McMaster University

Advanced Optimization Laboratory

AdvOL

Title:

Portfolio Credit-Risk Optimization

Authors:

Ian Iscoe, Alexander Kreinin, Helmut Mausser and
Oleksandr Romanko

Portfolio Credit-Risk Optimization

Ian Iscoe[a]**, Alexander Kreinin**[a]**, Helmut Mausser**[a]**, Oleksandr Romanko**[ab]

[a] *Quantitative Research Group, Algorithmics Inc., 185 Spadina Avenue, Toronto, ON, Canada*
[b] *Dept. of Computing and Software, McMaster University, 1280 Main St. West, Hamilton, ON, Canada*

December 29, 2009

Abstract

This paper evaluates several alternative formulations for minimizing the credit risk of a portfolio of financial contracts with different counterparties. Credit risk optimization is challenging because the portfolio loss distribution is typically unavailable in closed form. This makes it difficult to accurately compute Value-at-Risk (VaR) and expected shortfall (ES) at the extreme quantiles that are of practical interest to financial institutions. Our formulations all exploit the conditional independence of counterparties under a structural credit risk model. We consider various approximations to the conditional portfolio loss distribution and formulate VaR and ES minimization problems for each case. Two realistic credit portfolios are used to assess the in- and out-of-sample performance for the resulting VaR- and ES-optimized portfolios, as well as for those obtained by minimizing the variance or the second moment of the portfolio losses. We find that a Normal approximation to the conditional loss distribution performs best from a practical standpoint.

Keywords: credit risk, optimization, portfolio optimization, risk modelling, Value-at-Risk, expected shortfall

1 Introduction

For financial institutions, the benefits of managing (portfolio) credit risk include not only reduced monetary losses due to defaulted or downgraded obligations but also lower capital charges. While individual credit-risky positions can be hedged with credit derivatives such as credit default swaps, imperfectly correlated credit movements among counterparties also provide opportunities for mitigating credit risk at the portfolio level through diversification. In particular, using optimization techniques to restructure portfolios of credit-risky positions is an attractive possibility. However, such procedures face numerous challenges, foremost being the difficulty of representing the portfolio credit loss distribution with sufficient accuracy. This paper formulates several alternative optimization problems that derive from a structural (Merton) model of portfolio credit risk, and evaluates their effectiveness from the perspectives of risk mitigation and computational practicality.

Credit risk refers to the potential monetary loss arising from the default, or a change in the perceived likelihood of default, of a counterparty to a financial contract. Note that a reduction in the default probability, i.e., a transition to a more favourable credit state, results in a monetary gain. However, such gains are generally small relative to the losses that occur due to severe credit downgrades or default. Thus, the credit loss distribution (F) for a typical investment-grade portfolio is positively skewed, the long right tail being consistent with a small likelihood of substantial losses.

The complex relationships among asset prices, exposures and credit transitions preclude obtaining a closed-form representation of the actual credit loss distribution. Thus, for risk management purposes, it is necessary to replace F by some approximating distribution $\hat{F}$. The form of $\hat{F}$ varies depending on the underlying credit loss model. For example, reduced-form models (e.g., CreditRisk+ (Credit Suisse Financial Products, 1997)) provide $\hat{F}$ in closed form. However, their underlying assumptions may be viewed as overly simplistic in that they fail to capture the effects of credit state migrations and correlated movements of risk factors (de Servigny and Renault, 2004). In contrast, structural models (Gupton, Finger and Bhatia, 1997; Iscoe, Kreinin and Rosen, 1999) can provide a more realistic representation but typically require $\hat{F}$ to be an empirical distribution derived from Monte Carlo (MC) simulation.

Computing $\hat{F}$ from Monte Carlo simulation presents challenges for assessing credit risk because common risk measures, such as Value-at-Risk (VaR)[1] and expected shortfall (ES), involve extreme quantiles in the right tail. Thus, obtaining accurate risk estimates requires a huge number of samples, or scenarios. Initial attempts at minimizing credit risk relied exclusively on MC simulation and included the full set of loss scenarios in the formulation (Mausser and Rosen, 2000; Mausser and Rosen, 2001; Andersson, Mausser, Rosen and Uryasev, 2001; Zenios and Jobst, 2001; Zagst, Kehrbaum and Schmid, 2003). Clearly, a limitation of this approach is that the large size of the resulting optimization problem adversely affects computational performance. Subsequently, in Saunders, Xiouros and Zenios (2007) a large-portfolio approximation was used to obtain a more compact formulation.

More recently, variance-reduction techniques such as importance sampling have been shown (Tilke, 2006) to provide stable optimal solutions, with a relatively small number of scenarios. However, a potential problem with importance sampling, is that the required shift in distribution depends on the portfolio's risk, which of course changes with the portfolio's composition during the course of the optimization. Thus, it is not clear that the shift induced by the initial portfolio is also effective for the optimal portfolio.

Structural models infer a counterparty's credit state from its associated creditworthiness index, which depends on systemic risk factors in the form of credit drivers as well as a specific risk factor unique to each counterparty (Iscoe et al., 1999; Iscoe and Kreinin, 2000). Given a set of values for the credit drivers, credit transitions for all counterparties become independent. This conditional independence property can be exploited to obtain $\hat{F}$ in semi-analytical form, specifically, as a mixture of closed-form conditional loss distributions. Such representations are far more data-efficient than pure Monte Carlo sampling and the associated optimization problems are smaller as a result. In this paper, we evaluate the practicality of optimizing credit risk for three different representations of $\hat{F}$:

- Monte Carlo sampling;
- A mixture of Normal (Gaussian) conditional loss distributions;
- A mixture of conditional mean (expected) losses.

For comparison purposes, we also consider the performance of variance-based formulations as a way of reducing a portfolio's VaR and ES. Variance minimization, which dates back to the seminal work of Markowitz (1952), remains in widespread use for risk management purposes. It is well known that minimizing variance has the effect of also minimizing VaR and ES only for Normal distributions. Thus, in our context, minimizing variance effectively assumes that $\hat{F}$ is Normal. This is likely to be a poor approximation to the actual portfolio credit loss distribution. Nevertheless, its popularity makes variance minimization a useful benchmark when evaluating the performance of the structurally based formulations. Since variance only measures dispersion around the mean, as a second, related

---

[1]VaR is incorporated in the Basel II Capital Accord, which makes it an integral part of the regulatory requirements.

benchmark we also minimize the second moment of the credit loss distribution (which takes the mean into account).

Our formulations and computational experiments are intended to be consistent with managing the risk of a banking book. Since a typical banking book may contain thousands of counterparties, we allow for optimizing over groups of counterparties. Thus, a portfolio manager might elect to assign all counterparties from a given industry to the same group, for example, and then use the results of the optimization to restructure the portfolio at the industry level. Such an approach is much more practical than enacting changes to a large number of individual contracts, as might be suggested by optimizing at the counterparty level. We also limit the amount of trading to what can be implemented reasonably when rebalancing the banking book; namely short positions are not permitted and new groups may not be added to the existing portfolio. These limitations are enforced only to provide a realistic assessment of the optimization results; nothing precludes relaxing or eliminating such restrictions from a formulational standpoint. Finally, although we account for credit migration, it is assumed that exposures are deterministic, i.e., positions are not marked to market.

The rest of this paper is organized as follows. In Section 2 we introduce some notation and basic concepts, and describe relevant input data for modelling credit-risky instruments. Section 3 introduces the structural model for portfolio credit risk, for which future credit events may be simulated. In Section 4, several approximations are described for the loss distribution, F. Section 5 reviews the risk measures that will be optimized. The formulations of our credit-risk optimization problems follow in Section 6. We evaluate and analyze our computational results in Section 7. Finally, we present our conclusions, extensions and practical recommendations.

2   Portfolio Credit Losses

We are concerned with credit-risk modelling and optimization at the portfolio level only. Counterparty-level data is used as input to the portfolio-level models. Such data may be estimated from an internal model or provided by an external agency. We start by analyzing the input data for our portfolio level credit-risk optimization problems.

In this paper, we consider a single time period. At the end of the time period, each counterparty can migrate to a different credit state resulting in our losses ℓ_c^j, where $j = 1, \ldots, N_{\mathrm{CP}}$ indexes counterparties and c indexes credit states. There are C credit states available for counterparties, enumerated from $c = 0$ (default) through increasing credit ratings, to the highest credit rating $c = C - 1$. Note that negative losses (gains) are incurred if a counterparty migrates to a more favourable state. The probability of being in the state c at the end of the time period is p_c^j with p_0^j being the probability of default. Unadjusted exposure of counterparties at default corresponds to their values v^j ($v^j > 0$). In general, counterparty exposure is the economic loss that will be incurred on all outstanding transactions if a counterparty defaults, unadjusted by possible future recoveries. Exposures are computed subject to netting, mitigation and collateral. The recovery at default is assumed to be deterministic and recovery-adjusted exposures are equal to $v^j(1 - \gamma^j)$, where γ^j is the recovery rate.

A portfolio consists of a number, N_{CP}, of counterparties grouped into N_{G} groups. For instance, counterparties from the same country, the same industry and having the same credit rating can be grouped together. Changing positions in groups is more practical as it allows altering just the between-groups positions, leaving the problem of within-group rebalancing as a lower level problem. From the point of view of modelling and optimization, grouping decreases the number of decision

variables. Our modelling assumptions do not place any restrictions on the number of groups, group sizes or group compositions.

The value of the ith group, G_i, is $v_i = \sum_{j \in G_i} v^j$, and its loss is

$$\mathcal{L}_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{\text{CP } j \text{ is in credit state } c\}, \tag{1}$$

where $\mathbf{1}\{\}$ is the indicator function of the event in braces.

Let the decision variable, $x_i \geq 0$, $i = 1, 2, \ldots, N_{\mathrm{G}}$, denote the position in the ith group; $\boldsymbol{x} = (x_1, x_2, \ldots, x_{N_{\mathrm{G}}})^T$. Let $\boldsymbol{x}^0$ be the vector of positions in the initial portfolio. We set $x_i^0 = 1$ for all i, so that the positions are expressed as multiples of the initial holdings. The portfolio value is

$$v(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{G}}} v_i x_i. \tag{2}$$

The initial portfolio value is then $V = \sum_{i=1}^{N_{\mathrm{G}}} v_i$. The portfolio loss $\mathcal{L} \equiv \mathcal{L}(\boldsymbol{x})$ is defined as

$$\mathcal{L}(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{G}}} \mathcal{L}_i x_i. \tag{3}$$

3 Structural Model for Portfolio Credit Risk

The empirical distribution $\hat{F}$ of portfolio losses due to credit events, is obtained by simulation of an underlying structural model described below. Our simulation approach is based on the CreditMetrics framework (Gupton et al., 1997; Saunders and Allen, 2002) and the credit-risk portfolio framework from Iscoe et al. (1999), Iscoe and Kreinin (2000). The degree to which $\hat{F}$ approximates the true distribution F, and thus the quality of the associated risk estimates, depends on the number of samples. The effect of sample size is especially pronounced when estimating quantiles close to 1, as those lie in the extreme right tail of the distribution. It is well known that the variability of a (risk) estimate decreases as the number of samples increases.

A particular value of portfolio loss ℓ, is computed as a function of the sampled values of a set of risk factors (see (6)) that can be separated into two groups:

- Y denotes a set of systemic risk factors: credit drivers which are macroeconomic factors and sector indices;
- Z denotes a random vector of counterparty-specific, or idiosyncratic, credit-risk factors.

The joint distribution of default and migration events is described through the counterparties' *creditworthiness indices*. The creditworthiness index W_j determines the financial health of counterparty j, and is defined as

$$W_j = \beta^j Y_{n(j)} + \sqrt{1 - (\beta^j)^2} Z_j, \tag{4}$$

where Z_j is the idiosyncratic risk which is independent across counterparties and is normally distributed, $\mathcal{N}(0,1)$; $Y_{n(j)}$ is the counterparty's credit-driver, a standard normal random variable; credit drivers are correlated and normally distributed, $\mathcal{N}(0,\mathcal{C})$ with a given correlation matrix, $\mathcal{C}$; β^j is the factor loading parameter or the sensitivity of the counterparty j to its credit driver $Y_{n(j)}$. As a

result, the creditworthiness index W_j is normally distributed, $\mathcal{N}(0,1)$. Thus we are assuming that the creditworthiness index for each counterparty depends on one credit driver; i.e., the counterparty participates in only one sector (e.g., country-industry pair).

$P_c^j = \sum_{\varsigma \leq c} p_\varsigma^j$ is the cumulative probability of counterparty j being in credit state c or lower, so that $p_c^j = P_c^j - P_{c-1}^j$, with the convention, $P_{-1}^j \equiv 0$. The credit-state boundaries $\{B_c^j\}_{c=0}^{C-2}$ are defined as $B_c^j = \Phi^{-1}(P_c^j)$, $0 \leq c \leq C-2$ ($B_{-1}^j \equiv -\infty$, $B_{C-1}^j \equiv \infty$) due to $\mathbb{P}(W_j < B_c^j) = P_c^j$. A counterparty j is in credit state c at the time horizon if $B_{c-1}^j \leq W_j < B_c^j$, or equivalently:

$$\mathbb{P}(\text{CP } j \text{ is in credit state } c) = \mathbb{P}(B_{c-1}^j \leq W_j < B_c^j). \tag{5}$$

For this model, the group loss in (1) takes the specific form,

$$\mathcal{L}_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{B_{c-1}^j \leq \beta^j Y_{n(j)} + \sqrt{1-(\beta^j)^2} Z_j < B_c^j\}. \tag{6}$$

The key property of this model is *conditional independence*: given a value $\boldsymbol{y}$ of the credit drivers, Y, the creditworthiness indices are independent. Conditional independence allows us, in principle, to obtain the conditional loss distribution by convolution (Fast Fourier Transform). The downside of the convolution technique is that it is difficult to use it for optimization because of the large number of possible losses at the portfolio level. In the next section, we look at some practical alternatives.

4 Loss Distribution Approximations

Conditional independence gives rise to several variants of the credit-loss distribution model. Scenarios on Y are generated and then the *conditional* loss distribution is approximated by one of the methods. The *unconditional* loss distribution $\hat{F}$ is obtained as the mixture of the conditional loss distributions. Methods for approximating conditional loss distributions (Iscoe and Kreinin, 2008) include MC-sampling approximation (Section 4.1), Cental Limit Theorem (CLT) approximation (Section 4.2), and Law of Large Numbers (LLN) approximation (Section 4.3).

The number of scenarios can be greatly reduced if one makes simplifying assumptions about the loss distribution and/or the portfolio's composition. For example, in Saunders et al. (2007) an LLN approximation is also used in which the number of issuers is so large that their individual risks effectively "cancel out" statistically (i.e., the idiosyncratic risk is eliminated).

4.1 MC-Sampling Approximation

If we generate a sample $\boldsymbol{y}$ from the distribution of Y, the creditworthiness indices are conditionally independent given $Y = \boldsymbol{y}$. Since the idiosyncratic credit-risk factors are independent of the credit drivers, any number of samples $\boldsymbol{z}$ can be combined with the sample $\boldsymbol{y}$ while still preserving the required codependence structure.

Under the MC-sampling approximation, MK scenarios are generated as follows:

1. Generate a random sample $\boldsymbol{y}_s$, $s = 1, \ldots, M$ of systemic factors from the distribution of Y.
2. For each $s \in \{1, 2, \ldots, M\}$, generate a random sample $\boldsymbol{z}_{sk}$, $k = 1, \ldots, K$ of idiosyncratic factors from the distribution of Z (independently, across s).

We denote the jth counterparty's loss in the (s, k)th scenario, $(\boldsymbol{y_s}, \boldsymbol{z}_{sk})$, by

$$\ell_{sk}^j := \sum_{c=0}^{C-1} \ell_c^j \cdot \mathbf{1}\{B_{c-1}^j \leq \beta^j (\boldsymbol{y_s})_{n(j)} + \sqrt{1 - (\beta^j)^2} (\boldsymbol{z}_{sk})_j < B_c^j\}, \tag{7}$$

so that the sampled value of the ith group's loss is (cf. (6))

$$\ell_{i,sk} = \sum_{j \in G_i} \ell_{sk}^j.$$

The sampled portfolio loss is

$$L_{sk}(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{G}}} \ell_{i,sk} x_i.$$

The MC-sampling approximation to F, the cumulative distribution function (cdf), of the portfolio losses, is computed as

$$\hat{F}^{\mathrm{MC}}(\ell; \boldsymbol{x}) = \frac{1}{MK} \sum_{s,k} \mathbf{1}\{L_{sk}(\boldsymbol{x}) \leq \ell\}. \tag{8}$$

4.2 CLT Sampling Approximation

Another approximation that we can use instead of full MC sampling is the conditional application of the Central Limit Theorem (CLT) which is valid if the number of counterparties is large and the contribution of each counterparty is relatively small (granularity or "smallness" condition). Fewer scenarios are required for this approximation.

Under the CLT approximation, conditional losses (for each systemic sample s) are approximately normally distributed $\mathcal{N}(\mu_s(\boldsymbol{x}), \sigma_s^2(\boldsymbol{x}))$, where $\mu_s(\boldsymbol{x}) = \mathbb{E}[\mathcal{L}(\boldsymbol{x}) \,|\, Y = \boldsymbol{y_s}]$ is the conditional mean of total portfolio loss and $\sigma_s(\boldsymbol{x})^2 = \mathrm{Var}[\mathcal{L}(\boldsymbol{x}) \,|\, Y = \boldsymbol{y_s}]$ is its conditional variance.

To compute $\mu_s(\boldsymbol{x})$ and $\sigma_s^2(\boldsymbol{x})$, first note that the conditional probability of a counterparty j being in credit state c, given that $Y = \boldsymbol{y_s}$, is

$$\mathbb{P}(\text{CP } j \text{ is in credit state } c \,|\, Y = \boldsymbol{y_s}) = \mathbb{P}(B_{c-1}^j \leq \beta^j Y_{n(j)} + \sqrt{1 - (\beta^j)^2} Z_j < B_c^j \,|\, Y = \boldsymbol{y_s})$$

$$= \mathbb{P}\left(\frac{B_{c-1}^j - \beta^j \boldsymbol{y_s}}{\sqrt{1 - (\beta^j)^2}} \leq Z_j < \frac{B_c^j - \beta^j \boldsymbol{y_s}}{\sqrt{1 - (\beta^j)^2}}\right) = \Phi\left(\frac{B_c^j - \beta^j \boldsymbol{y_s}}{\sqrt{1 - (\beta^j)^2}}\right) - \Phi\left(\frac{B_{c-1}^j - \beta^j \boldsymbol{y_s}}{\sqrt{1 - (\beta^j)^2}}\right) \equiv p_{c,s}^j,$$

where Φ is the standard Normal cdf.

Under a given systemic scenario s, the conditional mean and variance of the loss due to the initial position with counterparty j are given by

$$\mu_s^j = \sum_{c \geq 0} \ell_c^j p_{c,s}^j,$$

$$(\sigma_s^j)^2 = \sum_{c \geq 0} (\ell_c^j)^2 p_{c,s}^j - (\mu_s^j)^2,$$

The conditional mean and variance of loss from the ith group are

$$\mu_{i,s} = \sum_{j \in G_i} \mu_s^j, \qquad \sigma_{i,s}^2 = \sum_{j \in G_i} (\sigma_s^j)^2.$$
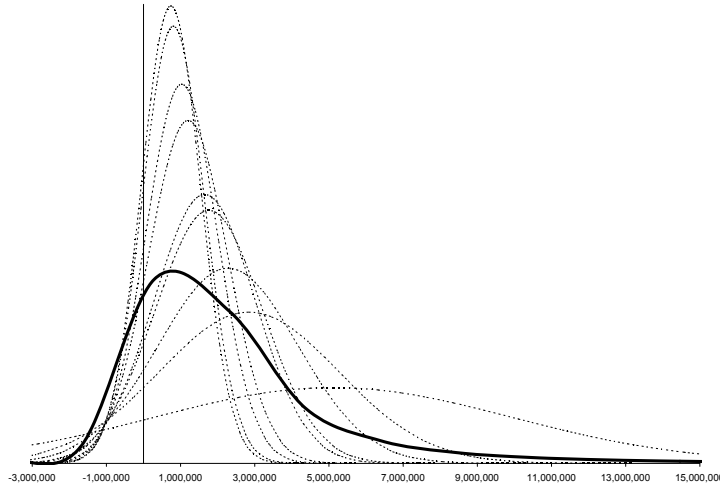
Figure 1: CLT Approximation of an Unconditional Loss Distribution

The portfolio's conditional mean and variance of loss are

$$\mu_s(\boldsymbol{x}) = \sum_{i=1}^{N_G} \mu_{i,s} x_i, \tag{9}$$

and

$$\sigma_s(\boldsymbol{x})^2 = \sum_{i=1}^{N_G} (\sigma_{i,s})^2 x_i^2. \tag{10}$$

The CLT approximation of the conditional portfolio loss distribution, is

$$\mathbb{P}_s(\mathcal{L}(\boldsymbol{x}) \leq \ell) \approx \Phi\left(\frac{\ell - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})}\right), \tag{11}$$

where $\mathbb{P}_s$ denotes the conditional probability measure with $\mathbb{P}_s(\mathcal{L}(\boldsymbol{x}) \leq \ell) = \mathbb{P}(\mathcal{L}(\boldsymbol{x}) \leq \ell \,|\, Y = \boldsymbol{y}_s)$.

The approximation to the unconditional distribution is then a mixture of Normal distributions and its cdf is equal to

$$\hat{F}^{\text{CLT}}(\ell; \boldsymbol{x}) = \frac{1}{M} \sum_{s=1}^{M} \Phi\left(\frac{\ell - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})}\right). \tag{12}$$

The resulting approximation for the unconditional loss distribution is illustrated in Figure 1. The dashed curves are the conditional Normal distributions of portfolio losses for $M = 9$ scenarios and the solid line is the unconditional loss distribution.

4.3   LLN Sampling Approximation

For a portfolio with a very large number of small counterparties we can use the Law of Large Numbers (LLN) to estimate conditional portfolio losses. In this case we assume that all specific risk

is diversified away so that the portfolio loss is the sum of expected losses. As a result, as $N_{\mathrm{CP}} \to \infty$ the conditional loss distribution is dominated by the mean loss over that scenario. As with the CLT conditional approximation, fewer scenarios are required for this approximation than for the MC approximation.

The LLN-approximation[2] of the conditional distribution of losses, is completely described by its mean, given by equation (9). The unconditional loss distribution is approximated by

$$\hat{F}^{\mathrm{LLN}}(\ell; \boldsymbol{x}) = \frac{1}{M} \sum_{s=1}^{M} \mathbf{1}\{\mu_s(\boldsymbol{x}) \leq \ell\}. \tag{13}$$

5 Risk Measures

Our primary interest is in quantile-based risk measures such as Value-at-Risk and expected shortfall (Conditional Value-at-Risk). For a detailed review of these risk measures we refer the reader to Szegö (2004). In addition to minimizing VaR and ES directly, we also consider minimizing variance and the second moment of losses, as alternative, indirect ways to reduce quantile-based risk. Our goal is to compare the performance of credit-risk optimization when we optimize over different risk measures and evaluate VaR and ES for the resulting optimal portfolios.

The estimated vector of unconditional mean losses is $\mathbb{E}[\mathcal{L}(\boldsymbol{x})]$ and the unconditional variance of losses is $\mathrm{Var}[\mathcal{L}(\boldsymbol{x})]$. Mean-variance (moment-based) optimization problems minimize variance or a combination of mean and variance. Another risk measure that can be used instead of variance is the *second moment*, which is $\mathbb{E}[\mathcal{L}(\boldsymbol{x})^2] = \mathrm{Var}[\mathcal{L}(\boldsymbol{x})] + \mathbb{E}[\mathcal{L}(\boldsymbol{x})]^2$.

Value-at-Risk is the maximum loss of a portfolio over a given time period and at a given level of probability. The Value-at-Risk function $\ell_\alpha(\boldsymbol{x})$ is the α-quantile of the loss distribution and is given by

$$\ell_\alpha(\boldsymbol{x}) = \min\{\ell \in \mathbb{R} : \mathbb{P}(\mathcal{L}(\boldsymbol{x}) \leq \ell) \geq \alpha\}.$$

Then for our structural model we have

$$\alpha = \mathbb{P}(\mathcal{L}(\boldsymbol{x}) \leq \ell_\alpha(\boldsymbol{x})) = \int \mathbb{P}_{\boldsymbol{y}}(\mathcal{L}(\boldsymbol{x}) \leq \ell_\alpha(\boldsymbol{x})) \, \mathrm{d}\varphi(\boldsymbol{y}) \approx \frac{1}{M} \sum_{s=1}^{M} \mathbb{P}_s(\mathcal{L}(\boldsymbol{x}) \leq \ell_\alpha(\boldsymbol{x})), \tag{14}$$

where φ is the distribution of the vector of credit drivers.

Expected shortfall $\mathrm{ES}_\alpha(\boldsymbol{x})$ is defined as the expected loss exceeding VaR and it can be written as

$$\mathrm{ES}_\alpha(\boldsymbol{x}) = \frac{1}{1-\alpha} \mathbb{E}\left[\mathcal{L}(\boldsymbol{x}) \cdot \mathbf{1}\{\mathcal{L}(\boldsymbol{x}) \geq \ell_\alpha(\boldsymbol{x})\}\right]. \tag{15}$$

For our structural model

$$\mathbb{E}\left[\mathcal{L}(\boldsymbol{x}) \cdot \mathbf{1}\{\mathcal{L}(\boldsymbol{x}) \geq \ell_\alpha(\boldsymbol{x})\}\right] \approx \frac{1}{M} \sum_{s=1}^{M} \mathbb{E}_s\left[\mathcal{L}(\boldsymbol{x}) \cdot \mathbf{1}\{\mathcal{L}(\boldsymbol{x}) \geq \ell_\alpha(\boldsymbol{x})\}\right].$$

---

[2]The usual interpretation of the LLN, would be that, under $\mathbb{P}_s$, $\lim_{N_{\mathrm{CP}} \to \infty} \mathcal{L}(\boldsymbol{x})/N_{\mathrm{CP}} = \lim_{N_{\mathrm{CP}} \to \infty} \mu_s(\boldsymbol{x})/N_{\mathrm{CP}}$ (assuming the latter limit exists), with $\mu_s(\boldsymbol{x})$ given by (9). We are using the term, LLN-approximation, to simply mean that under $\mathbb{P}_s$, $\mathcal{L}(\boldsymbol{x}) \approx \mu_s(\boldsymbol{x})$.

where $\mathbb{E}_s$ denotes the conditional expectation operator, $\mathbb{E}[\,\cdot\,|\,Y=\boldsymbol{y}_s]$.

The goal of minimizing the risk measures discussed above, leads us to the general formulation of the optimization problem. All the optimization formulations we are going to discuss in Section 6 will have the following general form:

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & g\left[\mathcal{L}(\boldsymbol{x})\right] \\
\text{s.t.} \quad & \boldsymbol{x} \in \Omega.
\end{aligned}
\tag{16}
$$

where $g\left[\mathcal{L}(\boldsymbol{x})\right]$ is the risk measure ($\mathrm{Var}[\mathcal{L}(\boldsymbol{x})]$, $\mathbb{E}[\mathcal{L}(\boldsymbol{x})^2]$, $\ell_\alpha(\boldsymbol{x})$ or $\mathrm{ES}_\alpha(\boldsymbol{x})$) and Ω denotes a (convex) feasible region, defined by a set of linear constraints.

6 Credit Risk Optimization Formulations

In this section we formulate the optimization problems for minimizing risk measures described in Section 5 within the framework introduced in Sections 3 and 4. These two dimensions, the risk measure and the approximation of the loss distribution, define the taxonomy of the optimization problems that we consider.

For ease of presentation, from here onwards we suppress the subscript G on N_G, the number of groups, and write simply N.

6.1 Moment-Based Formulations

Minimizing variance or the second moment of the loss distribution does not minimize the quantile-based risk measures (VaR and ES), unless the distribution of losses is Normal. So, it is expected that the moment-based formulations may not perform well for all quantiles.

For the moment-based formulations, we require the vector, $\boldsymbol{\mu}$, of unconditional expected credit losses of the groups, and the unconditional variance-covariance matrix, Q, of the groups' credit losses. The ith component of $\boldsymbol{\mu}$ (indexed by the groups) is the sum over the counterparties in the ith group, of the counterparties' mean losses:

$$
\mu_i = \sum_{j \in G_i} \sum_{c=0}^{C-1} \ell_c^j p_c^j.
$$

The (i_1, i_2) component of Q is the sum of the covariances of the counterparty losses, with the counterparties ranging over the i_1-th and i_2-th groups. MC approximation or semi-analytically, using the conditional independence of counterparty losses combined with a systemic MC simulation for unconditioning.

The variance $\mathrm{Var}[\mathcal{L}(\boldsymbol{x})]$ minimization problem is:

$$
\begin{aligned}
\min_{\boldsymbol{x} \in \mathbb{R}^N} \quad & \boldsymbol{x}^T Q \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{x} \in \Omega.
\end{aligned}
$$

The second moment $\mathbb{E}[\mathcal{L}(\boldsymbol{x})^2]$ minimization problem is:

$$
\begin{aligned}
\min_{\boldsymbol{x} \in \mathbb{R}^N} \quad & \boldsymbol{x}^T [Q + \boldsymbol{\mu}\boldsymbol{\mu}^T] \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{x} \in \Omega.
\end{aligned}
$$

6.2 ES and VaR Minimization with MC-Sampling Approximation

6.2.1 Optimization Formulation for ES Minimization

It was shown in Rockafellar and Uryasev (2000) and Rockafellar and Uryasev (2002) that minimization of expected shortfall $\text{ES}_\alpha(\boldsymbol{x})$ can be reduced to minimizing the function

$$\ell + \frac{1}{1-\alpha}\mathbb{E}([\mathcal{L}(\boldsymbol{x}) - \ell]^+), \tag{17}$$

where $a^+ = \max(0, a)$.

The function (17) is convex in ℓ; it is also convex in $\boldsymbol{x}$ if the function of losses, $\mathcal{L}(\boldsymbol{x})$, is convex in $\boldsymbol{x}$.

Having MK scenarios with corresponding probabilities $1/MK$ of occurring, we can approximate the expectation in (17):

$$\mathbb{E}([\mathcal{L}(\boldsymbol{x}) - \ell]^+) \approx \frac{1}{MK}\sum_{s,k}[L_{sk}(\boldsymbol{x}) - \ell]^+. \tag{18}$$

The ES optimization problem becomes:

$$\min_{\boldsymbol{x},\ell} \quad \ell + \frac{1}{(1-\alpha)}\frac{1}{MK}\sum_{s,k}^{MK}[L_{sk}(\boldsymbol{x}) - \ell]^+$$
$$\text{s.t.} \quad \boldsymbol{x} \in \Omega.$$

The problem is convex as the loss functions $L_{sk}(\boldsymbol{x})$ are linear and the set Ω is defined by linear equalities and inequalities. The ES minimization problem can be reduced to the linear problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^N, \boldsymbol{u}\in\mathbb{R}^{MK}, \ell\in\mathbb{R}} \quad \ell + \frac{1}{(1-\alpha)}\frac{1}{MK}\sum_{s,k}u_{sk},$$
$$\text{s.t.} \quad u_{sk} \geq L_{sk}(\boldsymbol{x}) - \ell, \ u_{sk} \geq 0, \ s = 1,\ldots,M, \ k = 1,\ldots,K \tag{19}$$
$$\boldsymbol{x} \in \Omega.$$

where $\{u_{sk}\}$ are auxiliary variables.

Note that to optimize the expectation in (17), we can use a number of approaches. Instead of formulating the problem as a large-scale linear optimization problem described in this section, it can be solved as a non-smooth, nonlinear optimization problem. In general, using nonsmooth optimization techniques is prohibitively slow. Instead, we can use the smoothing technique in Alexander, Coleman and Li (2006) to solve the problem with a smoothed ES function.

6.2.2 Sequential Algorithm for VaR Minimization

For the MC-sampling formulation, directly minimizing VaR requires integer programming. We use a heuristic technique, developed in Larsen, Mausser and Uryasev (2002), that is less computationally intensive. The algorithm minimizes VaR by solving a sequence of ES minimization problems while progressively fixing scenarios in the tail of the loss distribution.

In our setting, the algorithm is described below:

Step 0. Initialization
    1.   Set $\alpha_0 = \alpha$, $m = 0$, $H_0 = \{(s,k) : s = 1,\ldots,M, \ k = 1,\ldots,K\}$.

    2.   Assign a value to the parameter, ε, for discarding scenarios; $0 < \varepsilon < 1$.

Step 1. Optimization subproblem

    1.   Minimize α_m-ES

$$\min_{\boldsymbol{x},\boldsymbol{u},\ell,\gamma} \quad \ell + \nu_m \frac{1}{MK} \sum_{(s,k) \in H_m} u_{sk}$$

$$\begin{aligned}
\text{s.t.} \quad & L_{sk}(\boldsymbol{x}) \leq \ell + u_{sk}, \ u_{sk} \geq 0 && (s,k) \in H_m, \\
& L_{sk}(\boldsymbol{x}) \leq \gamma && (s,k) \in H_m, \\
& L_{sk}(\boldsymbol{x}) \geq \gamma && (s,k) \notin H_m, \\
& \boldsymbol{x} \in \Omega,
\end{aligned}$$

        where $\nu_m = 1/((1 - \alpha_m))$. Denote the optimal solution of this problem by $\boldsymbol{x}_m^*$.

    2.   Denote the order statistics of the losses $L_{sk}(\boldsymbol{x}_m^*)$, $s = 1, \ldots, M$, $k = 1, \ldots, K$ by L_n, $n = 1, \ldots, MK$. Also, denote the sorting order by writing $n(s,k) = n$ if (s,k) is the n-th index in the sorting order.

Step 2. Estimating VaR

    Calculate VaR estimate $\ell_m = L_{n_\alpha}$, where $n_\alpha = \min\{n : n/MK \geq \alpha\}$.

Step 3. Stopping and re-initialization

    1. $m = m + 1$.

    2. $b_m = \alpha + (1 - \alpha)(1 - \varepsilon)^m$ and $\alpha_m = \alpha/b_m$.

    3. $H_m = \{(s,k) \in H_{m-1} : n(s,k)/MK \leq b_m\}$.

    4.   If $H_m = H_{m-1}$ then stop the algorithm and return the estimate of the VaR-optimal portfolio $\boldsymbol{x}_m^*$ and VaR estimate ℓ_m, otherwise go to Step 1.

6.3 VaR and ES Minimization Based on CLT Approximation

Applying the CLT approximation to the conditional loss distribution, from (11) we obtain

$$\mathbb{P}_s(\mathcal{L}(\boldsymbol{x}) \leq \ell_\alpha(\boldsymbol{x})) \approx \Phi\left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})}\right).$$

From (14), we derive:

$$\frac{1}{M} \sum_{s=1}^{M} \Phi\left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})}\right) = \alpha. \tag{20}$$

Equation (20) determines the value of the objective function, $\ell_\alpha(\boldsymbol{x})$, implicitly as function of the decision vector, $\boldsymbol{x}$: for fixed $\boldsymbol{x}$, the VaR value $\ell_\alpha(\boldsymbol{x})$ is the solution of the equation (20) which must be solved numerically; for instance using bisection or Newton-type methods.

This leads to the following formulation of the VaR optimization problem with the conditional normal approximation:

$$\begin{aligned}
\min_{\boldsymbol{x} \in \mathbb{R}^N} \quad & \ell_\alpha(\boldsymbol{x}) \\
\text{s.t.} \quad & \frac{1}{M} \sum_{s=1}^{M} \Phi\left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})}\right) = \alpha, \\
& \boldsymbol{x} \in \Omega.
\end{aligned} \tag{21}$$

Turning to ES, under scenario s, $\mathcal{L}(\boldsymbol{x}) \sim \mathcal{N}(\mu_s(\boldsymbol{x}), \sigma_s^2(\boldsymbol{x}))$. Now, the ES for a general $\mathcal{N}(\mu, \sigma^2)$-distributed random variable, X, is well known to be

$$\frac{1}{1 - \alpha}\left[\mu\overline{\Phi}\left(\frac{x_\alpha - \mu}{\sigma}\right) + \sigma\phi\left(\frac{x_\alpha - \mu}{\sigma}\right)\right],$$

where x_α is the α-quantile of X, ϕ denotes the standard normal pdf, and $\overline{\Phi} = 1 - \Phi$.

Therefore, the expected shortfall $\mathrm{ES}_\alpha(\boldsymbol{x})$ at the quantile level α is

$$\mathrm{ES}_\alpha(\boldsymbol{x}) = \frac{1}{1-\alpha} \frac{1}{M} \sum_{s=1}^{M} \left[\mu_s(\boldsymbol{x}) \overline{\Phi} \left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})} \right) + \sigma_s(\boldsymbol{x}) \phi \left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})} \right) \right]. \qquad (22)$$

The resulting ES optimization problem with CLT approximation is:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^N} \quad & \mathrm{ES}_\alpha(\boldsymbol{x}) \\ \text{s.t.} \quad & \frac{1}{M} \sum_{s=1}^{M} \Phi \left(\frac{\ell_\alpha(\boldsymbol{x}) - \mu_s(\boldsymbol{x})}{\sigma_s(\boldsymbol{x})} \right) = \alpha, \\ & \boldsymbol{x} \in \Omega, \end{aligned} \qquad (23)$$

where $\mathrm{ES}_\alpha(\boldsymbol{x})$ is defined by equation (22).

To find numerical solutions for the nonlinear optimization problems (21) and (23) we compute by implicit differentiation the gradient and the Hessian for the VaR, from (20), and for the ES function, from (22).

6.4 ES and VaR Minimization with LLN Approximation

The linear optimization problem for ES minimization was defined by (19). The only difference between the MC and LLN approximation formulations is the loss function $\mathcal{L}(\boldsymbol{x})$. Using the loss function $\mathcal{L}_s(\boldsymbol{x})$ for the LLN approximation, we get the following ES optimization problem:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^N, \ \boldsymbol{u} \in \mathbb{R}^M, \ \ell \in \mathbb{R}} \quad & \ell + \frac{1}{(1-\alpha)} \frac{1}{M} \sum_{s=1}^{M} u_s \\ \text{s.t.} \quad & u_s \geq \mu_s(\boldsymbol{x}) - \ell, \ u_s \geq 0, \ s = 1, \ldots, M \\ & \boldsymbol{x} \in \Omega, \end{aligned} \qquad (24)$$

where $\mu_s(\boldsymbol{x})$ is the mean loss vector from LLN scenarios $(s = 1, \ldots, M)$, computed from (9).

The algorithm described in Section 6.2.2 is suitable for VaR minimization with the conditional LLN approximation. The modification of it for the LLN approximation is straightforward.

6.5 Taxonomy of Optimization Problems and Data Requirements

The taxonomy of optimization problems, that we have described in the previous subsections, is presented in Table 1.

Data requirements for optimization by each formulation can be summarized as follows:

- MC Sampling - M systemic scenarios, K specific scenarios for each systemic, N groups: KMN data points;
- CLT Approximation - M systemic scenarios, 1 mean and 1 variance for each systemic, N groups: $2MN$ data points;
- LLN Approximation - M systemic scenarios, 1 mean for each systemic, N groups: MN data points.

It is not surprising that the LLN formulation uses the least amount of data points for the optimization, as it is the most restrictive formulation, relying on the portfolio being very large and highly granular. If $K > 2$, the MC-sampling formulation requires the largest number of data points for optimization.

Table 1: The Taxonomy of Optimization Problems

		Risk Measure		
		VaR	**Expected Shortfall**	**Variance Second Moment**
Optimization Formulation	**CLT Approximation**	*Non-convex nonlinear problem* Direct VaR minimization with CLT sampling (Section 6.3)	*Convex nonlinear problem* Direct expected shortfall minimization with CLT sampling (Section 6.3)	
	LLN Approximation	*Linear problem (heuristic)* Successive expected shortfall minimization heuristics with LLN sampling (Section 6.4)	*Linear problem* Direct expected shortfall minimization with LLN sampling (Section 6.4)	
	Monte Carlo Sampling	*Linear problem (heuristic)* Successive expected shortfall minimization heuristics with Monte-Carlo sampling (Section 6.2.2)	*Linear problem* Direct expected shortfall minimization with Monte-Carlo sampling (Section 6.2.1)	
	Unconditional			*Convex quadratic problem* Vector of mean losses and variance-covariance matrix of losses are computed from Monte-Carlo sampling (Section 6.1)

7 Computational Tests

7.1 Constraints

In our computational tests, the feasible region, Ω, is defined by the following constraints:

- the value of the portfolio remains at its initial value (V);
- the expected return on the portfolio is at least that of the initial portfolio (R);
- the group positions are bounded by constants ($\underline{x}_i \leq \overline{x}_i$, for the ith group).

This leads to the following class of optimization problems:

$$
\begin{aligned}
\min_{x} \quad & g\left[\mathcal{L}(\boldsymbol{x})\right] \\
\text{s.t.} \quad & \sum_{i=1}^{N_{\mathrm{G}}} v_i x_i = V, \\
& \sum_{i=1}^{N_{\mathrm{G}}} r_i v_i x_i \geq RV, \\
& \underline{x}_i \leq x_i \leq \overline{x}_i, \ i = 1, \ldots, N_{\mathrm{G}},
\end{aligned}
\tag{25}
$$

Other types of constraints are possible, such as a total budget constraint (e.g., trading budget of 5% of portfolio value, including transaction costs) or limits on the total positions in certain categories, e.g., a particular credit rating or geographic region.

7.2 Performance Metrics

For the LLN, CLT and MC formulations, there are two sources of error affecting the quality of an optimal solution: (i) systemic sampling error due to the choice of M; and (ii) modelling (or formulation) error due to approximating the conditional loss distributions. (For the MC formulation, the choice of K is the influential parameter.)

To isolate and quantify these errors, we measure:

- *Out-of-sample actual risk*: the risk measure for the optimal portfolio (under a given formulation), evaluated by MC with a large number of out-of-sample scenarios. The out-of-sample actual risk is assumed therefore to be based on the true loss distribution.

- *In-sample actual risk*: the risk measure for the optimal portfolio (under a given formulation), evaluated by MC with the same systemic scenarios that are used for optimization and a large number of idiosyncratic samples for each systemic sample. *In-sample actual risk* is assumed therefore to be based on the true conditional loss distributions. The approximation error due to systemic sampling, is thus isolated and gauged by the difference between *out-of-sample actual risk* and *in-sample actual risk*.

- *In-sample reported risk*: the objective function value reported for the risk measure for the optimal portfolio (under a given formulation). The approximation error due to the choice of optimization formulation (equivalently, the choice of approximation in the conditional loss distributions) or model error, is thus isolated and gauged by the difference between *in-sample actual risk* and *in-sample reported risk*. The difference between *out-of-sample actual risk* and *in-sample reported risk*, contains the combined effect of systemic sampling error and formulation approximation error.

Note that for the variance and second-moment formulations, only out-of-sample actual risk is applicable.

7.3 Specific Problem Settings

Unless specific mention is made otherwise, the optimization problems were run with the following default settings:

- positions are between zero and twice the initial position: $0 \leq x_i \leq 2x_i^0$;
- $M = 10,000$ optimization scenarios (systemic);
- VaR and ES are calculated for the 99.9th percentile; i.e., $\alpha = 0.999$;
- the results are the averages over 5 trial runs (different sets of systemic scenarios are used for each trial).

In our computational experiments, we test different group sizes and group compositions in order to draw conclusions about their effects.

We have used the following optimization packages for implementing our formulations:

- IPOPT - nonlinear solver (version 3.5.5) (Wächter and Biegler, 2006);
- CPLEX - linear, quadratic, second-order conic solver (version 11.2) (ILOG Inc., 2008);
- MOSEK - linear, quadratic, second-order conic, convex nonlinear solver (version 5) (MOSEK ApS, 2008).

7.4 Portfolio Data Sources

The formulations described in the previous sections were tested and compared using two portfolios. The first portfolio is adapted from Rutter Associates LLC (2006). It consists of 3000 counterparties depending on 50 credit drivers and graded with 8 credit states (credit-state migrations). The second (proprietary) portfolio consists of 7470 counterparties with 34 credit drivers and only 2 (default or nondefault) credit states.
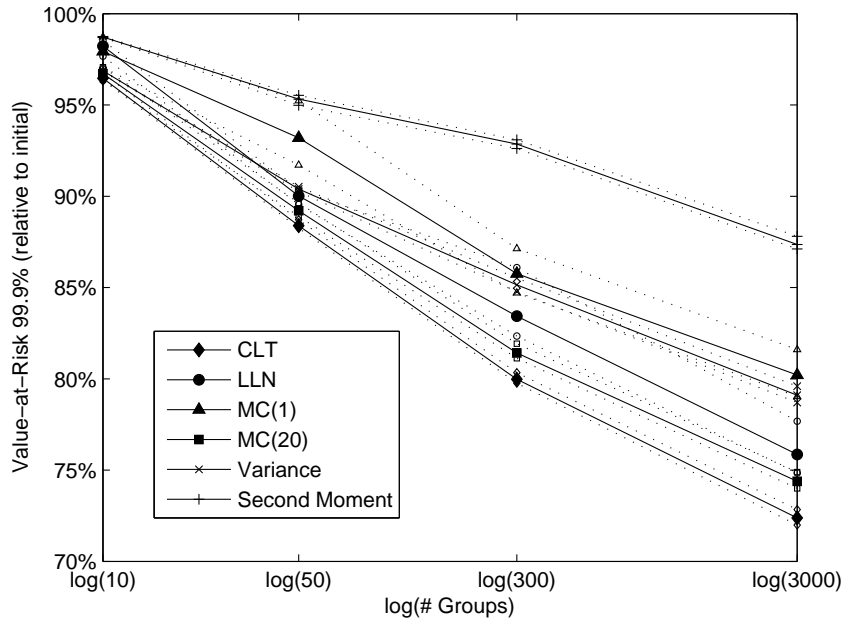
Figure 2: Out-of-Sample VaR

7.5 Results: 3000 Counterparties and Credit-State Migrations

Credit drivers in the portfolio are region-sector indices and each counterparty depends on one credit driver with $0.42 \leq \beta \leq 0.65$.

Grouping is done at random with four grouping schemas: 10 groups of 300 counterparties, 50 groups of 60 counterparties, 300 groups of 10 counterparties, and 3000 groups of 1 counterparty.

For *out-of-sample actual risk*, $M = 6,000,000$ and $K = 1$. For *in-sample actual risk*, $M = 10,000$ and $K = 150$. For *in-sample reported risk*, $M = 10,000$ and, for MC formulation MC(1), $K = 1$, while for MC formulation MC(20), $K = 20$.

7.5.1 *Out-of-sample, In-sample Comparisons*

The out-of-sample evaluation of the optimization results are presented in Figure 2 for VaR and in Figure 3 for ES. These figures show reductions in VaR and ES relative to the initial portfolio (100%). Horizontal axes display the group sizes (logarithmic scale). The solid lines are the averages over 5 trials and the dotted lines are the best and the worst results for the same 5 trials. The CLT formulation always has the best performance followed by MC sampling with $K = 20$ idiosyncratic samples per systemic scenario, which in turn is much better than MC sampling with $K = 1$. Extrapolating, we expect that MC sampling with very large K would match and possibly exceed the performance of CLT. However, increasing K leads to a larger, more computationally demanding problem.

There is little difference among the formulations' results when the number of groups is small because there is less opportunity to restructure the portfolio. The differences become more pronounced as more trading flexibility (i.e., smaller groups) is introduced. Formulations that show relatively
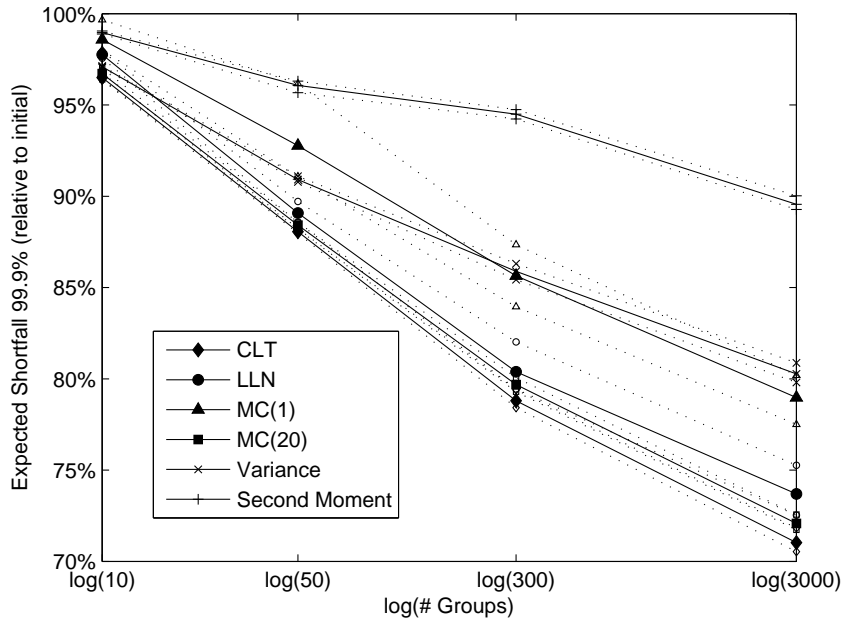
Figure 3: Out-of-Sample Expected Shortfall

good performance for all grouping schemas are the CLT, MC(20) and LLN. The LLN formulation performance is acceptable due to the fine granularity of the optimal portfolios (the narrow trading limits preserve the fine granularity of the initial portfolio). We investigate the issue of granularity in more detail later in this section. The CLT, MC(20) and LLN formulations are quite separated from each other with minimal overlaps when we look at the best and worst performance over 5 trials. It is more pronounced for the VaR models, partially due to the heuristic algorithm of Section 6.2.2 that we utilize for the MC and LLN approximations. That also explains why the performance of the MC and LLN formulations is not as good when minimizing VaR, as compared to the ES minimization.

The next step in evaluating the computational performance, is to look at the approximation quality of optimization formulations, which is summarized in Figure 4 for VaR and Figure 5 for ES. These figures display out-of-sample actual risk and the in-sample actual and reported risks for the various formulations.

The systemic sampling error is about the same for all formulations, as they all use the same number of systemic scenarios, M. The CLT approximation has the smallest formulation error, while the formulation error of MC(1) is the largest. For MC, the formulation error decreases when group-size increases. This observation can be explained by the way in which the tail of the conditional group-loss distribution is represented in each of the formulations. For MC sampling with only a few specific scenarios, the tail may be severely underestimated if the conditional counterparty default probabilities are very small. The optimizer will take advantage of this possibility by overinvesting in groups which, by chance, experience a small number of defaults, even though they may be more riskier theoretically. Thus, for a large number of groups (implying a small number of counterparties per group, in our experiment), the number of these chance opportunities will tend to be large, leading to a high formulation error. In contrast, the LLN and CLT formulations do not provide the optimizer
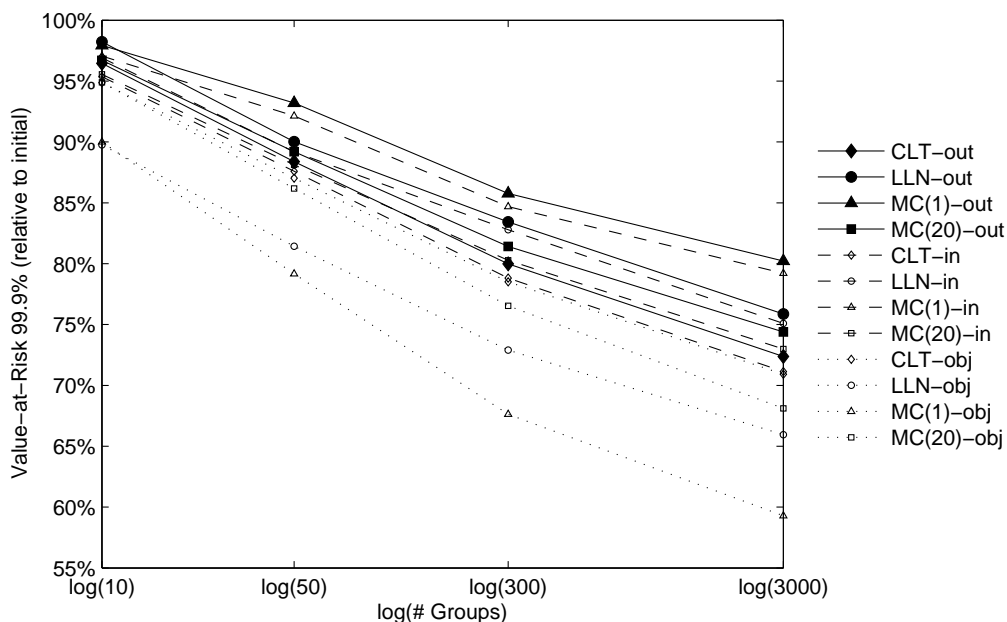
Figure 4: Approximation Quality for the VaR Optimization Problems

such opportunities, as the conditional group-loss distributions are derived analytically rather than empirically.

One observation from our computational experiments is that variance minimization in the mean-variance context tends to improve the quantile-based measures (Figures 2 and 3). One possible reason is the fact that counterparties with high probability of default have high variance of monetary losses (as we use a Bernoulli mixture model from Saunders et al. (2007) for defaults and migrations) and, as a result, variance minimization reduces VaR and ES measures, as the counterparties with high default probabilities are removed from the portfolio.

7.5.2 *Granularity*

It is known that the quality of the LLN and CLT approximations deteriorates as the portfolio becomes more coarsely granular (i.e., exposures are increasingly concentrated in a small number of dominant groups). To test granularity effects, we modified the initial portfolio (for this test only) to be the most finely grained possible, with the original portfolio value, V, distributed equally among all counterparties in the portfolio. The counterparties were left in the original grouping scheme for 50 equally sized groups and optimization problems were solved with five sets of progressively wider trading limits, starting with a no-trading policy (resulting in the original portfolio): [1, 1], [0, 2], [0, 15], [0, 30] and [0, 50]. As the trading limits become larger, there is a greater chance for obtaining coarsely grained portfolios and thus, more potential for poor results from the LLN and CLT approximations. As a granularity measure, we use the *Herfindahl-Hirschman Index* (HHI) which is the sum of squared group-weights[3] (by definition, the HHI converges to zero for infinitely

---

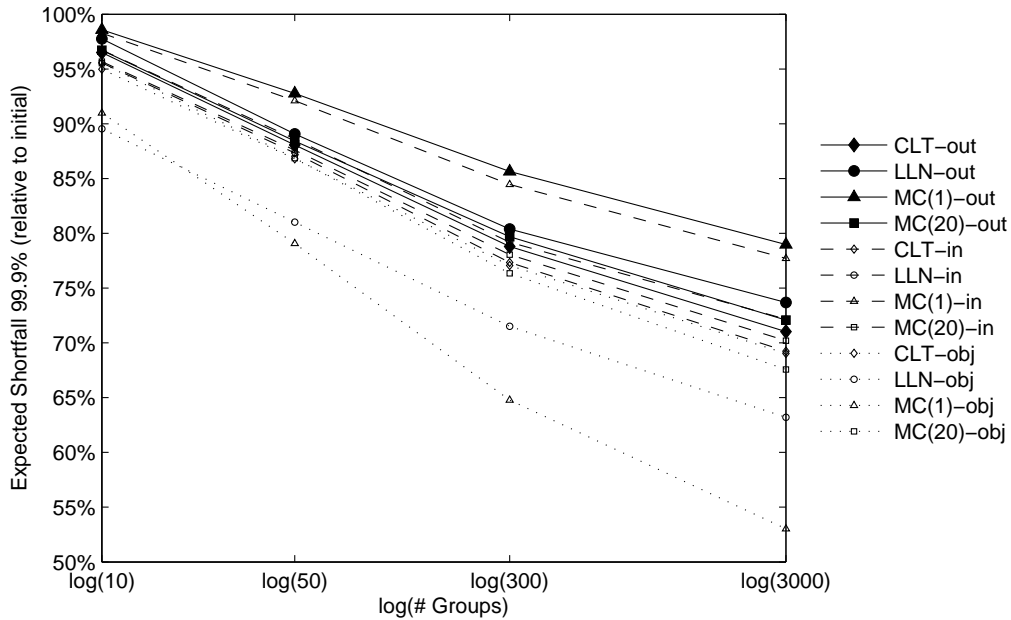[3]The group-weight for G_i, is $x_i v_i / V$.

Figure 5: Approximation Quality for the Expected Shortfall Optimization Problems

finely grained portfolios, i.e., when the number of groups grows indefinitely). In our case, the initial HHI equals 0.02, which is the lowest possible value for a portfolio comprising 50 groups.

Figure 6 plots formulation error (i.e., the relative difference between in-sample actual risk and in-sample reported risk) against granularity for the optimal portfolios obtained by the LLN, CLT, MC(1) and MC(20) formulations. The results suggest that the LLN approximation is far more susceptible to the adverse effects of granularity than the CLT formulation. While wider trading limits tend to produce more granular portfolios in general, the optimal MC and CLT portfolios maintain some degree of diversification (the HHI does not exceed 0.25). In contrast, LLN portfolios become progressively more concentrated and exhibit increasingly large formulation errors.[4] This is explained by the fact that, to minimize VaR or ES, the LLN formulation tries to reduce only the portfolio's conditional expected losses while the CLT formulation considers both the conditional expectation and the conditional variance. Reducing the variance of the portfolio losses entails smaller positions (i.e., greater diversification). It was also observed (not shown here) that the relative rankings of the sampled group losses across the systemic scenarios were much more variable than those of the expected group losses.[5] The greater degree of diversification present in the optimal MC portfolios, relative to those of LLN, is a reflection of this fact.

---

[4]The large formulation errors associated with MC(1) are due more to the relatively small sample size than to granularity effects.

[5]That is, while the loss of a particular group might be among the smallest losses sampled in one scenario and among the largest losses sampled in another, its expected loss tends to rank at approximately the same level in all scenarios.
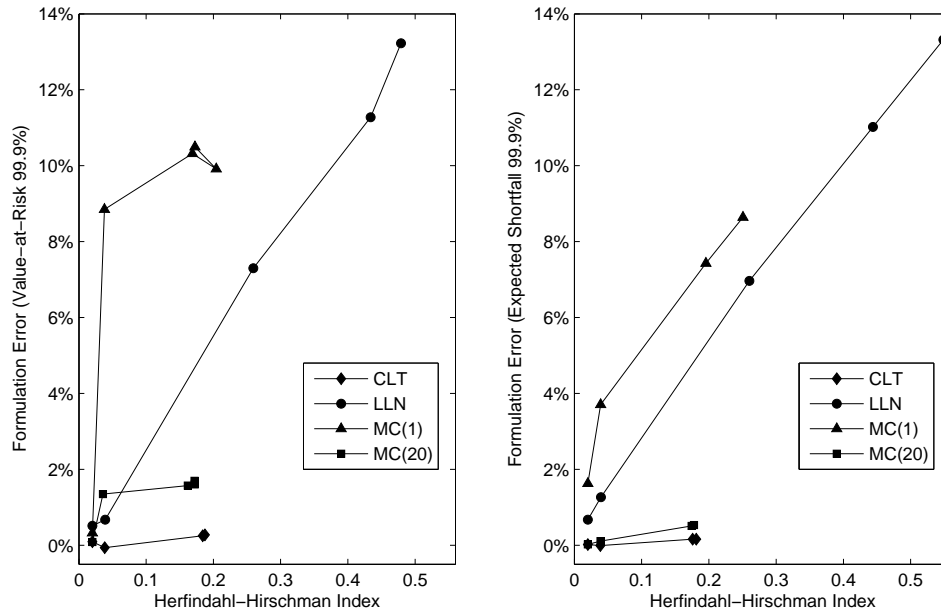
Figure 6: Granularity Effect

7.5.3 *Sample Size*

To see the effect of increasing the number of systemic samples for use in optimization, we compared our formulations with $M = 10,000$ and $M = 50,000$ systemic scenarios. For the MC approximation, in addition to the MC(1) formulation, in the first case we use $K = 20$ (MC(20)), and in the second case we use $K = 4$ (MC(4)), resulting in a total of $200,000$ MC scenarios in each case. Table 2 describes the systemic sampling effect and shows how out-of-sample performance is affected by the number of systemic samples. Increasing the number of systemic samples by a factor of 5, produces slight improvement for the MC(1) formulation and has a negligible effect for the others. It also turns out that CLT with $M = 10,000$ systemic samples, performs better than the other formulations with $M = 50,000$ systemic samples.

7.5.4 *Stability*

Analyzing the optimal group-weights for different random samples, makes it possible to compare the stability of formulations. Figure 7 shows the ranges of optimal group-weights for 5 trials when minimizing ES where, for each trial, different sets of (systemic) scenarios were used for optimization. Similarly to the sensitivity of the out-of-sample objective function value in Figure 3, stable group-weights are produced by the CLT formulation, while MC formulations require a larger number of scenarios to reach stability of the weights. For the moment-based formulations, the variance-covariance matrix of losses and the vector of mean losses are estimated from the $M = 10,000$ with $K = 150$ samples (1.5 million in-sample scenarios) and, consequently, are quite stable. This explains why moment-based formulations produce the most stable weights among all the formulations that were considered.

Table 2: Systemic Sampling Effect

VaR$_{99.9\%}$	10,000 Systemic Samples			VaR$_{99.9\%}$	50,000 Systemic Samples		
	10 Groups	50 Groups	300 Groups		10 Groups	50 Groups	300 Groups
CLT	96.5%	88.4%	80.0%	CLT	96.3%	88.3%	79.6%
MC(20)	96.7%	89.2%	81.4%	MC(4)	96.6%	89.2%	81.3%
LLN	98.2%	90.0%	83.4%	LLN	97.4%	89.5%	82.1%
MC(1)	97.9%	93.2%	85.8%	MC(1)	97.1%	90.4%	82.9%

ES$_{99.9\%}$	10,000 Systemic Samples			ES$_{99.9\%}$	50,000 Systemic Samples		
	10 Groups	50 Groups	300 Groups		10 Groups	50 Groups	300 Groups
CLT	96.5%	88.1%	78.8%	CLT	96.4%	87.9%	78.4%
MC(20)	96.7%	88.4%	79.7%	MC(4)	96.6%	88.5%	79.3%
LLN	97.8%	89.1%	80.4%	LLN	97.6%	89.4%	79.7%
MC(1)	98.6%	92.8%	85.6%	MC(1)	96.9%	89.5%	80.8%

7.5.5 *Quantile Level*

We also evaluated the sensitivity of ES to the quantile level (Figure 8). We used 300 groups and $M = 50,000$ for optimization. For lower quantiles (95% or less), the performance of all formulations is very similar. For high quantiles (99.97%), the performance of CLT becomes more distinguishable from the other formulations. As Figure 8 demonstrates, the CLT formulation has a 41% improvement over variance minimization (optimization results are evaluated for ES, out-of-sample) for the 99.97% quantile, but the variance minimization result is within 2% of the CLT formulation for the 95% quantile. It is known that moment-based formulations perform well when the quantile moves towards the center of a distribution, because the Normal approximation becomes more appropriate. MC sampling formulations require more scenarios to capture losses in the extreme tail of the distribution and, as a result, they perform less effectively than for lower quantiles. The same reasoning applies for the LLN formulation which better approximates the tail for high quantiles than MC formulations for the sample sizes used in our tests; but as we already know, the LLN formulation is very sensitive to the granularity effect (as tests in Figure 8 are performed for narrow trading bounds, the granularity effect is limited).

7.5.6 *Small Portfolio*

We tested the performance of the formulations when the portfolio comprises a very small number of counterparties (10 groups with 1 counterparty per group). This case is expected to be the worst for the CLT and LLN approximations, as the portfolio is no longer large and some counterparties can dominate within it. We conclude that LLN fails for this setup. CLT becomes unreliable for portfolios with a small number of counterparties. MC sampling, MC(20), always works, but requires a large number of idiosyncratic scenarios per systemic scenario. It is necessary to point out that this setting is unrealistic from the practical point of view, as real credit portfolios cannot be that small.

7.5.7 *Speed*

Table 3 summarizes running times for different optimization formulations. Evidently, solving the nonlinear problems for CLT is competitive with solving the linear problems in the MC-sampling
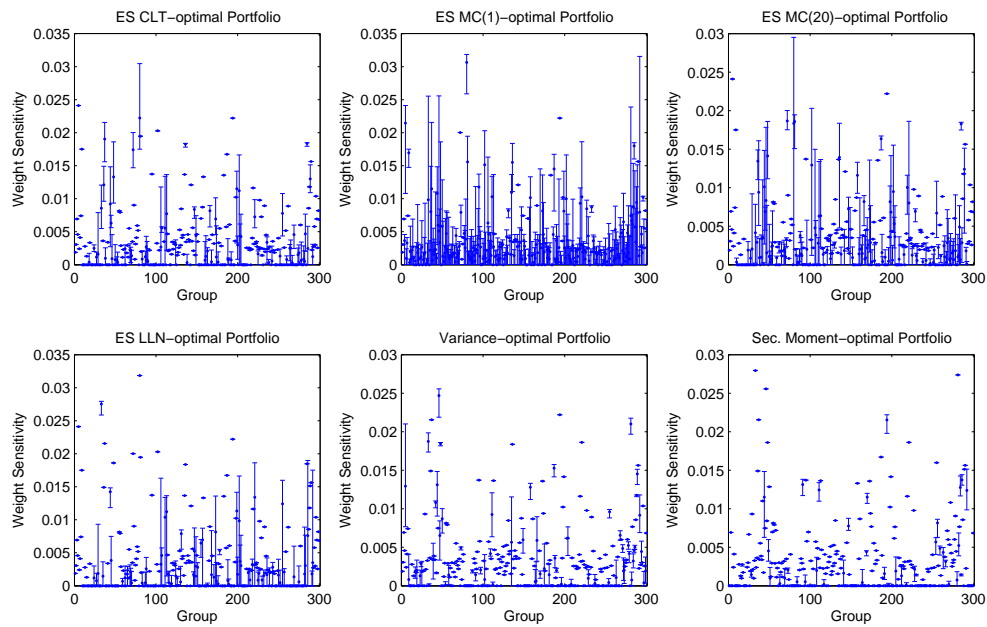
Figure 7: Sensitivity of Group-Weights to Different Sets of Optimization Scenarios

formulations, at a comparable level of (out-of-sample) accuracy. The computational tests were run on a server with 8 x Opteron 885 CPU, 16 cores (jobs run on 1 core), 64 Gb RAM. VaR optimization for MC(20) was run in parallel mode on 4 threads.

Table 3: Optimization Problem Performance: Elapsed Time (seconds)

Model	Solver	$\mathbf{ES}_{99.9\%}$			
		10 Groups	50 Groups	300 Groups	3000 Groups
CLT	IPOPT	4–8	6–8	14–83	81–1090
LLN	CPLEX	1	1–2	6–8	73–86
MC(1)	CPLEX	1	1–2	6–10	14–115
MC(20)	CPLEX	137–155	233–279	461–578	1050–1280

Model	Solver	$\mathbf{VaR}_{99.9\%}$			
		10 Groups	50 Groups	300 Groups	3000 Groups
CLT	IPOPT	4–25	5–7	14–55	400–1643
LLN	CPLEX	2–3	6–8	46–50	791–1025
MC(1)	CPLEX	2–3	8–10	46–69	2436–3312
MC(20)	CPLEX	3620–4080	2382–2777	6522–8563	39273–86383

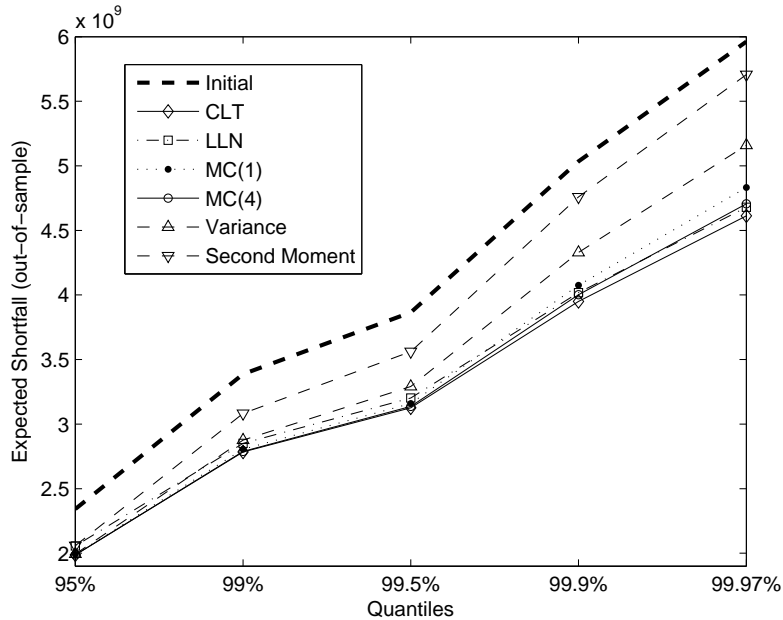| Model	Solver	**Variance**			
		10 Groups	50 Groups	300 Groups	3000 Groups
Uncond.	MOSEK	< 1	< 1	1	682–719

Figure 8: Formulation Performances for Different Quantiles

7.5.8 *Groupings*

We also ran computational tests with more realistic counterparty groupings. Counterparties depending on the same credit driver and which are in the same credit state at the beginning of the time period belong to the same group, resulting in 301 groups with a variable number of counterparties in each of them. Table 4 summarizes out-of-sample performance (relative to the performance of the initial portfolio) of the optimization formulations for $0 \leq x_i \leq 2x_i^0$ and $0.8x_i^0 \leq x_i \leq 1.2x_i^0$ trading limits. The narrow trading limits are more realistic when practical portfolio rebalancing is performed. The CLT formulation exhibits the best performance for both trading limits.

7.6   Results: 7470 Counterparties and Default/Nondefault Credit States

To confirm our findings, in more generality, we ran some of the computational tests for the portfolio with 7470 counterparties and only 2 credit states (default or no default). As with the previous portfolio, credit drivers are region-sector indices and each counterparty depends on one credit driver, but now $0.30 \leq \beta \leq 0.95$. Similar to the results in Table 4, we used realistic counterparty groupings, based on common credit driver and initial credit state, resulting in 277 groups with a variable number of counterparties in each of them. Table 5 summarizes out-of-sample performance of the optimization formulations for this portfolio.

The out-of-sample performance of optimization formulations for the [0.8, 1.2] and [0, 2] trading limits confirms the previous findings: the CLT formulation outperforms all the other optimization formulations when the number of counterparties in a portfolio is relatively large. Our numerical experiments show that allowing credit-state migrations or just default/nondefault credit states does not have any observable effect on performance of the optimization formulations, but the discrepancy between the

Table 4: Out-of-Sample Performance for the Portfolio with 3000 Counterparties

Model	Risk Measure	10000 Systemic Scenarios, 301 Groups	
		Trading Limits $[0.8, 1.2]$	Trading Limits $[0, 2]$
CLT	$ES_{99.9\%}$	93.62%	73.19%
	$VaR_{99.9\%}$	93.92%	74.53%
MC(20)	$ES_{99.9\%}$	93.63%	73.47%
	$VaR_{99.9\%}$	94.78%	75.31%
LLN	$ES_{99.9\%}$	93.69%	74.98%
	$VaR_{99.9\%}$	95.00%	77.29%
MC(1)	$ES_{99.9\%}$	94.94%	78.77%
	$VaR_{99.9\%}$	96.50%	79.86%
Variance	$ES_{99.9\%}$	95.92%	83.38%
	$VaR_{99.9\%}$	95.64%	82.05%
Second Moment	$ES_{99.9\%}$	97.89%	91.61%
	$VaR_{99.9\%}$	97.53%	89.55%

Table 5: Out-of-Sample Performance for the Portfolio with 7470 Counterparties

Model	Risk Measure	10000 Systemic Scenarios, 277 Groups	
		Trading Limits $[0.8, 1.2]$	Trading Limits $[0, 2]$
CLT	$ES_{99.9\%}$	89.84%	52.67%
	$VaR_{99.9\%}$	90.27%	54.40%
MC(20)	$ES_{99.9\%}$	89.84%	52.89%
	$VaR_{99.9\%}$	90.86%	55.15%
LLN	$ES_{99.9\%}$	89.87%	52.87%
	$VaR_{99.9\%}$	91.37%	55.64%
MC(1)	$ES_{99.9\%}$	90.39%	55.31%
	$VaR_{99.9\%}$	93.12%	58.61%
Variance	$ES_{99.9\%}$	90.01%	54.16%
	$VaR_{99.9\%}$	90.30%	55.20%
Second Moment	$ES_{99.9\%}$	91.26%	59.61%
	$VaR_{99.9\%}$	91.50%	60.19%

formulations gets smaller. Note that optimization results from the moment-based formulations are much closer to those of other formulations. This raises the question of whether the "convergence" of the different formulations is due to a feature of the model – default/nondefault vs. migration – or due to some feature of the initial portfolio. We ran the 3000-counterparty portfolio in default/nondefault mode (i.e., all nondefault credit states are amalgamated) and found that the differences between formulations were sustained. Therefore, we conclude that the convergence is due to features of the portfolios themselves, rather than using a migration model vs. a default-only model.

8   Conclusions and Future Directions

In this paper we analyzed different approaches to the portfolio credit-risk optimization problem within a conditional independence framework. We compared optimization of quantile-based risk measures with the traditional Markowitz mean-variance (moment-based) optimization formulation. Optimization of quantile-based risk measures using the MC approximation, requires a large number

of scenarios. If the number of instruments in the portfolio is relatively large, an efficient LLN or especially CLT approximation of the conditional distribution of portfolio losses can be utilized in the conditional independence framework. These two approximations can be used to significantly reduce in the number of scenarios required for portfolio optimization.

The formulation that we recommend for practical implementation, is the CLT approximation. The CLT approximation is attractive for optimization due to:

- Producing consistently better results than MC sampling with only a fraction of the data (10% in our experiments);
- Acceptable performance and solution quality when solving the nonlinear formulation;
- Being relatively robust to violations of the portfolio granularity condition.

The other two formulations commonly used in practice are the MC-sampling linear formulation and the moment-based quadratic formulation. The MC linear formulation is competitive with the CLT formulation only if the number of groups is very small (less than 10) and if groups consist of a very small number of counterparties. Moment-based formulations are not competitive, for extreme quantiles.

The optimization problems that we investigated, can be extended to the multiperiod setting, if required. One of the binding factors in that case is the optimization problem size, as it can become too large to be practically solved in reasonable time.

Our optimization formulations can incorporate importance sampling. Conditional distribution approximations, considered in this paper, significantly reduce the number of scenarios to be generated for idiosyncratic credit-risk factors, while importance sampling can reduce the data requirements for systemic risk factors. Both techniques can serve as a tool for reducing optimization problem size and improving out-of-sample optimal solution quality and stability.

Future work also includes incorporating stochastic exposures and stochastic recovery rates, and credit derivatives into the formulations. It is also worthwhile to investigate worst-case (robust) formulations for the optimization problems that we considered. From the implementation point of view, we would like to improve the VaR optimization heuristic for MC sampling. Finally, investigating which portfolio features (sensitivities of counterparties to credit drivers, portfolio size, etc.) may deteriorate or improve the moment-based formulation performance is a subject of future research.

References

Alexander, S., Coleman, T. F. and Li, Y. (2006). Minimizing CVaR and VaR for a portfolio of derivatives, *Journal of Banking and Finance* **30**(2): 583–605.

Andersson, F., Mausser, H., Rosen, D. and Uryasev, S. (2001). Credit risk optimization with Conditional Value-at-Risk criterion, *Mathematical Programming, Series B* **89**(2): 273–291.

Credit Suisse Financial Products (1997). CreditRisk+: A credit risk management framework, *Technical report*, Credit Suisse First Boston International, London, UK. `http://www.csfb.com/creditrisk/`.

de Servigny, A. and Renault, O. (2004). *Measuring and Managing Credit Risk*, McGraw-Hill, New York, NY.

Gupton, G. M., Finger, C. C. and Bhatia, M. (1997). CreditMetrics – technical document, *Technical report*, RiskMetrics Group, J.P.Morgan, New York.

ILOG Inc. (2008). *ILOG CPLEX User's Manual.* CPLEX 11.2.

Iscoe, I. and Kreinin, A. (2000). Default boundary problem, *Technical report ARPS 00-03*, Algorithmics Inc., Toronto, Canada.

Iscoe, I. and Kreinin, A. (2008). Large pool approximations for credit loss distributions, *Technical report ARPS 08-03*, Algorithmics Inc., Toronto, Canada.

Iscoe, I., Kreinin, A. and Rosen, D. (1999). An integrated market and credit risk portfolio model, *Algo Research Quarterly* **2**(3): 21–37.

Larsen, N., Mausser, H. and Uryasev, S. (2002). Algorithms for optimization of Value-at-Risk, *in* P. M. Pardalos and V. K. Tsitsiringos (eds), *Financial Engineering, E-Commerce and Supply Chain*, Vol. 70 of *Applied Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, chapter 8, pp. 129–157.

Markowitz, H. (1952). Portfolio selection, *The Journal of Finance* **7**(1): 77–91.

Mausser, H. and Rosen, D. (2000). Efficient risk/return frontiers for credit risk, *The Journal of Risk Finance* **2**(1): 66–78.

Mausser, H. and Rosen, D. (2001). Applying scenario optimization to portfolio credit risk, *The Journal of Risk Finance* **2**(2): 36–48.

MOSEK ApS (2008). *The MOSEK Optimization Tools Manual. Version 5.0.* Revision 105.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk, *The Journal of Risk* **2**(3): 21–41.

Rockafellar, R. T. and Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions, *Journal of Banking and Finance* **26**: 1443–1471.

Rutter Associates LLC (2006). Convergence of credit capital models, *Technical report*, Rutter Associates LLC. Sponsored by the International Association of Credit Portfolio Managers (IACPM) and International Swaps and Derivatives Association (ISDA). `http://www.isda.org/c_and_a/pdf/IACPM-ISDA-ConvergenceCreditCapModelsFeb2006.pdf`.

Saunders, A. and Allen, L. (2002). *Credit Risk Mesurement: New Approaches to Value at Risk and Other Paradigms*, 2nd edn, John Wiley & Sons, New York, USA.

Saunders, D., Xiouros, C. and Zenios, S. A. (2007). Credit risk optimization using factor models, *Annals of Operations Research* **152**(1): 49–77.

Szegö, G. (ed.) (2004). *Risk Measures for the 21st Century*, Wiley Finance Series, John Wiley & Sons, Chichester, UK.

Tilke, S. (2006). Reducing asset weights' volatility by importance sampling in stochastic credit portfolio optimization, *Technical report 417*, Department of Economics, University of Regensburg, Regensburg, Germany.

Wächter, A. and Biegler, L. T. (2006). On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming, *Mathematical Programming* **106**(1): 25–57.

Zagst, R., Kehrbaum, J. and Schmid, B. (2003). Portfolio optimization under credit risk, *Computational Statistics* **18**(3): 317–338.

Zenios, S. A. and Jobst, N. J. (2001). The tail that wags the dog: Integrating credit risk in asset portfolios, *The Journal of Risk Finance* **3**(1): 31–43.