perturbation method. Furthermore, for all examples, we assume that there are 10000 simulated sample paths, and for each sample path, 100 inner-level simulation

replications are used to determine the values and the Greeks of the derivatives in the portfolio. Some of the results reported in Sections 7.1 and 7.3 are also available in Jiang et al. (2016). We add more results on the use of the enhancing techniques and the comparisons.

7.1. A Portfolio with 5 Underlying Assets

Consider a portfolio that longs 3 call options and 2 put options based on 5 different underlying assets, which are mutually independent and driven by geometric Brownian motions (GBMs). Specifically, let $\mathbf{S}(t) = (S_1(t), \ldots, S_5(t))^\top$, and the portfolio $\Phi(t) = V_1^c(t) + V_2^c(t) + V_3^c(t) + V_4^p(t) + V_5^p(t)$. At time T, if the portfolio $\Phi(T) \le Q$, i.e., the loss $\Phi(0) - \Phi(T) \ge \Phi(0) - Q$, we consider the portfolio in default. Let the initial values of the five underlying assets $\mathbf{S}(0) = (50, 50, 60, 60, 70)^\top$, the drifts of the GBMs $\boldsymbol{\mu} = (0.05, 0.06, 0.07, 0.06, 0.05)^\top$, the volatilities $\boldsymbol{\sigma} = (0.1, 0.1, 0.1, 0.1, 0.1)^\top$, and the strikes of the options $\mathbf{K} = (40, 40, 45, 80, 85)^\top$. Let the risk-free interest rate $r_f = 0.02$ and the portfolio loss evaluation time $T = 0.3$. Notice that there are two probability measures in our problem. The price dynamics of $\mathbf{S}(t)$ with $0 \le t \le T$ are simulated under the real probability measure where $\boldsymbol{\mu}$ are used as the drifts of the GBMs, while the options are evaluated under the risk neutral measure where r_f is used as the drift of the GBMs. Suppose that $[0, T]$ is discreted into $N = 30$ equal-length intervals, we denote them by $0 = t_0 \le t_1 \le \cdots \le t_N = T$. Let the maturities of all the options be the same $\tau = 1$, and let $Q = 65$.

Even though the option prices and their deltas (i.e., price sensitivities w.r.t. $\mathbf{S}(T)$) may be calculated easily by Black-Scholes formula, we still use simulation to price them to mimic the more complicated situations. In our logistic regression model, we choose the basis function $\mathbf{X}(\mathbf{S}(t)) = (1, \mathbf{S}(t))^\top$. Furthermore, when applying the perturbation method, we use Equations (15) and (16) to approximate the perturbed risk factors and the corresponding options prices, respectively.

Because the distribution of $\mathbf{S}(T)$ is directly available given $\mathbf{S}_r(t)$ and the portfolio value $\Phi(T)$ may be calculated easily using Black-Scholes formula given $\mathbf{S}(T)$, we may evaluate the true value $p(\mathbf{S}_j(t))$ accurately using a large number of scenarios (e.g., 10^4 in our calculation) conditioning on any $\mathbf{S}_r(t)$. We can then use the true value as the benchmark to evaluate the performance of our proposed methods.

We plot the standard deviations of the all six parameter estimators in the logistic regression model, with and without the perturbation method, in Figure 4. We see that the perturbation method clearly reduces the variances. However, variance reduction diminishes as t increases. Furthermore, we see that the variances of the estimators of both methods may increase as t approaches T. This is because, when t is close to T, the exceedance probabilities are likely to be close to either 0 or 1, leading the parameters of the logistic regression models to very large values, thus increasing the variances of the estimators. Indeed, this phenomenon exists not only in this example, it exists in all three examples that we consider. In Figure 4, we did not give the standard deviations of the lasso and lasso+perturbation estimators because the number of risk factors in this example is small and lasso does not provide benefits.

Figure 4 provides useful insight. However, it does not tell us how good our proposed method is for risk estimation and classification. To evaluate the quality of risk estimation (e.g., predicting exceedance

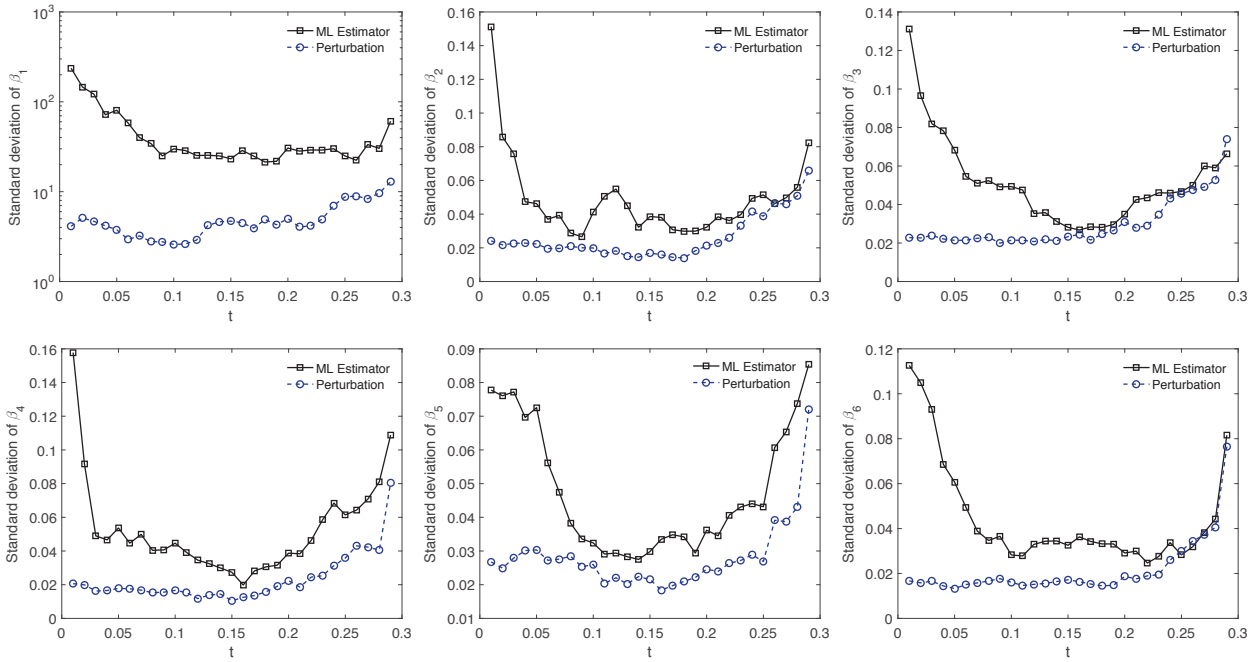

Figure 4 Standard deviations of the ML and perturbation estimators in Example 1

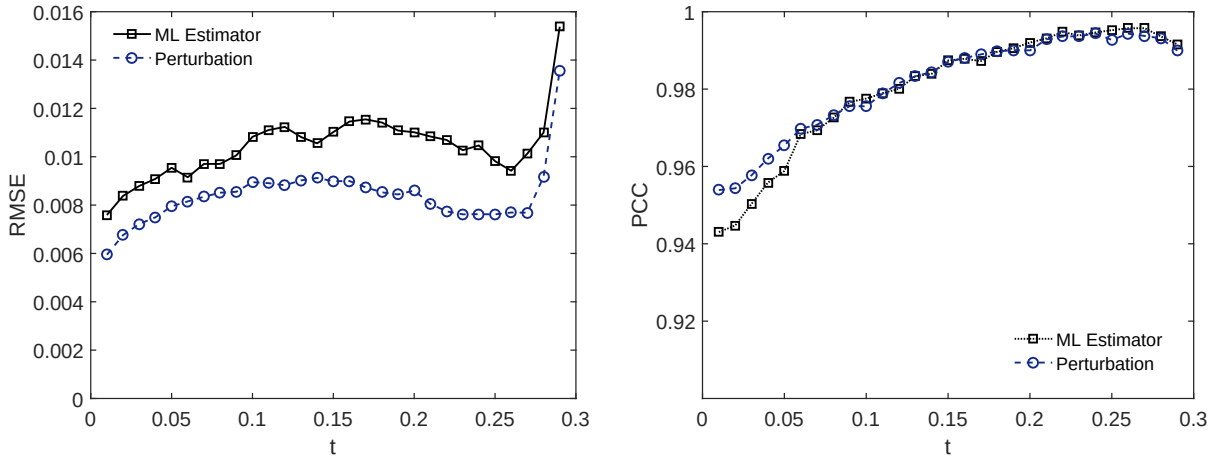

Figure 5 RMSE (left panel) and PCC (right panel) of the ML and perturbation estimators in Example 1

probabilities), we use the root-mean-square error (RMSE) as the criterion for the estimated probabilities at different time points. The RMSE is calculated as

$$\text{RMSE}(t) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\frac{1}{J}\sum_{j=1}^{J}[\hat{p}_l(\mathbf{S}_j(t)) - p(\mathbf{S}_j(t))]^2},$$

where L is the number of training sets and J is the number of testing sets for each training set, i.e., for each training set l, we get an estimated probability function $\hat{p}(\cdot)$, and we generate J testing sets with M sample paths to calculate the RMSE for this estimated probability throughout $t_i, i = 1, \ldots, N-1$, then we replicate L times to calculate the average. The RMSE basically tells us the average errors of the probability

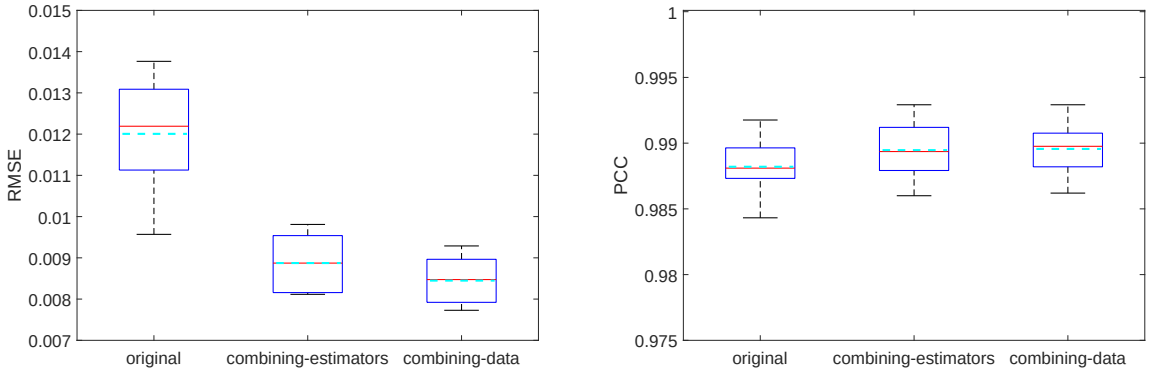

Figure 6 **Box plots of RMSEs and PCCs for the estimators of the original data, the combining-estimators method and the combining-data method based on** 40 **replications.** $u = 0.20$**,** $u' = 0.16$

estimates. To evaluate the quality of risk classification, we use the probability of correct classification (PCC) as the criterion. In the numerical study, we set $L = J = 10$, and $M = 500$, and plot the RMSEs and PCCs in Figure 5. From the figure, we see that the RMSEs are small (about 0.01) and the PCCs are large (mostly above 0.95) for all proposed methods, indicating that the risk estimations and classifications may be done precisely. Moreover, the perturbation method can improve the quality of risk estimation and classification, especially when t is small.

7.2. Additional Simulation

We consider the same example in Section 7.1. In addition, we generate some new sample paths $\tilde{\mathbf{S}}(u), u \in [u', T]$ from $u' \in (0, T)$ based on the observed values of $\tilde{\mathbf{S}}(u')$. We let $\tilde{n} = 1000$, which is 10% of the original sample size n.

We first consider the case where $u' = 0.16$ and $u = 0.20$, with all other parameters remaining the same as in Section 7.1. Notice that $u > u'$, which means the additional data were generated before the prediction time u. We plot the results in Figure 6. In this figure, the left, middle and right box plots of both panels are from the estimators of the original data, the combining-estimators method and the combining-data method, respectively. This figure shows that, with only 10% additional sample paths, the qualities of risk estimators are significantly improved, while the qualities of risk classifiers are slightly improved. Moreover, the estimators of the two methods with additional data have comparable performances.

Further, we study the improvement at different prediction time u with $u' = u - 0.04$, and summarize the results in Table 1. In this table, we see that both of the estimators with additional data have similar performances, and both outperform the one with only the original data, especially in risk estimations.

Next, we consider the case $u = u'$. By setting different values of u, we report the results in Table 2. Similar to the results in Table 1, both of the estimators with additional data have similar performances, and both significantly outperform the one with only the original data.

These numerical results show that both methods with additional data have comparable performances in risk estimations and classifications. However, the method of combining data is in general easier to implement than the method of combining estimators. Moreover, it can be extended easily to situations where

Table 1 RMSE and PCC for different parameter u with $u' = u - 0.04$

	original		combining estimators		combining data	
	RMSE	PCC	RMSE	PCC	RMSE	PCC
$u = 0.05$	0.0098	0.962	0.0087	0.965	0.0087	0.965
$u = 0.10$	0.0119	0.974	0.0093	0.977	0.0090	0.976
$u = 0.15$	0.0130	0.982	0.0096	0.984	0.0091	0.985
$u = 0.20$	0.0120	0.988	0.0089	0.990	0.0084	0.990
$u = 0.25$	0.0112	0.988	0.0083	0.989	0.0074	0.989

Table 2 RMSE and PCC for different u

	original		combining estimators		combining data	
	RMSE	PCC	RMSE	PCC	RMSE	PCC
$u = 0.05$	0.0096	0.964	0.0063	0.971	0.0063	0.969
$u = 0.10$	0.0105	0.979	0.0062	0.981	0.0063	0.981
$u = 0.15$	0.0131	0.983	0.0066	0.987	0.0066	0.986
$u = 0.20$	0.0109	0.987	0.0057	0.988	0.0059	0.987
$u = 0.25$	0.0107	0.990	0.0066	0.988	0.0068	0.989

multiple batches of data are added at different time points, as we only need to combine all data together and conduct logistic regression once. Therefore, we prefer and recommend the method of combing data.

7.3. Adding Perfectly Hedged Positions

In this example, we add another 20 underlying assets to Example 1 in Section 7.1. For each of the new underlying assets, we long one call option, and short one put option and the corresponding underlying asset. Then, by put-call parity (Hull 2014), the newly added positions are perfectly hedged and have a deterministic value. We further add a proper position of cash so that the value of the new portfolio is the same as that of Example 1. Therefore, the exceedance probability in this example is exactly same as that in Example 1 and it is not affected by the newly added underlying assets. We apply our proposed methods on all underlying assets, old and new, and our goal is to test the performance of the lasso and lasso+perturbation estimators.

We model the dynamics of all underlying assets as GBMs. More specifically, we now call the five assets used in Example 1 as $\mathbf{S}_1(t)$ and the new assets as $\mathbf{S}_2(t)$, and $\Phi(t) = \Phi_1(t) + \Phi_2(t)$. The parameters of $\mathbf{S}_2(t)$ (the initial values, drifts, volatilities, and correlations) and the strike prices $K_{2,i}$ are given in Table 1 in the online supplement. All other parameters remain the same as that in Example 1. Therefore, the exceedance probabilities remain the same. In this example we let the basis function $\mathbf{X}(\mathbf{S}(t)) = (1, \mathbf{S}(t))^\top = (1, S_{1,1}, \ldots, S_{1,5}(t), S_{2,1}(t), \ldots, S_{2,20}(t))^\top$.

In Figure 7 we plot the standard deviations of the estimators of $\beta_1, \ldots, \beta_6$, and in Figure 8 we plot the RMSEs of $\beta_7, \ldots, \beta_{26}$, because we know their true values are 0. From these figures, we see that the perturbation method reduces variances of all estimators. However, lasso reduces the variances of the intercept β_1 and $\beta_i, i = 7, \ldots, 26$, without reducing the variances of $\beta_2, \ldots, \beta_6$.

We also plot the RMSEs and PCCs in Figure 9. When compared to Figure 5, we find that the RMSEs and PCCs of the ML estimators are higher in this example than in Example 1, indicating that the added risk factors introduce more noise and make the risk estimation and classification more difficult. However, lasso,

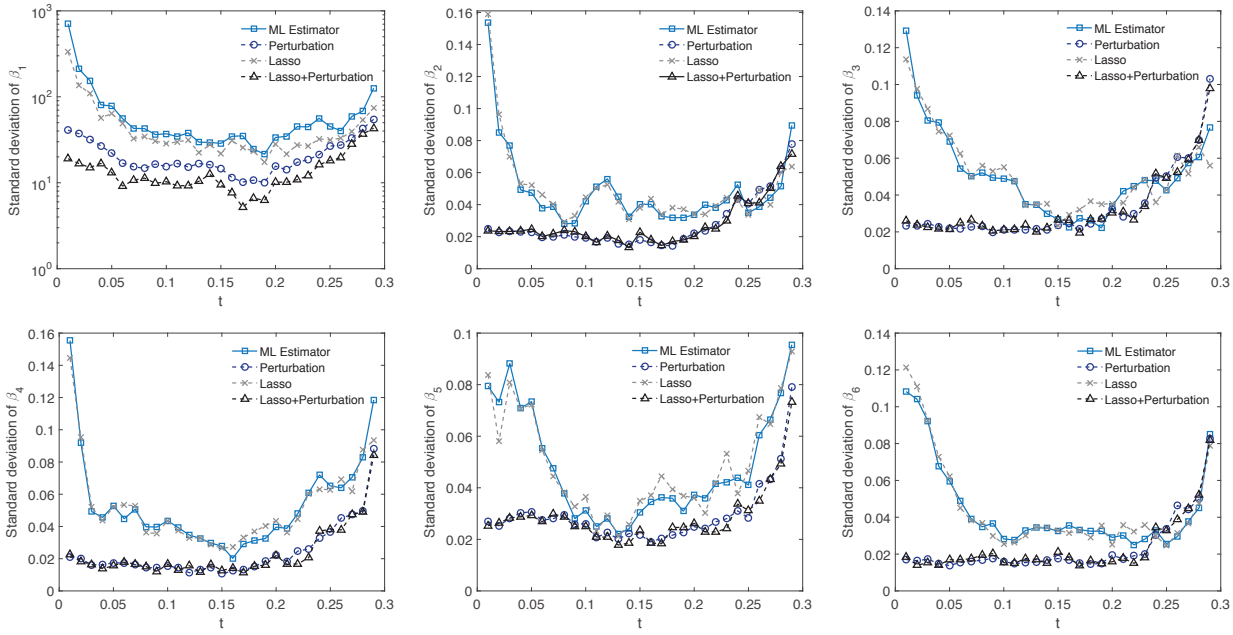

Figure 7 **Standard deviations of the ML, lasso, perturbation and lasso+perturbation estimators of $\beta_1, \ldots, \beta_6$ in Example 2**

perturbation and lasso+perturbation estimators can all improve the RMSEs and PCCs. In particular, the lasso+perturbation estimators perform very well, making the RMSEs and PCCs almost as good as those in Example 1.

7.4. A Portfolio with 80 Risk Factors

In this example we consider a portfolio with 60 underlying assets, and some of them are correlated. The portfolio has two groups of products. The first group longs one call option and one put option for each of the first 40 underlying assets, which are modeled by GBMs. The second group longs one share of the asset for each of the remaining 20 underlying assets, which are modeled by Heston's stochastic volatility models, i.e., the asset price $S(t)$ is modeled by

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sqrt{v(t)}S(t)dW_1(t) \\ dv(t) &= \kappa(\theta - v(t))dt + \eta\sqrt{v(t)}dW_2(t) \\ dW_1(t)dW_2(t) &= \rho' dt, \end{aligned}$$

where $W_1(t)$ and $W_2(t)$ are standard Brownian motions, κ is the rate of mean reversion, θ is the long-term mean of variance, η is the volatility of the volatility, μ is the drift of the stock, and ρ' is the correlation of two Brownian motions. Notice that, to make sure the volatility is always positive, we need the condition $2\kappa\theta > \eta^2$. Specifically, let $\mathbf{S}_1(t) = (S_{1,1}(t), \ldots, S_{1,40}(t))$ and $\mathbf{S}_2(t) = (S_{2,1}(t), \ldots, S_{2,20}(t))$ denote the underlying assets in the first and second group, respectively. Let $\Phi_1(t) = \sum_{i=1}^{40}\{V_i^c(t) + V_i^p(t)\}$, and $\Phi_2(t) = \sum_{i=1}^{20} S_{2,i}(t)$. Similar to previous two examples, we let $T = 0.3$ and the number of discretization points $N = 30$. Other parameters are given in Table 2 in the online supplement, and the maturity times of all the options are $\tau = 1$.

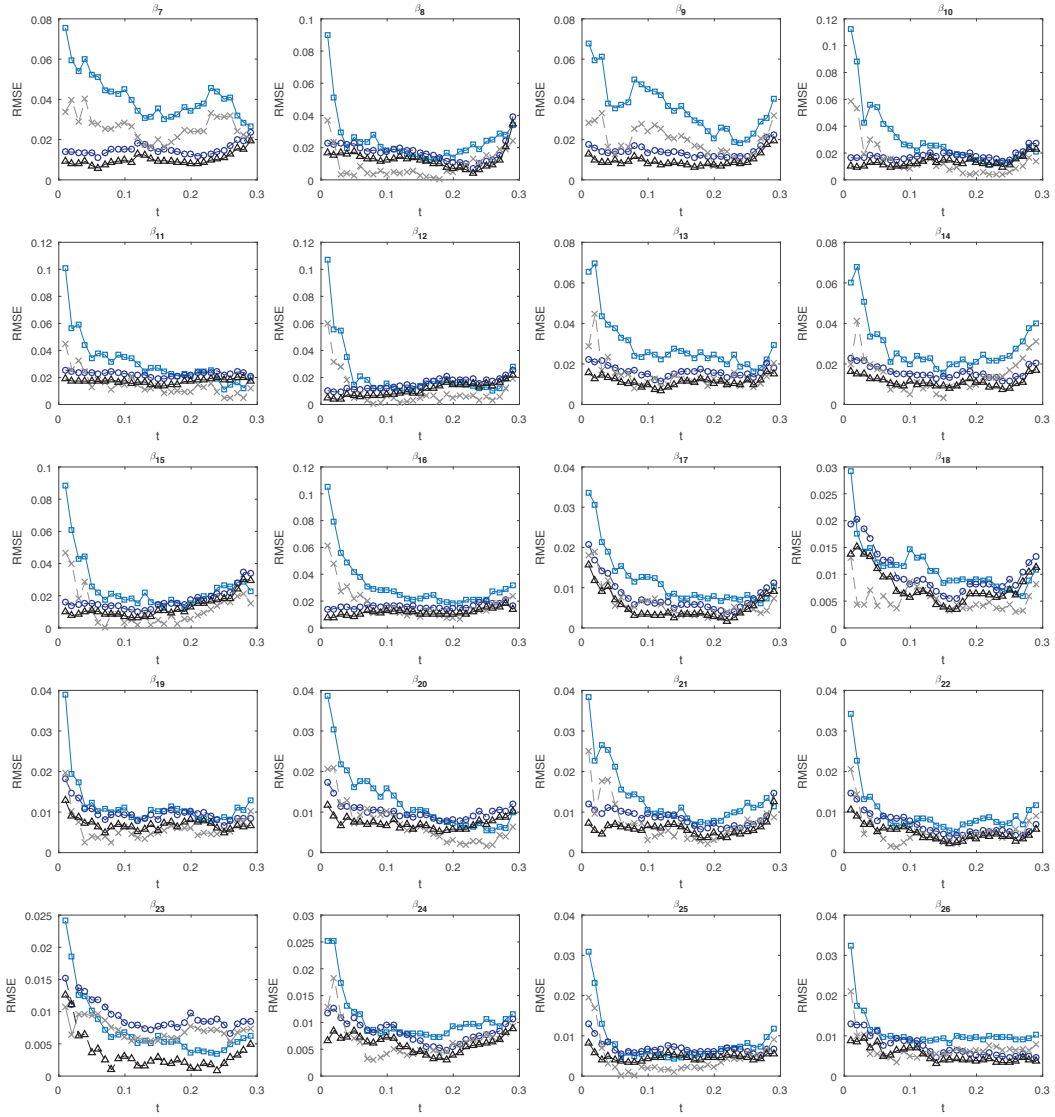

Figure 8 **RMSEs of the ML, lasso, perturbation and lasso+perturbation estimators of $\beta_7, \ldots, \beta_{26}$ in Example 2 (legends are in Figure 7)**

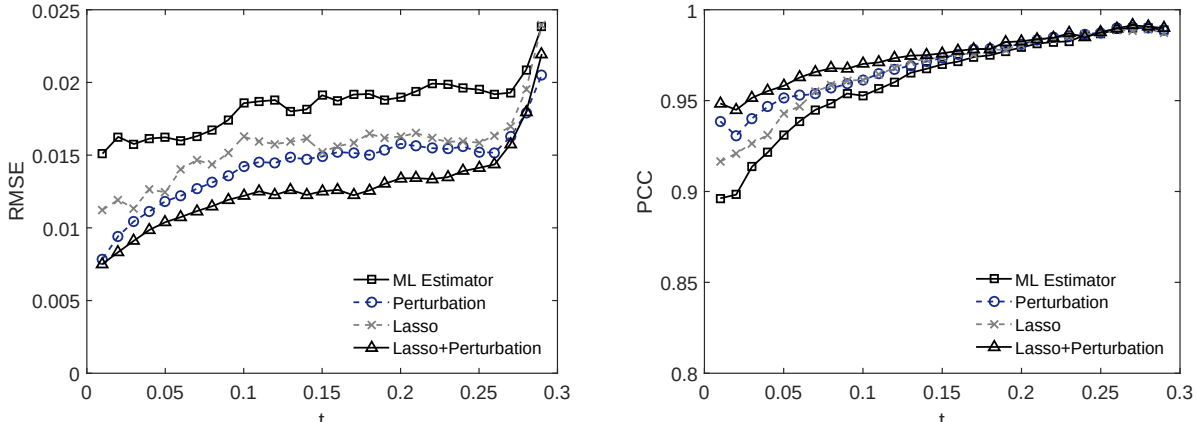

Figure 9 **RMSE (left panel) and PCC (right panel) of the ML, lasso, perturbation and lasso+perturbation estimators in Example 2**

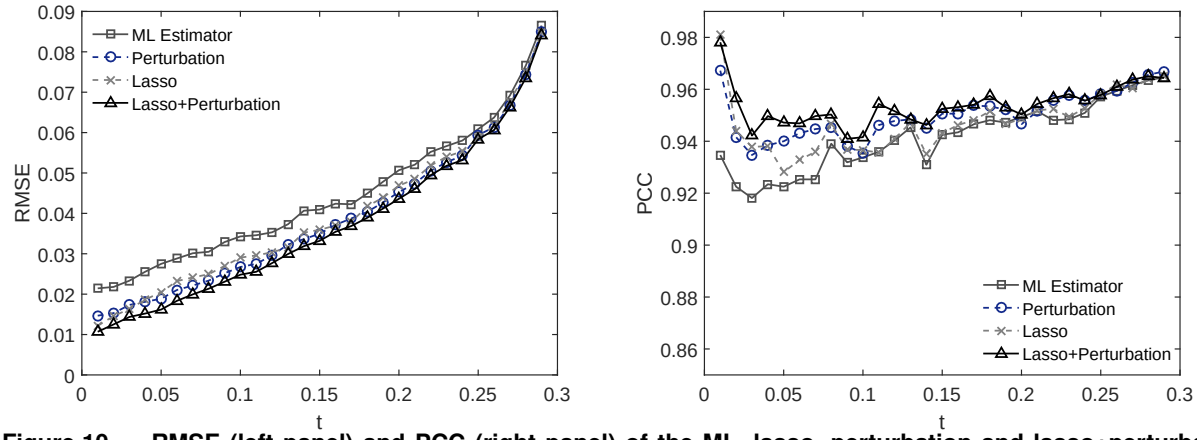

Figure 10 RMSE (left panel) and PCC (right panel) of the ML, lasso, perturbation and lasso+perturbation estimators in Example 3

This is a challenging example because there are 80 risk factors and only 10000 sample paths. We want to use this example to test the performance of the proposed methods for high-dimensional problems.

We evaluate the true exceedance probabilities by running 10000 observations at each time point, as in the first two examples, and use them as the benchmark to evaluate the RMSEs and PCCs, based on 100 testing paths with $L = J = 10$ and $M = 100$. At any time t, notice that there are two groups of risk factors, denoted by $\mathbf{S}(t)$ and $\mathbf{v}(t)$, where $\mathbf{v}(t) = (v_1(t), \ldots, v_{20}(t))^\top$ are the volatilities in Heston models at time t. We first include all the linear terms in the basis function, i.e., $\mathbf{X}(\mathbf{S}(t)) = (1, \mathbf{S}(t), \mathbf{v}(t))^\top$, and plot the RMSEs and PCCs in Figure 10. From the left panel, we see that the risk estimation is not as precise as in previous two examples, especially when the time is close to T. This may be because the number of risk factor is quite large in this example and the logistic regression model with only linear terms may be insufficient, i.e., having a high level of bias. However, from the right panel of the figure, we see that the quality of risk classification is still quite good, which also supports our argument that risk classification is typically an easier problem than risk estimation. In this example, the logistic model with only the linear terms of risk factors appears inadequate in approximating the real exceedance probability, especially for large t. This is a misspecification problem, which is a common issue for many statistical learning tools. To alleviate the problem, we add

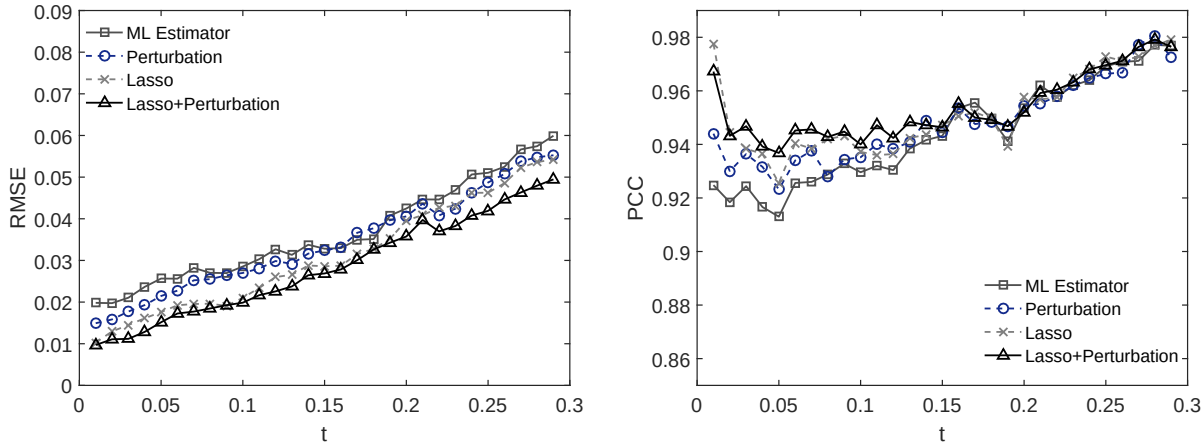

Figure 11 RMSE (left panel) and PCC (right panel) of the ML, lasso, perturbation and lasso+perturbation estimators in Example 3 with square terms

the square terms of the individual risk factors, i.e., $\mathbf{X}(\mathbf{S}(t)) = (1, \mathbf{S}(t), \mathbf{v}(t), \mathbf{S}^2(t), \mathbf{v}^2(t))^\top$, where $\mathbf{S}^2(t) = (S^2_{1,1}(t), \ldots, S^2_{1,40}(t), S^2_{2,1}(t), \ldots, S^2_{2,20}(t))^\top$ and $\mathbf{v}^2(t) = (v_1^2(t), \ldots, v_{20}^2(t))^\top$, and plot of the RMSEs and PCCs in Figure 11. The results show that the risk estimators are more precise (i.e., the RMSEs are smaller) and the risk classifications remain the same (i.e., the PCCs are similar). In both cases, we see that the lasso+perturbation works the best. This example demonstrates that the proposed estimators, especially the lasso+perturbation estimator, work well even for high-dimensional problems (notice that $\mathbf{X}(\mathbf{S}(t))$ is of 161 dimensions) with a reasonable sample size ($n = 10000$).

8. Conclusions and Future Research

In this paper we consider how to use retained simulation sample paths to estimate the exceedance probabilities and classify risk levels of a financial portfolio in real time. We propose various methods to solve the problem, study their asymptotic properties, and test their performance numerically on realistic examples. These methods belong to a new class of techniques, known as simulation analytics, which apply data mining and data analytics tools to estimate conditional statements. We also show that knowing the simulation model gives us advantages and allows us to develop methods, such as the perturbation method, that are more efficient than typical data analytics tools.

The perspective of simulation analytics creates many opportunities in financial risk management. For instance, we may consider how to simulate from different $\mathbf{S}(0)$, instead of perturbing a single initial value, so that the online risk monitoring problems may be solved more effectively, or how to conduct online risk monitoring if the sample paths are simulated using importance sampling, a tool often used for variance reduction in risk measurement. We may also consider how to handle portfolios that have path-dependent derivatives such as Asian options and barrier options, and investigate how to use nonlinear statistical learning tools, such as tree-based methods and neural networks, to improve the accuracy of the risk estimators and classifiers.

Acknolwedgement

The research reported in this paper is partially supported by Hong Kong Research Grants Council Grant Number GRF 16203214 and by National Science Foundation Grant Number CMMI-1537060.

References

Alexander C (2001) *Market Models: A Guide to Financial Data Analysis* (Chichester: John Wiley & Sons Ltd).

Bassamboo A, Juneja S, Zeevi A (2008) Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research* 56(3):593–606.

Belomestny D, Reib M (2006) Spectral calibration of exponential Lévy models. *Finance and Stochastics* 10(4):449–474.

Bonnans JF, Shapiro A (2000) *Perturbation Analysis of Optimization Problems* (New York: Springer).

Broadie M, Du Y, Moallemi CC (2011) Efficient risk estimation via nested sequential simulation. *Management Science* 57(6):1172–1194.

Broadie M, Du Y, Moallemi CC (2015) Risk estimation via regression. *Operations Research* 63(5):1077–1097.

Broadie M, Glasserman P (1996) Estimating security price derivatives using simulation. *Management Science* 42(2):269–285.

Fahrmeir L, Kaufmann H (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics* 13(1):342–368.

Fan J, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96(456):1348–1360.

Feng M, Staum J (2017) Green simulation: Reusing the output of repeated experiments. *ACM Transactions on Modeling and Computer Simulation* 37(4):Article 23.

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1):1–22.

Fu JC, Li G, Zhao LC (1993) On large deviation expansion of distribution of maximum likelihood estimator and its application in large sample estimation. *Annals of the Institute of Statistical Mathematics* 45:477–498.

Glasserman P, Heidelberger P, Shahabuddin P (2000) Variance reduction techniques for estimating value-at-risk. *Management Science* 46(10):1349–1364.

Glasserman P, Heidelberger P, Shahabuddin P (2002) Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* 12(3):239–269.

Glasserman P, Kang W, Shahabuddin P (2008) Fast simulation of multifactor portfolio credit risk. *Operations Research* 56(5):1200–1217.

Glasserman P, Li J (2005) Importance sampling for portfolio credit risk. *Management Science* 51(11):1643–1656.

Gordy MB, Juneja S (2010) Nested simulation in portfolio risk measurement. *Management Science* 56(10):1833–1848.

Hastie T, Tibshirani R, Friedman J (2011) *The Elements of Statistical Learning* (New York: Springer-Verlag), second edition.

Hoadley B (1971) Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics* 42(6):1977–1991.

Hong LJ, Hu Z, 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(4):Article 22.

Hong LJ, Juneja S, Liu G (2017) Kernel smoothing for nested estimation with application to portfolio risk measurement. *Operations Research* 65(3):657–673.

Hosmer DW, Lemeshow S (2004) *Applied Logistic Regression* (New York: John Wiley & Sons).

Hull JC (2014) *Options, Futures, and Other Derivatives* (New Jersey: Pearson Prentice Hall), 9th edition.

Iscoe I, Kreinin I (2010) Large pool approximations. Cont R, ed., *Encyclopedia of Quantitative Finance*, volume 3 (Chichester: John Wiley & Sons).

James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning* (New York: Springer).

Jiang G, Hong LJ, Nelson BL (2016) A simulation analytics approach to dynamic risk monitoring. Roeder TMK, Frazier PI, Szechtman R, Zhou E, Huschka T, Chick SE, eds., *Proceedings of the 2016 Winter Simulation Conference*, 437–447.

Jorion P (2006) *Value at Risk: The New Benchmark for Managing Financial Risk* (New York: McGraw-Hill), third edition.

Lee SH, Glynn PW (2003) Computing the distribution function of a conditional expectation via monte carlo: Discrete conditioning spaces. *ACM Transactions on Modeling and Computer Simulation* 13(3):238–258.

Liu G, Hong LJ (2011) Kernel estimation of the greeks for options with discontinuous payoffs. *Operations Research* 59(1):96–108.

Liu M, Nelson BL, Staum J (2010) Simulation on demand for pricing many securities. Johansson B, Jain S, Montoya-Torres J, Hugan J, Yucesan E, eds., *Proceedings of the 2010 Winter Simulation Conference*, 2782–2789.

Liu M, Staum J (2010) Stochastic kriging for efficient nested simulation of expected shortfall. *Journal of Risk* 12(3):3–27.

Longstaff FA, Schwartz ES (2001) Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies* 14(1):113–147.

Nelson BL (2016) 'Some tactical problems in digital simulation' for the next ten years. *Journal of Simulation* 10(1):2–11.

Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. Engle R, McFadden D, eds., *Handbook of Econometrics*, volume 4, chapter 36, 2111–2245 (New York: Elsevier).

Niederreiter H (1988) Low-discrepancy and low-dispersion sequences. *Journal of Number Theory* 30(1):51–70.

Rosenbaum I, Staum J (2015) Database Monte Carlo for simulation on demand. Yilmaz L, Chan WKV, Moon I, Roeder TMK, Macal C, Rossetti MD, eds., *Proceedings of the 2015 Winter Simulation Conference*, 679–688.

Rudin W (1991) *Functional Analysis* (New York: McGraw-Hill), second edition.

Sirignano J, Giesecke K (2018) Risk analysis for large pools of loans. *Management Science* In Advance.

Sun Y, Apley DW, Staum J (2011) Efficient nested simulation for estimating the variance of a conditional expectation. *Operations Research* 59(4):998–1007.

Tibshirani RJ (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58(1):267–288.

Van der Vaart AW (2000) *Asymptotic Statistics* (Cambridge, MA: Cambridge University Press).

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50(1):1–25.