

# TUNING THE PARAMETERS OF A TWO-LEVEL SIMULATION PROCEDURE WITH SCREENING

Hai Lan

Northwestern University

---

Parameters that govern the behavior of the simulation procedures are important to the effectiveness of sophisticated simulation. A parameter tuning method for a two-level simulation with screening is discussed in this paper. A special procedure is introduced to predict the behavior of the two-level simulation based on historical data or a pilot simulation. A hybrid method of Grid search and nonlinear convex local optimization techniques is adopted to find suitable input parameters to optimize the forecasted performance of the two-level simulation. Experiments show that the input parameters suggested by this method are close to the global optima.

Categories and Subject Descriptors: I.6.8 [**Simulation and Modeling**]: Type of Simulation – Monte Carlo

General Terms: Simulation Optimization, Risk Management

Additional Key Words and Phrases: Expected Shortfall, Tail Conditional Expectation, Conditional Tail Expectation, Conditional Value at Risk, Newton Method, Grid Searching, Ranking and Selection

---

## 1. INTRODUCTION

In a stochastic simulation, the input parameters can be divided into two categories: the modeling parameters, which are part of the description of the underlying model, and the procedure parameters, which are part of the simulation procedure. For example, the arrival rate of customers in a queuing simulation is a modeling parameter. The number of samples that the experimenter would like to take for each alternative system design in order to identify the best one (this is a typical ranking and selection problem) is a procedure parameter. Modeling parameters are the prior knowledge the experimenters have from the calibration of raw data or theoretical analysis. Procedure parameters stand for the way the simulation is designed or the policies by which the experimental resources are allocated. Usually they are controllable and highly related to the effectiveness, and sometimes even the validity, of the simulation procedure.

In practice, the procedure parameters are determined either by the experimenters' experience or from empirical rules suggested from repeated experiments with different procedure parameters. However, for these sophisticated simulation procedures, which are too expensive to run repeatedly, the tuning of procedure parameters needs special consideration. It is desirable to determine the optimal or suboptimal procedure parameters with a modest cost.

When facing an entirely new simulation problem, the experimenters might do a pilot simulation with a modest cost then tune the procedure parameters based on information from the pilot simulation. More frequently, the experimenters might face a problem similar to one seen before; for example, today's problem may be similar to yesterday's problem, but with slight changes to some of the modeling parameters. If so, the procedure parameters that were optimal for yesterday's problem should still perform very well when used in simulating today's problem. We refer to the information collected from the pilot simulation or historical

---

This research was done under the direction of Barry L. Nelson and Jeremy Staum, with the support of the National Science Foundation under Grant No. DMI-0555485.

Author's address: Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3119; email: h-lan@northwestern.edu.

Permission to make digital/hard copy of all or part of this material without fee for personal or classroom use provided that the copies are not made or distributed for profit or commercial advantage, the ACM copyright/server notice, the title of the publication, and its date appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

© 20YY ACM 0000-0000/20YY/0000-0001 \$5.00

simulation as pilot information. In this article, we will discuss the procedure parameter tuning based on pilot information by taking a two-level simulation with screening as an example.

In our prior articles [2007; 2008], we proposed two-level simulation procedures to calculate risk measures of any portfolio: with an outer level to simulate the market scenarios, and an inner level to calculate the corresponding loss of the portfolio conditioning on each market scenario. For large financial institutions with complex investment portfolios of multiple assets, the computation of market risk presents real challenges. The computation of risk measures for real problems may take days or weeks, yet the value of such computation may vanish within hours. This demands that computational efficiency be especially considered for any applicable procedure of risk measurement. We have partially addressed the efficiency of two-level simulation in the following aspects. In Lan et al. [2007] we have proposed a screening algorithm to save the computational budget on those market scenarios which are not important for the computation of loss. An adaptive procedure was also discussed to find a suitable number of scenarios  $k$  with the sample size of the screening algorithm,  $n_0$ , and a pre-specified total number of simulated payoffs,  $C$ . Furthermore we have proposed a much more efficient confidence interval procedure for expected shortfall<sup>1</sup> based on two-level simulation and introduced a procedure which can adapt itself to accomplish the simulation with given time  $T$  (wall clock time) in Lan et al. [2008]. Within this article, we improve the efficiency of two-level simulation by tuning the input procedure parameters.

The architecture of the two-level simulation is shown in Figure 1. The vertical axis shows the work of the outer level simulation: generating  $k$  scenarios. The horizontal axis demonstrates the inner level simulation: calculating payoffs conditioning on each market scenario. Furthermore, the inner level simulation is divided into two stages: the first stage is used to generate  $n_0$  payoffs for each scenario and screen out those scenarios which seem unimportant to the calculation of risk (as shown in Figure 1, scenario  $Z_2$  and  $Z_4$  are such unimportant scenarios); In the second stage,  $N_i$  payoffs are sampled for each market scenario  $i \in I$ , where  $I$  is the set of scenarios retained after screening, and the confidence interval of the expected shortfall are computed. The two parameters, the number of scenarios generated in the outer level  $k$  and the sample size of screening  $n_0$ , are the most important parameters to the performance of the two-level simulation. The number of scenarios  $k$  represents the effort spent on predicting the market movement in some future time (or, say, scenario). The uncertainty about the future of the market is the source of the risk. Usually we want  $k$  to be as large as possible.

The sample size of screening (also called the first stage sample size),  $n_0$ , is critical to the performance of the screening algorithm, and is strongly related to the second stage sample size  $N_i$ ,  $i \in I$ . The second stage sample size  $N_i$  stands for the effort we spent on estimating the conditioned value of portfolio on the scenarios that are important to the computation of loss. Again,  $N_i, \forall i \in I$  are desired to be as large as possible. Yet  $N_i$  is not directly controllable; instead, it is determined mainly by the number of scenarios  $k$  and the first stage sample size  $n_0$ . Roughly speaking,  $N_i$ 's are decreasing in  $k$ , yet not monotone in  $n_0$ . In all, the two parameters ( $k, n_0$ ) determine the effort spent on exploring market movement and on "exploiting" the generated scenarios. It is natural to improve the performance of the two-level simulation by tuning ( $k, n_0$ ) such that exploration of new scenarios and "exploitation" of generated ones are well balanced.

The setting of the sample size of the screening algorithm,  $n_0$ , is a traditional topic in the literature of simulation. There are different procedures for different purpose. Stein [1945] proposed to determine the sample size for fixed width confidence interval estimation through a two-stage procedure. Rinott [1978] computed the sample size of the second stage of a two-stage procedure to guarantee the probability of correct selection of the best. When fixed total number of samples and all alternatives are given, Chen et al. [2000] proposed to compute the sample size of each alternative that maximizes the probability of correct selection of the best. In our research, the number of alternatives, or, say, the number of scenarios,  $k$  and the sample size,  $n_0$ , are chosen together to minimize the forecasted confidence interval width of expected shortfall through the two-level simulation. The sample size  $n_0$  is selected to help screen out most inferior scenarios, while at the same time it is important to keep  $n_0$  as small as possible to save computational budget

---

<sup>1</sup>Mostly the expected shortfall can be taken as the negative conditional expectation of the left tail of the distribution of the portfolio's value; for rigorous definition, please refer to Acerbi and Tasche [2002]

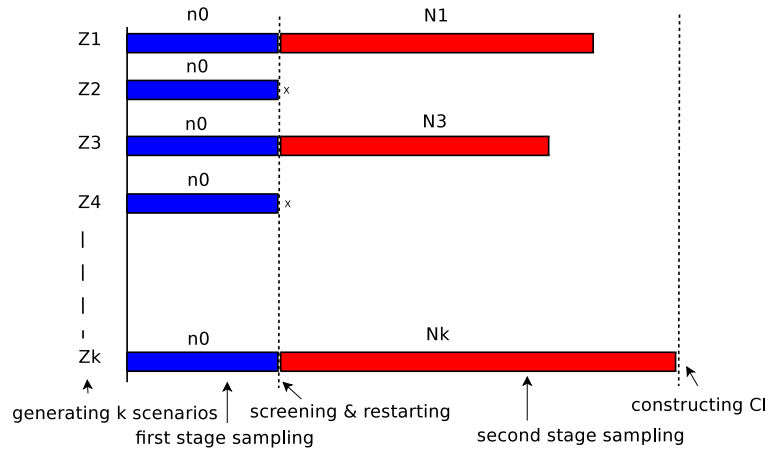


Fig. 1. Two-Level Simulation

for exploration of scenarios.

We have shown how to find a satisfactory number of scenarios  $k$  when the total number of simulated payoffs  $C$ , and the first stage sample size,  $n_0$ , are fixed [Lan et al. 2007]. The key idea is to set up the relationship between the market scenario number  $k$  and the forecasted confidence interval width  $W$  and then find the satisfactory  $k$  by nonlinear optimization techniques, which still applies in current research.

Section 3 introduces the approximation of  $W$  as a nonlinear function of  $(k, n_0)$ ; the prediction of the behavior of the two-level simulation with screening on the pilot information is discussed in Section 4. Section 5 will introduce the procedure to find satisfactory  $(k, n_0)$  when the total computational budget  $C$  is fixed. A similar procedure with total computational time  $T$  (wall time) fixed is discussed in Section 6. Experimental results are shown in Section 7 to verify the effect of the procedures proposed here. Some conclusions and suggestions for future work are discussed in Section 8.

## 2. REVIEW OF TWO-LEVEL SIMULATION

Any computational model for market risk measurement needs to address the following two problems: the model of scenarios (interest rates, stock prices etc.) in the future and the mapping from scenarios to the values of the portfolios studied. Naturally, the two-level simulation is adopted to calculate the risk measures. We can estimate the risk measure  $T(F_V)$  by a two-level simulation, in which the outer level of simulation generates scenarios  $Z_1, Z_2, \dots, Z_k$  and the inner level estimates each  $V_i := V(Z_i)$  by simulating  $V$  conditionally on  $Z_i$ .

To improve the computational efficiency, screening is adopted to eliminate the scenarios that do not affect the risk measure before spending too much computation on them. More specifically, the inner level is divided into two stages. At the first stage,  $n_0$  payoffs are simulated conditionally on each market scenario, then the screening procedure is applied at the end of the first stage. Let  $I$  stand for the subset retained after screening. Then at the second stage, for each market scenario  $Z_i$ ,  $i \in I$ , stratification-like budget allocation strategy is adopted; that is, the size of simulated payoffs for  $Z_i$  will be proportional to its variance which is estimated from the first stage sample variance  $S_i^2(n_0)$ . Figure 1 demonstrates how a computational budget is assigned through the whole two-level simulation.

Define  $\gamma := \{\pi_V(1), \pi_V(2), \dots, \pi_V(\lceil kp \rceil)\}$  as the set of scenarios we wish to keep after screening. The event of correct screening is  $\{\gamma \subseteq I\}$ , and the screening procedure guarantees  $\Pr\{\gamma \subseteq I\} \geq 1 - \alpha_s$ . The number of pairwise comparisons between  $\gamma$  and all other scenarios is  $(k - \lceil kp \rceil)\lceil kp \rceil$ . Therefore, for each ordered pair  $(i, j)$  we consider a hypothesis test that  $V_i \leq V_j$  at level  $\alpha_s / ((k - \lceil kp \rceil)\lceil kp \rceil)$ . If the hypothesis is rejected, we say  $Z_i$  is “beaten” by  $Z_j$ . For each  $i = 1, 2, \dots, k$ , let  $X_{i1}, X_{i2}, \dots, X_{in_0}$  be an i.i.d. sample drawn from the conditional distribution of  $X$  given  $Z_i$ , which satisfies  $E(X|Z_i) = V_i$  and  $E(X^2|Z_i) < \infty$ . Let  $\bar{X}_i(n_0)$  be its sample average. For each  $i, j = 1, 2, \dots, k$ , let  $S_{ij}^2(n_0)$  be the sample variance of

$X_{i1} - X_{j1}, X_{i2} - X_{j2}, \dots, X_{in_0} - X_{jn_0}$ . We screen out all risk factors that are beaten at least  $\lceil kp \rceil$  times:

$$I = \left\{ i : \sum_{i \neq j} \mathbf{1} \left\{ \bar{X}_i(n_0) > \bar{X}_j(n_0) + d \frac{S_{ij}(n_0)}{\sqrt{n_0}} \right\} < \lceil kp \rceil \right\} \quad (1)$$

where  $\mathbf{1}\{\cdot\}$  is an indicator function and

$$d = t_{n_0-1, 1-\alpha_s / ((k - \lceil kp \rceil) \lceil kp \rceil)} \quad (2)$$

is the  $1 - \alpha_s / ((k - \lceil kp \rceil) \lceil kp \rceil)$  quantile of the t-distribution with  $n_0 - 1$  degrees of freedom.

We adopt empirical likelihood and the extension of the Banerjee Theorem [Lan et al. 2008] to construct a confidence interval of the expected shortfall of the portfolio. Empirical likelihood involves assigning a weight vector  $\mathbf{w}$  to the scenarios  $Z_1, Z_2, \dots, Z_k$ . This vector  $\mathbf{w}$  must belong to the set

$$\mathcal{S}(k) := \bigcup_{\ell=1}^k \mathcal{S}_\ell(k) \text{ where } \mathcal{S}_\ell(k) := \left\{ \mathbf{w} : \mathbf{w} \geq 0, \sum_{i=1}^k w_i = 1, \sum_{i=1}^{\ell} w_i = p, \prod_{i=1}^k w_i \geq c k^{-k} \right\}, \quad (3)$$

where  $c$  is a critical value derived from a chi-squared distribution. Each  $\mathbf{w} \in \mathcal{S}(k)$  belongs to  $\mathcal{S}_\ell(k)$  for a unique integer  $\ell$ . There are integers  $l_{\min}$  and  $l_{\max}$  such that  $\mathcal{S}_\ell(k)$  is empty if  $\ell < l_{\min}$  or  $\ell > l_{\max}$ ; we therefore need only consider a limited range of  $\ell$ . Although  $\ell$  depends on  $\mathbf{w}$ , while  $l_{\min}$  and  $l_{\max}$  depend only on  $k$  and  $c$ , to lighten notation we do not make this dependence explicit. Because  $\text{ES}_{1-p}$  involves an average over the left tail containing probability  $p$ , we also define a transformed weight vector  $\mathbf{w}'$ :

$$w'_i := \begin{cases} -w_i/p, & i = 1, 2, \dots, \ell \\ 0, & \text{otherwise.} \end{cases}$$

The result, derived in Lan et al. [2008], is that the lower confidence limit is

$$\hat{L} := \min \left\{ \left\{ \min_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w'_i \bar{X}_{\pi_0(i)}(N_i) - z_{\text{lo}}(\ell) B_0(\ell) \right\} : \ell = \lceil kp \rceil, \lceil kp \rceil + 1, \dots, l_{\max} \right\} \quad (4)$$

and the upper confidence limit is

$$\hat{U} := \max \left\{ \left\{ \max_{\mathbf{w} \in \mathcal{S}_\ell(k)} \sum_{i=1}^{\ell} w'_i \bar{X}_{\pi_1(i)}(N_i) + z_{\text{hi}} B_S(\ell) \right\} : \ell = l_{\min}, l_{\min} + 1, \dots, \lceil kp \rceil \right\}, \quad (5)$$

where

$$\begin{aligned} B_0(\ell) &:= \underline{s}(\ell) \Delta(\ell), \\ s_i &:= \sqrt{S_i^2(N_i)/N_i}, \\ \underline{s}(\ell) &:= \max_{i=1,2,\dots,\ell} s_{\pi_0(i)}, \\ \Delta(\ell) &:= \sqrt{\max_{w \in \mathcal{S}(\ell)} \sum_{i=1}^{\ell} w_i'^2}, \\ z_{\text{lo}}(\ell) &:= t_{1-\alpha_{\text{lo}}, N_{\text{lo}}(\ell)-1}, \\ N_{\text{lo}}(\lceil kp \rceil) &:= \min_{i=1,2,\dots,\lceil kp \rceil} N_{\pi_0(i)}, \\ B_S(\ell) &:= \bar{s}(\ell) \Delta(\ell), \\ \bar{s}(\ell) &:= \max_{i=1,2,\dots,\ell} s_{\pi_s(i)}, \\ z_{\text{hi}} &:= t_{1-\alpha_{\text{hi}}, N_{\text{hi}}-1} \\ N_{\text{hi}} &:= \min\{N_{\pi_1(1)}, N_{\pi_1(2)}, \dots, N_{\pi_1(|I|)}\}. \end{aligned}$$

### 3. APPROXIMATION OF $W$

To demonstrate how to establish the mapping between the expected confidence interval width and the two parameters  $(k, n_0)$ , take the given simulated payoffs number  $C$  procedure as an example. For given wall clock time  $T$ , essentially it is very similar except that wall clock time  $T$  needs to be translated to the number of simulated payoffs.

We decompose  $W$  into  $W = W_1 + W_2 + W_3$ , as follows.

$$\begin{aligned}
 W &= \mathbf{E} \left\{ \hat{U} - \hat{L} \right\} \\
 &= \mathbf{E} \left\{ \left[ \sum_{i=1}^{\bar{l}} \tilde{w}'_i \bar{X}_{\pi_{1i}}(N_i) + z_{\text{hi}} B_S(\bar{l}) \right] - \left[ \sum_{i=1}^{\tilde{l}} \tilde{w}'_i \bar{X}_{\pi_{0i}}(N_i) + z_{\text{lo}}(\tilde{l}) B_0(\tilde{l}) \right] \right\} \\
 &= \mathbf{E} \left\{ \underbrace{\sum_{i=1}^{\bar{l}} \tilde{w}'_i \bar{X}_{\pi_{1i}}(N_i) - \sum_{i=1}^{\tilde{l}} \tilde{w}'_i \bar{X}_{\pi_{1i}}(N_i)}_{W_1} \right\} \tag{6}
 \end{aligned}$$

$$+ \mathbf{E} \left\{ \underbrace{z_{\text{hi}} B_S(\bar{l}) + z_{\text{lo}}(\tilde{l}) B_0(\tilde{l})}_{W_2} \right\} \tag{7}$$

$$+ \mathbf{E} \left\{ \underbrace{\sum_{i=1}^{\tilde{l}} \tilde{w}'_i \bar{X}_{\pi_{1i}}(N_i) - \sum_{i=1}^{\tilde{l}} \tilde{w}'_i \bar{X}_{\pi_{0i}}(N_i)}_{W_3} \right\} \tag{8}$$

where

$$\begin{aligned}
 \tilde{w} &:= \arg \min_{w \in \mathcal{S}} \sum_{i=1}^k w'_i X_{\pi_{0i}}(N_i) \\
 \tilde{l} &:= \arg \min_{l=[kp]}^{l_{\max}} \min_{w \in \mathcal{S}(l)} \sum_{i=1}^l w'_i X_{\pi_{0i}}(N_i) \\
 \underline{w} &:= \arg \min_{w \in \mathcal{S}} \sum_{i=1}^k w'_i X_{\pi_{1i}}(N_i) \\
 \underline{l} &:= \arg \min_{l=[kp]}^{l_{\max}} \min_{w \in \mathcal{S}(l)} \sum_{i=1}^l w'_i X_{\pi_{1i}}(N_i) \\
 \bar{w} &:= \arg \max_{w \in \mathcal{S}} \sum_{i=1}^k w'_i X_{\pi_{1i}}(N_i) \\
 \bar{l} &:= \arg \max_{l=l_{\min}}^{[kp]} \max_{w \in \mathcal{S}(l)} \sum_{i=1}^l w'_i X_{\pi_{1i}}(N_i)
 \end{aligned}$$

We can take  $W_1, W_2$  and  $W_3$  separately as the representation for the uncertainty of the outer level, uncertainty of the inner level, and bias of the tail.

We then need to solve a stochastic nonlinear integer optimization problem:

$$\begin{aligned} \min \quad & W_1 + W_2 + W_3 \\ \text{s.t.} \quad & kn_0 + \sum_{i=1}^{|I|} N_i \leq C \\ & k, n_0 \text{ positive integers} \end{aligned} \quad (9)$$

which, if there exists any solution, is very complex and time-consuming. Instead, by approximating  $W_1, W_2$  and  $W_3$  with  $\hat{W}_1, \hat{W}_2$  and  $\hat{W}_3$ , we set up another nonlinear optimization problem,

$$\begin{aligned} \min \quad & \hat{W}_1 + \hat{W}_2 + \hat{W}_3 \\ \text{s.t.} \quad & kn_0 + \sum_{i=1}^{|I|} N_i \leq C \\ & k > 0, n_0 > 0 \end{aligned} \quad (10)$$

which, though potentially difficult, is tractable.

Based on the following analysis, we can approximate  $W$  by

$$\begin{aligned} \hat{W}(k, n_0) := & \frac{E_o}{\sqrt{k}} + \\ & \left( \sigma \sqrt{\frac{K_1(k, n_0)}{C_1}} + \sigma(\sigma)z_{1-\frac{1}{K_1(k, n_0)}} \right) z_{1-\alpha_{lo}} \underline{\Delta}(k) + \\ & \left( \sigma \sqrt{\frac{K_1(k, n_0)}{C_1}} + \sigma(\sigma)z_{1-\frac{1}{\lceil k p \rceil}} \right) z_{1-\alpha_{hi}} \bar{\Delta}(k) + \\ & \frac{1}{\bar{l}} \sum_{(i,j)} \left\{ \frac{\sigma_{ij}^*}{\sqrt{2\pi}} \exp\left\{-\frac{d_{ij}^2}{2\sigma_{ij}^{*2}}\right\} + d_{ij} \left( \Phi\left(\frac{d_{ij}}{\sigma_{ij}^*}\right) - \Phi\left(\frac{d_{ij}}{\sigma_{ij}'}\right) \right) \right\}. \end{aligned} \quad (11)$$

### 3.1 Approximation of $W_1$

The width of the outer-level empirical likelihood confidence interval  $W_1$  is a function of  $k$ . Let  $E_o(k_0)/\sqrt{k_0}$  be the width of the outer-level empirical likelihood confidence interval in  $\mathcal{T}(k_0, n_{00})$ . Because empirical likelihood confidence interval widths are typically  $\mathcal{O}(k^{-1/2})$ , the width when there are  $k$  scenarios will be approximately  $E_o/\sqrt{k}$ . We approximate  $E_o$  by

$$\sqrt{k_0} \left( \max_{\mathbf{w} \in \mathcal{S}(k_0)} \sum_{i=1}^{k_0} w'_i \bar{X}_{(i)}(n_{00}) - \min_{\mathbf{w} \in \mathcal{S}(k_0)} \sum_{i=1}^{k_0} w'_i \bar{X}_{(i)}(n_{00}) \right)$$

where  $\bar{X}_{(i)}(n_{00})$  is the  $i$ th order statistic of the sample averages in  $\mathcal{T}(k_0, n_{00})$ .

### 3.2 Approximation of $W_2$

To deal with  $W_2$ , first we must make some simplifications regarding sample variances. Suppose that all sample variances in the future experiment will be the same at the first and second stages:  $S_i^2(N_i) \approx S_i^2(n_0)$ . In our two-level simulation,  $N_i$  is chosen to let  $s_i^2 := S_i^2(N_i)/N_i \approx \sum_{j \in I} S_j^2(n_0)/C_1$  be approximately the same for all  $i \in I$ , where  $C_1$  is the number of payoffs sampled in the second stage.

Further suppose that the average variance of the payoff conditional on scenarios that survive screening will be about the same as it was in  $\mathcal{T}(k_0, n_{00})$ :  $\sum_{j \in I} S_j^2(n_0)/|I| \approx \sigma^2$ , where  $\sigma^2$  is our estimate of this average variance from  $\mathcal{T}(k_0, n_{00})$ .

Suppose that  $|I|$  will be  $K_1(k, n_0)$ , an expected number of scenarios that survive screening, which depends on both the initial number of scenarios  $k$  and the sample size  $n_0$  used for screening. If the context permits, sometimes we also simplify  $K_1(k, n_0)$  as  $K_1$  (see Section 4.2 for the estimation of  $K_1$ ).

We can also assume that

$$s_i \sim N\left(\sigma\sqrt{K_1(k, n_0)/C_1}, \sigma(s)\right) \quad (12)$$

for all  $i \in I$ , where  $\sigma(s)$  is the standard deviation of sample deviation due to the fact that a chi-squared distribution can be asymptotically approximated by a Normal distribution when the degrees of freedom go up to infinite.

Due to the additivity and scalability of maximum operation,  $W_2$  can be approximated by

$$\left(\sigma\sqrt{\frac{K_1(k, n_0)}{C_1}} + \sigma(s)z^{(1)}(K_1(k, n_0))\right) z_{1-\alpha_{lo}}\underline{\Delta}(k) + \left(\sigma\sqrt{\frac{K_1(k, n_0)}{C_1}} + \sigma(s)z^{(1)}(\lceil kp \rceil)\right) z_{1-\alpha_{hi}}\bar{\Delta}(k) \quad (13)$$

where  $z^{(1)}(n)$  is the expectation of the maximum of  $n$  i.i.d standard normal random variables,

$$\underline{\Delta}(k) := \sqrt{\underline{w}^2 + \frac{(1-\underline{w})^2}{\lceil kp \rceil - 1}} \quad \text{and} \quad \bar{\Delta}(k) := \sqrt{\bar{w}^2 + \frac{(1-\bar{w})^2}{l_{\min} - 1}},$$

and  $\underline{w}, \bar{w} \in (0, p)$  are such that the weight vectors

$$\left(\underline{w}, \frac{p-\underline{w}}{\lceil kp \rceil - 1}, \dots, \frac{p-\underline{w}}{\lceil kp \rceil - 1}, \frac{1-p}{k-\lceil kp \rceil}, \dots, \frac{1-p}{k-\lceil kp \rceil}\right)$$

and

$$\left(\bar{w}, \frac{p-\bar{w}}{l_{\min} - 1}, \dots, \frac{p-\bar{w}}{l_{\min} - 1}, \frac{1-p}{k-l_{\min}}, \dots, \frac{1-p}{k-l_{\min}}\right)$$

are on the boundary of the feasible region  $\mathcal{S}(k)$ . To calculate  $\underline{w}$  or  $\bar{w}$ , we find the root of the nonlinear constraint in  $\mathcal{S}(k)$  and then multiply it by  $1/p$ .

### 3.3 Approximation of $W_3$

The effect of increasing the first stage sample size  $n_0$  can be seen in Equation (13): it allows screening to eliminate less pertinent scenarios so that the number of scenarios retained after screening,  $K_1(k, n_0)$ , decreases, while also decreasing the remaining budget  $C_1 = C - kn_0$ . However, there is another important effect from increasing  $n_0$ : it makes the first-stage ordering  $\pi_0$  more accurate and similar to the second-stage ordering  $\pi_1$ . This is the effect captured by the term  $W_3$ , which is unfortunately more difficult to approximate than  $W_1$  and  $W_2$ . For simplification,  $W_3$  is taken as the sum of the effects of all pairs of  $(i, j)$  where  $\pi_1(j) = \pi_0(i)$   $i \in \{\pi_0(1), \dots, \pi_0(\underline{l})\}$ ,  $j \in \{\pi_0(\underline{l}+1), \dots, \pi_0(|I|)\}$ , and such pairs of  $(i, j)$  are treated independently, where  $\underline{l}$  is the number of scenarios which are selected to construct the lower bound of the confidence interval. Assume  $\underline{l} = \bar{l} = \check{l}$  and  $\underline{w} = \bar{w}$ . Each  $\underline{w}'_i$ ,  $i \leq \underline{l}$  is simplified as  $\frac{1}{2}$ .

Let us assume that  $\bar{X}_i(n_0) \sim N(\mu_i, \sigma_i^2/n_0)$  and  $\bar{X}_i(N_i) \sim N(\mu_i, \sigma_i^2/N_i)$ . For notational simplicity, denote  $d_{ij} = \mu_i - \mu_j$ ,  $\sigma'_{ij} = \sigma_{ij}\sqrt{1/n_0}$ ,  $\sigma_{ij}^* = \sqrt{\sigma_i^2/N_i + \sigma_j^2/N_j}$ ,  $D_1(i, j) = \bar{X}_i(n_0) - \bar{X}_j(n_0)$  and  $D_2(i, j) = \bar{X}_i(N_i) - \bar{X}_j(N_j)$ . Obviously  $D_1(i, j) \sim N(d_{ij}, \sigma_{ij}'^2)$  and  $D_2(i, j) \sim N(d_{ij}, \sigma_{ij}^{*2})$ .

Then the effect of each pair  $(i, j)$  where  $\pi_1(j) = \pi_0(i)$   $i \in \{\pi_0(1), \dots, \pi_0(\underline{l})\}$ ,  $j \in \{\pi_0(\underline{l}+1), \dots, \pi_0(|I|)\}$  is

$$\begin{aligned} & \frac{1}{\underline{l}} \mathbb{E}[|D_2(i, j)| \cdot \mathbf{1}\{D_1(i, j)D_2(i, j) < 0\}] \\ &= \frac{1}{\underline{l}} \mathbb{E}[D_2(i, j)\mathbf{1}\{D_2(i, j) > 0\}\mathbf{1}\{D_1(i, j) < 0\}] \\ & \quad + \frac{1}{\underline{l}} \mathbb{E}[-D_2(i, j)\mathbf{1}\{D_2(i, j) < 0\}\mathbf{1}\{D_1(i, j) > 0\}] \\ &= \frac{1}{\underline{l}} \frac{\sigma_{ij}^*}{\sqrt{2\pi}} \exp\left\{-\frac{d_{ij}^2}{2\sigma_{ij}^{*2}}\right\} + \frac{1}{\underline{l}} d_{ij} \left( \Phi\left(\frac{d_{ij}}{\sigma_{ij}^*}\right) - \Phi\left(\frac{d_{ij}}{\sigma_{ij}'}\right) \right), \end{aligned}$$

where  $\Phi(\cdot)$  is the c.d.f of the standard normal distribution. As the equation indicates,  $W_3$  is not monotone in  $n_0$  either. It may then not be optimal to choose so large an  $n_0$  that the number of risk scenarios after screening  $|I|$  is extremely close to  $l_{\max}$ .

To simplify the computation, let us assume that for any  $j = \pi_0(\underline{l} + 1), \dots, \pi_0(|I|)$ , if there exist any  $i \leq \underline{l}$  such that  $\pi_1(i) = \pi_0(j)$ ,  $i$  must satisfy  $i = \arg \min_{i \in \{\pi_0(1), \dots, \pi_0(\underline{l})\}} d_{ij} / \sigma'_{ji}$ . Then  $W_3$  can be approximated by

$$\hat{W}_3(k, n_0) := \frac{1}{\underline{l}} \sum_{(i,j)} \left\{ \frac{\sigma_{ij}^*}{\sqrt{2\pi}} \exp\left\{-\frac{d_{ij}^2}{2\sigma_{ij}^{*2}}\right\} + d_{ij} \left( \Phi\left(\frac{d_{ij}}{\sigma_{ij}^*}\right) - \Phi\left(\frac{d_{ij}}{\sigma'_{ji}}\right) \right) \right\}$$

We have made a lot of assumptions and simplifications on the approximation of  $W_3$ , which may decrease the accuracy of the approximation especially when  $K_1 \gg l_{\max}$ . However, with reasonable computational budget  $C$  or computational time  $T$ , we would like to choose  $n_0$  large enough to keep our screening work efficient, so that  $K_1$  is usually close to  $l_{\max}$ . Besides, in such cases, the true value of  $W_3$  does not impact the choice of  $(k, n_0)$  too much as both  $\sigma^*$  and  $\sigma'$  are very small. The term  $W_3$  in the objective function works more like a penalty function to avoid choosing an  $n_0$  too small or too large. In this case,  $\hat{W}_3$  does not need to be a very good approximation to  $W_3$ .

#### 4. MEASUREMENT AND PREDICTION

In Equation (11), besides  $X_{(i)}(n_{00})$ ,  $\sigma$ ,  $\sigma(s)$  etc., which can be directly estimated or approximated from the pilot information,  $\mathcal{T}(k_0, n_{00})$ , the number of risk scenarios retained after screening  $K_1(k, n_0)$  and the number of allocated simulated payoffs at the second stage  $C_1$  need specific consideration. When the computational budget is given in total simulated payoffs  $C$ , we have that  $C_1$  can be easily calculated by  $C_1 = C - kn_0$ .

##### 4.1 When Computational Effort is Given by $T$

In this case, the number of allocated simulated payoffs for the second stage  $C_1$  has to be calculated by estimating the time left for the second stage sampling  $T_1$  and the unit time required for the sampling of each payoff  $R_{\text{sam}}$ . First we need to adopt a high precision timer provided by the operation system and assume that the running environment of our procedure is unchanged between the pilot simulation (or historical experiment) and the experiment we would like to run. Then we need to analyze the computational complexity of the two-level simulation.

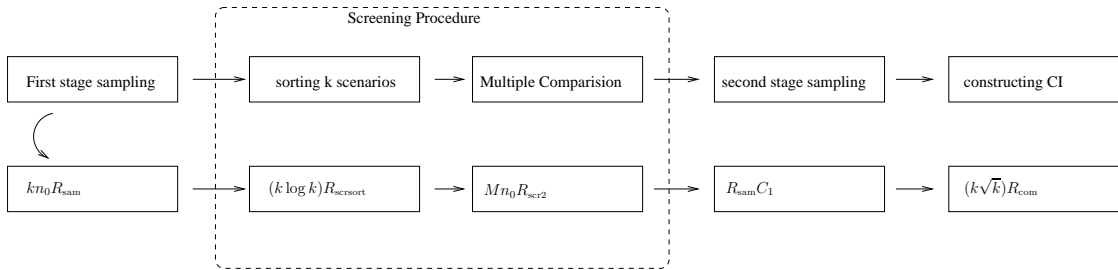


Fig. 2. Time Expensive Steps of Two-Level Simulation

In Figure 2, the first row of the diagram shows the main time-expensive steps of the two-level simulation. The second row shows the corresponding approximation formula of time spent on each step. Some of these are derived from the expected computation complexity. For example, the time spent on “sorting” has the expected computation complexity of  $\mathcal{O}(k \log k)$  as Quick Sort is adopted, and we may then approximate the time spent on “sorting” as  $(k \log k) R_{\text{scrsort}}$ , where  $R_{\text{scrsort}}$  is a constant multiplier. A similar reason accounts for the first stage sampling and second stage sampling approximation formulas, where  $R_{\text{sam}}$  is the average time spent on the simulation of unit payoff. The time spent on constructing the confidence interval (short as CI) is  $(k \log k) R_{\text{com}}$ , where  $R_{\text{com}}$  is the constant multiplier related with the construction of CI. A rough intuition is that the computation of the lower or upper bound limit of  $\sum_{w \in S(l)} w' \bar{X}_{\pi, i}(N_i)$  is  $\mathcal{O}(k)$

$\forall l \in [l_{\min}, l_{\max}]$ , and  $l_{\max} - l_{\min}$  is asymptotically  $\mathcal{O}(\sqrt{k})$ . The time spent on the multiple comparisons of the screening procedure is a function of  $k$  and  $n_0$ , even monotone on  $k$ . But the relationship of the time spent on the multiple comparisons and the first stage sample size  $n_0$  is complex, due to the fact that increasing  $n_0$  will help to make the statistical inference on each market scenario easier or to keep it unchanged if  $n_0$  has already been large enough. What this means is that, the number of comparisons need to be done, denoted as  $M$ , decreases or remains unchanged; meanwhile, for each comparison, the computational effort on computing the variance of difference  $S_{ij}^2(n_0)$ , which is dominant in the total computational effort of each comparison, is a linearly increasing function of  $n_0$ . We leave the estimation of  $M$  to Section 4.2. If the number of comparisons  $M(k, n_0)$  (shorten as  $M$ ) is already known, the time spent on multiple comparisons of the screening procedure can be approximated as  $Mn_0R_{\text{scr}2}$ , where  $n_0R_{\text{scr}2}$  stands for average time on the calculation of each  $S_{ij}^2(n_0)$ , and  $R_{\text{scr}2}$  is a constant multiplier. All the constant multipliers (the  $R$ 's) are determined from the observed computation time at the historical or pilot simulation.

Based on the approximations above, we can estimate  $C_1$  as following:

$$C_1 = \frac{T_1}{R_{\text{sam}}} = \frac{T - kn_0R_{\text{sam}} - kR_{\text{scr}} - (k \log k)R_{\text{scrsort}} - (k\sqrt{k})R_{\text{com}} - Mn_0R_{\text{scr}2}}{R_{\text{sam}}}, \quad (14)$$

where the number of comparisons  $M$  is still unknown. The number of scenarios retained after screening,  $K_1(k, n_0)$ , is also an unknown variable that needs to be predicted.

#### 4.2 Predicting the Behavior of Screening

Figure 4.2 shows the problem we discuss in this section: predicting the number of scenarios after screening  $K_1(k, n_0)$  (and the number of comparisons to be done  $M(k, n_0)$ ) when the computational effort is given by simulated payoffs  $C$  (or wall clock time  $T$ ). The predictor of the screening procedure is proposed as

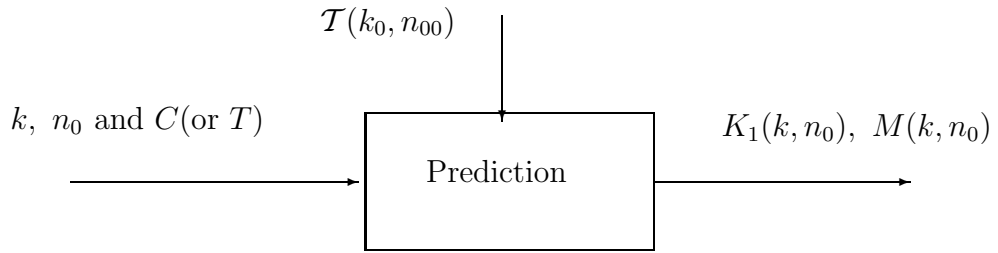
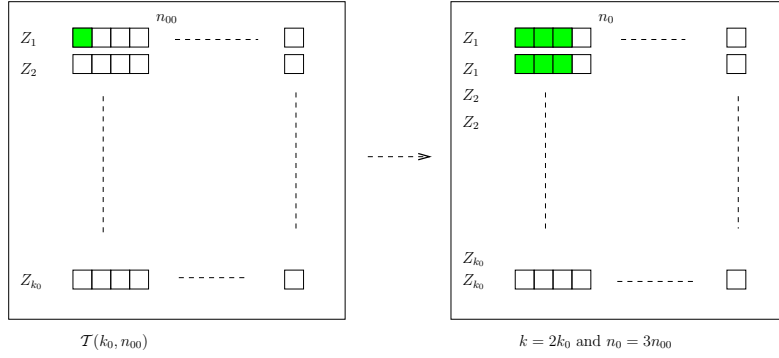


Fig. 3. Function of the Predictor of Screening

a function that takes  $\mathcal{T}(k_0, n_{00})$  and any pair of  $(k, n_0)$  as inputs. The outputs are  $K_1(k, n_0)$  and  $T_1$ . For simplicity and computational efficiency, based on  $\mathcal{T}(k_0, n_{00})$  for any pair of  $(k, n_0)$ , suppose all the  $k$  scenarios are re-sampled from  $\mathcal{T}(k_0, n_{00})$  equally; i.e., each  $Z_i \in \mathcal{T}(k_0, n_{00})$  repeats  $k/k_0$  times.<sup>2</sup> Similarly, all payoffs conditioned on each market scenario  $Z_i, i = 1, \dots, k$  are re-sampled from  $\mathcal{T}(k_0, n_{00})$ . Each  $X_{ij} \in \mathcal{T}(k_0, n_{00}), i = 1, \dots, k_0, j = 1, \dots, n_{00}$  is assumed to repeat  $(k/k_0)(n_0/n_{00})$  times in the predicted sampling with  $k$  scenarios and  $n_0$  simulated payoffs conditioned on each market scenario. Figure 4 shows that when  $k = 2k_0$  and  $n_0 = 3n_{00}$ , according to our re-sampling sketch, each payoff in  $\mathcal{T}(k_0, n_{00})$  repeats 6 times, as the “colored” boxes indicate. It is obvious that the above resampling does not really needed to be done; indeed, we only need to follow the sketch to value  $\bar{X}_i(n_0), S_i^2(n_0), i = 1 \dots, k$  etc..

<sup>2</sup>It is against the extreme value theorem to assume that  $Z_{\pi_V(0)}$  and  $Z_{\pi_V(k)}$  are unchanged as  $k$  is increasing. Yet those scenarios are already easy to be screened out when we assume they are unchanged as  $k$  increases. Even though we can use more sophisticated techniques to emulate  $Z_{\pi_V(0)}$  and  $Z_{\pi_V(k)}$  more precisely, the improvement of the prediction of  $K_1$  or  $M$  nearly vanishes.

Fig. 4. Re-sampling based on  $\mathcal{T}(k_0, n_{00})$  for  $k = 2k_0$  and  $n_0 = 3n_{00}$ 

By assuming the above re-sampling sketch, we can predict  $M$  as the following expression.

$$K_1'(k, n_0) = \left\{ \sum_{i=1}^{k_0} \frac{k}{k_0} \mathbf{1} \left\{ \sum_{j=1, j \neq i}^{k_0} \frac{k}{k_0} A_{ji}(n_0) < \lceil kp \rceil \right\} \right\} \sqrt{l_{\max}} \quad (15)$$

$$M = \frac{k^2}{k_0^2} M' \quad (16)$$

where  $A_{ji}(n_0) = \mathbf{1} \left\{ \bar{X}_i(n_{00}) > \bar{X}_j(n_{00}) + d \sqrt{\frac{S_{ij}^2(n_{00})}{n_0}} \right\}$  is an indicator function,  $M'$  is the number of comparisons needed to be done in order to calculate  $K_1'(k, n_0)$ . Rather than use  $K_1'(k, n_0)$ , we now outline a better approach to predict  $K_1(k, n_0)$ .

We estimate  $K_1(k, n_0)$  through a procedure similar to the prediction of  $M$ . The only difference is that  $A_{ji}(n_0)$  is not assumed to be a determined value, but instead a Bernoulli random variable, which indicates that  $S_{ij}(n_0) = S_{ij}(n_{00})$  is not required in the prediction of  $K_1(k, n_0)$ . This provides a better prediction than  $K_1'(k, n_0)$ . By assuming

$$\frac{\bar{X}_{i'}(n_0) - \bar{X}_{j'}(n_0)}{\sqrt{\frac{S_{i'j'}^2(n_0)}{n_0}}} \sim N \left( \frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{\sqrt{\frac{S_{ij}^2(n_{00})}{n_0}}}, 1 \right), \quad \bar{X}_{i'}(n_0) = \bar{X}_i(n_{00}) \text{ and } S_{i'}^2(n_0) = S_i^2(n_{00})$$

where  $i = \lceil \frac{i' k_0}{k} \rceil$  and  $j = \lceil \frac{j' k_0}{k} \rceil$ , we have  $\Pr\{A_{ji}(n_0) = 1\} = \Phi \left( d - \frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{\sqrt{\frac{S_{ij}^2(n_{00})}{n_0}}} \right)$  for  $\forall i \neq j, 1 \leq i, j \leq k_0$ , we then have

$$\mathbb{E} \left[ \mathbf{1} \left\{ \sum_{j=1, j \neq i}^{k_0} \frac{k}{k_0} A_{ji}(n_0) < \lceil kp \rceil \right\} \right] \approx \Phi \left( \frac{\lceil kp \rceil - 0.5 - \mu_B(i)}{\sigma_B(i)} \right),$$

where

$$\begin{aligned} \mu_B(i) &= \sum_{j=1, j \neq i}^{k_0} \frac{k}{k_0} \Pr\{A_{ji}(n_0)\} \\ \sigma_B^2(i) &= \sum_{j=1, j \neq i}^{k_0} \frac{k}{k_0} \Pr\{A_{ji}(n_0)\} (1 - \Pr\{A_{ji}(n_0)\}). \end{aligned}$$

Thus, we have the following approximation,

$$K_1(k, n_0) \approx \sum_{i=1}^{k_0} \frac{k}{k_0} \Phi \left( \frac{[kp] - 0.5 - \mu_B(i)}{\sigma_B(i)} \right) \quad (17)$$

The following procedure describes how  $K_1(k, n_0)$  and  $M'$  are actually calculated according to Equation (16) and (17). For time efficiency and numerical stability, we approximate  $\Phi(x) = 1$  and  $\Phi(-x) = 0$  when  $x \geq 2$ .

- (1) Feed  $\bar{X}_i(n_{00}), S_i^2(n_{00}), \pi_0(i) \forall i = 1, \dots, k_0$  and  $S_{ij}^2(n_{00})$  for some  $i, j$  from  $\mathcal{T}(k_0, n_{00})$ .
- (2) Take inputs  $k$  and  $n_0$ .
- (3) Initialize  $K_1 = [kp]$ ,  $M' = 0$ ,  $d = t_{n_0-1, 1-\alpha_s / ((k-[kp])[kp])}$ ,  $\tilde{S}^2(n_{00}) = \max_{i=0}^{[kp]} S_{\pi_0(i)}^2(n_{00})$  and  $i \leftarrow k_0$ .
  - (a) *Pre-screening*: If  $\bar{X}_{\pi_0(i)}(n_{00}) > \bar{X}_{\pi_0([kp])}(n_{00}) + d\sqrt{(S_{\pi_0(i)}^2(n_{00}) + \tilde{S}^2(n_{00}))/n_0}$ , scenario  $\pi_0(i)$  is pre-screened out: go to Step (3c).
  - (b) *Screening*: Initialize  $c \leftarrow 0$ ,  $b \leftarrow 0$ ,  $\mu_B \leftarrow 0$ ,  $\sigma_B^2 \leftarrow 0$ ,  $e \leftarrow 0$  and  $j \leftarrow 1$ .
    - i. If  $S_{\pi_0(i)\pi_0(j)}^2$  is unassigned, compute the sample variance  $S_{\pi_0(i)\pi_0(j)}^2$  of  $X_{\pi_0(i)1} - X_{\pi_0(j)1}, X_{\pi_0(i)2} - X_{\pi_0(j)2}, \dots, X_{\pi_0(i)n_0} - X_{\pi_0(j)n_0}$ .
    - ii. If  $c < [kp]$ , set  $M' \leftarrow M' + 1$ .
    - iii. If  $\bar{X}_{\pi_0(i)}(n_{00}) > \bar{X}_{\pi_0(j)}(n_{00}) + (d+2)S_{\pi_0(i)\pi_0(j)}(n_{00})/\sqrt{n_0}$ , scenario  $\pi_0(j)$  beats scenario  $\pi_0(i)$  for sure: set  $b \leftarrow b + \frac{k}{k_0}$ .  
 Else if  $\bar{X}_{\pi_0(i)}(n_{00}) > \bar{X}_{\pi_0(j)}(n_{00}) + (d-2)S_{\pi_0(i)\pi_0(j)}(n_{00})/\sqrt{n_0}$ , scenario  $\pi_0(j)$  beats scenario  $\pi_0(i)$  with probability  $a \leftarrow \Phi \left( d - \frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{\sqrt{\frac{S_{ij}^2(n_{00})}{n_0}}} \right)$ : set  $\mu_B \leftarrow \mu_B + \frac{k}{k_0}a$  and  $\sigma_B^2 \leftarrow \sigma_B^2 + \frac{k}{k_0}a(1-a)$ .  
 Otherwise, scenario  $\pi_0(j)$  cannot beat scenario  $\pi_0(i)$  for sure: set  $e \leftarrow e + \frac{k}{k_0}$ .
    - iv. If  $\bar{X}_{\pi_0(i)}(n_{00}) > \bar{X}_{\pi_0(j)}(n_{00}) + dS_{\pi_0(i)\pi_0(j)}(n_{00})/\sqrt{n_0}$ , set  $c \leftarrow c + \frac{k}{k_0}$ .
    - v. If  $b \geq [kp]$ , scenario  $\pi_0(i)$  is screened out: go to Step (3c).  
 Otherwise, set  $j \leftarrow j + 1$ .
    - vi. If  $j < i$ , go to Step (3b)i).  
 Otherwise, go to Step (3c).
  - (c) If  $b \geq [kp]$ , scenario  $\pi_0(i)$  is screened out for sure: set  $K_1 \leftarrow K_1$ .  
 Else if  $(j-1)\frac{k}{k_0} - e < [kp]$ , scenario  $\pi_0(i)$  can not be screened out for sure: set  $K_1 \leftarrow K_1 + \frac{k}{k_0}$ .  
 Otherwise, scenario  $\pi_0(i)$  is retained with probability  $\Phi \left( \frac{[kp] - 0.5 - \mu_B(i)}{\sigma_B(i)} \right)$ : set  $K_1 \leftarrow K_1 + \frac{k}{k_0} \Phi \left( \frac{[kp] - 0.5 - \mu_B(i)}{\sigma_B(i)} \right)$ .  
*Loop*: Set  $i \leftarrow i - 1$ . If  $i > [k_0p]$ , go to Step (3a).
- (4) set  $K_1 \leftarrow K_1 \vee l_{\max}$ .
- (5) set  $M \leftarrow \frac{k^2}{k_0^2} M'$

For experiments in which the computational effort is measured by the total number of simulated payoffs,  $C$ , only  $K_1(k, n_0)$  is significant. For given computational time  $T$ , both  $K_1(k, n_0)$  and  $M$  are important.

### 4.3 Predicting the Behavior of Screening for Difficult Case

Due to the fact that optimizing  $n_0$  will make it as small as possible (but still large enough to screen out a lot of inferior scenarios at the end of the first stage). This will push the optimized parameters  $(k, n_0)$ , say  $(k^*, n_0^*)$ , to a highly risky position sensitive to the quality of the approximation,  $\hat{W}(k, n_0)$ . That is,  $n_0^*$  may be slightly small yet it may lead to dramatically bad performance for some runs of two-level simulations. Although the expectation of the output is taken as the objective, for the reason of simplification, the width of predicted confidence interval is approximated by the determined function of predicted averages of the underlying random variables, such as  $M$ ,  $K_1(k, n_0)$  and so on. Such approximation will retain better precision when those random variables are more stable. From the viewpoint of experimenters, It is desirable to have a very good chance of good performance for only one trial of simulation considering the limited simulation efforts.

That is, both the goodness and robustness of the output are required. Meanwhile, we cannot afford to set up and solve a nonlinear robust optimization or chance constraint problem. A special step called “robustness” is adopted right after the optimized  $(k^*, n_0^*)$  is achieved. Tailor  $(k^*, n_0^*)$  a little bit to increase the stability of the underlying random variables then further improve the prediction of the confidence interval width and stabilize the output performance of the simulation. To accomplish this goal, the unfavorable instead of average behavior of the screening procedure, in which the unimportant scenarios are difficult to be screened out, needs to be predicted. Without specification, all assumptions we made in Section 4.2 are still applicable here.

From Section 4.2, the number of scenarios after screening  $K_1(k, n_0)$  can be taken as the sum of Bernoulli random variables with probability of success equal to  $\Phi\left(\frac{\lceil kp \rceil - 0.5 - \mu_B(i)}{\sigma_B(i)}\right)$  for every  $i' = 1, \dots, k$  ( $i = \lceil \frac{i'k_0}{k} \rceil$ ). Shown in Equation (17), the number of scenarios retained after screening  $K_1(k, n_0)$  is predicted as the expectation of the sum of the Bernoulli random variables. Then we can take the 95% quantile of  $K_1(k, n_0)$ , denoted as  $K_1(k, n_0)^{(0.95)}$ , as the typical output when the unimportant scenarios are difficult to be screened out, or say, with confidence of 95%, the number of scenarios retained after screening is no more than  $K_1(k, n_0)^{(0.95)}$ <sup>3</sup>. Similar to Equation (17), the 95% quantile  $K_1(k, n_0)^{(0.95)}$  can be approximated by a 95% quantile of a normal distribution with mean,  $\sum_{i=1}^{k_0} \frac{k}{k_0} \Phi\left(\frac{\lceil kp \rceil - 0.5 - \mu_B(i)}{\sigma_B(i)}\right)$  and variance,  $\sum_{i=1}^{k_0} \frac{k}{k_0} \Phi\left(\frac{\lceil kp \rceil - 0.5 - \mu_B(i)}{\sigma_B(i)}\right) \left(1 - \Phi\left(\frac{\lceil kp \rceil - 0.5 - \mu_B(i)}{\sigma_B(i)}\right)\right)$ .

For the number of comparisons needed to be done,  $M$ , a similar method is applicable. Let  $M = \sum_{i'=\lceil kp \rceil+1}^k M_{\pi_0(i')}$ , where  $M_{\pi_0(i')}$  is the number of comparisons needed to be done for the decision of market scenario  $\pi_0(i')$ . From Section 4.2, each comparison can be taken as a Bernoulli random variable, then  $M_{\pi_0(i')}$  is the smaller one of  $i' - 1$  or the number of comparisons needed to be done until the total number of successful comparisons is  $\lceil kp \rceil$ . For simplification, we assume all the Bernoulli random variables are i.i.d with the probability of success equal to  $\frac{\mu_B(i)}{k}$  ( $i = \lceil \frac{i'k_0}{k} \rceil$ )<sup>4</sup>. Then the number of comparisons  $M$  can be estimated as

$$\mathbb{E}[M] = \sum_{i'=\lceil kp \rceil+1}^k \mathbb{E}[M_{\pi_0(i')}] \quad (18)$$

where,

$$\mathbb{E}[M_{\pi_0(i')}] = \begin{cases} \lceil kp \rceil, & \text{when } b > \lceil kp \rceil; \\ i' - 1, & \text{when } i' - e - 1 < \lceil kp \rceil; \\ b + \frac{\lceil kp \rceil - b}{\bar{p}_B} \text{NB}_{\bar{p}_B}(\lceil kp \rceil - b + 1, i' - e - \lceil kp \rceil - 2) \\ \quad + (1 - \text{NB}_{\bar{p}_B}(\lceil kp \rceil - b, i' - e - \lceil kp \rceil - 1))(i' - b - 1). & \text{o.w.} \end{cases} \quad (19)$$

and  $\text{NB}_{\bar{p}_B}(x, y)$  is the c.d.f. of the Negative Binomial Distribution with probability of success  $\bar{p}_B$ , the number of successful trials  $x$  and the number of failed trails  $y$ . Appendix A gives further details about how the above equation is set up and how  $\bar{p}_B$  is computed.

## 5. FIXED $C$ PROCEDURE

By approximating  $W$  by  $\hat{W}$ , we simplify Problem (9) to Problem (20)

<sup>3</sup>This also indicates the two-level simulation works well in a probability of 95% if it could work well in such unfavorable case, without consideration of the possible bias from  $\mathcal{T}(k_0, n_{00})$  and the prediction procedures.

<sup>4</sup>As usually the comparisons with high success probability will be done first, then the statistical inference about  $\pi_0(i')$  could be done with fewer comparisons. This simplification will make  $M_{\pi_0(i')}$  overestimated. This is the reason that this method is adopted here instead of in Section 4.2.

$$\min \hat{W}(k, n_0) \quad (20)$$

$$\text{s.t. } C - kn_0 = C_1 \quad (21)$$

$$k \geq 40/p \quad (22)$$

$$n_0 \geq 30 \quad (23)$$

where Constraint (21) expresses the allocation of computational budget between the outer level and inner level, constraint (22) stands for the minimal empirical requirement for  $k$  in order to keep the validity from Lan et al. [2007], and constraint (23) tries to satisfy the normality assumption for simulated payoffs (see Lesnevski et al. [2008]). These are just some rough guidelines for large enough sample sizes to get adequate coverage of the estimated confidence interval, which is guaranteed only as  $k \rightarrow \infty$  and  $n_0 \rightarrow \infty$ .

The property of  $\hat{W}(k, n_0)$  depends on the pilot information  $\mathcal{T}(k_0, n_{00})$  we have, then essentially on the portfolio studied. When  $n_0$  is fixed,  $\hat{W}(\cdot, n_0)$  is decreasing then increasing in  $k$  and can be taken as a quasi-convex function of  $k$ . When  $k$  is fixed,  $\hat{W}(k, \cdot)$  is decreasing overall then increasing in  $n_0$ . Yet locally on the decreasing side,  $\hat{W}(k, \cdot)$  usually looks zigzag because the number of scenarios after screening  $K_1(k, n_0)$  is discrete, which means that even overall increasing  $n_0$  can help to reduce  $K_1(k, n_0)$ , not every increment of  $n_0$  will decrease  $K_1(k, n_0)$ . The properties of  $\hat{W}(k, n_0)$  require special consideration in regards to global convergence when multivariate convex minimum techniques such as the cyclic coordinate method [Bazaraa et al. 2005] are adopted to solve Problem (20) - (23).

For the  $k$  coordinate searching step, the Golden Point Intersection method [Bazaraa et al. 2005] is adopted. For  $n_0$  coordinate searching, considering the complexity of  $\hat{W}(k, n_0)$ , a grid search hybrid with the Armijio method [Bazaraa et al. 2005] is adopted here. The grid search method simply determines the objective value at each grid point. It may offer some protection against local minima but it is not very efficient. The Armijio method is adopted to search around the best grid point to hasten the convergence speed. The search methods we adopt here can only guarantee a local minimum. If  $\hat{W}(k, n_0)$  is strictly quasi-convex, the solution is also global. Unfortunately,  $\hat{W}(k, n_0)$  in general does not have such a property. The application of the grid search method helps to protect against local minima if the grid points are dense enough. Our experiments show that dozens of grid points hybrid with the Armijio method can help to find the global minimum very quickly.

With the optimization techniques described above and the prediction procedure in Section 4.2, local optimal parameters  $(k^*, n_0^*)$  are archived for the optimization of  $\hat{W}(k, n_0)$ . Then the ‘‘robustness’’ step is adopted to find tuning  $(k^*, n_0^*)$  locally so that it makes the two-level simulation work well and stably.

The  $(k, n_0)$  tuning procedure for a given total number of simulated payoffs  $C$  is described as follows.

(1) **Initialization**

Set the minimum of  $n_0$ ,  $n_{0,\min}$ , as 30 and set  $n_0^*$  as  $n_{00}$  and  $k$  as  $k_0$ .

Set the minimum of  $k$ ,  $k_{\min}$ , as  $40/p$  and calculate  $k_{\max}$  from Constraint (21) under given  $n_0^*$ .

Let  $a \leftarrow 0$  and  $\hat{W}_N^* \leftarrow \hat{W}(k_0, n_{00})$ .

(2) **DO (with the prediction procedure in Section 4.2):**

(a) **Searching for  $k^*$**

Golden Method to search the optimal  $k^* \in [k_{\min}, k_{\max})$  under given  $n_0^*$ .

Let  $k_{\text{old}} \leftarrow k$  and  $k = k^*$ .

(b) **Updating  $n_0^*$**

i. Let  $n_{\text{old}} \leftarrow n_0^*$ .

ii. If  $a = 0$ , Generate grids of  $n_0$ :

Generate the guesses of  $n_0$  by computing  $\frac{d^2 S_{ij}(n_{00})}{(X_i(n_{00}) - X_j(n_{00}))^2}$  for certain pairs of  $i$  and  $j$ ;

Calculate their corresponding  $\hat{W}(k^*, n_0)$ ;

Find the  $n_0^*$  that has the minimal  $\hat{W}(k^*, n_0)$  among all grid points;

Take Armijo search along the decreasing direction of  $n_0^*$ , keep on updating  $n_0^*$  such that  $\hat{W}(k^*, n_0^*)$  is the minimum.

Set  $a \leftarrow 1$ .

- iii. locally optimizing  $n_0^*$ :  
Perform a climbing-hill search on both sides of  $n_0^*$ , keep updating  $n_0^*$ .
- iv. Update  $k_{\min}$  or  $k_{\max}$ :  
if  $n_0^* > n_{\text{old}}$ , let  $k_{\max} \leftarrow k^*$ ; else let  $k_{\min} \leftarrow k^*$ .
- v. Update  $\hat{W}_N^* \leftarrow \hat{W}(k, n_0)$   
While  $(|n_{\text{old}} - n_0^*| > 0.1n_{\text{old}} \vee 10$  or  $|k_{\text{old}} - k^*| > 10/p)$ .

### (3) Robustness Step

- (a) Set  $\hat{W}_D = \hat{W}(k^*, n_0^*)$  with the prediction procedure in Section 4.3,  $a \leftarrow 0$ .  
Set  $k_{\min} = 40/p \vee k^*/4$  and  $k_{\max} = 4 * k^*$ .  
Set  $k^R = k^*$  and  $n_0^R = n_0^*$ .
  - (b) DO (with the prediction procedure in Section 4.3):
    - i. Set  $n_{\text{old}} \leftarrow n_0^R$  and perform a climbing-hill search on both sides of  $n_0^R$ , keep updating  $n_0^R$ .
    - ii. Set  $k_{\text{old}} \leftarrow k^R$  and use the Golden Method to search the optimal  $k^R \in [k_{\min}, k_{\max})$  under given  $n_0^R$ .
    - iii. Update  $k_{\min}$  or  $k_{\max}$ :  
if  $n_0^R > n_{\text{old}}$ , let  $k_{\max} \leftarrow k^R$ ; else let  $k_{\min} \leftarrow k^R$ .
    - iv. Update  $\hat{W}_D^* \leftarrow \hat{W}(k, n_0)$   
While  $(|n_{\text{old}} - n_0^R| > 0.1n_{\text{old}} \vee 10$  or  $|k_{\text{old}} - k^R| > 10/p)$  and  $a < 3$ .
  - (c) If  $\hat{W}_D^R > 1.1\hat{W}_N^*$ , set  $k^R = k^*/1.2$  and  $n_0^R = 1.2n_0^*$ .
- (4) Output:  $(k^R, n_0^R)$ .

## 6. FIXED $T$ PROCEDURE

Similarly, by approximating  $W$  by  $\hat{W}$ , we simplify Problem (9) to Problem (24)

$$\min \hat{W}(k, n_0) \tag{24}$$

$$\text{s.t. } C_1 = \frac{T - kn_0R_{\text{sam}} - kR_{\text{scr}} - (k \log k)(R_{\text{scrsort}} + R_{\text{com}}) - Mn_0R_{\text{scr2}}}{R_{\text{sam}}} \tag{25}$$

$$\begin{aligned} k &\geq 40/p \\ n_0 &\geq 30 \end{aligned}$$

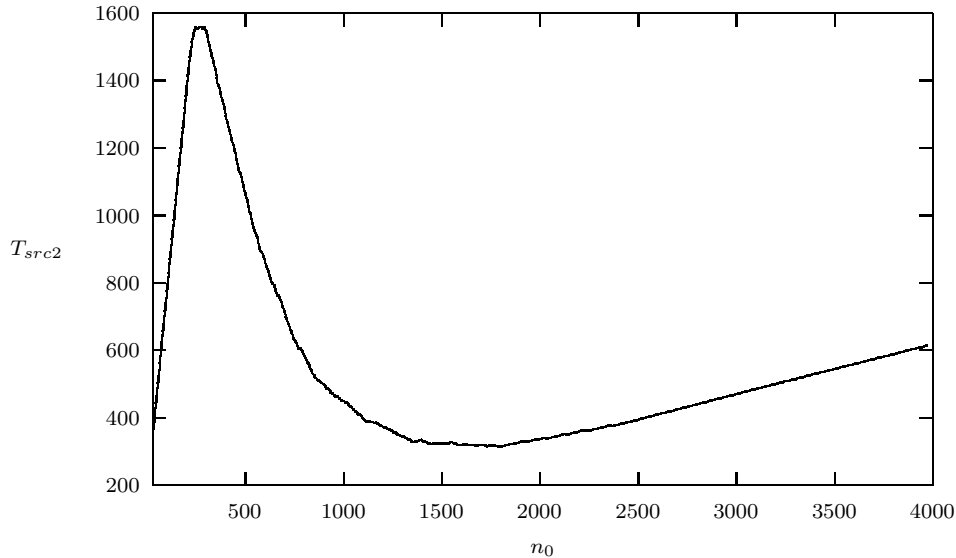
where Constraint 25 expresses the allocation of computational time between the outer level and inner level. The other constraints have the same meaning as Constraint (22) - (23).

The total simulated payoffs at the second stage,  $C_1$ , decreases on  $k$ ; however, in most applications, it is not monotone on  $n_0$  due to the fact that  $Mn_0R_{\text{scr2}}$  (denoted as  $T_{\text{scr2}}$ ) goes up, down and then up as  $n_0$  increases, as shown in Figure (6). As the first stage sample size  $n_0$  increases, the computation of each comparison increases, while the total comparisons need to be done,  $M$ , may decrease. When  $n_0$  is very small, the increment of each comparison overcomes the decrement of the number of total comparisons  $M$ , then  $T_{\text{scr2}}$  increases; later along with the increment of  $n_0$ , the number of total comparisons  $M$  has already been decreased to a modest number and the decreasing rate might be even larger, then  $T_{\text{scr2}}$  decreases. Finally, when  $n_0$  is large enough,  $M$  remains almost unchanged as  $n_0$  increases, then  $T_{\text{scr2}}$  increases again.

The procedure of given  $T$  is almost the same as the procedure described in Section 5. The only difference is that  $k^R$  is set to be  $\frac{k^*}{1.1}$  in the ‘‘robustness’’ step when  $\hat{W}_D^R > 1.1\hat{W}_N^*$ .

## 7. EXPERIMENTS

In this section, we test the ‘‘goodness’’ of the pair of  $(k, n_0)$  suggested by our procedures. By ‘‘goodness’’ we mean that  $(k, n_0)$  is close to the optimal choice and robust to different runs of the two-level simulation. As the user usually does not rerun the same experiment of two-level simulation with different random seeds to obtain the average output, it is important to have the pair  $(k, n_0)$  suggested by our procedures work robustly as well as efficiently.

Fig. 5. Relation of  $T_{scr2}$  and  $n_0$ 

The test beds we used are the experiments of selling a single put option, and holding a portfolio of options on two stocks discussed in Lan et al. 2008. For simplicity, we call them Problem 1 and Problem 2. More experiments are shown in Lan 2009.

We tested the simulation procedures by producing a 90% confidence interval (CI) for  $ES_{0.99}$  of Problem 1 and Problem 2. The  $ES_{0.99}$  of Problem 1 is \$3.39 and that of Problem 2 is \$32.4. Unless specified, the number of macro-replication is 20. The decomposition of the total allowed error probabilities,  $\alpha = \alpha_o + \alpha_s + \alpha_{hi} + \alpha_{lo}$ , is far less important to the performance of the two-level simulation than the two parameters  $(k, n_0)$ , where the terms on the right side are error probabilities allocated to an outer-level confidence interval, screening, and inner-level lower and upper confidence limits, respectively. The composition of  $\alpha_o = \alpha/2$ ,  $\alpha_s \approx \alpha/4$ ,  $\alpha_{hi} = \alpha_{lo} = (\alpha - \alpha_o - \alpha_s)/2$  works well in all our experiments.

We ran the experiments on a PC with a 2.4 GHz CPU and 4 GB memory under 64-bit Red Hat Linux. The code was written in C++ and compiled by gcc 3.4.6.

We set  $k_0 = \frac{40}{p}$  and  $n_{00}$  automatically by the procedure in Appendix B.1 to ensure that  $\mathcal{T}(k_0, n_{00})$  is accurate enough for the prediction procedures to work well.

### 7.1 Effectiveness and Robustness of a Given $C$ Procedure

To show the effectiveness of our procedure, we compare the estimated confidence interval width simulated with the chosen  $(k, n_0)$  of our procedure and those pairs of  $(k, n_0)$  selected arbitrarily. For problem 1, the number of scenarios  $k$  and first stage sample size  $n_0$  are retrospectively selected from  $250 \leq k \leq 128000$  and  $10 \leq n_0 \leq 800$  as  $C = 16000000$ . A total of 66 pairs of  $(k, n_0)$  were arbitrarily selected. With some pairs of  $(k, n_0)$  like  $(k, n_0) = (256000, 30)$ , from which the two-level simulation is extremely time-consuming and the confidence intervals estimated are obviously far from the best. For such a case, we only ran a small number of macro replications, such as 5 or 10. Otherwise 100 replications were tested for each pair of  $(k, n_0)$  selected. A total of 4877 macro replications were run. The experimental results are listed in Lan 2009. Four typical pairs of  $(k, n_0)$  and their simulation results are shown in Figure 6. As displayed in the figure, within a very wide range of  $(k, n_0)$ , the variances of the estimated confidence interval width generated by different runs of the simulation are relatively small, which means that the two-level simulation works stably with such parameters and it is both feasible and meaningful to find a good pair of  $(k, n_0)$  within such a range. Meanwhile, there are some pairs of  $(k, n_0)$ , like  $(k, n_0) = (256000, 30)$ , the two-level simulation has

a chance to generate very wide confidence intervals although sometimes it works very well. Such pairs of  $(k, n_0)$  should be avoided.

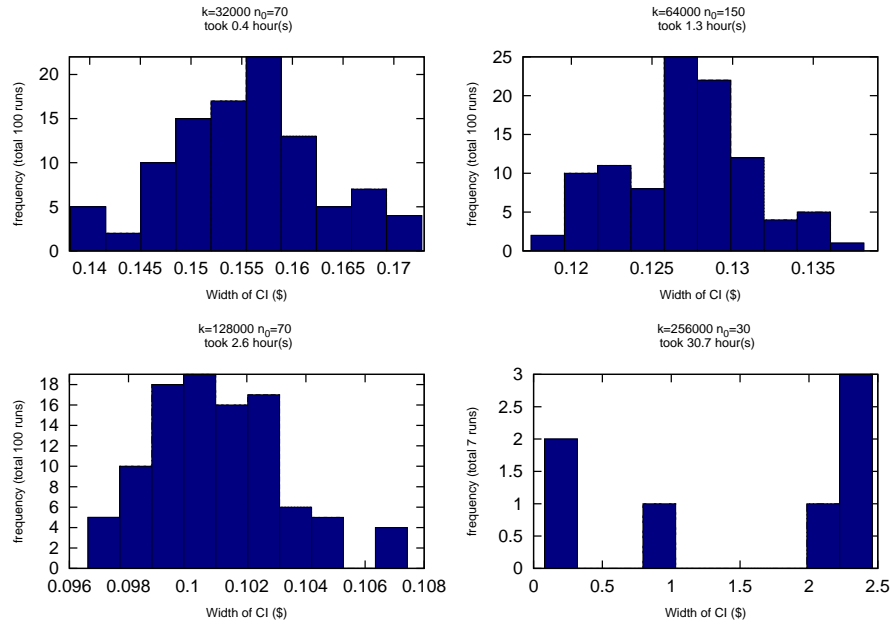


Fig. 6. Histogram of simulation performance of Problem 1 with different pairs of  $(k, n_0)$  (arbitrarily selected)

With  $C = 16000000$ , the procedure to find the good parameters was run 20 times with same  $k_0$  but different  $\mathcal{T}(k_0, n_{00})$ . For each chosen  $(k, n_0)$ , 20 replications of the two-level simulation were run. The results are shown in Lan 2009. The experiment shows that our procedure can find a pair of  $(k, n_0)$  with which the two-level simulation works almost the best, and meanwhile stably. In other hand, the pairs of  $(k, n_0)$  suggested from our procedure are stable across different  $\mathcal{T}(k_0, n_{00})$ . Figure 7 confirms the above description vividly. The cross points are the  $(k, n_0)$  arbitrarily selected, almost evenly partitioning the feasible region with denser designs around the optima. The star points are the  $(k, n_0)$  suggested by our procedure, which are shown to be efficient in terms of the precision of the estimated confidence intervals and robust across multiple replications of the two-level simulation and also in terms of the different  $\mathcal{T}(k_0, n_{00})$  that we used.

For a different computational budget  $C$ , similar experiments have been done. All results (in Lan 2009) confirm that our procedure can find efficient and robust pair of parameters  $(k, n_0)$  for the two-level simulation.

## 7.2 Effectiveness and Robustness of a Given $T$ Procedure

The efficiency and robustness of our procedure for given computational time  $T$  can be illustrated in a similar way. We define the accomplish rate as the ratio of the number of replications of the two-level simulation accomplished within a given time  $T$  to the total number of replications we run. This is different from the total computational budget given in the total number of payoffs simulated  $C$ , in which case, it is nature to tell if a pair of  $(k, n_0)$  is feasible or not without any prior information about the behavior of the two-level simulation. When the total computational budget is given in the wall clock time  $T$ , it becomes random. By assuming that the running environments remain almost the same between the two-level simulation and the test run to generate  $\mathcal{T}(k_0, n_{00})$ , our procedure can find a pair of  $(k, n_0)$  not only feasible for most times but also close to the best attainable in very similar running environments.

Figure 8 shows the experiment of the two-level simulation with Problem 2, in which  $T = 240$  seconds and  $(k, n_0)$  are arbitrarily selected. It indicates that there is no simple rule to choose  $(k, n_0)$  for a given  $T$  as the relationships of  $W$  and  $(k, n_0)$  are not monotone, and the two procedure parameters,  $k$  and  $n_0$ ,

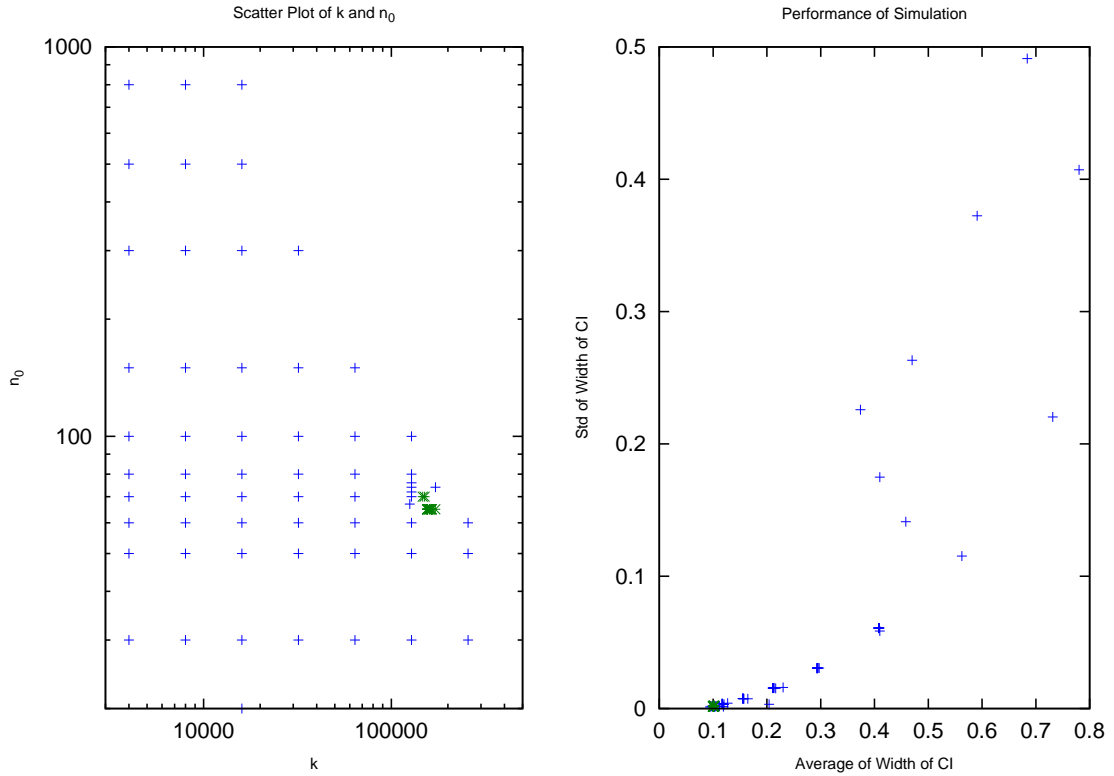


Fig. 7. Simulation performance of Problem 1 with  $(k, n_0)$  suggested by our procedure and arbitrarily selected

depend on each other. The pair,  $(k, n_0) = (6000, 6000)$  is very risky and should not be suggested by our procedure, although sometimes the two-level simulation can generate very narrow confidence intervals. As

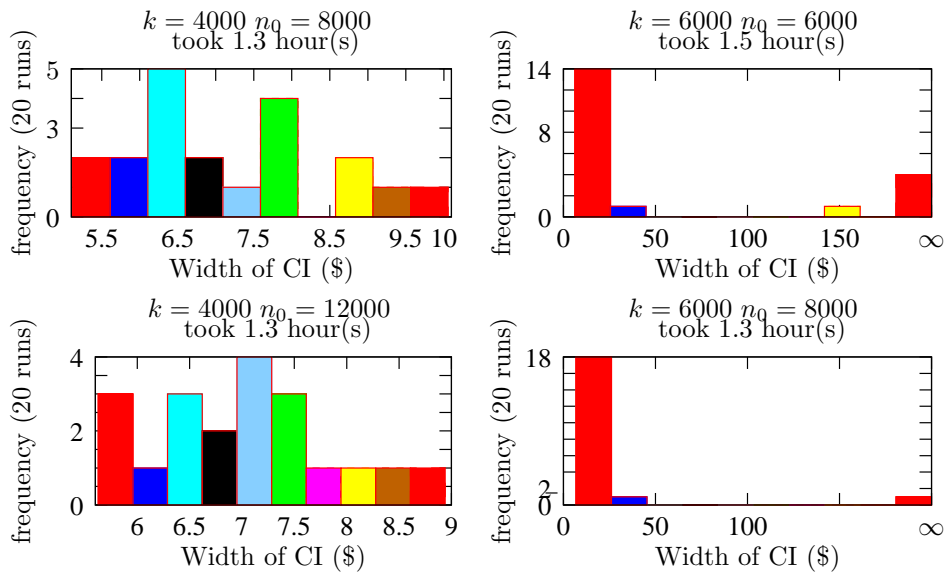


Fig. 8. Histogram of simulation performance of Problem 2 with different pairs of  $(k, n_0)$  (arbitrarily selected)

shown in Figure 9, compared to an arbitrarily selected  $(k, n_0)$ , the pair of  $(k, n_0)$  suggested by our procedure is demonstrated to be effective and robust. The experiment was done with  $T = 240$  seconds and a total of 20 pairs of  $(k, n_0)$  are found with different pilot information  $\mathcal{T}(k_0, n_{00})$ . Twenty replications of the two-level simulation were run to test the effectiveness and robustness of each pair. The accomplish rate of all the experiments with  $(k, n_0)$  suggested by our procedure is 100%. We confirm again that the  $(k, n_0)$  suggested by our procedure are almost the most efficient and robust procedure parameters; and our procedure is also robust to the pilot information.

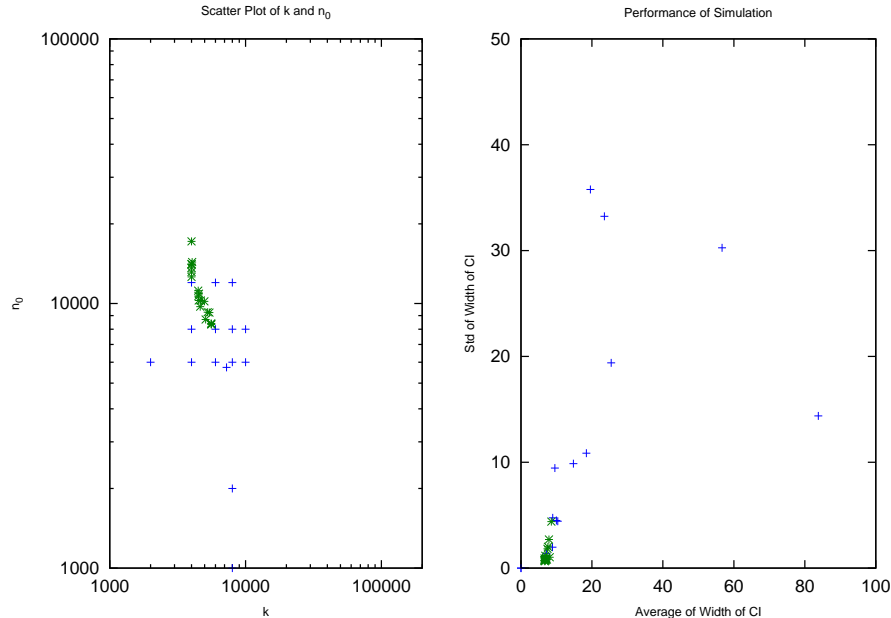


Fig. 9. Simulation performance of Problem 2 with  $(k, n_0)$  suggested by our procedure and arbitrarily selected

## 8. CONCLUSION AND FUTURE WORK

In this paper, we have proposed algorithms to predict the performance of a two-level simulation screening based on information retrieved from historical data or a test run. Furthermore, we introduced procedures to choose good values of the number of scenarios to explore and the number of payoffs to generate at the first stage for a given total number of simulated payoffs or given total allowed experiment time (wall clock). Experimental results indicate that by tuning the parameters through the procedures we provided, the two-level simulation works efficiently and stably. Compared with arbitrarily selected experimental parameters, we have a very good chance to generate much accurate estimation of the risk measure, the expected shortfall.

In the prediction algorithms, the sum of dependent Bernoulli random variables is approximated by a normal random variable by ignoring the dependency. Yu and Zelterman [2002] proposed a new distribution to describe the sum of dependent Bernoulli random variables. By introducing their work into the prediction algorithm, a better prediction may be attained.

Another possible future direction is to replace the screening procedure with a more efficient one, such as a derivative of the adaptive or multi-stages screening algorithm in Lesnevski et al. [2008]; Kim and Nelson [2006]. By taking the total number of simulated payoffs or the total time for screening as the decision variable and allowing the actual used simulation resources to vary a little bit around the designed value, the overall performance could be further improved. The optimization of the parameters of adaptive procedures has seldom been studied in prior literature.

## REFERENCES

- ACERBI, C. AND TASCHE, D. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 1487–1503.
- BAZARAA, M. S., SHERALI, H. D., AND SHETTY, C. 2005. *Nonlinear Programming: theory and algorithms*, 3rd ed. Wiley-Interscience.
- CHEN, C.-H., LIN, J., YÜCESAN, E., AND CHICK, S. E. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems* 10(3), 251–270.
- KIM, S.-H. AND NELSON, B. L. 2006. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. *Operation Research* 54, 475–488.
- LAN, H. 2009. Two-level simulation of expected shortfall: Confidence intervals, efficient simulation procedures and high-performance computing (in preparation). Ph.D. thesis, Northwestern University.
- LAN, H., NELSON, B. L., AND STAUM, J. 2007. a confidence interval for tail conditional expectation via two-level simulation. In *Proceedings of the 2007 Winter Simulation Conference*, S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, Eds. IEEE Press, 949–957.
- LAN, H., NELSON, B. L., AND STAUM, J. 2008. A confidence interval procedure for expected shortfall via two-level simulation. Northwestern University Working Paper.
- LESNEVSKI, V., NELSON, B. L., AND STAUM, J. 2008. An adaptive procedure for estimating coherence risk measures based on generalized scenarios. *Journal of Computational Finance* 11, 4, 1–31.
- RINOTT, Y. 1978. On two-stage selection procedures and related probability-inequalities. *communications in Statistics - Theory and Methods* A8, 799–811.
- STEIN, C. 1945. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* 16(3), 243–258.
- YU, C. AND ZELTERMAN, D. 2002. Sums of dependent bernoulli random variables and disease clustering. *Statistics & Probability Letters* 57, 4 (May), 363–373.

## A. PROOF OF EQUATION (19)

The number of comparisons needed to be done  $M$  can be predicted as:

$$E[M] = E\left[\sum_{i'=\lceil kp \rceil+1}^k M_{\pi_0(i')}\right].$$

As shown in Figure 10, by assuming the probability of the event that the market scenario  $\pi_0(i')$  is beaten by  $\pi_0(j')$  is given by following step function

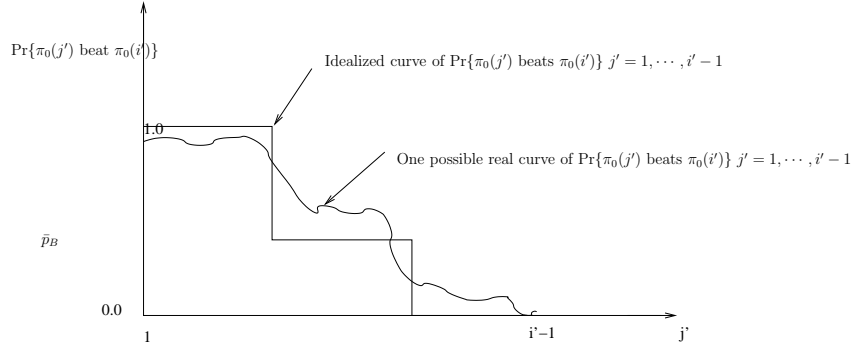
$$\Pr\{\pi_0(j') \text{ beats } \pi_0(i')\} = \begin{cases} 1, & j' \leq b; \\ \bar{p}_B, & b < j' < i' - e - 1; \\ 0, & i' - e - 1 \leq j' \leq i' - 1. \end{cases} \quad (26)$$

we can simplify the calculation of  $E[M_{\pi_0(i')}]$  while keeping the approximation close to reality. In Equation (26),  $b$  and  $e$  depend on  $i'$ . For simplicity of notation, we do not explicitly show such dependence. Variables  $b$  and  $e$  can be computed through the procedure described in Section 4.2, as can the average probability  $\bar{p}_B$ .

Under all the above assumptions, we have

$$E[M_{\pi_0(i')}] = \begin{cases} \lceil kp \rceil, & \text{when } b > \lceil kp \rceil; \\ i' - 1, & \text{when } i' - e - 1 < \lceil kp \rceil; \\ b + \sum_{h=\lceil kp \rceil-b}^{i'-1-e-b} h \binom{h-1}{\lceil kp \rceil-b-1} \bar{p}_B^{\lceil kp \rceil-b} (1-\bar{p}_B)^{h+b-\lceil kp \rceil} & \text{o.w.} \\ + (1 - \text{NB}_{\bar{p}_B}(\lceil kp \rceil - b, i' - e - \lceil kp \rceil - 1))(i' - b - 1). & \end{cases} \quad (27)$$

Since  $h \binom{h-1}{\lceil kp \rceil-b-1} \bar{p}_B^{\lceil kp \rceil-b} (1-\bar{p}_B)^{h+b-\lceil kp \rceil} = \frac{\lceil kp \rceil-b}{\bar{p}_B} \binom{h}{\lceil kp \rceil-b} \bar{p}_B^{\lceil kp \rceil-b+1} (1-\bar{p}_B)^{h+b-\lceil kp \rceil}$ , we can further simplify Equation (27) to Equation (19).

Fig. 10. Probability of  $\pi_0(j')$  beating  $\pi_0(i')$ 

## B. PARAMETERS $K_0$ AND $N_{00}$ OF THE TEST RUN

Until now, we have assumed that  $\mathcal{T}(k_0, n_{00})$  contains information necessary for the prediction of  $k$  and  $n_0$ . Here we discuss how to determine the  $k_0$  and  $n_{00}$  of the test run upon which  $k$  and  $n_0$  are predicted.

For  $k_0$ , we have demonstrated that when the number of scenarios  $k \geq 40/p$ , the coverage rate of the estimated confidence interval of the expected shortfall  $ES_{1-p}$  is no less than the nominal confidence level [Lan et al. 2007]. That is, when  $k \geq 40/p$ , the asymptotic correctness of our procedure holds. We then also require that  $k_0 \geq 40/p$ . In our experiments,  $p = 0.01$ , we took  $k_0 = 4000$ , which worked well for all tested problems.

The behavior of screening procedure under  $(k, n_0)$  is predicted by using  $\frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{S_{ij}(n_{00})}$ ,  $i, j = 1, \dots, k_0$  repeatedly. We need to set  $n_{00}$  large enough such that  $\frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{S_{ij}(n_{00})}$  is close to its true value. Suppose that  $\frac{\tau_{ij}^2}{n_{00}}$  is the variance of  $\frac{\bar{X}_i(n_{00}) - \bar{X}_j(n_{00})}{S_{ij}(n_{00})}$ , then we can determine  $n_{00}$  by

$$n_{00} \geq \frac{\tau_{ij}^2}{\frac{(\bar{X}_i(n'_{00}) - \bar{X}_j(n'_{00}))^2}{S_{ij}^2(n'_{00})} a^2} \quad \forall i, j = 1, \dots, k_0,$$

where  $a$  stands for the ratio of standard variance to mean value and is set to 0.05 in all experiments,  $n'_{00}$  is the size of samples used to estimate  $n_{00}$ , usually chosen as 160 or even larger because all the data can be reused in the test run. Let  $Y_h(i, j) = X_{i,h} - X_{j,h}$   $h = 1, \dots, n'_{00}$ . We often drop  $(i, j)$  and write  $Y_h(i, j)$  as  $Y_h$  when the context is clear. Suppose  $y_1, y_2, \dots$  are the moments of  $Y_h$ . We denote  $\hat{y}_1, \hat{y}_2, \dots$  as the ordinary estimators of  $y_1, y_2, \dots$ , that is,  $\hat{y}_1 = \sum_{h=1}^{n'_{00}} Y_h$ ,  $\hat{y}_2 = \sum_{h=1}^{n'_{00}} Y_h^2$  and so on. From the  $\Delta$ -method, we can estimate  $\tau_{ij}^2$  by

$$\hat{\tau}_{ij}^2 = \frac{1}{(\hat{y}_2 - \hat{y}_1^2)^3} \left\{ \hat{y}_1^2 \hat{y}_4 - \frac{1}{4} \hat{y}_1^2 \hat{y}_2^2 - \hat{y}_1 \hat{y}_2 \hat{y}_3 + \hat{y}_2^3 \right\}.$$

Then we can estimate the lower bound of  $n_{00}$  by

$$\max_{i,j} \frac{400}{(\hat{y}_2 - \hat{y}_1^2)^2 \hat{y}_1^2} \left\{ \hat{y}_1^2 \hat{y}_4 - \frac{1}{4} \hat{y}_1^2 \hat{y}_2^2 - \hat{y}_1 \hat{y}_2 \hat{y}_3 + \hat{y}_2^3 \right\}. \quad (28)$$

Usually  $i = 1, j \in \{\pi_0(k_0 p), \pi_0(k_0)\}$  is taken instead of the whole possible set of  $i, j$  to save computation. See Appendix B.1 for the procedure on how to determine  $n_{00}$  adaptively and generate the test run data  $\mathcal{T}(k_0, n_{00})$ .

### B.1 Procedure of $\mathcal{T}(k_0, n_{00})$

We introduce the procedure used to generate the test run data  $\mathcal{T}(k_0, n_{00})$  adaptively.

#### (1) Initialization:

Set  $n_{00} \leftarrow 240$  (default value),  $k_0 \leftarrow 40/p$  and  $n_{\max} = 10^9/k_0$  (limited by the memory the computer can

provide).

Set  $a \leftarrow 0$ .

(2) **Generating Scenarios:**

Generate  $k_0$  scenarios independently and calculate the unit time used on generating market scenario  $R_{\text{sce}}$ .

(3) **Simulation of Portfolio Payoffs**

(a) Set  $n_{\text{old}} \leftarrow 0$

(b) DO:

i. Generate additional  $n_{00} - n_{\text{old}}$  payoffs for each market scenario with Common Random Numbers.

ii. Set  $n_{\text{old}} \leftarrow n_{00}$ .

iii. Update  $n_{00}$  according to Equation (28).

iv. Set  $n_{00} \leftarrow n_{00} \wedge n_{\text{max}}$  and  $a \leftarrow a + 1$ .

While( $n_{00} > n_{\text{old}}$  and  $a < 3$ )

(c) Set  $n_{00} \leftarrow n_{\text{old}}$ .

(d) Calculate the unit time spent on simulation of one payoff  $R_{\text{sam}}$ .

(4) **Screening:**

Perform the screening operation on all the  $k_0$  scenarios.

Record  $\bar{X}_i(n_{00})$ ,  $S^2(n_{00})$ ,  $\pi_0(\cdot)$  and  $S_{ij}^2(n_{00})$ .

Calculate the time related coefficients  $R_{\text{scr}1}$ ,  $R_{\text{scr}2}$  and  $R_{\text{scr}3}$ .

(5) **Computing the Empirical Likelihood Estimator:**

Compute the Empirical Likelihood estimation with data  $\bar{X}_{\pi_0(1)}(n_{00}), \dots, \bar{X}_{\pi_0(I)}(n_{00})$ .

Calculate coefficient  $E_{\text{out}}$  and  $R_{\text{com}}$ .