Importance Sampling and Quantile Estimation for Concentration Credit Risk

Efficient algorithms for assessing concentration in credit portfolios

by

Quirijn van Hattem

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Monday July 7, 2025 at 11:00 AM.

Student number:	5113792		
Project duration:	November 2024 – July 2025		
	Dr. ir. L. E. Meester	TU Delft	Daily Supervisor
	Ir. R. Vedder	Triodos Bank	External Supervisor
	Prof. dr. A. Papapantoleon	TU Delft	Committee Member
	Dr. ir. J. Bierkens	TU Delft	Committee Member
Faculty:	Applied Mathematics	TU Delft	
Affiliation:	Triodos Bank		

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Triodos Bank

Importance Sampling and Quantile Estimation for Concentration Credit Risk

Quirijn van Hattem

July 10, 2025

Preface

Completing this thesis concludes my Master's in Applied Mathematics at TU Delft. Over the past years, I have learned a lot about the theory of stochastic processes. I have always enjoyed the puzzles of probability and statistics and loved reading through elegant proofs. This work was exciting to apply the theory in practice, and I am grateful for the opportunity to write this in collaboration with Triodos Bank.

Efficient simulation of credit risk is a well-known problem, where mathematical theory and practice blend together in a beautiful way. Since it is a very broadly addressed problem, many great minds have already worked on this. During the process of this thesis, I thought some answers were very intuitive, but as it turned out, proofs are very complex to construct. This made me truly appreciate the work of my predecessors.

I have got a habit of diving in head-first — trying things out, hitting walls, then iterating. This "learning by doing" can give great insights, but it also means I sometimes find myself scrambling against deadlines — just as my supervisor warned me about.

My special thanks go to Ludolf Meester of TU Delft, who supervised me throughout this assignment. Beyond helping me with mathematical theory, Ludolf taught me a highly structured way of working and encouraged me to be more critical of my own work. Although this one-year period is short to fully develop these skills, I will carry these lessons with me for the rest of my life.

I also wish to express my gratitude to Rick Vedder, who was my supervisor from Triodos. Rick was always ready to help me out and taught me a lot about the bank's operations. I thank Matthew Byrne, who helped me implement the new algorithm; Gerben de Klerk; and other colleagues from Triodos for their support and the enjoyable working environment. My gratitude goes to Antonis Papapantoleon and Joris Bierkens for reading my report and serving on the committee.

Quirijn van Hattem
Delft, July 2025

Summary

This thesis investigates efficient Monte Carlo methods for estimating the 99.9% Value-at-Risk of concentrated credit portfolios modelled through a normal copula framework. Crude Monte Carlo simulation is inefficient when estimating extreme loss levels. To address this inefficiency, variance reduction techniques are applied, including importance sampling and its adaptive variant. This thesis provides empirical evidence that these techniques make extreme quantile estimation computationally feasible in a realistic portfolio credit risk setting.

Multiple methods from the literature are reviewed to approach an appropriate importance sampling proposal. Bernoulli tilting conditional on the common factors is an effective method when obligors are weakly correlated. However, when large losses are mainly driven by common factors in the credit portfolio, shifting their means turns out to be more effective. This has to be done with care, since an improper shift may increase the variance. To this end, a deterministic method that approaches the mode of the zero-variance distribution is discussed. Furthermore, adaptive importance sampling is studied, which aims to minimize the variance directly by iteratively updating the shift based on past samples. This thesis introduces a method that uses the generalized Poisson–Binomial distribution, together with an inverse Fourier transform, to approximate the zero-variance distribution nearly exactly for simple portfolio cases.

A deterministic mean-shifted importance sampling (M-IS) and an adaptive importance sampling (AIS) approach are applied on stylized portfolios. These portfolios are constructed such that they mimic real-world conditions. The M-IS method achieves an estimated variance reduction factor of 175 on a large portfolio with 20 000 obligors. Performance on the actual Triodos Bank portfolio delivered similar performance, but detailed results are withheld for reasons of confidentiality. For a moderately sized portfolio consisting of 1 000 obligors, M-IS is no more effective than crude Monte Carlo because of the higher impact of idiosyncratic concentrations. In this case, AIS finds a better balance between the various kinds of concentrations. With conservative parameters, AIS achieves an estimated variance reduction factor of 5 on the moderately sized portfolio. Bernoulli tilting could further increase the variance reduction in these portfolio scenarios. Moreover, as the number of samples increases, AIS further reduces variance. Finally, this thesis investigates how to construct asymptotic confidence intervals for the extreme quantile estimates of these stylized portfolios and demonstrates via experiments that they exhibit asymptotic normal behaviour despite the discreteness of the loss distribution.

Nomenclature

Abbreviation	Definition
ASRF	Asymptotic Single Risk Factor model
AIS	Adaptive Importance Sampling
B-IS	Bernoulli-tilting Importance Sampling
CMC	Crude Monte Carlo simulation
CLT	Central Limit Theorem
GPB	Generalized Poisson–Binomial distribution
IRB	Internal Ratings-Based
IS	Importance Sampling
M-IS	Mean-shifted Importance Sampling
SLLN	Strong Law of Large Numbers
SLSQP	Sequential Least Squares Quadratic Programming
VaR	Value-at-Risk

Contents

List of Figures

List of Tables

1

Introduction

If an individual or organization borrows money from a bank, there is a risk they may fail to repay the loan. Credit risk represents this potential financial loss when obligors default on their loans. If a bank fails to manage credit risk properly, it can cause liquidity issues. To protect banks, their clients, and, more importantly, the stability of the global financial system, the Basel Accords were developed, outlined in the Basel framework (see [5]). These internationally recognized sets of agreements aim to strengthen the regulation, supervision, and risk management of banks and other financial organizations. Accordingly, banks should hold enough capital to absorb losses arising from extreme but potential credit events. Specifically, they focus on the total annual credit loss at the 99.9th percentile of its probability distribution. This quantile is commonly referred to as the 99.9% Value-at-Risk (VaR).

Estimating this level of risk is complex, as it depends on many correlated elements, such as macroeconomic, sectoral, regional factors, and individual loan exposures. Dependency structures between obligors are modelled through the normal copula model, originally associated with J.P. Morgan's widely used CreditMetrics system (see Gupton, Finger, et al. [18]). This thesis investigates efficient Monte Carlo methods for estimating the 99.9% Value-at-Risk of such concentrated credit portfolios.

The *Asymptotic Single Risk Factor* (ASRF) model, based on the foundational work of Merton [26], Vasicek [31], and Gordy [17], is a simplified credit risk model used by Basel II/III to set minimum capital charges for credit risk under the IRB (Internal Ratings-Based) approach. It assumes a single underlying common risk factor and an infinitely granular portfolio, meaning that individual exposures are infinitesimally small relative to the total exposure of the portfolio. Since it reduces the high-dimensional problem to a one-dimensional problem, it allows for closed-form capital rules rather than costly high-dimensional simulation. The 99.9% VaR calculated with this model is referred to as the *macroeconomic VaR*. However, this is an approximation of real-world credit portfolio interactions. In practice, a bank's credit portfolio is not infinitely granular, and obligors are correlated through more than a single common factor.

A *multifactor model* extends the ASRF model by incorporating multiple common factors that each influence different subsets of obligors, thereby capturing more realistic dependency structures within portfolio subgroups. These concentrated groups of exposures induce a so-called concentration risk: the presence of correlation between obligors increases the potential for simultaneous defaults, leading to higher losses. Furthermore, the multifactor model also considers the idiosyncratic concentration risk, which arises when large exposures increase the impact of specific obligors. In such cases, the default of a single obligor can lead to significant losses, which the ASRF model will underestimate. The extra capital needed to account for concentration risk is called the *concentration risk add-on*. The macroeconomic VaR together with this add-on forms the overall 99.9% VaR estimate of the portfolio credit loss distribution.

It is generally difficult to compute the VaR of credit portfolios, because of the high dimensional nature of the problem. As a result, these quantiles are estimated using Monte Carlo methods; computational algorithms that use random sampling to produce statistical estimates. Estimating extreme quantiles via Monte Carlo simulation is computationally intensive, especially for large credit portfolios. Consequently, developing algorithms for efficient estimation of credit concentration risk remains a significant and

ongoing challenge.

A common approach to improving the efficiency of Monte Carlo estimates is the use of variance reduction techniques. One well-known method is *Importance Sampling* (IS), which improves simulation efficiency by drawing samples from a modified distribution, which generates more samples in a region of interest. Glynn et al. [16] apply IS to extreme quantile estimation and establishes that significant variance reduction can be achieved under suitable conditions.

In the context of the normal copula model for credit risk modelling, IS offers a promising solution, see e.g., Arvanitis and Gregory [4], Kalkbrener, Lotter, et al. [22], Merino and Nyfeler [25], and Morokoff [27]. The main challenge for IS is to find an appropriate IS proposal distribution such that the variance of the quantiles is minimized. Glasserman and Li [15] introduce a two-step IS approach. Conditional on the common factors, they use Bernoulli tilting through *exponential tilting* and establish asymptotic optimality for portfolios with low-correlated obligors. For applying IS on the common factors, they propose several approximations to shift the common factors towards the mode of the zero-variance distribution. Glasserman, Kang, et al. [14] extend this work by introducing a *mixed IS* procedure that generalizes the two-step method from a single common factor to multiple common factors and establish conditions for asymptotic optimality. Egloff, Leippold, et al. [12] propose an *Adaptive Importance Sampling* (AIS) method applied on portfolio credit risk estimation, using a Robbins-Monro approximation scheme to iteratively update the shift vector. Difficulties in the AIS method lie in guaranteeing the convergence of the quantile estimate. Egloff and Leippold [11] provide further theoretical support for quantile estimation under AIS for general distributions and establish almost sure convergence for adaptive quantile estimators for non-unique quantiles.

Constructing confidence intervals for extreme quantile estimates is challenging, especially when sampling from an IS proposal. Chu and Nakayama [9] establish a Central Limit Theorem (CLT) for IS quantile estimators and provide consistent variance estimators. He, Jiang, et al. [21] Show that quantile estimates under AIS are asymptotically normal.

Triodos Bank, headquartered in the Netherlands, focuses on sustainable banking and aims to support social, environmental, and cultural development. Currently, the bank uses a crude Monte Carlo approach to simulate losses from a multifactor model through the normal copula framework to estimate concentration risk add-ons. Since the model does not make use of variance reduction techniques, it is computationally intensive. Moreover, due to the inherent randomness of Monte Carlo sampling, repeated simulations can yield varying results. Given these challenges, it is desired to decrease computational cost and at the same time increase the precision of estimates. Furthermore, developing methods to quantify confidence intervals is important to better quantify the uncertainty of estimates.

In collaboration with Triodos Bank, this thesis aims to address this challenge. The goal is to enhance the computational efficiency of concentration risk add-on estimation while quantifying the uncertainty through confidence intervals. To achieve this goal, the impact of approximation methods, optimization strategies, and variance reduction techniques — specifically, IS and AIS — are studied. Various techniques from the literature for approximating the optimal shift are evaluated on their effectiveness. Furthermore, this thesis introduces a method to approximate the zero-variance distribution nearly exactly for simple portfolios using the generalized Poisson–Binomial distribution and an inverse Fourier transform. Additionally, the extent to which asymptotic confidence intervals align with empirical results is examined. Finally, the collaboration with Triodos Bank also involves implementing and testing the algorithm on real-world data.

This thesis starts with reviewing the basics of modelling portfolio credit risk in Chapter 2. In Chapter 3, the foundations of Monte Carlo methods and the principles of IS, AIS and quantile estimation are introduced. Next, in Chapter 4, we discuss methods to select suitable IS proposals for extreme quantile estimation in the credit portfolio setting, substantiated with experiments. These experiments give valuable insights about the methods before they are applied to more complex test portfolios mimicking real-world cases, described in Chapter 5. In Chapter 6, numerical results across these portfolio scenarios are presented. Finally, Chapter 7 concludes this thesis.

2

Portfolio Credit Risk Modelling

In this chapter, we introduce the mathematical framework for modelling portfolio credit risk. Consider a credit portfolio exposed to m obligors. Let D_k be a default indicator variable that takes the value 1 if obligor k defaults within a one-year time horizon, and 0 otherwise. The total annual credit loss L of the bank is given by

$$L = \sum_{k=1}^{m} \ell_k D_k, \tag{2.1}$$

where ℓ_k is the exposure of the kth obligor.

Let F_L denote the cumulative distribution function of L. We are interested in the VaR at the 99.9% level. If F_L is continuous, the VaR at level α is defined as the quantile q_α of L, i.e. the unique solution of

$$F_L(q_\alpha) = \alpha,$$

where $\alpha = 0.999$. However, in our setting, F_L is not continuous, since ℓ_k is deterministic, leading to discrete jumps in the distribution. Therefore, we adopt the more general definition of VaR:

$$\mathrm{VaR}_L(\alpha) = \inf\{x : F_L(x) \geq \alpha\}. \tag{2.2}$$

Assuming each exposure and default probability as deterministic, the main difficulty in simulating L remains modelling the dependence among obligors. Below, we first explain how L is modelled through the normal copula model, which offers a framework for capturing default dependencies in a credit portfolio. Next, we review the ASRF model, which provides an analytical approximation of $\mathrm{VaR}_L(\alpha)$. Finally, we discuss the multifactor framework along with its concentration risk add-ons.

2.1. Dependence Modelling in Portfolio Credit Risk

In the normal copula model, dependence between the m obligors is modelled through a multivariate normal random vector $(X_1, X_2, \ldots, X_m)$. The latent variable X_k is a latent variable associated with the kth obligor. It defaults the obligor if below a certain threshold:

$$D_k = \begin{cases} 1, & \text{if } X_k < \Phi^{-1}(p_k), \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Here, Φ is the standard normal distribution function, and p_k is the *marginal probability of default*. This implies that

$$\mathbb{P}(D_k = 1) = \mathbb{P}(X_k < \Phi^{-1}(p_k)) = p_k. \tag{2.4}$$

Assume there are d common factors. The dependence between the latent variables X_k is modelled using the following factor structure:

$$X_k = \mathbf{r}_k^\top \mathbf{Y} + b_k Z_k, \tag{2.5}$$

where

- $\mathbf{r}_k \in \mathbb{R}^d$ is a vector of factor loadings satisfying $0 \leq \|\mathbf{r}_k\| \leq 1$;
- $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ is a d-dimensional standard normal random vector representing mutually independent common factors;
- $Z_k \sim \mathcal{N}(0,1)$ are independent standard normal random variables representing the idiosyncratic factors;
- $b_k = \sqrt{1 - \mathbf{r}_k^\top \mathbf{r}_k}$ ensures that $X_k \sim \mathcal{N}(0,1)$.

The factor loading vector $\mathbf{r}_k$ determines the correlation between the kth obligor and the common factors. Despite the induced dependencies by the normal copula model, since X_k is standard normally distributed, the kth obligor still has p_k probability to default.

2.2. The Asymptotic Single Risk Factor Model

There are multiple studies that suggest the existence of a common factor influencing the default probability of all obligors (see, for example, Hatchett and Kühn [20] and Koju, Koju, et al. [23]). We refer to this common factor as the macroeconomic factor.

The ASRF model assumes the existence of this macroeconomic factor and models it within the normal copula framework as a standard normal random variable, denoted by Y_M. In the single-factor case where $d = 1$, the factor model in (2.5) reduces to

$$X_k = r_{kM} Y_M + \sqrt{1 - r_{kM}^2}\, Z_k.$$

Here, r_{kM} is the factor loading of the kth obligor with respect to the macroeconomic factor Y_M.

The IRB approach for corporate exposures (see Banking Supervision [5], Section 31.5) specifies the correlation between the kth obligor and the macroeconomic factor as

$$\rho(p_k) = 0.12 \cdot \frac{1 - e^{-50 p_k}}{1 - e^{-50}} + 0.24 \cdot \left(1 - \frac{1 - e^{-50 p_k}}{1 - e^{-50}}\right). \tag{2.6}$$

Equation (2.6) implies an inverse relationship between p_k and the correlation with the macroeconomy. This reflects obligors with low default probability tending to default only in macroeconomic downturns, whereas riskier obligors default more on an individual basis. In Chapter 5, we also use this correlation to compute the factor loading as the square root of $\rho(p_k)$:

$$r_{kM} = \sqrt{\rho(p_k)}. \tag{2.7}$$

Assuming infinite granularity, in the case of the existence of a single common factor, Vasicek [31, page 3] derives an expression for the distribution function of portfolio losses in an:

$$F_{L_M}(x) = \Phi\left(\frac{\sqrt{1 - r_{kM}^2}\, \Phi^{-1}(x) + \Phi^{-1}(\alpha)}{r_{kM}}\right). \tag{2.8}$$

This result is further used to compute the VaR at level α. We refer to this as the *macroeconomic VaR*, given by

$$\mathrm{VaR}_{L_M}(\alpha) = \sum_{k=1}^{m} \ell_k\, \Phi\left(\frac{\Phi^{-1}(p_k) + r_{kM}\, \Phi^{-1}(\alpha)}{\sqrt{1 - r_{kM}^2}}\right), \tag{2.9}$$

where L_M denotes the loss only accounting for the macroeconomic factor.

2.3. The Multifactor Model and Concentration Risk Add-Ons

Unlike the ASRF framework, which models only a single common factor, the multifactor generalizes to the case where $d > 1$. Specifically, we assume that each obligor is exposed to three factors: a macroeconomic factor, a sector-specific factor, and a geographic-region-specific factor. Let there be d_S sectors and d_G geographic regions. These are modelled through $\mathbf{Y} \in \mathbb{R}^d$, with $d = 1 + d_G + d_S$. Accordingly, the

factor loading vector $\mathbf{r}_k$ has at most three non-zero components: r_{kM} for the macroeconomic factor, and one each for the sectoral and geographic factors, which are calculated based on empirical observations of correlations between exposures within the same geography and sector. We assume non-negative correlations, so we restrict the factor loading vector to be non-negative component-wise: $\mathbf{r}_k \geq 0$. The factor loading vector for the kth obligor is expressed as

$$\mathbf{r}_k = \begin{pmatrix} r_{kM} & 0 & \dots & 0 & r_{kg(k)} & 0 & \dots & 0 & r_{ks(k)} & 0 & \dots \end{pmatrix}^\top, \quad \text{for } k \in \{1, \dots, m\}, \tag{2.10}$$

where $g(k) \in \{2, \dots, k_G+1\}$ and $s(k) \in \{d_G+2, \dots, d_G+d_S+1\}$ denote the indices of the geographic and sectoral factors to which the kth obligor is exposed. For notational convenience, we write $r_{kg(k)} := r_{kg}$ and $r_{ks(k)} := r_{ks}$.

Total Concentration Risk Add-on

The total concentration risk add-on represents the additional capital needed to account for all modelled common factors beyond the macroeconomic factor. We denote this add-on by

$$C_{\text{total}} := \text{VaR}_{L_{MGSI}}(\alpha) - \text{VaR}_{L_M}(\alpha), \tag{2.11}$$

where L_{MGSI} is the loss accounting for macroeconomic, geographic, sectoral, and idiosyncratic factors, and L_M includes only the macroeconomic factor.

Frequently, banks want to understand the impact of modelling concentration, or the contribution of specific sectors or geographies. Therefore, marginal and specific concentration add-ons are computed.

Marginal Concentration Risk Add-Ons

The marginal concentration risk add-ons refer to the extra capital required when extending the model from only modelling the macroeconomic factor to also include a subset of the common factors and/or the idiosyncratic factors. These marginal add-ons are computed with losses similar to (2.1), but differ in the fact that now certain factors are excluding the corresponding common factors. We provide two examples, including the geographical and idiosyncratic risk.

If the model only includes macroeconomic and idiosyncratic components, the loss can be expressed as

$$L_{MI} = \sum_{k=1}^{m} \ell_k \mathbf{1}\left\{ \frac{\Phi^{-1}(p_k) - r_{kM} Y_M}{\sqrt{1 - r_{kM}^2}} < Z_k \right\}, \tag{2.12}$$

where $\mathbf{1}\{x < c\}$ denotes the indicator function that is equal to 1 if $x < c$ and 0 otherwise.

If only macroeconomic and geographic components are modelled, the expected loss conditional on the common factors becomes deterministic. The loss is expressed as

$$L_{MG} = \sum_{k=1}^{m} \ell_k \Phi\left(\frac{\Phi^{-1}(p_k) - r_{kM} Y_M - r_{kg} Y_g}{\sqrt{1 - r_{kM}^2 - r_{kg}^2}} \right). \tag{2.13}$$

The use of the indicator function in (2.12) versus the normal distribution in (2.13) reflects this presence or absence of idiosyncratic risk. When idiosyncratic risk is present, loss realization is stochastic.

The corresponding marginal add-ons are denoted as

$$\begin{aligned} C_{MI} &:= \text{VaR}_{L_{MI}}(\alpha) - \text{VaR}_{L_M}(\alpha), \\ C_{MG} &:= \text{VaR}_{L_{MG}}(\alpha) - \text{VaR}_{L_M}(\alpha). \end{aligned}$$

Since concentration increases the tail heaviness of the loss distribution, concentration risk add-ons should be positive to reflect the heightened risk. These marginal add-ons indicate how much additional capital a bank should hold to account for a particular type of concentration.

Specific Concentration Risk Add-On

The specific concentration risk add-on quantifies the capital impact of including or excluding a particular subset of exposures within a portfolio. It is defined as the difference between the total concentration risk add-on with and without that subset.

Suppose we are interested in computing this specific add-on for a sub-portfolio A, where A denotes an index set of obligors. Let $A \cup A^c$ represent the full portfolio. The total concentration risk add-on excluding sub-portfolio A is denoted by $C_{\text{total}}^{A^c}$.

The specific concentration risk add-on for subset A is computed as

$$C_{\text{total}}^{\Delta A} = C_{\text{total}}^{A \cup A^c} - C_{\text{total}}^{A^c},$$

where $C_{\text{total}}^{A \cup A^c} = C_{\text{total}}$ is the total concentration risk add-on for the full portfolio.

This measure captures the impact of a sub-portfolio A on the overall concentration risk. A positive value of $C_{\text{total}}^{\Delta A}$ indicates that A increases concentration risk (i.e., reduces diversification), while a negative value implies that A contributes to diversification and lowers the total concentration risk of the portfolio.

3

Monte Carlo Methods for Quantile Estimation

In this chapter, we study quantile estimation with Monte Carlo methods. Below, we first introduce some of the fundamental principles of Monte Carlo methods and review the basic concepts of variance reduction techniques, specifically IS and its adaptive variant, AIS. Next, we discuss methods based on order statistics for estimating quantiles and describe how exact confidence intervals can be constructed in this setting. To estimate extreme quantiles, we focus on tail probability estimation, which makes it easier to apply IS and AIS methods for more efficient estimation.

3.1. Monte Carlo Methods and Variance Reduction Techniques

Monte Carlo methods are a class of computational algorithms that use random sampling to obtain numerical results. They are especially useful for estimating outcomes dependent on high-dimensional variables, where standard numerical methods become inefficient or intractable. Classical numerical integration techniques suffer from the curse of dimensionality, while the convergence rate of Monte Carlo estimates is independent of the dimension. This dimensional robustness makes Monte Carlo methods attractive for such problems. As a result, they are widely used in fields such as physics, machine learning, and finance.

We begin by stating two fundamental theorems from probability theory that form the foundation for why Monte Carlo methods are effective in many applications. These results ensure that averages computed from random samples converge to the true expected values, and they quantify the uncertainty of these estimates.

The first theorem proves convergence of independent and identically distributed (i.i.d.) random variables, showing that the sample mean converges to the expected value. This is known as the *Strong Law of Large Numbers* (SLLN).

Theorem 1 (Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu < \infty$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu, \quad \text{as } n \to \infty, \tag{3.1}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

The next theorem describes the distributional behaviour of the sample mean. It states that the sampling distribution of the sample average of i.i.d. observations converges to a normal distribution as the sample size grows. This result is known as the *Central Limit Theorem* (CLT).

Theorem 2 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

denote the sample mean. Then,

$$\sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \text{as } n \to \infty, \tag{3.2}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Proofs of both theorems can be found in, e.g., Gut [19], Chapters 6-7.

These theorems enable the evaluation of integrals via random sampling. We begin with a simple example of Monte Carlo integration in one dimension.

Example 1. *Consider the problem of evaluating the integral*

$$\int_0^1 h(x)\mathrm{d}x,$$

where h is an integrable function. Suppose we can generate independent random variables $U_1, U_2, \ldots, U_n$, each uniformly distributed over the interval $[0, 1]$. By Theorem 1, the sample average of the function values converges almost surely to the value of the integral:

$$\frac{1}{n}\sum_{i=1}^{n} h(U_i) \xrightarrow{a.s.} \int_0^1 h(x)\mathrm{d}x \quad \text{as } n \to \infty.$$

Moreover, if h is square-integrable on $[0, 1]$, the sample mean satisfies a central limit theorem by Theorem 2.

Monte Carlo integration naturally extends to higher-dimensional integrals. According to (3.2), independent of the dimension, the error decreases at the rate of $n^{-1/2}$. But, if increasing the number of samples becomes computationally infeasible, variance reduction techniques offer an alternative. These methods aim to reduce the variability in simulation outputs, allowing a desired level of precision to be achieved with fewer samples. Common variance reduction techniques include control variates, antithetic variates, and stratified sampling (see, e.g., Glasserman [13], Chapter 4). We focus on IS and AIS and review their principles below.

3.1.1. Importance Sampling

With IS, we aim to reduce the variance by producing more important observations. This is done by sampling paths from an alternative sampling distribution, known as the proposal distribution. To correct for sampling from this proposal instead of the original distribution, the estimate is re-weighted.

To be more concrete, let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$ with distribution F, and let $h : \mathbb{R}^d \to \mathbb{R}$ be a function integrable with respect to F. Our goal is to compute the expectation

$$\mathbb{E}_F[h(\mathbf{X})] = \int_{\mathbb{R}^d} h(\mathbf{x})F(\mathrm{d}\mathbf{x}),$$

where $\mathbb{E}_F$ denotes expectation with respect to the distribution F, meaning that the random vector $\mathbf{X}$ is distributed according to F. If we choose a proposal distribution G such that $F \ll G$, it implies the existence of a density p such that

$$\mathrm{d}F(\mathbf{x}) = p(\mathbf{x})\mathrm{d}G(\mathbf{x}), \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \tag{3.3}$$

The function p is also referred to as the *Radon–Nikodym derivative* or *likelihood ratio*. The relationship in (3.3) yields the standard IS identity.

$$\mathbb{E}_F[h(\mathbf{X})] = \int_{\mathbb{R}^d} h(\mathbf{x})\mathrm{d}F(\mathbf{x}) = \int_{\mathbb{R}^d} h(\mathbf{x})p(\mathbf{x})\mathrm{d}G(\mathbf{x}) = \mathbb{E}_G\left[h(\mathbf{X})p(\mathbf{X})\right], \tag{3.4}$$

where $\mathbb{E}_G$ is the expectation with respect to G. Given $\mathbf{X}_1, \mathbf{X}_2 \ldots, \mathbf{X}_n$ drawn i.i.d. from G, the IS estimator is

$$\frac{1}{n}\sum_{i=1}^{n} h(\mathbf{X}_i)p(\mathbf{X}_i), \tag{3.5}$$

where the computed likelihood ratio $p(\mathbf{X}_i)$ is used to correct for drawing $\mathbf{X}_i$ from the proposal distribution. Henceforth, we will refer to these evaluated likelihood ratios as IS weights. Assuming $\mathrm{Var}_G\left(h(\mathbf{X})p(\mathbf{X})\right) < \infty$, then, by the SLLN,

$$\frac{1}{n}\sum_{i=1}^{n} h(\mathbf{X}_i)p(\mathbf{X}_i) \xrightarrow{\text{a.s.}} \mathbb{E}_G\big[h(\mathbf{X})p(\mathbf{X})\big], \quad \text{as } n \to \infty.$$

Since the estimator is unbiased, variance comparisons reduce to comparing second moments. Under IS, the second moment becomes

$$\mathbb{E}_G\left[\left(h(\mathbf{X})p(\mathbf{X})\right)^2\right] = \mathbb{E}_F\left[h(\mathbf{X})^2 p(\mathbf{X})\right], \tag{3.6}$$

which may be larger or, ideally, smaller than $\mathbb{E}_F[h(\mathbf{X})^2]$. This shows that the effectiveness of IS for variance reduction depends on choosing the right proposal distribution.

The optimal proposal G^* is the one that minimizes the second moment in (3.6), and must therefore satisfy

$$\mathrm{d}G^*(\mathbf{x}) \propto |h(\mathbf{x})|\mathrm{d}F(\mathbf{x}).$$

More precisely, assuming $\mathbb{E}_F[|h(\mathbf{X})|] > 0$, the unique proposal distribution G^* (up to F-almost everywhere) under which the IS estimator has zero variance is given by

$$\mathrm{d}G^*(\mathbf{x}) = \frac{|h(\mathbf{x})|}{\mathbb{E}_F[|h(\mathbf{X})|]}\mathrm{d}F(\mathbf{x}). \tag{3.7}$$

The corresponding density $p^*(\mathbf{x}) := |h(\mathbf{x})|/\mathbb{E}_F[|h(\mathbf{X})|]$ is known as the *zero-variance density*, as it yields an IS estimator with zero variance. For a derivation, see, e.g., Boks [6] (Section 1.4).

In practice, since the normalizing constant $\mathbb{E}_F[|h(\mathbf{X})|]$ is typically unknown (and often the very quantity being estimated), one must approximate G^* by choosing a tractable family of densities that mimics the shape of $|h(\mathbf{x})|\mathrm{d}F(\mathbf{x})$.

3.1.2. Adaptive IS

In adaptive importance sampling (AIS), samples are drawn from a sequence of proposal distributions. The goal is to reduce the variance of the estimate by iteratively improving the proposal distribution. For now, we assume the existence of a black-box function $\mathcal{I}$ which updates a proposal distribution by using the latest available information. In practice, such a function can be difficult to obtain, since it could depend on parameters we aim to estimate. We return to this issue in Section 4.3.1, where we discuss an $\mathcal{I}$ in the portfolio credit risk setting.

Let $G_1, G_2, \ldots, G_n$ be proposal distributions such that $F \ll G_i$ for $i = 1, 2 \ldots, n$. Then, as in ordinary IS, this implies the existence of the sequence of densities $p_1, p_2, \ldots, p_n$ such that, for each $i = 1, 2, \ldots, n$,

$$\mathrm{d}F(\mathbf{x}) = p_i(\mathbf{x})\mathrm{d}G_i(\mathbf{x}), \quad \text{for all } \mathbf{x} \in \mathbb{R}^d.$$

Suppose that for each iteration $i = 1, \ldots, n$, we draw a sample $\mathbf{X}_i$ from a proposal distribution G_i. Following each draw, the proposal G_{i+1} is updated in accordance with $\mathcal{I}$.

Given each draw $\mathbf{X}_i$, we define the corresponding IS weight as

$$W_i := p(\mathbf{X}_i), \quad \text{for } i = 1, \ldots, n,$$

This procedure yields a sequence of weighted samples $\{(\mathbf{X}_i, W_i)\}_{i=1}^n$. Similarly to IS, the AIS estimator for $\mathbb{E}_F[h(X)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} h(\mathbf{X}_i)W_i.$$

But there is an important difference between the IS and AIS estimators. In the AIS case, the samples $\mathbf{X}_i$ are drawn from a sequence of proposal distributions that evolve over time and depend on past samples. Consequently, the data pairs $(\mathbf{X}_i, W_i)$ are neither independent nor identically distributed. As a result, the SLLN from Theorem 1 cannot be directly applied, and alternative techniques are required to establish the convergence of this estimate towards its true value.

In some cases, a martingale structure arises, allowing the application of martingale convergence theorems to establish the almost sure convergence of the AIS estimator under appropriate conditions, see, e.g, Arouna [3]. We return to the question of convergence in Section 4.3, where we use AIS in the portfolio credit risk setting.

3.2. Quantile Estimation

In this section we revisit the problem of computing $\mathrm{VaR}_L(\alpha)$, defined in (2.2) in the portfolio credit risk setting. We assume that the losses $L_1, L_2, \ldots, L_n$ are generated according to a multifactor model described in (2.5). Below, we discuss quantile estimation using order statistics and the estimation of multiple tail probabilities. We address both methods because the first provides exact confidence intervals, while the second offers advantages when IS is applied to estimate extreme quantiles.

3.2.1. Order Statistics

To estimate q_α, we order the samples $L_1, L_2, \ldots, L_n$ such that

$$L_{1:n} \leq L_{2:n} \leq \cdots \leq L_{n:n}.$$

Then the α-quantile estimator can be written as

$$\hat{q}_\alpha = L_{\lceil \alpha n \rceil : n}, \tag{3.8}$$

where $\lceil s \rceil$ is the ceiling function, which rounds any real number s up to the nearest integer. Interpolation methods between the order statistics can be applied in order to reduce the bias of the estimate.

An attractive property of order statistics is that we can create an exact confidence interval for the true quantile q_α. The number of observations less than or equal to a fixed quantile threshold follows a binomial distribution. Consequently, this can be used to construct an exact confidence interval. This is useful because it does not require knowledge of the distribution F_L. Since this method is directly applicable to constructing confidence intervals, we state the result.

Theorem 3 (Exact Confidence Interval)**.** *Let $L_{1:n} \leq \cdots \leq L_{n:n}$ be order statistics from i.i.d. samples from a distribution F. Then for $\alpha \in (0,1)$, the probability that the quantile q_α lies in the interval $[L_{k_1:n}, L_{k_2:n})$ is*

$$\mathbb{P}(L_{k_1:n} < q_\alpha < L_{k_2:n}) = \sum_{i=k_1}^{k_2-1} \binom{n}{i} \alpha^i (1-\alpha)^{n-i}. \tag{3.9}$$

Although this result is likely known, we provide a brief proof in Appendix A. Equation (3.9) yields a confidence interval for q_α with exact finite-sample coverage probability. The values k_1 and k_2 can be chosen such that this coverage probability approaches the desired confidence level.

3.2.2. Estimating Tail Probability

Rather than estimating q_α directly, we can estimate the tail probability, defined as

$$P_x := \mathbb{P}(L > x),$$

for multiple values of x near q_α. By identifying the x for which this probability is closest to the desired level $1-\alpha$, we can recover the corresponding quantile.

To make this concrete in our credit loss setting, observe that the credit loss is a mapping from the common and idiosyncratic factors to the reals, i.e., $L : (\mathbf{Y}, \mathbf{Z}) \to \mathbb{R}$, where $(\mathbf{Y}, \mathbf{Z}) \in \mathbb{R}^v$ is multivariate standard normally distributed, with $v = d + m$. The tail probability can be expressed as

$$P_x = \mathbb{E}[\mathbf{1}\{L > x\}] = \int_{\mathbb{R}^v} \mathbf{1}\{L > x\} \mathrm{d}\mathbf{\Phi}(\mathbf{y}, \mathbf{z}),$$

where $\mathbf{\Phi}$ denotes the multivariate standard normal distribution function. If we draw n i.i.d. vectors $(\mathbf{Y}_1, \mathbf{Z}_1), (\mathbf{Y}_2, \mathbf{Z}_2), \ldots, (\mathbf{Y}_n, \mathbf{Z}_n)$ from $\mathcal{N}(\mathbf{0}, \mathbf{I}_v)$, the Crude Monte Carlo (CMC) estimator is given by

$$\hat{P}_{x,n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{L_i > x\}.$$

This estimate satisfies the CLT:

$$\sqrt{n}\big(\hat{P}_{x,n} - P_x\big) \xrightarrow{d} \mathcal{N}\big(0,\, \hat{\psi}\big), \quad n \to \infty,$$

where

$$\hat{\psi} = P_x(1 - P_x). \tag{3.10}$$

A consistent estimator for $\tilde{\psi}$ is

$$\hat{\psi}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{L_i > x\} - \hat{P}_{x,n}^2. \tag{3.11}$$

In Section 3.3 we present a similar CLT for the case of IS-based estimation of P_x.

The corresponding quantile estimate is

$$\hat{q}_{\alpha,n} = \inf\left\{x : \hat{P}_{x,n} \leq 1 - \alpha\right\}. \tag{3.12}$$

For unique quantiles — that is, when q_α is the unique solution of $P_x \leq \alpha \leq P_{x^-}$ — Serfling [29, page 75] shows that $\hat{q}_{x,n}$ converges to q_α with probability 1. In our case, the discreteness of F_L means quantiles are not unique. In this case, the lower and upper α-quantiles, defined as

$$q_\alpha = \inf\{x : \mathbb{P}(L \leq x) \geq \alpha\}$$
$$q^\alpha = \sup\{x : \mathbb{P}(L \leq x) \leq \alpha\}$$

are distinct. If F_L has a jump at level α, then

$$q_\alpha < q^\alpha,$$

so that the set of α-quantiles is the entire interval $[\,q_\alpha,\, q^\alpha]$.

In this situation, although the number of samples n goes to infinity, the quantile estimates in (3.12) may instead oscillate across the interval $[\,q_\alpha,\, q^\alpha]$ infinitely often.

3.3. Extreme Quantile Estimation with IS

Since the 99.9% VaR of the loss is located in the far tail of the distribution, the available observations in the neighbourhood of the true quantile are sparse, resulting in high variability across repeated quantile estimations. Simply generating more observations is inefficient, as most of them are not located in the tail. IS offers a solution by generating more observations in the region of interest. We apply IS on tail probabilities since they can be expressed as expectations.

Below, we use the results from Glynn et al. [16]. They show that applying IS to estimating quantiles with level α close to 0 or 1 can yield substantial variance reduction. Under standard regularity conditions, they prove a CLT for these quantile estimators. Below, we discuss such IS quantile estimators used for our portfolio credit risk setting.

Let us assume there exists a distribution $\mathbf{G}$ such that $\Phi \ll \mathbf{G}$. Then, there exists a density p satisfying

$$\mathrm{d}\mathbf{\Phi}(\mathbf{y}, \mathbf{z}) = p(\mathbf{y}, \mathbf{z})\mathrm{d}\mathbf{G}(\mathbf{y}, \mathbf{z}), \quad \text{for all } (\mathbf{y}, \mathbf{z}) \in \mathbb{R}^v.$$

If we draw $(\mathbf{Y}_1, \mathbf{Z}_1), (\mathbf{Y}_2, \mathbf{Z}_2), \ldots, (\mathbf{Y}_n, \mathbf{Z}_n)$ from $\mathbf{G}$, the unbiased estimator for P_x is given by

$$\tilde{P}_{x,n} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{L_i > x\}\, W_i, \quad \text{where } W_i := p(\mathbf{Y}_i, \mathbf{Z}_i). \tag{3.13}$$

Henceforth, we use a tilde on our estimates to indicate that samples are drawn from IS proposal distributions rather than their original distributions.

We write the asymptotic variance as

$$\tilde{\psi}^2 := \mathrm{Var}_{\mathbf{G}}\left(\mathbf{1}\{L > x\}\, W_i\right). \tag{3.14}$$

If $\tilde{\psi}^2 < \infty$, then, since $\tilde{P}_{x,n}$ is a sum of i.i.d. random variables, a CLT is satisfied:

$$\sqrt{n}\big(\tilde{P}_{x,n} - P_x\big) \xrightarrow{d} \mathcal{N}\big(0,\, \tilde{\psi}^2\big), \quad \text{as } n \to \infty.$$

A consistent estimator for $\tilde{\psi}$ is

$$\tilde{\psi}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{L_i > x\} W_i^2 - \tilde{P}_{x,n}^2. \tag{3.15}$$

The corresponding quantile estimator is

$$\tilde{q}_{\alpha,n} = \inf\left\{x : \tilde{P}_{x,n} \le 1-\alpha\right\}. \tag{3.16}$$

This quantile estimate is computed by first ordering the losses such that $L_{1:n} \le L_{2:n} \le \cdots \le L_{n:n}$. Denote $W_{(i:n)}$ as the IS weight corresponding to the ordered sample $L_{i:n}$. Then, by selecting the largest integer k such that $\frac{1}{n}\sum_{i=k}^{n} W_{(i:n)} \ge 1-\alpha$ holds, the IS quantile estimate is chosen as $\tilde{q}_{\alpha,n} := L_{k:n}$. For shorthand, we denote this mapping as

$$\Pi_\alpha : \{(L_i, W_i)\}_{i=1}^{n} \to \tilde{q}_{\alpha,n} \tag{3.17}$$

We arrive at the most challenging part. It remains to find such a proposal $\mathbf{G}$ in our portfolio credit risk setting. We further discuss this in Chapter 4.

3.4. Asymptotic Confidence Intervals

Throughout this section, we temporarily drop discreteness assumptions on L and assume it has a continuous and strictly positive density f_L at q_α. Below, we first discuss CMC and IS confidence intervals under continuity, then we discuss what happens if L is discrete.

CMC setting

We start with the CMC setting, as it introduces the challenge corresponding to asymptotic confidence intervals for extreme quantiles in the basic setting. Under the regulatory assumptions named above, $\hat{q}_{\alpha,n}$ satisfies a CLT:

$$\sqrt{n}(\hat{q}_{\alpha,n} - q_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f_L(q_\alpha)^2}\right), \quad \text{as } n \to \infty. \tag{3.18}$$

For a proof, see e.g., Serfling [29, page 77]. This allows for the construction of a confidence interval of the form

$$\left(\hat{q}_{\alpha,n} \pm z_{p/2} \cdot \frac{\sqrt{\alpha(1-\alpha)}}{f_L(q_\alpha)\sqrt{n}}\right),$$

where $z_{p/2} = \Phi^{-1}(1-p/2)$ denotes the $(1-p/2)$-quantile of the standard normal distribution. However, for extreme quantiles $f_L(q_\alpha)$ may be very close to zero, leading to very high variance. Moreover, because constructing the interval requires evaluating the quantile density $\phi_\alpha := 1/f_L(q_\alpha)$, this approach is impractical. Fortunately, in the CMC case, we can avoid this step by using the exact confidence interval in (3.9).

IS setting

In the IS case, we cannot use the exact confidence interval. To construct a confidence interval based on a CLT, we state and discuss results from Chu and Nakayama [9, 10] below. Their work offers a detailed analysis and construction of asymptotic confidence intervals under variance reduction techniques. Furthermore, they provide a consistent estimator for ϕ_α.

We summarize some of their results as a theorem, where we use $\tilde{F}_n(x) := 1 - \tilde{P}_{x,n}$ and $\tilde{\psi}_n$ defined in (3.15):

Theorem 4. *(Quantile CLT Chu and Nakayama) Suppose that $F'(q_\alpha) > 0$, and suppose there exist $\varepsilon > 0$ and $\delta > 0$ such that $\mathbb{E}_{\mathbf{G}}\left[\mathbf{1}\{L > q_\alpha - \delta\}W^{2+\varepsilon}\right] < \infty$. Then,*

$$\frac{\sqrt{n}}{\tilde{\psi}_n \tilde{\phi}_n(\kappa)} \left(\tilde{q}_{\alpha,n} - q_\alpha\right) \xrightarrow{d} \mathcal{N}(0,1), \quad \text{as } n \to \infty$$

where $\tilde{\phi}_n$ is defined as

$$\tilde{\phi}_n(\kappa) := \frac{\tilde{F}_n^{-1}(\alpha + h_n(\kappa)) - \tilde{F}_n^{-1}(\alpha - h_n(\kappa))}{2h_n(\kappa)}, \quad \text{with } h_n(\kappa) := \kappa n^{-1/2}. \tag{3.19}$$

For the proof, we refer to Chu and Nakayama [10] (Theorem 4.2).

Theorem 4 enables us to construct a $(1-p) \times 100\%$ confidence interval for the α-quantile:

$$\left(\tilde{q}_{\alpha,n} \pm z_{p/2} \cdot \frac{\tilde{\psi}_n \tilde{\phi}_n(\kappa)}{\sqrt{n}}\right), \tag{3.20}$$

When selecting the parameter κ, it must be chosen such that $0 \leq \alpha \pm h_n(\kappa) \leq 1$, which ensures that the quantile estimates remain within the distribution's support and are based on distinct order statistics. Chu and Nakayama [9] also proves consistency for linear combinations for quantile density estimators—such as Richardson extrapolation—to reduce bias.

Furthermore, we find it worthwhile to mention that He, Jiang, et al. [21] establish asymptotic normality of the quantile estimates under AIS under similar conditions, but do not provide a consistent estimator for the quantile density.

Discussing the Discrete Case

Having discussed the continuous case, we go further with the discrete case of L. For discrete distributions, it is infeasible to construct asymptotic confidence intervals, also in the non-IS case. Take, for example, an extreme portfolio consisting of 100 obligors, all with exposure 1, except for a single obligor, who has exposure 1 000. Possible annual losses are $1\,000, 1\,001, \ldots$. However, losses strictly between 99 and 1 000 cannot occur. Hence, constructing a confidence interval around the points 99 and 1 000 would be impracticable. Despite this lack of continuity, it may happen that the distribution seems to behave continuously in the case of portfolios consisting of large numbers of obligors. In large portfolios with evenly spread exposures, L can take many more distinct values. Consequently, this reduces big jumps in F_L, approximating continuity. Alternatively, we could assume the underlying variable representing the exposure is continuous. Another technique would be artificial smoothing, where instead of modelling only L, we model $L + \epsilon$, where $\epsilon \sim \mathcal{N}(0,1)$ is independent of L. Since ϵ has a strict positive density everywhere, conditions are satisfied to construct an asymptotic confidence interval.

In Section 6.4 we test the construction of a CI with Theorem 4, as if continuity holds for the distributions describing portfolio credit loss.

Moreover, it is worthwhile to mention that Ma, Genton, et al. [24] and Wang and Hutson [32] use approaches based on mid-distribution functions and fractional order statistics to establish asymptotic normality, but with no incorporation of IS yet. This would be an interesting direction for further research, as time constraints prevent it for this thesis.

4

Optimizing Proposal Distributions

In this chapter, we focus on finding suitable IS proposal distributions that minimize the variance of extreme quantile estimators. Thereby, we concentrate on the portfolio credit loss setting, where losses are modelled using a multifactor model described in (2.5).

Below, we start with reviewing the methods from Glasserman and Li [15]. They propose Bernoulli tilting conditional on the common factors. We study this method through an experiment involving a portfolio in scenarios of both high and low correlation among assets. Next, we discuss how IS can be applied through a shifted mean of the common factors. For this, we use the results from Glasserman and Li [15] who propose, among other approximations, the constant approximation method. As a final method, we discuss an AIS method developed by Egloff, Leippold, et al. [12] and Egloff and Leippold [11].

Furthermore, introduce a method to compute a zero-variance distribution for simple portfolios. This involves using the Generalized Poisson Binomial (GPB) distribution. We conclude this chapter with an experiment on a specific portfolio scenario.

4.1. Bernoulli Tilting Conditional on the Common Factors

Glasserman and Li [15] state that, conditional on the common factors, the credit loss L becomes a sum of independent Bernoulli variables. By tilting the Bernoulli probabilities via an *exponential twist*, they construct an effective proposal distribution to estimate P_x at large values of x. They show that Bernoulli tilting is effective for portfolios with low correlation among obligors, but may offer no efficiency gains over CMC for highly correlated portfolios.

We start with discussing their theoretical foundation. Next, we demonstrate and compare the effectiveness of Bernoulli tilting on two specific portfolio scenarios that differ in correlation structure with an experiment.

4.1.1. Theoretical Foundation

Conditioned on $\mathbf{Y}$, the indicators D_k of obligor defaults become independent Bernoulli variables. The conditional default probability of the kth obligor is

$$p_k(\mathbf{y}) = \mathbb{P}(D_k = 1 \mid \mathbf{Y} = \mathbf{y}) = \mathbb{P}(X_i < \Phi^{-1}(p_i) \mid \mathbf{Y} = \mathbf{y}) = \Phi\left(\frac{\Phi^{-1}(p_k) - \mathbf{r}_k^\top \mathbf{y}}{b_k}\right). \tag{4.1}$$

Next, an exponential twist is applied to these probabilities by selecting the parameter $\theta > 0$, determining the tilted default probability as

$$p_k(\theta, \mathbf{y}) = \frac{p_k(\mathbf{y}) e^{\theta \ell_k}}{1 + p_k(\mathbf{y})(e^{\theta \ell_k} - 1)},$$

which increases the default probability. Moreover, the larger the exposure ℓ_k, the greater the increase in default probability. Using $\theta = 0$ corresponds to using the original probabilities.

Applying this modification for $k = 1, \ldots, m$, the likelihood ratio under these twisted probabilities simplifies to

$$\prod_{k=1}^{m} \left(\frac{p_k(\mathbf{y})}{p_k(\theta, \mathbf{y})} \right)^{D_k} \left(\frac{1 - p_k(\mathbf{y})}{1 - p_k(\theta, \mathbf{y})} \right)^{1 - D_k} = e^{-\theta L + \Lambda(\theta, \mathbf{y})}, \tag{4.2}$$

where

$$\Lambda(\theta, \mathbf{y}) = \log \mathbb{E}\left[e^{\theta L} \mid \mathbf{Y} = \mathbf{y} \right] = \sum_{k=1}^{m} \log\left(1 + p_k(\mathbf{y})(e^{\theta \ell_k} - 1)\right),$$

is the *cumulant generating function* of L. Substitution and direct computation confirm the equivalence in (4.2). The resulting unbiased estimator for P_x becomes

$$\mathbf{1}\{L > x\} e^{-\theta L + \Lambda(\theta, \mathbf{y})}. \tag{4.3}$$

It remains to choose an appropriate θ. For this, a θ is selected that minimizes the upper bound of the second moment. For all $\theta \geq 0$, an upper bound is given by

$$\mathbb{E}_\theta \left[\mathbf{1}\{L > x\} e^{-2\theta L + 2\Lambda(\theta, \mathbf{Y})} \mid \mathbf{Y} = \mathbf{y} \right] \leq e^{-2\theta x + 2\Lambda(\theta, \mathbf{y})},$$

where $\mathbb{E}_\theta$ denotes expectation under the θ-twisted probabilities. Thus, minimizing the upper bound reduces to maximizing the function $\theta x - \Lambda(\theta, \mathbf{y})$ over θ. If, for at least one k we have $\ell_k > 0$ and $0 < p_k(\mathbf{y}) < 1$, then, since $\Lambda(\theta, \mathbf{y})$ is strictly convex for $\theta > 0$, this maximum is attained at

$$\theta_x = \begin{cases} 0, & \text{if } x \leq \frac{\partial}{\partial \theta} \Lambda(\theta, \mathbf{y})\big|_{\theta=0}, \\ \text{the unique solution to } \frac{\partial}{\partial \theta} \Lambda(\theta, \mathbf{y}) = x, & \text{if } x > \frac{\partial}{\partial \theta} \Lambda(\theta, \mathbf{y})\big|_{\theta=0}. \end{cases} \tag{4.4}$$

To interpret this choice of θ_x, the standard property of exponential twisting is used:

$$\begin{aligned} \mathbb{E}_\theta \left[L \mid \mathbf{Y} = \mathbf{y} \right] &= \sum_{k=1}^{m} p_k(\theta, \mathbf{y}) \cdot \ell_k \\ &= \sum_{k=1}^{m} \frac{p_k(\mathbf{y}) e^{\theta \ell_k}}{1 + p_k(\mathbf{y})(e^{\theta \ell_k} - 1)} \cdot \ell_k \\ &= \frac{\partial}{\partial \theta} \Lambda(\theta, \mathbf{y}). \end{aligned}$$

Thus, in the first case of (4.4), the conditional expectation of L is greater than x and the event $\{L > x\}$ is not rare, in which twisting the probabilities is not necessary. In the second case, we have $\mathbb{E}_{\theta_x} \left[L \mid \mathbf{Y} = \mathbf{y} \right] = x$. Hence, this choice of θ_x shifts the conditional expectation of the loss to the quantile of interest x. Moreover Glasserman and Li [15] (Theorem 2) establishes the conditions under which (4.3) yields an asymptotic optimality, meaning that as the event of interest becomes increasingly rare, the second moment decreases at the fastest possible rate among all unbiased estimators. However, as they mention, this optimality in the asymptotic sense fails if obligors are highly correlated, and Bernoulli tilting conditional on $\mathbf{Y}$ may not be more effective than CMC simulation.

4.1.2. Algorithm for Bernoulli Tilting

The procedure to generate loss samples under Bernoulli tilting and compute $\tilde{q}_{\alpha,n}$ is outlined in Algorithm 1. We refer to this algorithm as the Bernoulli tilting IS (B-IS) algorithm. In step 5 of the algorithm, we solve θ_x such that (4.4) is satisfied. Moreover, when $\mathbb{E}\left[L \mid \mathbf{Y} = \mathbf{y} \right] = x$, we use the bisection method to find $\frac{\partial}{\partial \theta} \Lambda(\theta, \mathbf{y}) = x$, see Appendix B. The computation of $\tilde{q}_{\alpha,n}$ is done by the mapping Π_α defined in (3.17).

4.1.3. Bernoulli Tilting: High- vs Low-Correlated Obligors

In this experiment, we assess the performance of Bernoulli tilting for quantile estimation in portfolios with different levels of correlation between obligors. Specifically, we are interested in the quantile q_α, with $\alpha = 0.99$.

Algorithm 1: B-IS

Input : Number of replications n; Probability level α; Threshold x; Portfolio data $\{\ell_k, p_k, \mathbf{r}_k, b_k\}_{k=1}^m$

Output : Estimated quantile $\tilde{q}_{\alpha,n}$

for $i = 1$ **to** n **do**
- Sample $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$
- **for** $k = 1$ **to** m **do**
  - Compute $p_k(\mathbf{Y}) \leftarrow \Phi\left(\frac{\Phi^{-1}(p_k) - \mathbf{r}_k^\top \mathbf{Y}}{b_k}\right)$
- Solve θ_x such that (4.4) is satisfied
- Compute $\Lambda \leftarrow \sum_{k=1}^m \log\left(1 + p_k(\mathbf{Y})(e^{\theta_x \ell_k} - 1)\right)$
- **for** $k = 1$ **to** m **do**
  - Compute $p_k(\theta_x, \mathbf{Y}) \leftarrow \dfrac{p_k(\mathbf{Y}) e^{\theta_x \ell_k}}{1 + p_k(\mathbf{Y})(e^{\theta_x \ell_k} - 1)}$
  - Sample $D_k \sim \text{Bernoulli}(p_k(\theta_x, \mathbf{Y}))$
- Compute $L_i = \sum_{k=1}^m \ell_k D_k$
- Compute $W_i = \exp\left(-\theta_x L_i + \Lambda\right)$

Compute $\tilde{q}_{\alpha,n} = \Pi_\alpha\left(\{(L_i, W_i)\}_{i=1}^n\right)$

return $\tilde{q}_{\alpha,n}$

We consider a simple portfolio with 1 000 obligors, where each obligor is assumed to have a default probability $p_k = 0.01$ and an exposure $\ell_k = 1$. The latent variable X_k is modelled through a single factor model following the framework described in (2.5). In the first scenario, we consider a large factor loading with the common factors for all obligors, inducing highly correlated obligors within the portfolio, given by

$$X_k^{\text{high}} = 0.3\, Y + \sqrt{1 - 0.3^2}\, Z_k, \quad \text{for } k = 1, 2, \ldots, 1\,000,$$

In the second scenario, we consider a low factor loading, inducing low-correlated obligors within the portfolio, given by

$$X_k^{\text{low}} = 0.01\, Y + \sqrt{1 - 0.01^2}\, Z_k, \quad k = 1, 2, \ldots, 1\,000,$$

Before we test Bernoulli tilting, we perform a CMC simulation on both cases, where we generate 100 000 replications. These samples are shown in Figure 4.1. It is worth noticing that the increased correlation induces a heavier tail, clearly demonstrating the effect of concentration. The CMC 99% quantile estimates are

$$\hat{q}_\alpha^{\text{high}} = 45, \quad \text{and } \hat{q}_\alpha^{\text{low}} = 18.$$

Next, we apply the B-IS procedure with Algorithm 1. For the required threshold x to tailor θ_x to, we use our CMC estimates, i.e., $x^{\text{high}} = 45$ and $x^{\text{low}} = 18$. Glasserman and Li [15] demonstrate through numerical results that when the tilting parameter θ is tailored to a specific loss level x, it remains effective for estimating probabilities at larger — and even significantly larger — loss levels. This suggests that Algorithm 1 is relatively insensitive to the precise choice of x. So, using the CMC quantile estimates as the thresholds should be a more then reasonable approach. For both scenarios, we carry out 250 runs, each with $n = 100$ replications. Each run produces a CMC quantile estimate ($\hat{q}_{\alpha,n}$) and an IS quantile estimate ($\tilde{q}_{\alpha,n}$).

The results in the scenario high-correlated obligors are shown in Figure 4.2a and for the low-correlation scenario in Figure 4.2b. It is worth noting that the right-tail skew shows the bias in quantile estimation, which is particularly clear in small-sample experiments, like these with only 100 replications. For the high-correlation case, we obtain an estimated variance reduction of 1.54, versus a variance reduction of 26.85 for the low-correlation case.

As shown in Figures 4.2a and 4.2b, Bernoulli tilting can significantly reduce variance when the correlation between obligors is low. However, its effectiveness diminishes in high-correlation scenarios,

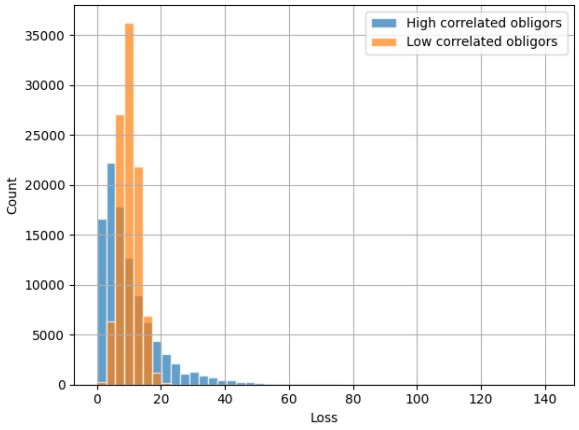
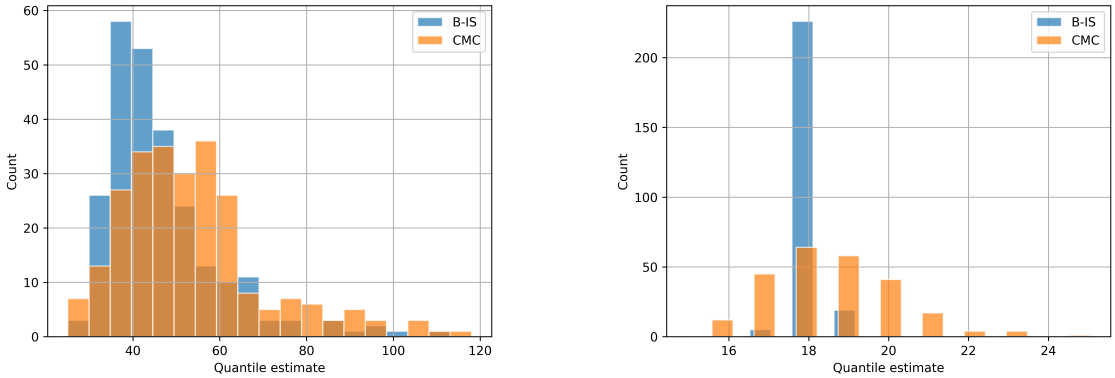


Figure 4.1: CMC simulation of 100 000 loss samples for both high- and low-correlation scenarios. The high-correlation scenario shows heavier tails, with the estimated 99% quantile at 45, compared to 18 in the low-correlation scenario.



(a) High correlated obligors within the portfolio.

(b) Low correlated obligors within the portfolio.

Figure 4.2: Histogram of 99% quantile estimates with IS ($\tilde{q}_{\alpha,n}$) compared to CMC ($\hat{q}_{\alpha,n}$), for the high- and low-correlation scenario. In the high-correlation scenario, IS via Bernoulli tilting achieves only modest variance reduction (factor 1.54). In the low-correlation scenario, it achieves a strong variance reduction (factor 26.85) relative to CMC.

where large losses are primarily driven by shifts in the common factor, consistent with the notions of Glasserman and Li [15].

Furthermore, we want to report the computational overhead induced by Bernoulli tilting: Glasserman and Li [15] mention a roughly two times slowdown per replication. This means that in the high-correlation scenario, it would be less efficient to use Bernoulli tilting, as the variance reduction is only 1.54.

4.2. Shifting the Common Factors

We continue with applying IS by shifting the mean of the common factors. If we draw $\mathbf{Y}$ from a multivariate normal $\mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_d)$, with a shifted mean $\boldsymbol{\mu}$, the unbiased estimator for P_x is

$$\mathbf{1}\{L > x\} \exp\left(-\mathbf{y}^\top \boldsymbol{\mu} + \frac{\|\boldsymbol{\mu}\|^2}{2}\right).$$

The IS weight (after the indicator) is derived from the likelihood ratio $\boldsymbol{\phi}(\mathbf{y})/\boldsymbol{\phi}(\mathbf{y}-\boldsymbol{\mu})$, where $\boldsymbol{\phi}$ is the multivariate standard normal density function. To identify a suitable shift $\boldsymbol{\mu}$, we first discuss the approach of Glasserman and Li [15], who propose selecting $\boldsymbol{\mu}$ by approximating the mode of the zero-variance distribution. Next, we discuss the AIS method, which uses a Robbins–Monro approximation scheme, as proposed by Egloff, Leippold, et al. [12] and Egloff and Leippold [11].

4.2.1. Approximating the Zero-Variance Mode

The variance of the tail probability estimator under IS can be expressed as a decomposition by the *law of total variance*:

$$\mathrm{Var}\left(\tilde{P}_{x,n}\right) = \mathbb{E}\left[\mathrm{Var}(\tilde{P}_{x,n} \mid \mathbf{Y})\right] + \mathrm{Var}\left(\mathbb{E}[\tilde{P}_{x,n} \mid \mathbf{Y}]\right). \tag{4.5}$$

Using Bernoulli tilting conditional on $\mathbf{Y}$ minimizes the first term on the right-hand side. The variance of $\mathbb{E}\left[\tilde{P}_{x,n} \mid \mathbf{Y}\right]$ is computed by integrating over all possible realizations of $\mathbf{Y}$. Therefore, its zero-variance density function to draw $\mathbf{Y}$ from is proportional to the mapping

$$\mathbf{y} \mapsto \mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})\boldsymbol{\phi}(\mathbf{y}) \propto \mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})e^{-\|\mathbf{y}\|^{\mathbf{2}}/\mathbf{2}}. \tag{4.6}$$

The normalization constant, required to turn this into a valid density, is unknown. To address this, Glasserman and Li [15] suggest using a multivariate normal whose mean coincides with the mode from the zero-variance density. In other words, the mean of a multivariate normal distribution is selected such that it is located at the maximum of (4.6). This leads to the following optimization problem:

$$\max_{\mathbf{y}} \mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})e^{-\|\mathbf{y}\|^{\mathbf{2}}/\mathbf{2}}. \tag{4.7}$$

Since solving this optimization is generally difficult, Glasserman and Li [15] proposes several approximation methods. One of those is the *constant approximation*.

Constant Approximation

To approximate the conditional probability $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})$, first, L is replaced by its conditional expectation $\mathbb{E}[L \mid \mathbf{Y} = \mathbf{y}]$. Since this is constant when conditioned on $\mathbf{Y}$, the probability of exceeding x is equal to 1 or 0, i.e., $\mathbf{1}\{\mathbb{E}[L \mid \mathbf{Y} = \mathbf{y}] > x\}$. Thus, the optimization in (4.7) simplifies to maximizing $\exp(-\|\mathbf{y}\|^2/2)$, subject to the condition that the conditional expectation is required to be greater than x, i.e.,

$$\min\left\{\|\mathbf{y}\|^2 : \mathbb{E}[L \mid \mathbf{Y} = \mathbf{y}] > x\right\}, \tag{4.8}$$

where

$$\mathbb{E}[L \mid \mathbf{Y} = \mathbf{y}] = \sum_{k=1}^{m} \ell_k \Phi\left(\frac{\Phi^{-1}(p_k) - \mathbf{r}_k\mathbf{y}}{b_k}\right). \tag{4.9}$$

Since the objective function $\|\mathbf{y}\|^2$ is convex and differentiable, and $\mathbb{E}[L \mid \mathbf{Y} = \mathbf{y}]$ monotonically increases as $\mathbf{y}$ shifts in the negative direction, the optimization can be efficiently solved with a gradient-based optimization algorithm. For this, we use a Sequential Least Squares Quadratic Programming (SLSQP) solver, see Appendix B. The solution to (4.8) shifts the expected loss towards the level of interest x.

However, shifting the expected loss closer to a specified quantile does not necessarily increase the frequency of observations in the neighbourhood of the target quantile. Moreover, Glasserman, Kang, et al. [14] (Section 4) noted a risk in approximations of the zero-variance distribution. In some scenarios, it could increase the variance of $\tilde{P}_{x,n}$. This effect arises because of new "important" rare events under the IS proposal. These events contribute to the estimate, but were actually more likely under the original distribution. We will demonstrate this in the experiment in Section 4.5.

4.2.2. Algorithm for IS on Common Factors

The procedure to generate loss samples and estimate $\tilde{q}_{\alpha,n}$ with common factors drawn from a shifted multivariate normal is outlined in Algorithm 2. It returns $\tilde{q}_{\alpha,n}$ estimated via the mapping Π_α as defined in (3.17). We refer to this algorithm as the shifted Mean IS (M-IS) algorithm.

Algorithm 2: M-IS

Input : Number of replications n; Probability level α; Shift $\boldsymbol{\mu}$; Portfolio data $\{\ell_k, p_k, \mathbf{r}_k, b_k\}_{k=1}^m$
Output : Estimated quantile $\tilde{q}_{\alpha,n}$

for $i = 1$ **to** n **do**
  Sample $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_k)$, $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_m)$
  for $k = 1$ **to** m **do**
    $X_k \leftarrow \mathbf{r}_k^\top \mathbf{Y} + b_k Z_k, \quad D_k \leftarrow \mathbf{1}\{X_k < \Phi^{-1}(p_k)\}$
  Compute $L_i = \sum_{k=1}^m \ell_k D_k, \quad W_i = \exp\left(-\boldsymbol{\mu}^\top \mathbf{Y} + \frac{1}{2}\|\boldsymbol{\mu}\|^2\right)$
Compute $\tilde{q}_{\alpha,n} = \Pi_\alpha\left(\{(L_i, W_i)\}_{i=1}^n\right)$
return $\tilde{q}_{\alpha,n}$

4.3. Iteratively Updating the Shift Vector via AIS

In this section, we discuss a method that iteratively updates the shift during a simulation to minimize variance. We introduced the basic idea of AIS in Subsection 3.1.2, where we mentioned the existence of a black box function $\mathcal{I}$ that still needed to be defined. To this end, we adopt the procedure from Egloff, Leippold, et al. [12] and Egloff and Leippold [11], who propose using a Robbins-Monro approximation scheme.

For shorthand, the second moment of the tail probability at loss level x is denoted as

$$m_x(\boldsymbol{\mu}) := \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbf{1}\{L > x\}\, w_{\boldsymbol{\mu}}(\mathbf{Y})^2\right], \tag{4.10}$$

where $\mathbb{E}_{\boldsymbol{\mu}}$ is the expectation under the shifted common factors $\mathbf{Y}$, and

$$w_{\boldsymbol{\mu}}(\mathbf{y}) := \exp\left(-\mathbf{y}^\top \boldsymbol{\mu} + \frac{\|\boldsymbol{\mu}\|^2}{2}\right) \tag{4.11}$$

the IS weight. We denote the unique minimizer of (4.10) as

$$\boldsymbol{\mu}^* = \arg\min_{\boldsymbol{\mu}} m_x(\boldsymbol{\mu}).$$

The proof of uniqueness is provided later, in subsection 4.3.2.

The Robbins-Monro scheme is a stochastic approximation used to update $\boldsymbol{\mu}$ towards a zero of the gradient $\nabla_{\boldsymbol{\mu}} m_x(\boldsymbol{\mu})$.

4.3.1. Robbins-Monro Approximation

The Robbins-Monro approximation reminds us of standard optimization- or machine learning methods. Those minimize a measured error by computing its derivative with respect to the parameter of interest. By updating that parameter in the negative direction of the gradient, the algorithm, hopefully, gets a smaller error in the next steps. In essence, the Robbins-Monro approximation does the same, but now variance is the objective to minimize, and we have a stochastic gradient that needs to be estimated. The procedure is as follows: During the nth iteration, a loss sample L_n is generated using $\mathbf{Y}_n$, drawn from $\mathcal{N}(\boldsymbol{\mu}_n, \mathbf{I}_d)$. The shift $\boldsymbol{\mu}_n$ is then updated via the rule defined by

$$\boldsymbol{\mu}_{n+1} = \boldsymbol{\mu}_n + \gamma_n H(\mathbf{Y}_n, \boldsymbol{\mu}_n, x), \tag{4.12}$$

where $H(\mathbf{Y}, \boldsymbol{\mu}, x)$ is a stochastic gradient estimate, given by

$$H(\mathbf{Y}, \boldsymbol{\mu}, x) = \mathbf{1}\{L > x\}\, w_{\boldsymbol{\mu}}(\mathbf{Y})^2 (\mathbf{Y} - \boldsymbol{\mu}), \tag{4.13}$$

which is determined by

$$\nabla_{\boldsymbol{\mu}} m_x(\boldsymbol{\mu}) = -\mathbb{E}_{\boldsymbol{\mu}}\left[H(\mathbf{Y}, \boldsymbol{\mu}, x)\right]. \tag{4.14}$$

The interchange of gradient and integral is justified shortly in Subsection 4.3.2. We refer to x as the gradient threshold, since H is non-zero in the event $\{L > x\}$. Ideally, $\boldsymbol{\mu}^*$ is tailored towards an x close to the target quantile. In Subsection 6.2.4 we test AIS with a gradient threshold far from the target quantile.

Moreover, the step sizes γ_n must satisfy the Robbins-Monro conditions:

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty, \tag{4.15}$$

The first condition ensures that the step size is big enough such that it allows the optimal $\boldsymbol{\mu}^*$ to be reached. The second condition ensures the step size is small enough such that random noise is controlled, necessary for eventual convergence to $\boldsymbol{\mu}^*$.

4.3.2. A Unique Optimum

To show that $\boldsymbol{\mu}^*$ is a unique minimizer, we discuss the results from Egloff and Leippold [11] (section 5.2). First, the gradient expression in (4.13) is calculated by taking the derivative inside the expectation using the likelihood ratio method, using that the IS weights $w_{\boldsymbol{\mu}}(\mathbf{y})$ can be bounded. To see this, we can use Young's inequality to obtain

$$-\mathbf{y}^\top \boldsymbol{\mu} \leq \frac{1}{2\varepsilon}\|\mathbf{y}\|^2 + \frac{\varepsilon}{2}\|\boldsymbol{\mu}\|^2, \quad \text{for any } \varepsilon > 0.$$

Substituting into the IS weight expression gives

$$\begin{aligned} w_{\boldsymbol{\mu}}(\mathbf{y}) &= \exp\left(-\mathbf{y}^\top \boldsymbol{\mu} + \frac{\|\boldsymbol{\mu}\|^2}{2}\right) \\ &\leq \exp\left(\frac{1}{2\varepsilon}\|\mathbf{y}\|^2\right) \exp\left(\left(1 + \frac{\varepsilon}{2}\right)\|\boldsymbol{\mu}\|^2\right), \quad \text{for any } \varepsilon > 0. \end{aligned} \tag{4.16}$$

Hence, the second moment is finite:

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbf{1}\{L > x\}\, w_{\boldsymbol{\mu}}(\mathbf{Y})^2\right] \leq \exp\left((2+\varepsilon)\,\|\boldsymbol{\mu}\|^2\right) \cdot \mathbb{E}_{\boldsymbol{\mu}}\left[\exp\left(\frac{1}{\varepsilon}\|\mathbf{y}\|^2\right)\right] < \infty,$$

where the bound in (4.16) and the fact that $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_d)$ has finite exponential moments is used. This justifies the interchange between differentiation and integration in (4.14). Furthermore, the chain rule can be used to obtain $H(\mathbf{Y}, \boldsymbol{\mu}, x)$:

$$\nabla_{\boldsymbol{\mu}} w_{\boldsymbol{\mu}}^2(\mathbf{Y}) = w_{\boldsymbol{\mu}}^2(\mathbf{Y})(\boldsymbol{\mu} - \mathbf{Y}).$$

Moreover, the Hessian of $m_x(\boldsymbol{\mu})$, given by

$$\nabla_{\boldsymbol{\mu}}^2 m_x(\boldsymbol{\mu}) = \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbf{1}\{L > x\} \cdot w_{\boldsymbol{\mu}}^2(\mathbf{Y})\left(\mathbf{I}_d + (\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top\right)\right], \tag{4.17}$$

is positive definite whenever $\mathbb{P}(L > x) > 0$. In this case, $m_x(\boldsymbol{\mu}) \to \infty$ as $\|\boldsymbol{\mu}\| \to \infty$, and together with strict convexity (from the positive definiteness of the Hessian), this guarantees that $m_x(\boldsymbol{\mu})$ admits a unique global minimizer $\boldsymbol{\mu}^*$.

4.3.3. Adaptive Truncation

However, it is difficult to determine the region that includes the optimal shift $\boldsymbol{\mu}^*$ before we start the AIS scheme. If we don't have prior knowledge on where $\boldsymbol{\mu}^*$, we don't have prior boundedness for the Robbins-Monro algorithm, which is necessary to guarantee convergence. The shift vector $\boldsymbol{\mu}_n$ can grow too large or fluctuate uncontrollably due to the growth of the stochastic gradient $H(\mathbf{Y}, \boldsymbol{\mu}_n, x)$. For this, we adopt a truncation scheme from Egloff, Leippold, et al. [12] and summarize their reasoning below.

The method corrects $\boldsymbol{\mu}_n$ by mapping it back to a chosen point $\boldsymbol{\mu}_0$ if jumped too far, out of the allowed region. The truncated update rule is defined as

$$\boldsymbol{\mu}_{n+1} = \left(\boldsymbol{\mu}_n + \gamma_n H(\mathbf{Y}_n, \boldsymbol{\mu}_n, x)\right)\mathbf{1}_{J_n} + \boldsymbol{\mu}_0 \mathbf{1}_{J_n^c}, \tag{4.18}$$

where J_n^c is the complement of the event J_n, and the event J_n is defined as

$$J_n := \left\{\|\boldsymbol{\mu}_n - \gamma_{n+1} H(\mathbf{Y}_{n+1}, \boldsymbol{\mu}_n, x)\| \leq R_{\tau_n}\right\}. \tag{4.19}$$

Here, $(R_n)_{n\geq 1}$ is a sequence of positive numbers diverging to infinity such that all possible parameters can be reached, and the sequence $(\tau_n)_{n\geq 1}$ is defined recursively by

$$\tau_0 := 0, \quad \tau_n := \sum_{i=0}^{n} \mathbf{1}_{J_n^c}.$$

The count τ_n keeps track of how often the update falls outside the acceptable bound, controlling the growth of the truncation radius. But to achieve convergence towards $\boldsymbol{\mu}^*$, the number of truncations should be finite. The convergence proof for the truncated Robbins-Monro algorithm (4.18) relies on the fact that

$$\limsup_{n\to\infty} \left\| \gamma_n \sum_{i=1}^{n} H(\mathbf{Y}_i, \boldsymbol{\mu}_i, x) - \mathbb{E}[H(\mathbf{Y}_i, \boldsymbol{\mu}_i, x) \mid \boldsymbol{\mu}_i] \right\| \leq \epsilon \tag{4.20}$$

almost surely, then the truncation is only invoked a finite number of times. This means that the weighted average of the approximated gradient converges to the true gradient. For this proof see Arouna [2] (Theorem 3), which uses martingale convergence theorems to prove this. Furthermore, if (4.20) holds for $\epsilon = 0$, then the algorithm (4.18) converges to the unique optimizer $\boldsymbol{\mu}^*$. For a full proof, see Chen, Guo, et al. [7] and Chen and White [8].

4.3.4. Choosing a Step Size

We now specify our choice for the sequence $(\gamma_n)_{n\geq 1}$. If γ_n decays too quickly, the updates in (4.12) become negligible, limiting adaptation. Egloff, Leippold, et al. [12, page 14] considers a step size that decreases only when the loss L_n exceeds a threshold, given by

$$\gamma_n = \frac{\eta}{\beta + \Delta \sum_{i=1}^{n} \mathbf{1}\{L_i > x\}}, \tag{4.21}$$

where $\eta, \beta, \Delta > 0$ are tuning parameters. Since $\mathbb{P}(L > x) > 0$, we have

$$\sum_{i=1}^{n} \mathbf{1}\{L_i > x\} \asymp n \cdot \mathbb{P}(L > x) \quad \text{almost surely as } n \to \infty, \tag{4.22}$$

where $\asymp$ means that the number of exceedances grows asymptotically at the same rate as n. Consequently, the sequence $(\gamma_n)_{n\geq 1}$ behaves like $\gamma_n \asymp n^{-1}$ almost surely, which satisfies the Robbins–Monro conditions in (4.15).

4.3.5. Convergence under AIS

This Robbins–Monro based scheme also builds on the theoretical foundations laid out by Arouna [3], who proves a SLLN and a CLT (Theorem 1 and 2) for the resulting estimator sequence under AIS, using classical martingale convergence theorems to establish consistency. In particular, if the shift sequence $(\boldsymbol{\mu}_n)_{n\geq 1}$ converges almost surely to the optimal shift $\boldsymbol{\mu}^*$, then the AIS estimator

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{L_i > x\}\, w_{\boldsymbol{\mu}_i}(\mathbf{Y}_i)$$

converges almost surely to the true tail probability P_x.

In addition to strong consistency, the AIS estimator also satisfies a CLT, as shown in Arouna [3]. Let $\tilde{P}_{x,n}$ denote the AIS estimator of P_x at iteration n:

$$\tilde{P}_{x,n} := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{L_i > x\}\, w_{\boldsymbol{\mu}_i}(\mathbf{Y}_i), \tag{4.23}$$

where each $\mathbf{Y}_i$ is drawn from $\mathcal{N}(\boldsymbol{\mu}_i, \mathbf{I}_d)$ and the IS weights are as defined in (4.11).

The CLT holds under the additional conditions that the second moment $m_x(\boldsymbol{\mu})$ is continuous at $\boldsymbol{\mu}^*$, and that there exists $a > 1$ such that the higher moment

$$\mathbb{E}_{\boldsymbol{\mu}} \left[\left(\mathbf{1}\{L > x\}\, w_{\boldsymbol{\mu}}(\mathbf{Y}) \right)^{2a} \right] < \infty.$$

These conditions are satisfied in our setting with analogous computations as in Subsection 4.3.2, where bounds on the IS weight are established by using Young's inequality and the finiteness of exponential moments under the shifted Gaussian proposal.

Thus, the estimator satisfies

$$\sqrt{n}\left(\tilde{P}_{x,n} - P_x\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{\psi}^2\right), \quad \text{as } n \to \infty. \tag{4.24}$$

Here, the asymptotic variance $\tilde{\psi}$ and its natural estimator $\tilde{\psi}_n$ are identical to the IS case and are given by (3.14) and (3.15), respectively.

Furthermore, the corresponding AIS estimate for q_α is identical to (3.16). Showing the consistency of $\tilde{q}_{\alpha,n}$ is more complex as this returns to the problem of non-uniqueness and therefore, oscillates between two quantile estimates. To this end, Egloff and Leippold [11] (Theorem 4.2 and Theorem 4.2) establishes almost sure convergence under suitable conditions (Assumption 4.1 and Assumption 4.2), by modifying the empirical distribution. Under Gaussian distributions, they show (Section 5.2) that these conditions hold.

4.3.6. Algorithm for AIS on Common Factors

In Algorithm 3 the method for using AIS is outlined. We want to highlight that it is possible to use an initial shift vector $\boldsymbol{\mu}_1$, such as the constant approximation method.

Algorithm 3: AIS

Input : Number of replications n; Threshold x; Probability level α; Initial shift $\boldsymbol{\mu}_1$, Step size parameters η, β, Δ; Truncation point $\boldsymbol{\mu}_0$; Region bound sequence $(R_n)_{n\geq 1}$; Portfolio data $\{\ell_k, p_k, \mathbf{r}_k, b_k\}_{k=1}^m$

Output : Estimated quantile $\tilde{q}_\alpha$

Set $\boldsymbol{\mu} = \boldsymbol{\mu}_1$
Set $\tau = 0$
for $i = 1$ **to** n **do**
  Sample $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_k), \quad \mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_m)$
  for $k = 1$ **to** m **do**
    Set $X_k \leftarrow \mathbf{r}_k^\top \mathbf{Y} + b_k Z_k, \quad D_k \leftarrow \mathbf{1}\{X_k < \Phi^{-1}(p_k)\}$
  Compute $L_i = \sum_{k=1}^m \ell_k D_k, \quad \text{and} \quad W_i = \exp\left(-\boldsymbol{\mu}_i^\top \mathbf{Y} + \frac{1}{2}\|\boldsymbol{\mu}_i\|^2\right)$
  Set $H \leftarrow \mathbf{1}\{L_i > x\} W_i^2 (\mathbf{Y} - \boldsymbol{\mu})$
  if $\|\boldsymbol{\mu} + \gamma H\| \geq R_\tau$ **then**
    Reset $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}_0$
    Update $\tau \leftarrow \tau + 1$
  else
    Update $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} + \gamma H$
  Update $\gamma \leftarrow \frac{\eta}{\beta + \Delta \sum_{j=1}^i \mathbf{1}\{L_j > x\}}$
Compute $\tilde{q}_\alpha = \Pi_\alpha\left(\{(L_i, W_i)\}_{i=1}^n\right)$
return $\tilde{q}_\alpha$

4.4. Approximating the Zero-Variance Distribution with the GPB

To compute the zero-variance distribution for simple portfolio cases, we introduce a method that uses the Generalized Poisson-Binomial (GPB) distribution to efficiently recover the function in (3.7), which is proportional to the zero-variance distribution.

To find the zero-variance distribution in (4.6) we need to evaluate $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})$ for different $\mathbf{y}$. We observe that, conditional on $\mathbf{Y}$, the loss L follows a GPB distribution, since the GPB is determined by a sum of scaled Bernoulli variables with differing success probabilities. Zhang, Hong, et al. [33] present a method for computing the GPB distribution using the *inverse discrete Fourier transform* (IDFT), which provides us an alternative approach for evaluating the conditional loss distribution.

Evaluating Conditional Tail Probability using the IDFT

Let us consider a special portfolio case with m obligors, where the kth obligor has default probability p_k and exposure ℓ_k, modelled through a factor model described in (2.5). For now, we assume that all

exposures ℓ_k are integer values: To apply the IDFT in its simplest form, the data points must lie in a uniform grid. Using grid size steps with integer values (or equal to the greatest common divisor of the sample points), ensures uniform spacing. While this may not always hold for real-world applications, we use the method described below only as a computational tool on a special case in subsection 4.5, where we only consider integer exposures.

To compute $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})$, we start by using the characteristic function of L conditioned on $\mathbf{Y}$, which can be expressed as a Fourier series. Then, after a few algebraic substitutions, we apply the IDFT to recover the probability mass function of L, enabling us to calculate $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})$.

Conditioned on $\mathbf{Y} = \mathbf{y}$, the characteristic function of L is given by

$$\begin{aligned}\varphi(t, \mathbf{y}) &= \mathbb{E}\left[\exp(\mathbf{i}tL) \mid \mathbf{Y} = \mathbf{y}\right] \\ &= \mathbb{E}\left[\prod_{k=1}^{m} \exp(\mathbf{i}tD_k) \mid \mathbf{Y} = \mathbf{y}\right] \\ &= \prod_{k=1}^{m} \left[(1 - p_k(\mathbf{y})) + p_k(\mathbf{y}) \exp(\mathbf{i}t\ell_k)\right],\end{aligned}$$

where $p(\mathbf{y})$ is the conditional probability defined in (4.1). The last equality follows from the conditional independence of the random variables D_k. Using the definition of the characteristic function, the left-hand side can be expressed as a Fourier series, leading to the equality

$$\sum_{\xi=0}^{\ell} f(\xi, \mathbf{y}) \exp(\mathbf{i}t\xi) = \prod_{k=1}^{m} \left[(1 - p_k(\mathbf{y})) + p_k(\mathbf{y}) \exp(\mathbf{i}t\ell_k)\right], \tag{4.25}$$

where $f(\xi, \mathbf{y})$ denotes the probability mass function of L conditional on $\mathbf{Y} = \mathbf{y}$. To numerically compute $f(\xi, \mathbf{y})$, we can use the IDFT. The IDFT allows us to recover the entire probability mass function $f(\xi, \mathbf{y})$ for all $\xi = 0, 1, \ldots, \ell$ by inverting the characteristic function.

Let $n = \ell + 1$, and define the fundamental frequency as $\omega = \frac{2\pi}{n}$. For $l = 0, \ldots, n-1$, we substitute

$$x_l(\mathbf{y}) := \varphi(\omega l, \mathbf{y}) = \prod_{k=1}^{m} \left[(1 - p_k(\mathbf{y})) + p_k(\mathbf{y}) \exp(\mathbf{i}\omega l\ell_k)\right].$$

in (4.25). Next, by applying the IDFT on (4.25), the conditional probability mass function is recovered as

$$f(\xi, \mathbf{y}) = \frac{1}{n} \sum_{l=0}^{n-1} \exp(-\mathbf{i}\omega\xi l)\, x_l(\mathbf{y}), \quad \text{for } \xi = 0, \ldots, \ell.$$

Here, $f(\xi, \mathbf{y})$ is equal to $\mathbb{P}(L = \xi \mid \mathbf{Y} = \mathbf{y})$. The conditional distribution can then be expressed as

$$\mathbb{P}(L \leq x \mid \mathbf{Y} = \mathbf{y}) = \sum_{\xi=0}^{\lfloor x \rfloor} f(\xi, \mathbf{y}). \tag{4.26}$$

Deriving the Zero-Variance Denisty

The method described allows us to calculate the conditional probability efficiently, since the computation has complexity $O(mn + \log m)$. This includes n evaluations of $x_\ell(\mathbf{y})$, and the fast Fourier transform enables computations in $O(\log m)$ time. Furthermore, it is possible to evaluate (4.6) over a chosen grid and by normalizing, we can recover the zero-variance distribution. We provide the algorithm in the Appendix B.

4.5. Comparing M-IS and AIS on a Special Portfolio Scenario

In this experiment, we compare the performance of M-IS and AIS on a special case adopted from Glasserman, Kang, et al. [14] (Section 4) with a portfolio of 1 000 obligors. Each obligor has a default probability $p_k = 0.05$ and an exposure $\ell_k = 1$, modelled through a two-factor model described by

$$\begin{aligned} X_k &= 0.7Y_1 + \sqrt{0.51}\, Z_k, && \text{for } 1 \leq k \leq 500 \\ X_k &= 0.65Y_2 + \sqrt{0.577}\, Z_k, && \text{for } 500 < k \leq 1\,000 \end{aligned}$$

Our goal is to estimate the tail probability P_x at loss level $x = 300$ and the quantile q_α at level $\alpha = 0.999$.

Before we find the optimal mode of the zero-variance density to shift the mean for M-IS, we use the GPB distribution and compute the zero-variance density as described in Section 4.4. We provide the code for this in the Appendix B. The result is shown in Figure 4.3, where we can see the maximum of the density is shifted more negatively for y_1 than for y_2. This is due to the small difference in factor loading with Y_1 and Y_2, since this difference causes $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y})$ to increase more rapidly as y_1 becomes more negative compared to y_2. Therefore, the maximum of $\mathbb{P}(L > x \mid \mathbf{Y} = \mathbf{y}) \exp(-\|\mathbf{y}\|^2/2)$ is shifted more negatively in the y_1 direction.

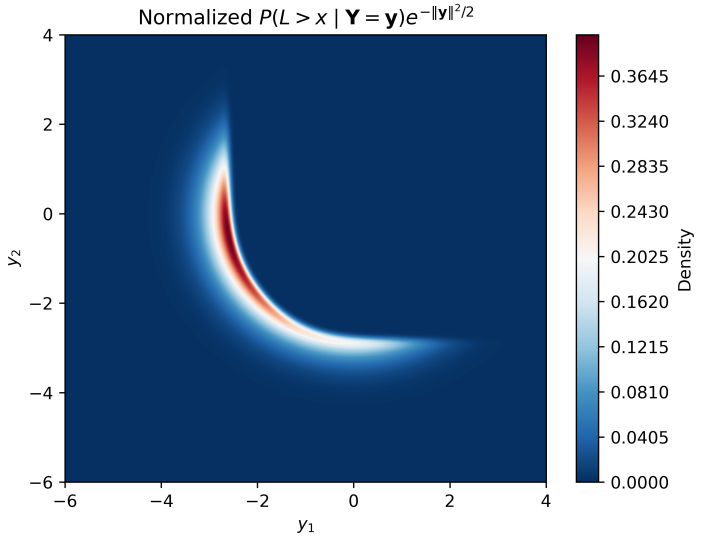


Figure 4.3: Normalized function proportional to the zero-variance density over the shift space $\mathbf{y} = (y_1, y_2)^\top$, calculated using the GPB. The maximum is shifted more negatively in the y_1 direction due to the difference in factor loading.

For the M-IS method, we shift $\boldsymbol{\mu}_{\text{M-IS}}$ towards this maximum, given by

$$\boldsymbol{\mu}_{\text{M-IS}} := (-2.605,\ -0.450)^\top.$$

Note that this is close to the shift used by Glasserman, Kang, et al. [14] (up to the sign as they use mirror default conditions).

For the AIS method, we set the initial shift for $\boldsymbol{\mu}_{\text{AIS}}$ equal to the zero vector. The following parameters are used for the step size $(\gamma_n)_{n\geq 1}$:

$$\eta = 20, \quad \beta = 100, \quad \Delta = 2.$$

We choose these parameters based on trial and error, where we aim for stable convergence of $\boldsymbol{\mu}_{\text{AIS}}$ (these parameter settings led to eventual stable convergence, see Figure 4.7).

We perform a single run of 100 000 replications, both with the M-IS and the AIS method. Their procedures are outlined in Algorithm 2 and in Algorithm 3. Every 100 replications, we record the quantile estimate $\tilde{q}_{\alpha,r}$ defined in (3.16), the tail probability estimate $\tilde{P}_{x,r}$ defined in (3.13) (for both M-IS and AIS), and use the consistent estimator $\tilde{\psi}_r^2$ defined in (3.15) for the sample variance of $\tilde{P}_{x,r}$. Here, r denotes the number of replications. We compare these estimates to estimates from a CMC run. We want to highlight that $\tilde{\psi}_r^2$ cannot be directly used as a measure for the variance of the quantile estimate. As we discussed in Section 3.4, to construct an asymptotic confidence interval, continuity conditions need to be satisfied. However, Theorem 4 shows that, in the continuous case, higher $\tilde{\psi}_r^2$ implies higher variance for the quantile estimate. In this experiment, we keep our analyses only to the sample variance of the tail probability and rather use it as an indication of the effectiveness of our estimates.

Figure 4.4 illustrates how these estimates evolve as the number of replications increases for the M-IS method, and Figure 4.5 illustrates this for the AIS method. For the M-IS method, we see that $\tilde{q}_{\alpha,r}$ and $\tilde{P}_{x,r}$ make large jumps, whereas the CMC estimates seem to converge. These sudden jumps also appear even more significantly for $\tilde{\psi}_r^2$, shown in the third figure. The final estimate for the sample

variance is equal to $\tilde{\psi}_r^2 = 0.142$, which is much higher than the CMC sample variance. Hence, M-IS is very ineffective in this case.

For the AIS method, we see that the estimates $\tilde{q}_{\alpha,r}$ and $\tilde{P}_{x,r}$ seem to converge faster than the CMC estimates, and we observe a decrease in the sample variance. Its final value, equal to $\tilde{\psi}_r^2 = 1.95 \times 10^{-3}$, further demonstrates the superior efficiency of the AIS approach.

We notice that, although the loss level $x = 300$ and the $\alpha = 0.99$ do not correspond to the exact same quantile ($P_x \approx 0.9885$), the way that $\tilde{q}_{\alpha,r}$ and $\tilde{P}_{x,r}$ vary over different r is very similar.

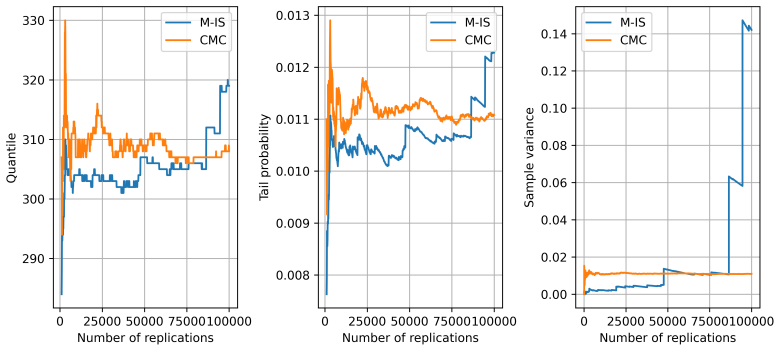


Figure 4.4: Comparing M-IS and CMC estimates in a single run of 100 000 replications. The estimated IS quantile ($\tilde{q}_{\alpha,r}$) is shown in the first figure, the estimated IS tail probability ($\tilde{P}_{x,r}$) in the second figure and the sample variance ($\tilde{\psi}_r^2$) is shown in the third figure, all as a function of the number of replications. We observe large variations of the estimates. Furthermore, the final value for $\tilde{\psi}_r^2$ is equal to 0.142, indicating very inefficient sampling.

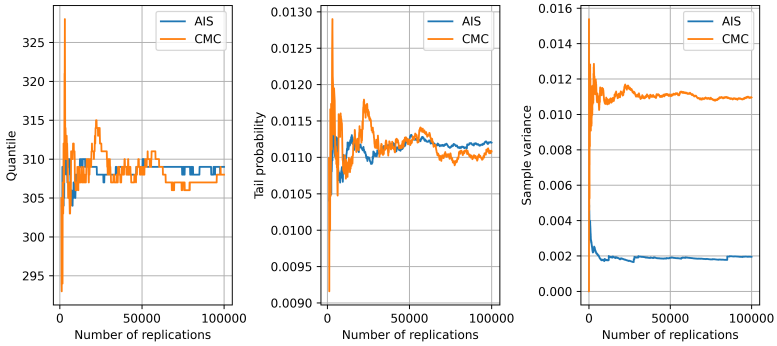
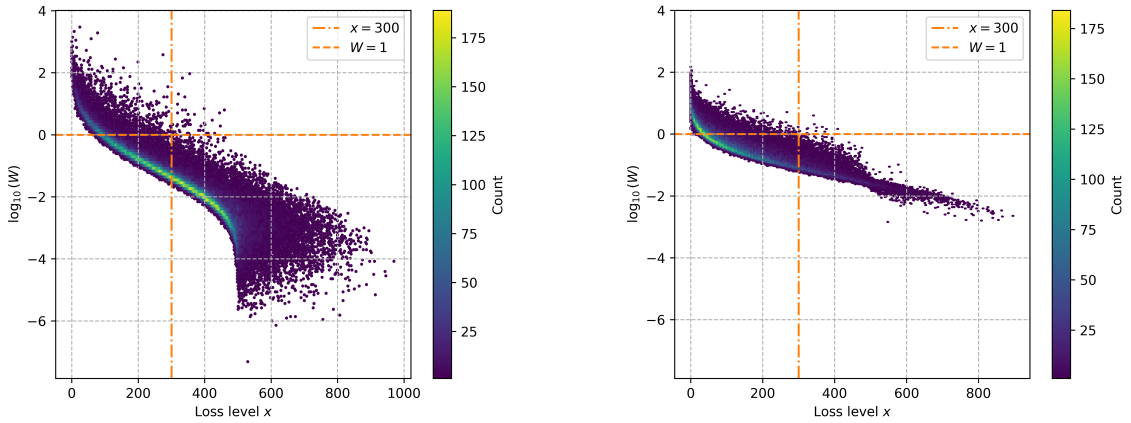


Figure 4.5: Comparing AIS and CMC estimates in a single run of 100 000 replications. The estimated IS quantile ($\tilde{q}_{\alpha,r}$) is shown in the first figure, the estimated IS tail probability ($\tilde{P}_{x,r}$) in the second figure and the sample variance ($\tilde{\psi}_r^2$) is shown in the third figure, all as a function of the number of replications. The AIS estimates seem to converge faster than the CMC estimates. Furthermore, the final value for $\tilde{\psi}_r^2$ is equal to 1.95×10^{-3}, indicating efficient sampling.

To explain the difference in performance of M-IS and AIS, we compare the IS weights and corresponding loss samples. In Figure 4.6, we show the losses against IS weights in a scatter plot. We plotted the horizontal line at the IS weight equal to 1. We plotted the vertical line $x = 300$ to clearly indicate the loss samples contributing to $\tilde{P}_{x,r}$. In the M-IS case, we observe that some loss samples above this threshold have very large IS weights (up to nearly 10^2). These large IS weights arise because the corresponding rare events are more probable under the original distribution than under the IS distribution. If these large IS weights correspond to losses above the threshold x, this leads to big jumps in the empirical distribution function estimated via (3.13), since the empirical distribution sums over IS weights for samples with losses greater than x. In the case of large IS weights, this directly increases the variation in $\tilde{P}_{x,r}$ and in $\tilde{q}_{\alpha,n}$. In the AIS case, we observe less skewed values for the IS weights. Moreover, we observe lower IS weights corresponding to the losses greater than $x = 300$, making $\tilde{P}_{x,r}$ and $\tilde{q}_{\alpha,n}$ more stable.

We show the evolution of the components of $\boldsymbol{\mu}_{\text{AIS}}$ in Figure 4.7 for each iteration (replication). The components of $\boldsymbol{\mu}$ seem to converge, with the final shift-vector equal to $\boldsymbol{\mu}_{\text{AIS}} = (-1.02, -0.88)^\top$. This balances the shift better between the factors. This is quite different compared to $\boldsymbol{\mu}_{\text{M-IS}}$, and together with the estimated sample variance for both methods, this shows that setting the mean of the multivariate normal equal to the mode of the zero-variance distribution is suboptimal.



(a) M-IS **(b)** AIS

Figure 4.6: Bin-scatter plots of IS weights against losses for a single run of 100 000 replications, comparing M-IS and AI. For the M-IS case we observe very skewed IS weights, with large weights above threshold $x = 300$. This results in less precise estimates for $\tilde{P}_x$.

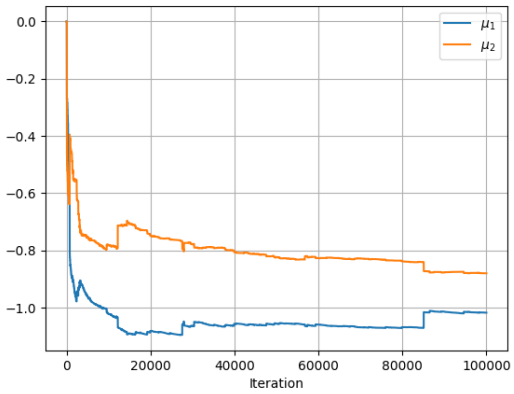
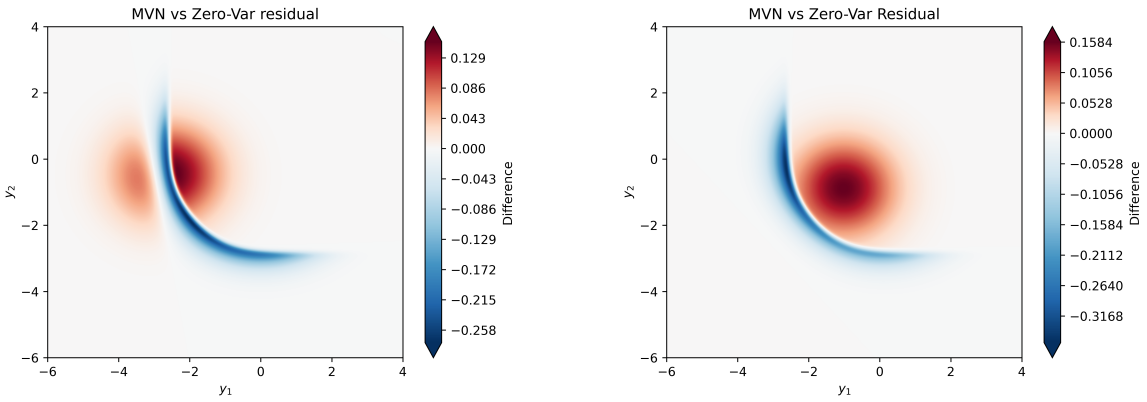


Figure 4.7: Evolution of the components of $\boldsymbol{\mu}_{\text{AIS}}$ over iterations, starting at $(0, 0)$ and converging to $(-1.02, -0.88)^\top$. This is a more balanced shift compared to the M-IS case.

To see the sub-optimality from M-IS more clearly, we return to the zero-variance distribution. In Figure 4.8, we illustrated the pointwise residuals between the normalized zero-variance and the multivariate normals (the distributions from which we sample the common factors) with their means located at $(-2.61, -0.45)^\top$, and $(-1.02, -0.88)^\top$ for the M-IS and the AIS (the shift corresponding to the final replication) case. In both cases, we observe that the normalized zero-variance density differs significantly from the multivariate normal. This does not immediately mean that it will increase variance, but it does get tricky when the difference is serious in the regions of interest. In Figure 4.8a we observe the multivariate normal extends beyond the zero variance density in the negative direction. Sampling from that region would generate large losses, very likely above 300, contributing to the estimates of $\tilde{P}_x$. These are the important regions. However, other areas can also give rise to large losses. In the M-IS case, only one subregion of interest is targeted, which makes events in other important regions even rarer, for example a sample (Y_1, Y_2) from $[0, \infty) \times [-3, -\infty)$. This would almost certainly create a loss above level 300, but is an extremely rare event under the new proposal, increasing variance. The AIS method seems to account for this and balances the shift better between kinds of concentration and prevents the emergence of new, rare yet important events.

Glasserman, Kang, et al. [14] proposes a clever method of using mixtures of IS proposals to address this problem. This method divides the portfolio into subgroups of the same type. Each subgroup that has a total exposure that could potentially exceed the loss threshold by itself is assigned an IS proposal.



(a) M-IS residual

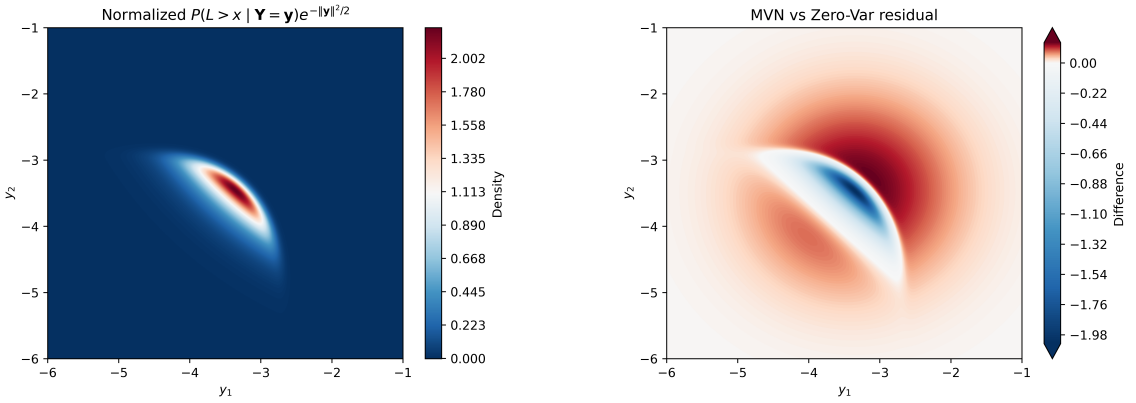
(b) AIS residual

Figure 4.8: Pointwise residuals between the normalized zero-variance density and multivariate normal densities with two shift vectors: (a) M-IS mean shift $\boldsymbol{\mu}_{\text{M-IS}} = (-2.61, -0.45)^\top$, and (b) AIS final shift $\boldsymbol{\mu}_{\text{AIS}} = (-1.02, -0.88)^\top$. In the M-IS case, only one subregion of interest is targeted, which can make events in other important regions even rarer. The AIS procedure produces a much more balanced shift.

In other words, IS is applied to the sub-portfolios with the largest impact. We choose to stick with the AIS method since our test portfolios described in Chapter 5 consist of many different types, potentially leading to the problem that no, or only some, subgroups can cross the target threshold.

High Threshold

Glasserman, Kang, et al. [14] also study the large threshold regime ($x = 800$) and observes a very low sample variance for the M-IS method. We compute the function proportional to the zero-variance distribution for this threshold and illustrate it in Figure 4.9a. We select the mean shift $\boldsymbol{\mu}_{\text{M-IS}}$ by setting it equal to the mode of the zero-variance, given by $(-3.36 - 3.46)^\top$. Looking at the residual between the zero-variance and the multivariate normal, we can observe a more balanced shift compared to the low threshold case. Now, the important region that creates losses above $x = 800$ is focused in one direction. Therefore, shifting towards the zero-variance mode will, with near certainty, avoid generating any new rare but important events and be very effective.



(a) Zero-variance distribution (high threshold)

(b) M-IS residual (high threshold)

Figure 4.9: The mode of the Zero-variance distribution is located in a far negative region for the high threshold $x = 800$. The M-IS mean shift set at its mode $\boldsymbol{\mu}_{\text{M-IS}} = (-3.36, -3.46)^\top$. Since the important region is focussed in one direction, driving the shift towards the mode of the zero-variance is likely effective.

To estimate a quantile close to that extreme loss level, we carry out an initial run and select the high level $\alpha = 1 - 5.5 \times 10^{-7}$. We perform a run with 100 000 replications, and show the results in Figure 4.10, with a final estimated sample variance of $\psi_r^2 = 2.92 \times 10^{-12}$, demonstrating very efficient sampling. In Figure 4.11, we observe very low IS weights for losses above $x = 800$, which results in very precise estimates $\tilde{P}_{x,n}$ for losses above 800. In fact, tail probabilities for losses above level 600 would be estimated with more precision compared to CMC.

To Conclude this Sub-Experiment

We observe that for the low threshold regime, the M-IS approach, in which we shifted the mean towards the mode of the zero-variance distribution, was ineffective. It shifts the common factors unevenly, while both are common factors are important drivers of losses above the chosen threshold. The AIS method

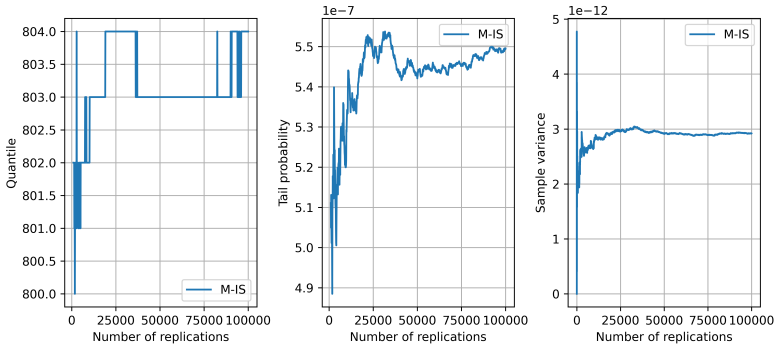


Figure 4.10: M-IS estimates for a large threshold $x = 800$ in a single run of 100 000 replications. The estimated IS quantile ($\tilde{q}_{\alpha,r}$) is shown in the first figure, the estimated IS tail probability ($\tilde{P}_{x,r}$) in the second figure and the sample variance ($\tilde{\psi}_r^2$) is shown in the third figure, all as a function of the number of replications. The final value for $\tilde{\psi}_r^2$ is equal to 2.92×10^{-12}, demonstrating super efficient sampling. Under CMC, there was no loss sample created above $x = 800$.

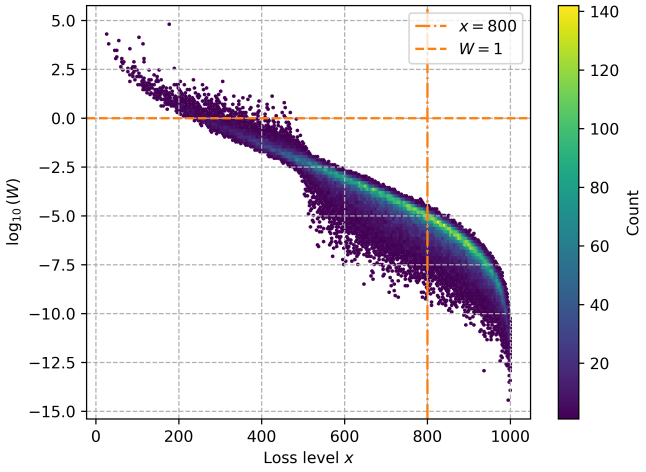


Figure 4.11: Bin-scatter plot of IS weights against losses for a single run of 100 000 replications for the M-IS case in the large threshold $x = 800$. We observe very skewed IS weights, but because the IS weights above $x = 800$ are very small, this results in very precise estimates for $\tilde{P}_x$.

seems to take this into account and balances more between the common factors in the low-threshold regime. Both common factors jointly drive large losses above $x = 800$, causing the important region to be more focused in one direction. Therefore, shifting the common factors towards the mode of the zero-variance density is effective in this case.

5

Test Portfolios

We evaluate our algorithms on three scenarios derived from a common base portfolio: (1) a large portfolio calibrated to mimic realistic dependence on the macroeconomic factor; (2) the same portfolio with dependency on the macroeconomic factor scaled down, so we can examine the effect of reduced dependence; and (3) a subset of portfolio 1 with fewer obligors. More specifically:

Scenarios

1. In the first scenario we consider a stylized portfolio of 20 000 obligors. Each obligor has a marginal default probability p_k and an exposure ℓ_k. We scaled exposures such that the macroeconomic VaR, defined in (2.9), is exactly 1 million. We assume each obligor is associated with a geographic-region and a certain sector. Table 5.2 presents the distribution of obligors over the different region-sector combinations. The table also shows the total exposure per subset of obligors, expressed as a percentage of the total exposure, alongside the average marginal default probability for each region-sector combination. The obligors are unevenly spread across region-sector combinations, each with varying levels of exposure and default probabilities. Furthermore, large exposures generally have a low probability of default, while smaller exposures usually have a higher probability of default. Figure 5.2 demonstrates this inverse relationship with a scatter-plot of p_k against ℓ_k. For each obligor, the factor loading vector $\mathbf{r}_k$, see (2.10), determines the correlation with the macroeconomic factor, the regional factors, and the sectoral factors. It follows the index ordering:

$$(M,\ A,\ B,\ C,\ D,\ S1,\ S2,\ S3,\ S4,\ S5,\ S6)^\top,$$

   where M denotes the macroeconomic factor, A, B, C, D the regions, and $S1, S2, S3, S4, S5, S6$ the sectors. The stylized region- and sector-specific factor loadings are shown in Tables 5.4 and 5.5, respectively. These are the same for all obligors within the same region or sector. To realize the negative correlation between the macroeconomic factor and the probability of default, the IRB corporate correlation factor loading in (2.7) is used. The total exposure and the macroeconomic VaR, calculated with (2.9), are shown in Table 5.1.

2. Since we discussed the impact of correlation structures in portfolios on variance reduction, we want to study the influence of changing the dependence through common factors. Therefore, in the second scenario, we use exactly the same portfolio as in Scenario 1, but with a reduced factor loading coefficient to assess the effect of a lower factor loading for the macroeconomic factor: for each kth obligor the macroeconomic factor loading is replaced by $\sqrt{\rho_k}/2$. In Table 5.1, we observe that this reduces the macroeconomic VaR compared to scenario 1.

3. To assess how performance changes when the algorithms are applied to smaller portfolios, we use a random subset of 1 000 obligors drawn from the first scenario. The distribution of obligors across region-sector combinations is shown in Table 5.3.

Quantity	Scenario 1	Scenario 2	Scenario 3
Total exposure in millions	8.197	8.197	0.388
Macroeconomic VaR in millions	1.000	0.392	0.070

Table 5.1: Total exposure and the macroeconomic VaR for each scenario. The macroeconomic VaR is calculated with (2.9).

Region	Sector	Count	Share of total exposure	Avg. default probability
A	S1	109	2.77%	2.29%
	S2	166	1.63%	3.22%
	S3	14	10.07%	0.03%
	S4	1	0.00%	1.00%
	S5	256	3.01%	2.73%
	S6	19	1.81%	2.74%
B	S1	207	7.53%	3.59%
	S2	500	7.51%	3.42%
	S3	10	10.19%	0.03%
	S4	13704	17.71%	1.08%
	S5	1043	5.37%	3.17%
	S6	124	8.14%	2.21%
C	S1	59	0.66%	4.46%
	S2	305	2.76%	3.44%
	S3	6	0.60%	0.06%
	S4	2242	4.35%	1.01%
	S5	1048	7.17%	3.66%
	S6	168	2.75%	3.03%
D	S3	3	5.93%	0.04%
	S4	16	0.02%	1.00%
All		**20 000**	**100%**	

Table 5.2: Cross table of region-sector combinations showing the number of obligors, their share of total exposure, and the average marginal default probability within each combination, corresponding to the portfolio in scenarios 1 and 2.

Region	Sector	Count	Share of total exposure	Avg. default probability
A	S1	2	1.53%	1.20%
	S2	11	3.14%	3.40%
	S3	2	4.19%	0.03%
	S5	17	0.88%	2.85%
	S6	1	0.06%	2.00%
B	S1	12	3.44%	2.20%
	S2	18	11.10%	3.67%
	S4	691	19.12%	1.11%
	S5	48	3.47%	2.75%
	S6	10	28.14%	1.74%
C	S1	4	0.32%	4.75%
	S2	14	7.63%	4.14%
	S4	105	3.57%	1.01%
	S5	54	13.08%	3.41%
	S6	9	0.25%	3.00%
D	S4	2	0.08%	1.00%
All		**1 000**	**100%**	

Table 5.3: Cross table of region-sector combinations showing the number of obligors, their share of total exposure, and the average marginal default probability within each combination, corresponding to the portfolio in scenario 3.

Region	A	B	C	D
Factor loading	0.10	0.05	0.10	0.00

Table 5.4: Regional-specific factor loadings. All obligors located in the same region share the corresponding geographic factor loading.

Sector	S1	S2	S3	S4	S5	S6
Factor loading	0.20	0.30	0.00	0.10	0.20	0.00

Table 5.5: Sector-specific factor loadings. All obligors in the same sector share the corresponding sector factor loading.

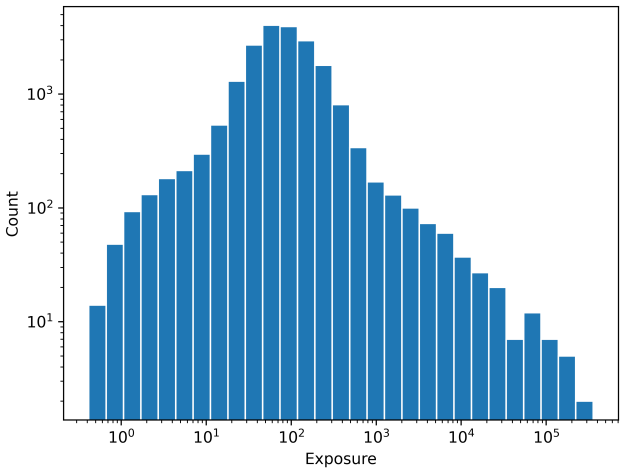


Figure 5.1: Distribution of ℓ_k from the portfolio corresponding to scenarios 1 and 2. Both the x- and y-axis are in log-scale.

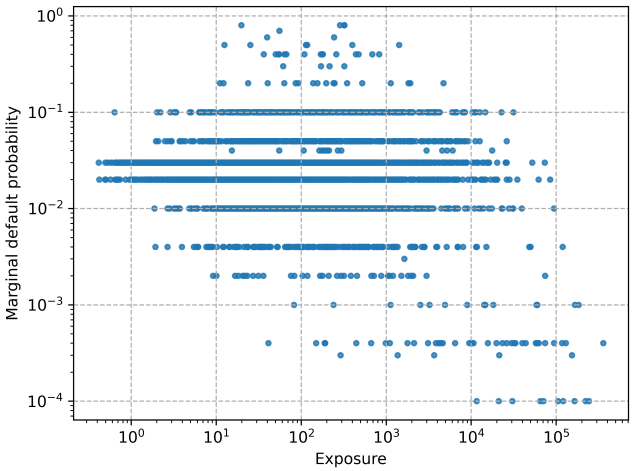


Figure 5.2: Scatter-plot of ℓ_k against p_k from the portfolio corresponding to scenarios 1 and 2. Both the x- and y-axis are in log-scale.

6 Results

In this chapter, we compare results using the M-IS and the AIS methods from Algorithms 2 and 3 respectively, against CMC for the three scenarios described in Chapter 5. Our goal is to estimate q_α, with $\alpha = 0.999$. The simulation code to generate the losses corresponding to these results is shown in Appendix B.

We begin with multiple CMC runs to construct an exact confidence interval as reference and obtain an initial estimate of the target quantile. Next, we carry out single-run experiments under different settings. Here we monitor the quantile estimate, tail probability estimates, and the sample variance of the estimated tail probability. These single runs are less computationally expensive. Although the sample variance of a tail probability estimate is not directly equivalent to the variance of a quantile estimate, we use it as an indication of the effectiveness of methods. After a discussion of the results from the single-run experiments, we select 2 experiments and carry out multiple runs to estimate the empirical variance of the quantile estimator.

From our experiment with Bernoulli tilting in Subsection 4.1.3, we observed low effectiveness on portfolios with high-correlated exposures. Consequently, we disregard Bernoulli tilting in our subsequent experiments.

6.1. Crude Monte Carlo

For every scenario, we carry out 100 CMC runs, each with 100 000 replications. Because we are using CMC, we can calculate the exact confidence using (3). Pooling all samples together, resulting in 10^7 samples per scenario, resulted in the following exact confidence intervals at a 99% confidence level for scenarios 1, 2 and 3 respectively:

$$(1.084, 1.094) \times 10^6$$
$$(5.192, 5.224) \times 10^5$$
$$(1.318, 1.326) \times 10^5$$

The code to calculate this confidence interval is shown in Appendix B. In Table 6.1 we show the sample average of the CMC quantile estimates for q_α over the 100 runs, denoted by $\bar{q}_\alpha$, and the sample standard deviation, denoted by $\hat{\sigma}_{q_\alpha}$.

Scenario	$\bar{q}_\alpha$	$\hat{\sigma}_{q_\alpha}$
1	1.088×10^6	1.994×10^4
2	5.202×10^5	7.281×10^3
3	1.321×10^5	1.751×10^3

Table 6.1: Sample average ($\bar{q}_\alpha$) and sample standard deviation ($\hat{\sigma}_{q_\alpha}$) from 100 CMC estimates for q_α. Each sample quantile was estimated on 100 000 replications.

6.2. Single-Run Experiments

In this section we carry out single-run experiments with 100 000 replications for each experiment. We start with the M-IS approach using the constant approximation to determine the shift. Then, we compare different AIS methods with different parameter settings. Specifically, we select different step size parameters for γ_n defined in (4.21), and the gradient threshold x for H defined in (4.13), both influencing the convergence rate of the shift vector.

In every run, we track the quantile estimate $\tilde{q}_{\alpha,r}$, the tail probability estimate $\tilde{P}_{x,r}$, and the sample variance of $\tilde{P}_{x,r}$, determined with the natural estimator $\tilde{\psi}_r^2$ described in (3.15), where r denotes the number of replications and x indicates the loss level of interest. We compare each estimate to its CMC counterpart from the first run (Section 6.1), using the same random-number seed.

We select the loss level of interest as the average 99.9% quantile obtained from the CMC outcomes in Table 6.1, specifically $x = \bar{q}_\alpha$ for the scenario in question. We divide the CMC sample variance $\hat{\psi}_r^2$ and the IS sample variance $\tilde{\psi}_r^2$ at the final replication to obtain an estimated variance reduction factor. For this, we use the unbiased sample variance estimators.

6.2.1. M-IS with the Constant Approximation

In this experiment, we evaluate the M-IS approach outlined in Algorithm 2. We apply the constant approximation method (see Section 4.2) to determine the mean shift for the common factors, with the corresponding code available in Appendix B. For this optimization, we use the threshold $x = \bar{q}_\alpha$ from the CMC results. This resulted in the following mean shifts $\boldsymbol{\mu}$:

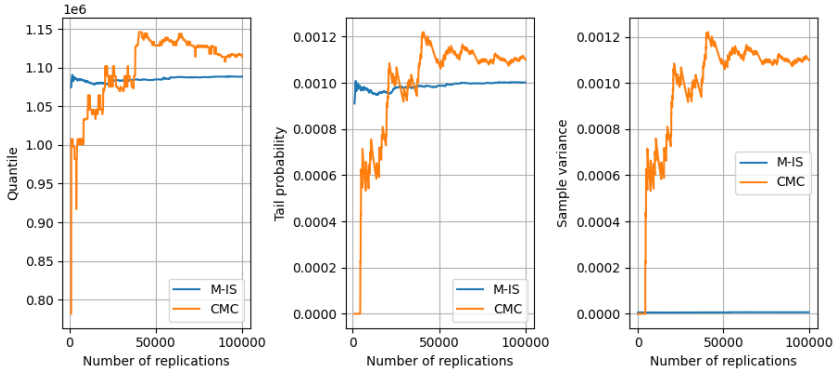
$$\boldsymbol{\mu}^1 = (-3.10,\ -0.11,\ -0.22,\ -0.19,\ 0,\ -0.18,\ -0.43,\ 0,\ -0.21,\ -0.39,\ 0)^\top$$
$$\boldsymbol{\mu}^2 = (-3.11,\ -0.24,\ -0.42,\ -0.45,\ 0,\ -0.35,\ -1.16,\ 0,\ -0.35,\ -0.95,\ 0)^\top$$
$$\boldsymbol{\mu}^3 = (-4.04,\ -0.07,\ -0.33,\ -0.27,\ 0,\ -0.10,\ -0.74,\ 0,\ -0.22,\ -0.37,\ 0)^\top$$

It is worth noting that since there is no correlation with region D and sectors S3 and S6, these resulted in a 0-value shift in $\boldsymbol{\mu}$. Furthermore, the second shift is much larger than the first one, while the dependence on the macroeconomic factor is lower. This occurs because it is necessary to shift the mean further to the negative side in order to align the conditional expected loss with the loss level x, since the impact of the macroeconomic factor is less than in the first scenario.

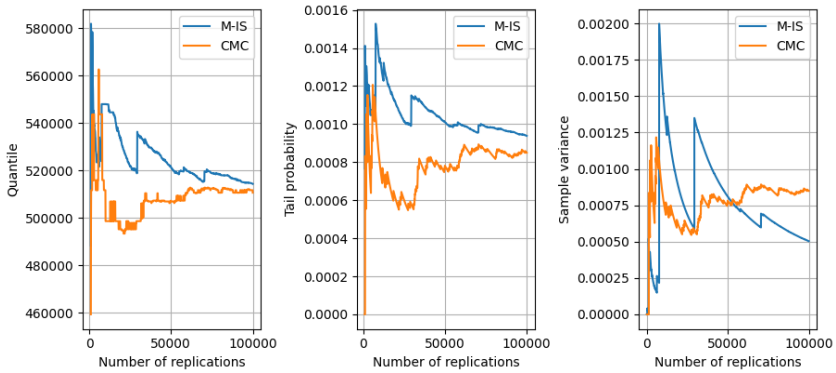
The results for the three scenarios are shown in Figure 6.1. The estimated variance reduction factors for scenarios 1–3 are **161.39**, **1.69**, and **0.97**, respectively (for the tail probability estimators).

We observe a significant variance reduction for scenario 1, but less or no variance reduction for the other scenarios. In scenario 2, the correlation with the macroeconomic factor is weaker than in scenario 1, so the relative impact of the shifted macroeconomy is reduced. In scenario 3, there is a rise in variance. This happens because of significant idiosyncratic factors: The default of a small number of obligors with high exposures in scenario 3 can trigger the event $\{L > x\}$, even if the common factors are low under the new IS distribution (causing large IS weights and thus high variance, similar to the dynamics between Y_1 and Y_2 discussed in Subsection 4.5). In Figure 6.1c we can observe such an event around the 70 000th replication.

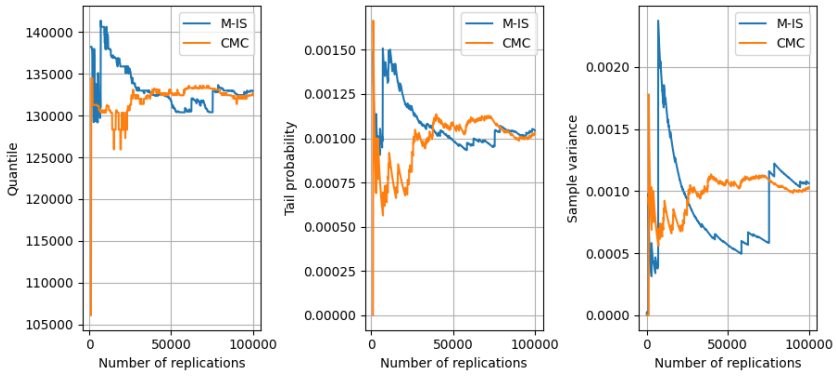
We plotted the IS weights versus the losses in Figure 6.2. We observe very skewed weights for all scenarios. However, for scenario 1, losses above the threshold correspond to IS weights below 1, resulting in a more precise estimate of $\tilde{P}_{x,n}$, while for the other scenarios, this is not the case, and some important losses correspond to large IS weights, decreasing the precision of $\tilde{P}_{x,n}$. Like we demonstrate in Subsection 4.5, M-IS is less effective if various concentration factors independently can generate impactfull losses.



(a) Scenario 1. Final $\tilde{\psi}_r^2 = 6.81 \times 10^{-6}$.



(b) Scenario 2. Final $\tilde{\psi}_r^2 = 5.03 \times 10^{-4}$.



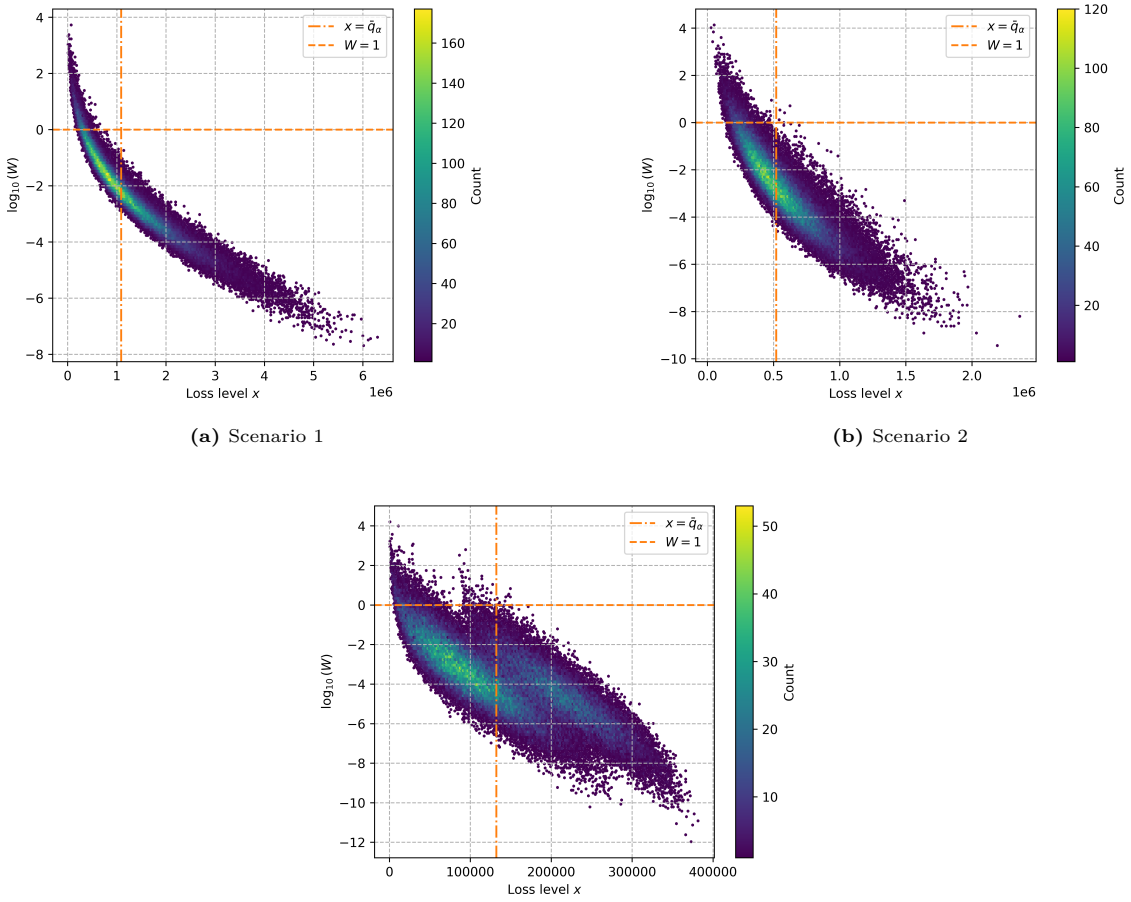
(c) Scenario 3. Final $\tilde{\psi}_r^2 = 1.06 \times 10^{-3}$.

Figure 6.1: Comparison of M-IS versus CMC estimates across Scenarios 1–3 in a single run of 100 000 replications. For each scenario the estimated IS quantile ($\tilde{q}_{\alpha,r}$), the estimated IS tail probability ($\tilde{P}_{x,r}$), and the sample variance ($\tilde{\psi}_r^2$) are shown as a function of r.

6.2.2. AIS with Slow Convergence

In this experiment, we evaluate the AIS approach outlined in Algorithm 3. We denote the shift vector after the nth update by $\boldsymbol{\mu}_n$, where n indexes the iteration. In our experiments, each iteration produces one sample, so the iteration index n is identical to the replication index r. For the threshold in the gradient (4.13), we set $x = \bar{q}_\alpha$. We choose a step size γ_n, as defined in (4.21), that decreases rapidly with the number of exceedances $\{L > x\}$ compared to our next experiment in Subsection 6.2.3. This yields more stable updates but slower convergence of $\boldsymbol{\mu}_n$. The step size parameters are:

$$\eta = 10, \quad \beta = 100, \quad \Delta = 1$$



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

Figure 6.2: Losses versus IS weights under M-IS for scenarios 1-3. The IS weights are very skewed due to the large shifts. In scenario 1 this leads to a precise estimate of $\tilde{P}_{x,n}$, since IS weights corresponding to losses above the threshold are small. Whereas, for the other scenarios this is not the case. The big shift in scenario 3 is due to a single obligor with an exposure of 80 000. This highlights the impact of idiosyncratic risk in smaller portfolios.

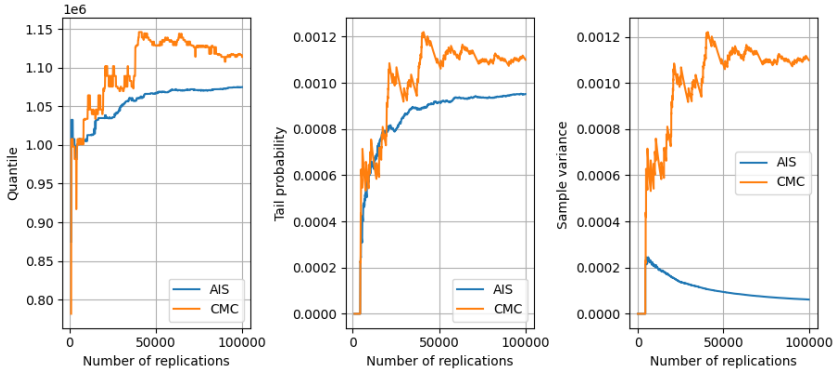
We define the region bound parameters as

$$R_n = R_0 + \log(\tau_n + 1), \quad \text{for } n = 1, 2, \ldots, 100\,000, \tag{6.1}$$

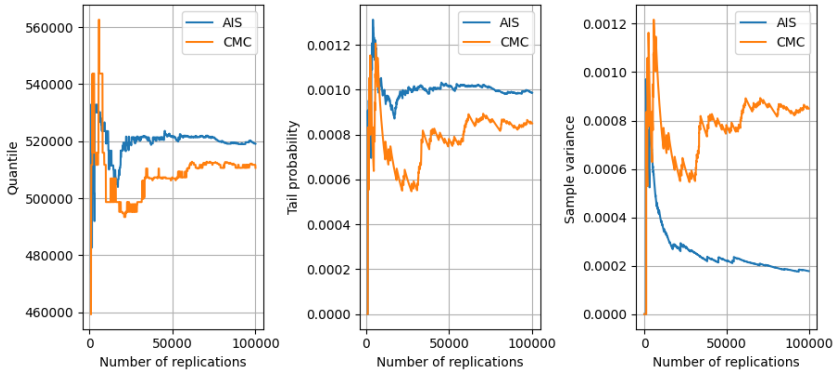
where $R_0 = 4$, and τ_n represents the count of boundary exceedances. We choose $R_0 = 4$ as a rough initial estimate based on the norms of $\{\boldsymbol{\mu}^j\}_{j=1}^3$ from the M-IS runs, since the initial shift should lie within the allowed region. We use logarithmic increments that cause the region to expand gradually, aligning with our assumption that the optimal shift $\boldsymbol{\mu}^*$ (for the scenario in question) is smaller than $\{\boldsymbol{\mu}^j\}_{j=1}^3$. Although difficult to prove, our assumption is supported by the results presented in Subsection 4.5, where AIS directs $\boldsymbol{\mu}_n$ to a more balanced shift compared to the mode of the zero-variance density, which the constant approximation tries to estimate.

The results for the three scenarios are shown in Figure 6.3. The evolution of $\boldsymbol{\mu}_n$ for the different scenarios is shown in Figure 6.4. The estimated variance reduction factors for scenarios 1–3 are **17.78**, **4.78**, and **6.15**, respectively.

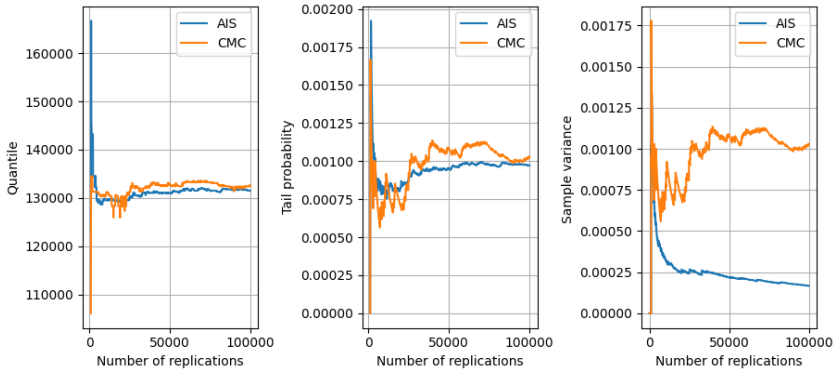
We observe that for scenario 1, the variance reduction is lower than in the M-IS case. This is because in the M-IS case, the shift was close to $\boldsymbol{\mu}^*$. With the current selected parameter settings, there are not enough iterations to drive the shift enough towards $\boldsymbol{\mu}^*$. This can be observed in Figure 6.4a, especially for the first component μ_1 corresponding to the macroeconomic factor, which seems to converge to a more negative value. Compared to the M-IS case, for scenarios 2 and 3 we do observe a variance reduction.



(a) Scenario 1. Final $\tilde{\psi}_r^2 = 6.18 \times 10^{-5}$.



(b) Scenario 2. Final $\tilde{\psi}_r^2 = 1.78 \times 10^{-4}$.



(c) Scenario 3. Final $\tilde{\psi}_r^2 = 1.67 \times 10^{-4}$.

Figure 6.3: Comparison of AIS versus CMC estimates across Scenarios 1–3 in a single run of 100 000 replications, with $\eta = 10$, $\beta = 100$, $\Delta = 1$ and $x = \bar{q}_\alpha$. For each scenario the estimated IS quantile ($\tilde{q}_{\alpha,r}$), the estimated IS tail probability ($\tilde{P}_{x,r}$), and the sample variance ($\tilde{\psi}_r^2$) are shown as a function of r.

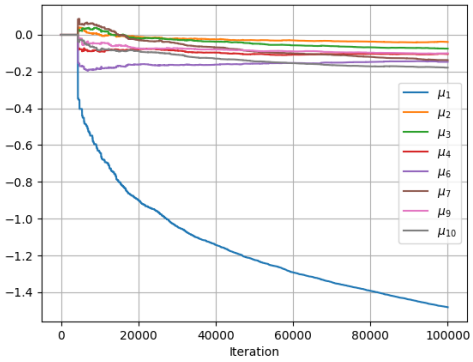
6.2.3. AIS with Fast Convergence

In this experiment, we use the same settings as in Subsection 6.2.2, but with step size parameters

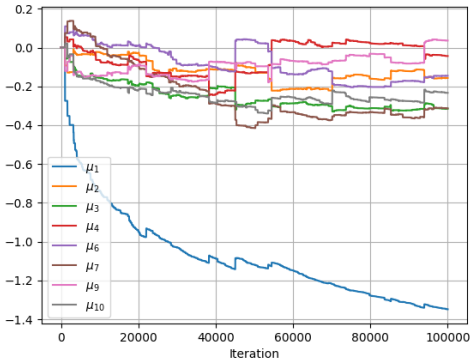
$$\eta = 200, \quad \beta = 1000, \quad \Delta = 1,$$

so γ_n is approximately two times the size as and decreases slower compared to the case in Subsection 6.2.2.This approach allows for faster convergence of $\boldsymbol{\mu}_n$, but the updates become less controlled. The results for the three scenarios are shown in Figure 6.5. The evolution of $\boldsymbol{\mu}_n$ for the different scenarios is shown in Figures 6.6. The estimated variance reduction factor for scenarios 1–3 is **27.93**, **4.50**, and **6.96**, respectively.

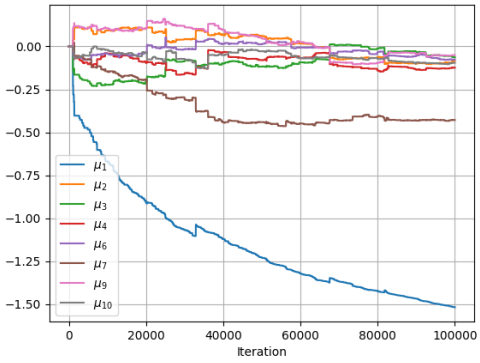
As expected, scenarios 1 and 3 show a larger variance reduction compared to the AIS case with a smaller step size, since the shift moves more quickly toward $\boldsymbol{\mu}^*$. In scenario 2 we observe a worsening



(a) Scenario 1



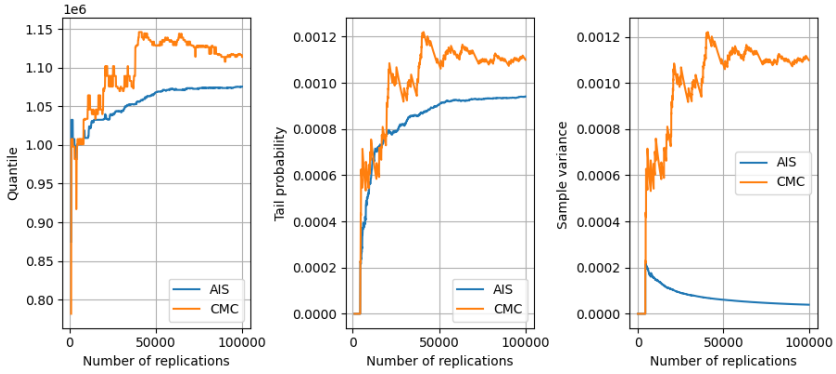
(b) Scenario 2



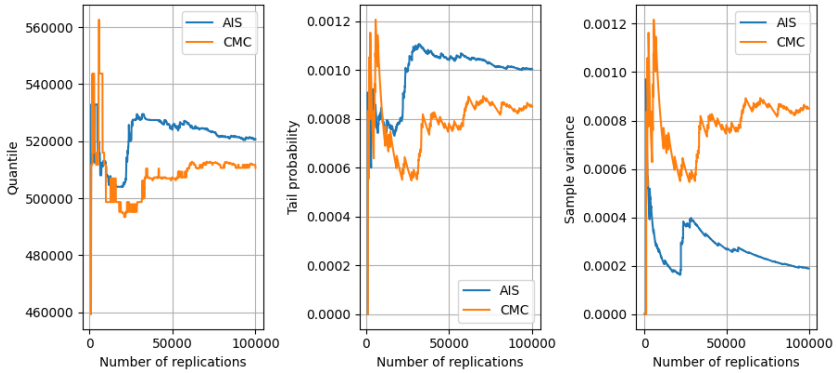
(c) Scenario 3

Figure 6.4: Evolution of the components of $\boldsymbol{\mu}_n$ as a function of n for Scenarios 1-3, with step size parameters $\eta = 10$, $\beta = 100$, $\Delta = 1$ and gradient threshold $x = \bar{q}_\alpha$.

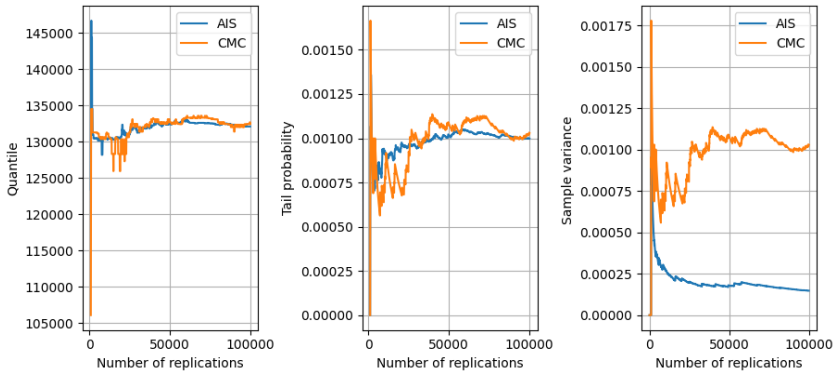
over the previous AIS case. This follows from the erratic behaviour of the shift vector in Figure 6.6b: just after the 20 000th replication, its components jump dramatically following a rare event. The large step size amplifies this jump, sending $\boldsymbol{\mu}_n$ into a suboptimal region. Consequently, subsequent replications sample less effectively for the estimates of interest.



(a) Scenario 1. Final $\tilde{\psi}_r^2 = 3.93 \times 10^{-5}$.



(b) Scenario 2. Final $\tilde{\psi}_r^2 = 1.89 \times 10^{-4}$.



(c) Scenario 3. Final $\tilde{\psi}_r^2 = 1.48 \times 10^{-4}$.

Figure 6.5: Comparison of AIS versus CMC estimates across Scenarios 1–3 in a single run of 100 000 replications, with $\eta = 200$, $\beta = 1000$, $\Delta = 1$ and $x = \bar{q}_\alpha$. For each scenario the estimated IS quantile ($\tilde{q}_{\alpha,r}$), the estimated IS tail probability ($\tilde{P}_{x,r}$), and the sample variance ($\tilde{\psi}_r^2$) are shown as a function of r.

6.2.4. AIS with low Gradient Threshold

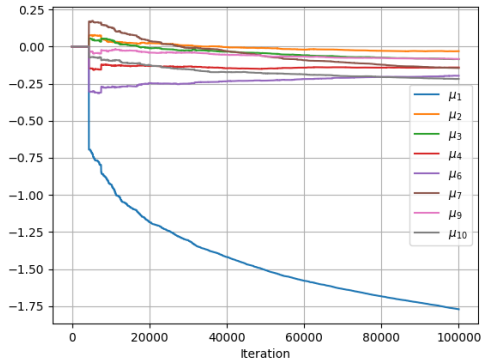
In this experiment, we use the same settings as in Subsection 6.2.2, but we set the gradient threshold in (4.13) equal to

$$x = \bar{q}_\alpha/2.$$

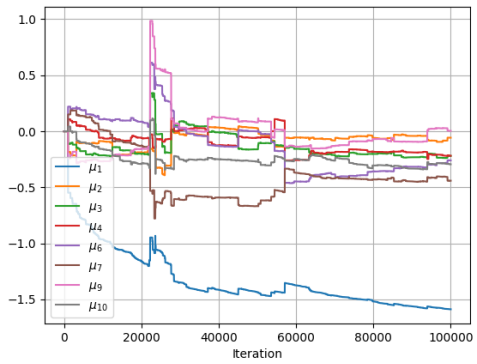
Since this causes the gradient defined to be non-zero at more iterations, $\boldsymbol{\mu}_n$ converges faster. Furthermore, we select a larger Δ, hence the step size parameters are

$$\eta = 10, \quad \beta = 100, \quad \Delta = 2.$$

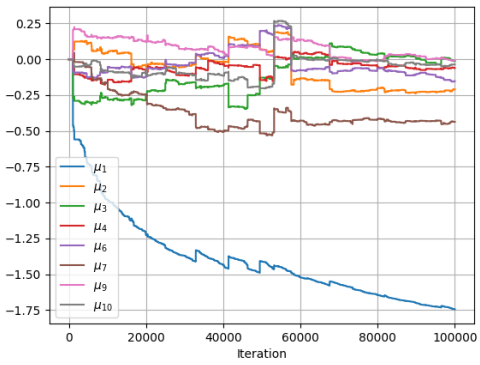
Thus, the decay of γ_n is greater, but the update of $\boldsymbol{\mu}_n$ is more controlled for as n increases. Since $\boldsymbol{\mu}_n$ is tailored towards a lower loss level further from the 99.9% quantile, it converges to a sub-optimal shift.



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Figure 6.6: Evolution of the components of $\boldsymbol{\mu}_n$ as a function of n for Scenarios 1-3, with step size parameters $\eta = 200$, $\beta = 1000$, $\Delta = 1$ and gradient threshold $x = \bar{q}_\alpha$.

The results for the three scenarios are shown in Figure 6.7. The evolution of $\boldsymbol{\mu}_n$ for the different scenarios is shown in Figure 6.8. The estimated variance reduction factors for scenarios 1–3 are **82,61**, **7.67**, and **3.81**, respectively.

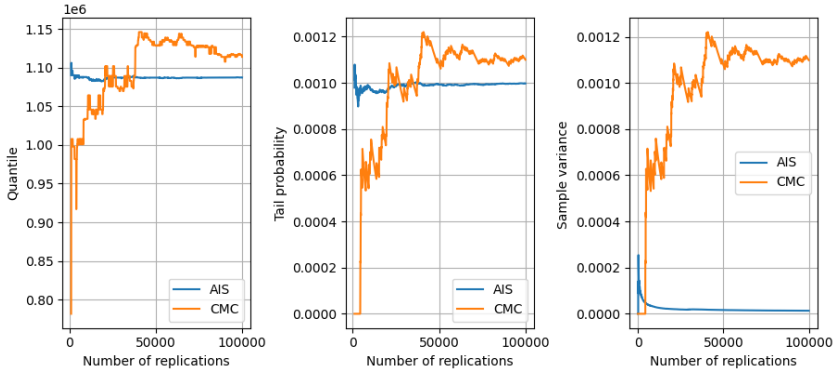
In scenario 1, we observe a large variance reduction compared to other AIS methods. The lower gradient threshold causes the gradient to be non-zero at more iterations, outweighing the lower variance reduction caused by convergence to a sub-optimal shift.

However, the runs on scenarios 2 and 3 highlight a risk of AIS. For these scenarios, we compute an estimated variance reduction in $\tilde{P}_{x,r}$, however, we observe that $\tilde{q}_{\alpha,n}$ and $\tilde{P}_{x,r}$ seem to be off compared to other experiments. Due to the low threshold, idiosyncratic factors cause $\boldsymbol{\mu}_n$ to behave very erratically, demonstrated in Figures 6.8b and 6.8c. The boundary R_n is even crossed a couple of times, resetting it to $\mathbf{0}$. Still, $\boldsymbol{\mu}_n$ is moved multiple times into sub-optimal regions within the boundaries, skewing the IS weights. This causes the current replications to appear low-variance, but as more samples are drawn, new rare events will eventually inflate the variance.

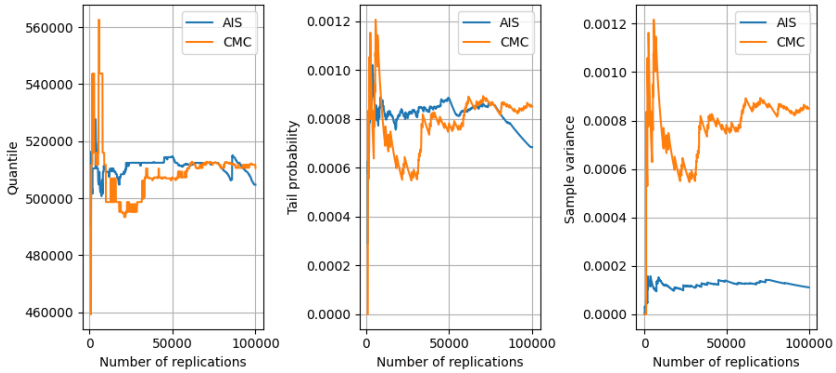
6.2.5. Conclusion on Single Run Experiments

Overall, M-IS approach with the constant approximation delivers the biggest estimated variance reduction when high losses are driven by common factors together, as we observe for scenario 1. In scenarios 2 and 3, idiosyncratic factors have a higher impact on large losses, causing a lower effect with M-IS. On the portfolios where idiosyncratic factors are more dominating, AIS is more effective. We tested the following parameter settings:

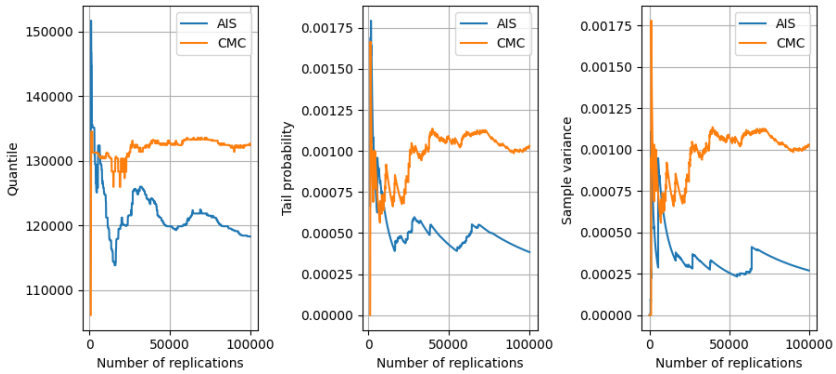
- *Slow convergence* ($\eta = 10, \beta = 100$) stabilizes updates at the cost of convergence speed .
- *Fast convergence* ($\eta = 200, \beta = 1000$) recovers more of the optimal shift in but increases the risk of jumping the shift towards sub-optimal regions.



(a) Scenario 1. Final $\tilde{\psi}_r^2 = 1.33 \times 10^{-5}$,.



(b) Scenario 2. Final $\tilde{\psi}_r^2 = 1.11 \times 10^{-4}$.

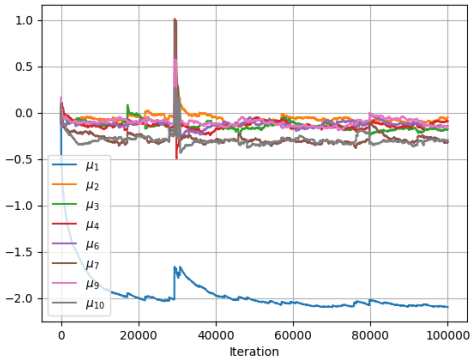


(c) Scenario 3. Final $\tilde{\psi}_r^2 = 2.39 \times 10^{-4}$.

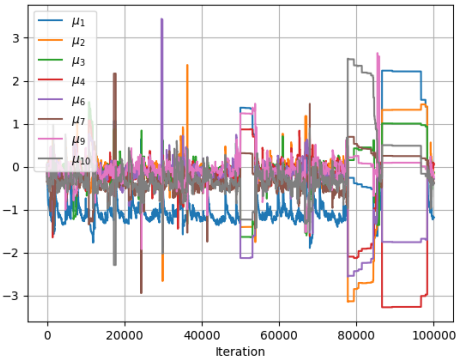
Figure 6.7: Comparison of AIS versus CMC estimates across Scenarios 1–3 in a single run of 100 000 replications, with with $\eta = 10$, $\beta = 100$, $\Delta = 2$ and gradient threshold $x = \bar{q}_\alpha/2$. For each scenario the estimated $\tilde{q}_{\alpha,r}$, $\tilde{P}_{x,r}$ and $\tilde{\psi}_r^2$ is shown as a function r.

- *Low gradient threshold* ($x = \bar{q}_\alpha/2$, $\Delta = 2$) accelerates convergence to a sub-optimal shift. Furthermore, a lower threshold also means idiosyncratic factors have higher impact, causing erratic updates.

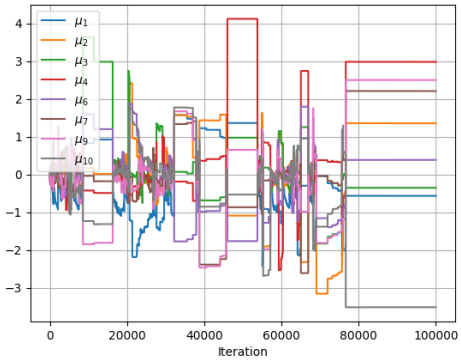
These findings are consistent with our findings in Chapter 4: M-IS shifts the common factors toward the mode of the zero-variance distribution and is particularly effective when those factors collectively drive large losses. AIS — with carefully tuned step-size parameters and gradient thresholds — is more effective when several concentration elements each can trigger large losses, since it balances the mean shift across these risk factors. This would result in a more effective shift in a scenario with high idiosyncratic risk.



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Figure 6.8: Evolution of the components of $\boldsymbol{\mu}_n$ as a function of n for Scenarios 1-3, with step size parameters $\eta = 10$, $\beta = 100$, $\Delta = 2$ and gradient threshold $x = \bar{q}_\alpha/2$. In scenarios 2 and 3 we see many erratic jumps because of high idiosyncratic impact.

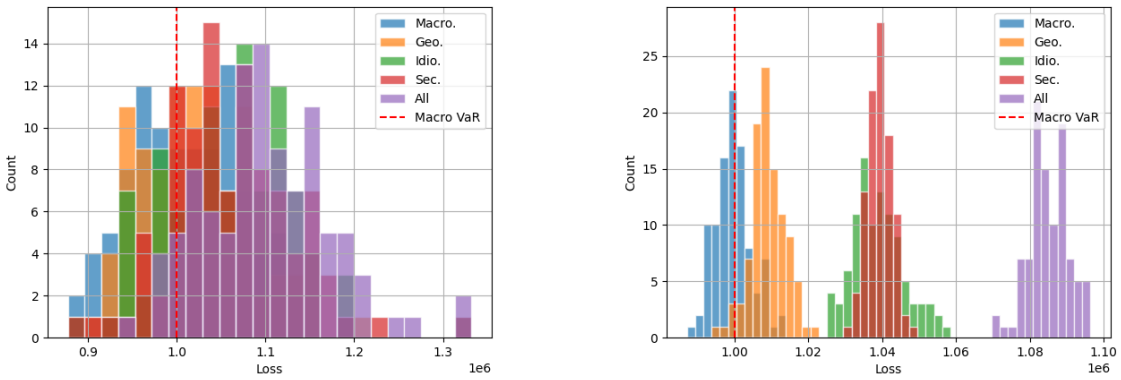
6.3. Final Results

In this section, we compare the sample variance of IS quantile estimates with that of their CMC counterparts. For this, we carry out multiple runs. Based on the most promising single-run experiments, we apply the M-IS method to Scenario 1 and the AIS method with slow, stable convergence to Scenario 3. Scenario 3 is chosen over Scenario 2 because its greater idiosyncratic impact makes it a more interesting case. Additionally, we compare the concentration risk add-ons and the marginal add-ons, both discussed in Section 2.3. To compute the concentration risk add-ons, we subtract the macroeconomic VaR from Table 5.1 from the estimated quantiles. To calculate the marginal add-ons, we first compute the marginal quantiles. These are calculated with loss samples which — together with the macroeconomic factor — include only geographic-region, idiosyncratic, or sectorial factors.

6.3.1. M-IS on Scenario 1

The setup for M-IS is identical to the one in Subsection 6.2.1 for Scenario 1, using the deterministic shift $\boldsymbol{\mu}^1$. We carry out 100 runs using Algorithm 2, each run consisting of 10 000 replications. The estimated total quantile (including all concentration factors) and the estimated marginal quantiles are shown in the histograms in Figure 6.9. Their sample means, the average concentration risk add-ons, together with their sample standard deviations, are shown in Table 6.2.

A significant reduction in variance is observed for the total and marginal quantile estimates. Specifically, the reduction in variance in the estimated marginal quantile with only sectorial factors is substantial. For the sectorial marginal add-ons, idiosyncratic factors are not included. Additionally, obligors are more correlated with sectorial factors than with regional factors, which explains their larger impact on the overall variance.



(a) CMC **(b)** M-IS

Figure 6.9: Comparison of the 100 quantile estimates from CMC and M-IS including total and marginal quantiles for scenario 1.

Type	CMC			M-IS			VR
	$\bar{q}_\alpha$	$\bar{C}$	$\hat{\sigma}_{q_\alpha}$	$\bar{q}_\alpha$	$\bar{C}$	$\hat{\sigma}_{q_\alpha}$	
All	1 101 963	101 963	70 830	1 085 104	85 104	5 357	**175**
Macroeconomic	1 011 467	11 467	62 595	999 552	−448	4 881	**164**
Geographical	1 022 049	22 049	62 972	1 008 947	8 947	4 698	**180**
Idiosyncratic	1 057 001	57 001	70 672	1 038 893	38 893	6 837	**107**
Sectorial	1 059 847	59 847	68 455	1 039 250	39 250	3 351	**417**

Table 6.2: Sample mean ($\bar{q}_\alpha$), average concentration risk add-on ($\bar{C}$), and sample standard deviation ($\hat{\sigma}_{q_\alpha}$) over 100 runs (10 000 replications each) for scenario 1, CMC vs. M-IS. The estimated variance reduction factor VR is the ratio of CMC to M-IS sample variances.

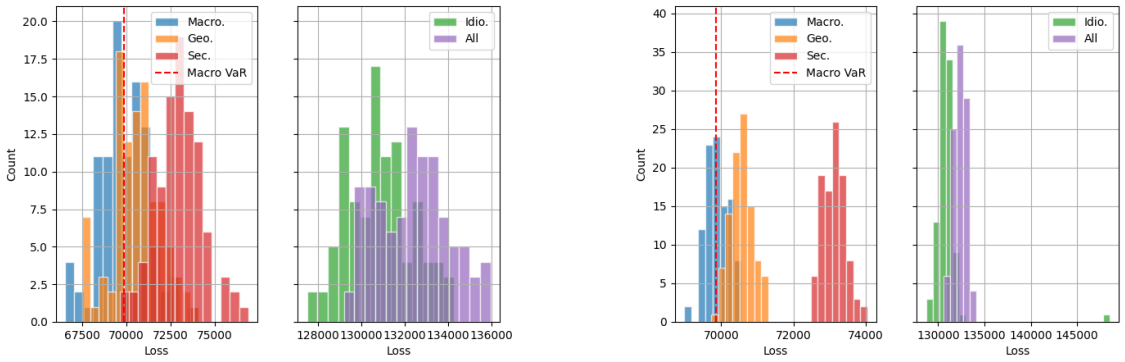
6.3.2. AIS on Scenario 3

The setup for AIS is identical to that in Subsection 6.2.2 for scenario 3. We execute 100 runs using Algorithm 3, each consisting of 100 000 replications. In this instance, we increased the number of replications per run compared to the experiment in Subsection 6.3.1, because AIS achieves better performance when additional replications enable it to drive $\boldsymbol{\mu}_n$ further towards the optimal shift. The estimated total quantiles and the estimated marginal quantiles are shown in the histograms in Figure 6.10. The sample mean, the average concentration risk add-in, and the sample standard deviation are shown in Table 6.3.

A moderate reduction in variance is observed for the estimated total quantiles and the estimated marginal quantiles, except for the idiosyncratic marginal quantiles. For these quantiles we observe a variance increase due to the high impact of idiosyncratic factors.

Type	CMC			AIS			VR
	$\bar{q}_\alpha$	$\bar{C}$	$\hat{\sigma}_{q_\alpha}$	$\bar{q}_\alpha$	$\bar{C}$	$\hat{\sigma}_{q_\alpha}$	
All	132 375	62 375	1 645	132 318	62 318	717	**5**
Macroeconomic	69 878	−122	1 373	69 902	902	310	**20**
Geographical	70 536	536	1 372	70 562	562	326	**18**
Idiosyncratic	130 856	60 856	1 478	130 891	60 891	1 886	**1**
Sectorial	72 941	2 941	1 328	73 123	3 123	311	**18**

Table 6.3: Sample mean ($\bar{q}_\alpha$), average concentration risk add-on ($\bar{C}$), and sample standard deviation ($\hat{\sigma}_{q_\alpha}$) over 100 runs (100 000 replications each) for scenario 3, CMC vs. AIS. The estimated variance reduction factor VR is defined as the ratio of CMC to AIS sample variances.



(a) CMC **(b)** AIS

Figure 6.10: Comparison of the 100 quantile estimates from CMC and AIS including total and marginal quantiles, for scenario 3.

Summary Final Results

We carried out 100 independent runs of M-IS on Scenario 1 and AIS on Scenario 3 and compared their performance against CMC. Our main findings are:

- On Scenario 1, deterministic mean-shifting (M-IS) achieves a variance reduction factor of 175 for the quantile estimate and even up to 417 for the sectorial marginal quantile estimate.
- On the more idiosyncratic Scenario 3, adaptive IS (AIS) delivers a 5 times reduction on the overall quantile estimate — driven by the large idiosyncratic influence — with up to 20 times reduction on marginal quantile estimates.
- In both cases, the sample quantile estimates align with the exact confidence intervals (level 99) from the CMC runs in Section 6.1.

Computation Time and Efficiency Gain

We run each simulation on an Intel® Core™ i7-8750H CPU @ 2.20 GHz. On scenario 1, with 20 000 obligors, the algorithm produces approximately 700-800 loss samples per second, excluding the computation of marginal loss samples. If marginal losses (macroeconomic, idiosyncratic, sectorial, and geographic-regional) are computed, the algorithm is approximately 3-4 times slower since these need to be calculated separately (see Appendix B).

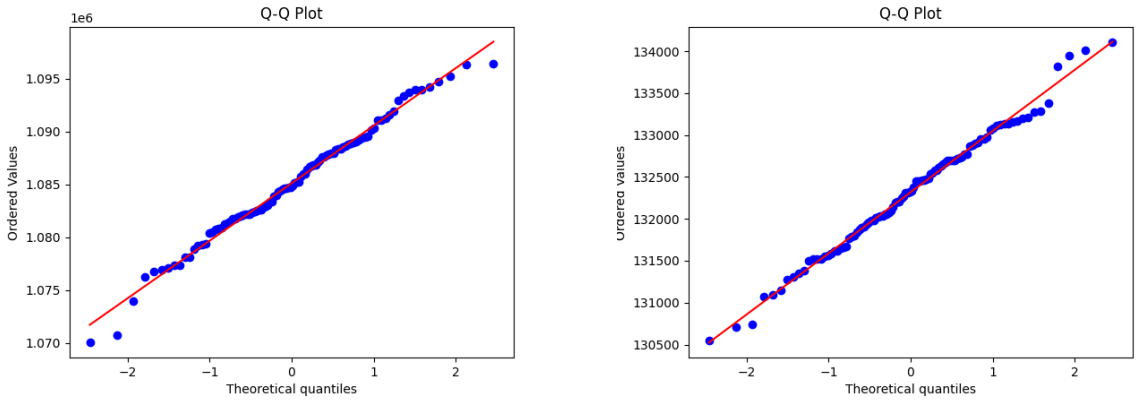
The M-IS scheme adds almost no extra cost during simulation — the constant approximation step is negligible (the optimization is solved within milliseconds). Likewise, the AIS scheme only adjusts the common factor weights at each iteration, adding minimal overhead. As a result, the observed variance reductions translate almost directly into efficiency gains.

6.4. Constructing Confidence Intervals under Continuity Assumptions

We use our 100 runs from M-IS on scenario 1 and AIS on scenario 3 to analyse the construction of asymptotic confidence intervals from Section 3.4. We keep in mind that, strictly speaking, continuity of the parent distribution is required (which fails in the portfolio credit risk setting), and for AIS, the consistency of the sample variance estimates still needs to be established—nevertheless, the results below align with our expectations.

We test normality of the 100 estimated quantiles via Q–Q plots in Figures 6.11a and 6.11b, which lie close to the 45° line. Furthermore, we use the Shapiro–Wilk test Shapiro and Wilk [30]. For Scenario 1 the test statistic is $W = 0.9891$ ($p = 0.5886$), and for Scenario 3 $W = 0.9915$ ($p = 0.7870$), providing no evidence against the assumption of normality.

For the first run, we plotted the estimated quantile density $\phi(\kappa)$ with (3.19) for different values of the bandwidth parameter κ, shown in Figures 6.12a and 6.12b. This figure highlights the importance of



(a) M-IS (scenario 1) **(b)** AIS (scenario 3)

Figure 6.11: Q-Q plots of the 100 estimated quantiles for M-IS (scenario 1) and AIS (scenario 3), confirming asymptotic Normality.

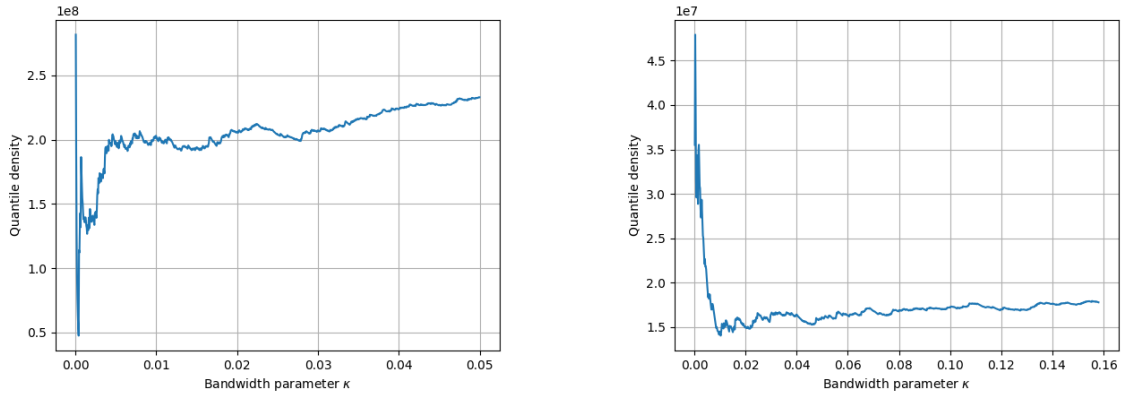
carefully choosing a bandwidth: A bandwidth that's too small yields highly varying density estimates, whereas one that's too large induces greater bias.

For each run, we calculated the quantile density $\tilde{\phi}(\kappa)$ with (3.19), with $\kappa = 0.01$, and $\tilde{\psi}_n$ with (3.15) to compute the sample standard deviation, given by their product. Figures 6.13a and 6.13b display histograms of the resulting standard deviations for each method (asymptotic in blue) against the sample standard deviation (red line). From these, we computed the average of the asymptotic standard deviation and compared it with the sample standard deviation:

$$\text{M-IS (scenario 1):} \quad \bar{\sigma}_{\text{asymp}} = 5537, \quad \hat{\sigma}_{q_\alpha} = 5357,$$

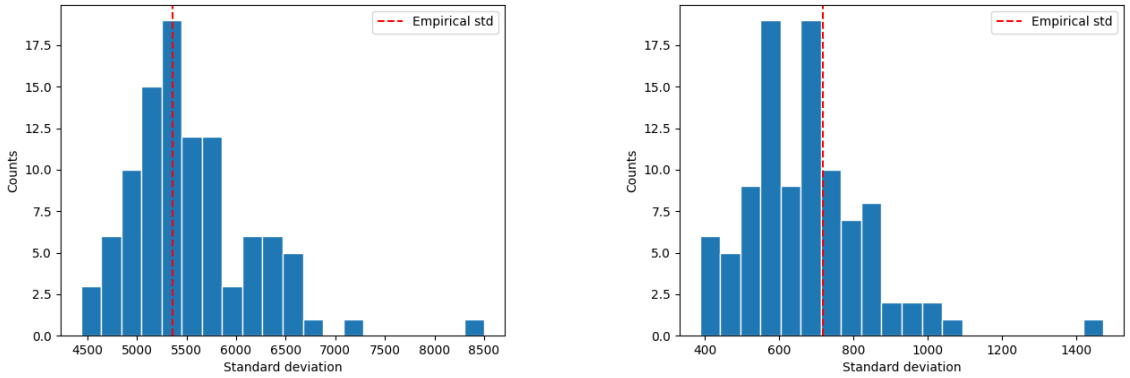
$$\text{AIS (scenario 3):} \quad \bar{\sigma}_{\text{asymp}} = 672, \quad \hat{\sigma}_{q_\alpha} = 717.$$

For both the M-IS and the AIS method, we see that the average of the estimated asymptotic standard deviation is close to the sample standard deviation. However, the asymptotic estimates show quite some dispersion, indicating that a larger number of replications per run is required to achieve more stable estimates. If stable standard deviation estimates can be obtained, then constructing $(1-p)$–confidence intervals using (3.20) will provide accurate coverage.



(a) M-IS (scenario 1) **(b)** AIS (scenario 3)

Figure 6.12: Estimated quantile density $\phi_n(\kappa)$ for M-IS (scenario 1) and AIS (scenario 3) for the first run as a function of the bandwidth parameter κ.



(a) M-IS (scenario 1)

(b) AIS (scenario 3)

Figure 6.13: Histograms estimated asymptotic standard deviations (blue) for M-IS (scenario 1) and AIS (scenario 3) with bandwidth parameter $\kappa = 0.01$, over 100 runs. The average of the estimated asymptotic standard deviations ($\bar{\sigma}_{\text{asymp}}$) is 5537 and 672, compared to the sample standard deviation ($\hat{\sigma}_{q_\alpha}$, red line) 5357 and 717, respectively for the M-IS and AIS case.

7 Conclusion

In this thesis, we apply IS and AIS algorithms to compute 99.9% VaR concentration-risk add-ons and quantify their statistical error. We tested them on stylized credit portfolios, constructed such that they mimic real-world conditions.

Methods

For applying IS effectively, an appropriate IS proposal distribution is needed. We compare three strategies to find the right IS proposals:

1. Bernoulli tilting on idiosyncratic defaults conditional on the common factors.
2. Deterministic mean shifts (M-IS) on the common factors.
3. Adaptive Importance Sampling (AIS), iteratively updating the mean shifts.

Bernoulli tilting is an effective method for obligors that are weakly correlated through common factors, but becomes less effective when obligors are highly correlated. In that case, shifting the common factors is more impactful. The deterministic shift is calculated using the constant-approximation method, which aims to find the mode of a zero-variance distribution. In the large-threshold regime for correlated portfolios, the common factors collectively drive the losses exceeding the threshold; hence, shifting them in that direction works well.

In contrast, in a low-threshold regime there are multiple concentration factors that can each independently push losses above the threshold. A single shift along one direction will miss the others, creating rare events in other important regions and actually increasing variance. The AIS method avoids this by rebalancing its shifts during the simulation whenever variance starts to rise. The trade-off is that AIS requires careful parameter tuning.

Finally, using the generalized Poisson–Binomial distribution and the IDFT, we can recover the conditional loss probabilities for simple portfolios. We used this as a tool to visualize the zero-variance distribution, confirming how crucial it is to balance shifts across all important regions.

Testing methods on Stylized Portfolios

On the basis of promising results from single run experiments, we tested the M-IS and AIS on two scenarios, with realistic portfolios. M-IS on the large portfolio leads to a significant variance reduction of factor 175. In the smaller portfolio, AIS provides an estimated variance reduction of factor 5. Here, we choose conservative parameter settings to force a more controlled convergence. Loosening the parameter settings or increasing the number of samples leads to a higher variance reduction, and Bernoulli tilting enhances it further — especially since idiosyncratic factors have a higher impact in smaller portfolios.

Both M-IS and AIS add almost no extra simulation cost, so the observed variance reductions translate directly into efficiency gains.

Constructing Asymptotic Confidence Intervals

Furthermore, we discussed asymptotic estimates for constructing confidence intervals, relaxing continuity assumptions, for both M-IS and AIS. For large portfolios, the estimated and empirical standard deviations matched closely. However, when discreteness increases (in the case of small portfolios), constructing asymptotic confidence intervals becomes infeasible.

Further Research

For further research, we identify some promising directions:

- **Comparing the optimal shift with mixed IS.** An interesting follow-up would compare the optimal shift, found via AIS (with multiple runs and parameter adjustments), to the mixed IS method of Glasserman, Kang, et al. [14]. Given that balancing region importance in the low-threshold regime motived the mixed IS method, this would reveal how a single optimal shift compares in effectiveness.
- **Tuning methods for AIS and its convergence.** Study appropriate tuning methods and find optimal convergence rates of AIS for complex distributions (see, e.g., Akyildiz and Míguez [1]).
- **Asymptotic confidence intervals for discrete distributions.** An interesting direction would be combining the approach of Wang and Hutson [32] with IS, allowing a more robust method to construct confidence intervals for discrete distributions under IS.
- **AI for computing the zero-variance distribution.** Last but not least, artificial intelligence is a hot topic nowadays. However, with strong regulation, the use of AI to directly estimate 99.9% VaR is not always allowed. Since IS gives an unbiased estimator, using AI to approximate the zero-variance distribution, or more simply, optimal shifts, would be very interesting and feasible with the right training data (see, e.g., Müller, McWilliams, et al. [28], which combines IS with neural nets).

To conclude, our implementations demonstrate that (adaptive) importance sampling can make 99.9% VaR estimation computationally feasible and statistically reliable in realistic portfolio credit risk settings, substantiated with empirical evidence.

References

[1] Ömer Deniz Akyildiz and Joaquín Míguez. "Convergence rates for optimised adaptive importance samplers". In: *Statistics and Computing* 31 (2021), pp. 1–17.

[2] Bouhari Arouna. "Robbins-Monro Algorithms and Variance Reduction in Finance". In: *The Journal of Computational Finance* 7.2 (2003). DOI: 10.21314/JCF.2003.111.

[3] Bouhari Arouna. In: *Monte Carlo Methods and Applications* 10.1 (2004), pp. 1–24. DOI: doi:10.1515/156939604323091180. URL: https://doi.org/10.1515/156939604323091180.

[4] Angelo Arvanitis and Jon Gregory. "Credit: The complete guide to pricing, hedging and risk management". In: *(No Title)* (2001).

[5] Basel Committee on Banking Supervision. *Basel Framework*. Tech. rep. Bank for International Settlements, 2024. URL: https://www.bis.org.

[6] Arnout Boks. "Maximally Adaptive Nonparametric Importance Sampling". Master's thesis. Delft, The Netherlands: Delft University of Technology, 2012. URL: https://resolver.tudelft.nl/uuid:ca591240-ffcf-4624-81bb-ec020b853255.

[7] Han-Fu Chen, Lei Guo, and Ai-Jun Gao. "Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds". In: *Stochastic Processes and their Applications* 27 (1987), pp. 217–231. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(87)90039-1. URL: https://www.sciencedirect.com/science/article/pii/0304414987900391.

[8] Xiaohong Chen and Halbert White. In: *Studies in Nonlinear Dynamics & Econometrics* 6.1 (2002). DOI: doi:10.2202/1558-3708.1000. URL: https://doi.org/10.2202/1558-3708.1000.

[9] Fang Chu and Marvin K Nakayama. "Confidence intervals for quantiles and value-at-risk when applying importance sampling". In: *Proceedings of the 2010 Winter Simulation Conference*. IEEE. 2010, pp. 2751–2761.

[10] Fang Chu and Marvin K. Nakayama. "Confidence intervals for quantiles when applying variance-reduction techniques". In: *ACM Trans. Model. Comput. Simul.* 22.2 (Mar. 2012). ISSN: 1049-3301. DOI: 10.1145/2133390.2133394. URL: https://doi.org/10.1145/2133390.2133394.

[11] Daniel Egloff and Markus Leippold. "Quantile estimation with adaptive importance sampling". In: *The Annals of Statistics* 38.2 (Apr. 2010). ISSN: 0090-5364. DOI: 10.1214/09-aos745. URL: http://dx.doi.org/10.1214/09-AOS745.

[12] Daniel Egloff, Markus Leippold, Stephan Jöhri, and Curdin Dalbert. "Optimal importance sampling for credit portfolios with stochastic approximation". In: *Available at SSRN 693441* (2005).

[13] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Stochastic Modelling and Applied Probability. Springer New York, 2003. ISBN: 9780387216171. URL: https://books.google.nl/books?id=aeAlBQAAQBAJ.

[14] Paul Glasserman, Wanmo Kang, and Perwez Shahabuddin. "Fast Simulation of Multifactor Portfolio Credit Risk". In: *Operations Research* 56.5 (2008), pp. 1200–1217. DOI: 10.1287/opre.1080.0558.

[15] Paul Glasserman and Jingyi Li. "Importance sampling for portfolio credit risk". In: *Management science* 51.11 (2005), pp. 1643–1656.

[16] Peter W Glynn et al. "Importance sampling for Monte Carlo estimation of quantiles". In: *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*. Citeseer. 1996, pp. 180–185.

[17] Michael B. Gordy. "A risk-factor model foundation for ratings-based bank capital rules". In: *Journal of Financial Intermediation* 12.3 (2003), pp. 199–232. ISSN: 1042-9573. DOI: https://doi.org/10.1016/S1042-9573(03)00040-8. URL: https://www.sciencedirect.com/science/article/pii/S1042957303000408.

[18] Greg Gupton, Christopher Finger, and Mickey Bhatia. “CreditMetrics - Technical Document”. In: (Apr. 1997).

[19] Allan Gut. *Probability: A Graduate Course*. 2nd ed. Springer Texts in Statistics. Published: 17 October 2012. Softcover reprint published: 09 November 2014. New York, NY: Springer, 2013, pp. XXVI+602. ISBN: 978-1-4614-4708-5. DOI: 10.1007/978-1-4614-4708-5. URL: https://doi.org/10.1007/978-1-4614-4708-5.

[20] J P L Hatchett and R Kühn. “Effects of economic interactions on credit risk”. In: *Journal of Physics A: Mathematical and General* 39.10 (Feb. 2006), pp. 2231–2251. ISSN: 1361-6447. DOI: 10.1088/0305-4470/39/10/001. URL: http://dx.doi.org/10.1088/0305-4470/39/10/001.

[21] Shengyi He, Guangxin Jiang, Henry Lam, and Michael C Fu. “Adaptive importance sampling for efficient stochastic root finding and quantile estimation”. In: *Operations Research* 72.6 (2024), pp. 2612–2630.

[22] Michael Kalkbrener, Hans Lotter, and Ludger Overbeck. “Sensible and efficient capital allocation for credit portfolios”. In: *Risk* 17.1 (2004), S19–S24.

[23] Lesh G. Koju, Ramesh Koju, and Shiguang Wang. “Macroeconomic determinants of credit risks: evidence from high-income countries”. In: *European Journal of Management and Business Economics* 29.1 (2020), pp. 41–53. DOI: 10.1108/EJMBE-02-2018-0032.

[24] Yanyuan Ma, Marc G Genton, and Emanuel Parzen. “Asymptotic properties of sample quantiles of discrete distributions”. In: *Annals of the Institute of Statistical Mathematics* 63 (2011), pp. 227–243.

[25] Sandro Merino and Mark A Nyfeler. “Applying importance sampling for estimating coherent credit risk contributions”. In: *Quantitative Finance* 4.2 (2004), p. 199.

[26] Robert C. Merton. “ON THE PRICING OF CORPORATE DEBT: THE RISK STRUCTURE OF INTEREST RATES”. In: *The Journal of Finance* 29.2 (1974), pp. 449–470. DOI: https://doi.org/10.1111/j.1540-6261.1974.tb03058.x.

[27] William J Morokoff. “An importance sampling method for portfolios of credit risky assets”. In: *Proceedings of the 2004 Winter Simulation Conference, 2004.* Vol. 2. IEEE. 2004, pp. 1668–1676.

[28] Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. “Neural importance sampling”. In: *ACM Transactions on Graphics (ToG)* 38.5 (2019), pp. 1–19.

[29] Robert J Serfling. *Approximation theorems of mathematical statistics*. John Wiley & Sons, 1980.

[30] Samuel Sanford Shapiro and Martin B Wilk. “An analysis of variance test for normality (complete samples)”. In: *Biometrika* 52.3-4 (1965), pp. 591–611.

[31] Vasicek. “THE DISTRIBUTION OF LOAN PORTFOLIO VALUE”. In: *Risk* (2002).

[32] Dongliang Wang and Alan D Hutson. “A fractional order statistic towards defining a smooth quantile function for discrete data”. In: *Journal of statistical planning and inference* 141.9 (2011), pp. 3142–3150.

[33] Man Zhang, Yili Hong, and Narayanaswamy Balakrishnan. “The generalized Poisson-binomial distribution and the computation of its distribution function”. In: *Journal of Statistical Computation and Simulation* (Feb. 2018), pp. 1–13. DOI: 10.1080/00949655.2018.1440294.

A

Additional proofs

Exact Confidence interval (Theorem 3)

Proof. Define the random variable

$$S = \sum_{i=1}^{n} \mathbf{1}\{L_i < x\},$$

which counts the number of observations less than x. Then

$$\begin{aligned}
\mathbb{P}(L_{k:n} < x) &= \mathbb{P}(S \geq k) \\
&= \sum_{i=k}^{n} \mathbb{P}(S = i) \\
&= \sum_{i=k}^{n} \binom{n}{i} \left[F_L(x^-)\right]^i \left[1 - F_L(x^-)\right]^{n-i},
\end{aligned}$$

and therefore

$$\begin{aligned}
\mathbb{P}(L_{k_1:n} < x < L_{k_2:n}) &= \mathbb{P}(k_1 \leq S < k_2) \\
&= \sum_{i=k_1}^{k_2-1} \binom{n}{i} \left[F_L(x^-)\right]^i \left[1 - F_L(x^-)\right]^{n-i}.
\end{aligned}$$

Hence, setting $x = q_\alpha$ and using

$$F_L(q_\alpha^-) \;\leq\; \alpha \;\leq\; F_L(q_\alpha)$$

gives the desired result.

□

B

Code

Bisection method to find θ_x 4.1.2

```
def p_cond(y, R, b, p):
    p_cond_ = sp.stats.norm.cdf((sp.stats.norm.ppf(p) - R*y) / b)
    return p_cond_

def psi(theta, y, R, b, p):
    p_cond_ = p_cond(y, R, b, p)
    psi_ = np.sum(np.log(1 + p_cond_ * (np.exp(theta) - 1)))
    return psi_

def solve_theta_x(y, R, b, p, x_target, tol=1e-4, theta_max=10):
    theta_low, theta_high = 0, theta_max
    p_cond_ = p_cond(y, R, b, p)
    while theta_high - theta_low > tol:
        theta_mid = 0.5 * (theta_low + theta_high)
        p_theta = p_cond_ * np.exp(theta_mid ) / (1 + p_cond_ * (np.exp(theta_mid) - 1)
            )
        L_mean = np.sum(p_theta)
        if L_mean < x_target:
            theta_low = theta_mid
        else:
            theta_high = theta_mid
    return 0.5 * (theta_low + theta_high)
```

Calculating exact confidence interval in CMC

Here we show how to construct an exact confidence interval. To address overflow in case n is very large, we work in logarithms to account for large numbers from the gamma function:

$$
\begin{aligned}
m! &= \Gamma(m+1), \\
\binom{n}{k} &= \frac{\Gamma(n+1)}{\Gamma(k+1)\,\Gamma(n-k+1)}, \\
\ln \binom{n}{k} &= \ln\Gamma(n+1) - \ln\Gamma(k+1) - \ln\Gamma(n-k+1), \\
\ln P(X=k) &= \ln \binom{n}{k} + k \ln p + (n-k)\ln(1-p), \\
P(X=k) &= \exp\big(\ln P(X=k)\big).
\end{aligned}
$$

```
def ci_quantile(data, q=0.999, conf=0.99):
    sd = sorted(data)
    n = len(sd)
    p, r = q, 1-q
    i0 = math.ceil(n*p) - 1
```

```
    logb0 = (math.lgamma(n+1)
             - math.lgamma(i0+1)
             - math.lgamma(n-i0+1)
             + i0*math.log(p)
             + (n-i0)*math.log(r))
    b0 = math.exp(logb0)
    bl = bh = b0
    kl = kh = i0
    mass = b0
    toggle = True
    while mass < conf:
        if toggle:
            bl = bl * (kl) / (n-kl) * (r/p)
            kl -= 1
            mass += bl
        else:
            bh = bh * (n-kh) / (kh+1) * (p/r)
            kh += 1
            mass += bh
        toggle = not toggle
    return sd[math.ceil(n*p)-1], sd[kl], sd[kh]
```

Constant approximation method 4.2.1

```
def Constant_Approximation(m,d,R,denum,l,threshold,mu_indx):

    j = R.shape[1]
    m = R.shape[0]

    def objective(x):
        return np.sqrt(np.sum(x**2))

    def constraint(x):

        mu = np.zeros(j)
        mu[mu_indx] = x

        rhs = (d - np.dot(R, mu)) / denum

        loss = np.sum(l * norm.cdf(rhs))

        return loss - threshold  # Constraint: must be >= threshold

    x0 = np.zeros(len(mu_indx))

    result = minimize(objective,
                      x0,
                      constraints=[{'type': 'ineq', 'fun': constraint}],
                      options={'ftol': 1e-15, 'disp': False})

    y_opt = result.x

    return y_opt
```

GPB for finding the computing the zero-variance distribution 4.4

```
from collections import Counter

def gpb_pmf(p_vec, ell_vec):
    n = int(np.sum(ell_vec)) + 1  # max possible discrete loss + 1
    omega = 2 * np.pi / n
    l = np.arange(n)  # Correct Fourier grid
    log_xl = np.zeros(n, dtype=complex)  # Must match the grid size!

    # Group identical (p, b) pairs to speed up
    pair_counts = Counter(zip(p_vec, ell_vec))

    for (p_k, ell_k), count in pair_counts.items():
        term = (1 - p_k) + p_k * np.exp(1j * omega * l * ell_k)
```

```
        log_xl += count * np.log(term)
    x_l = np.exp(log_xl)
    xi = np.fft.ifft(x_l).real[1:]
    xi = xi/np.sum(xi)

    # Ensure non-negativity due to numerical error
    xi[xi < 0] = 0
    return xi[::-1]  # normalize PMF
```

```
y1_space = np.linspace(4, -6, 250)
y2_space = np.linspace(4, -6, 250)
Y1, Y2 = np.meshgrid(y1_space, y2_space)

# --- Compute zero-variance weight grid Z ---
Z = np.zeros_like(Y1)
for i in range(Y1.shape[0]):
    for j in range(Y1.shape[1]):
        mu = np.array([Y1[i, j], Y2[i, j]])
        z = (stats.norm.ppf(p) - np.dot(R, mu)) / b
        p_adjusted = stats.norm.cdf(z)
        pmf_vals = gpb_pmf(p_adjusted, ell_vec)
        pmf = pmf_vals / pmf_vals.sum()
        cdf = np.cumsum(pmf)
        quantile_index = np.searchsorted(np.arange(len(pmf)), threshold)
        Z[i, j] = (1 - cdf[quantile_index]) * np.exp(-0.5 * mu.dot(mu))

dy1 = y1_space[1] - y1_space[0]
dy2 = y2_space[1] - y2_space[0]
bin_area = dy1 * dy2

# total mass under Z
total_mass = Z.sum() * bin_area

# normalized Z
Z = Z / total_mass
# --- Compute MVN density grid Z2 ---
```

Multifactor Model for Simulating the Test portfolios Chapter 6

Here we provide the code to simulate the losses with IS and AIS on the test portfolios. First, the model is initialized. Then, the following function describes the algorithm to really start generating the losses. It also calculates marginal losses.

Already a lot is optimized with vectorized calculations. For small portfolios, speed-up could be achieved by generating losses in chunks instead of using the for loop. Keep in mind that, in that case, updates with AIS have to be done vectorized. Such computation would mean that samples are realized, and shifts are computed over the iterations of one chunk.

Sampling specific losses (from sub-portfolios) is also possible, with the option to view corresponding IS weights. Here, speed up is possible with vectorization. For this, use segmentation matrix (sparse matrix) which, with an dot-product with the vector of obligors exposures would lead to an array of losses within segments.

```
class MultiFactorModel:

    def __init__(self,
                 l: np.array,
                 p: np.array,
                 R: np.array,
                 K: tuple,
                 sparse = False,
                 specific_add_ons = False,
                 unique_combos = None,
                 category_indices = None,
                 ):
        self.sparse = sparse


```

```
        self.m = len(l)
        self.K = K
        self.K_tot = np.sum(self.K)

        assert p.shape[0]==self.m
        assert R.shape[0]==self.m

        self.constant = 1/np.sqrt(2)
        self.l = l
        self.p = p
        self.x_d = sp.stats.norm.ppf(p)

        self.R_m = R[:,0]
        self.R_mT = self.R_m.T
        self.R_m_square = self.R_m**2
        self.R_m_sqrt = np.sqrt(1 - self.R_m_square)
        self.R_m_sqrt_numerator = 1/self.R_m_sqrt

        self.g_0 = self.K[0]
        self.g_1 = self.K[1]+1

        self.R_g = R[:,self.g_0:self.g_1]
        self.R_gT = self.R_g.T
        self.R_g_square = np.sum(self.R_g**2, axis=1)
        self.R_mg_sqrt = np.sqrt(1 - (self.R_m_square+self.R_g_square))
        self.R_mg_sqrt_numerator = 1/self.R_mg_sqrt

        self.s_0 = self.K[0]+self.K[1]
        self.s_1 = self.K[0]+self.K[1]+self.K[2]+1

        self.R_s = R[:, self.s_0: self.s_1]
        self.R_sT = self.R_s.T
        self.R_s_square = np.sum(self.R_s**2, axis=1)
        self.R_ms_sqrt = np.sqrt(1 - (self.R_m_square+self.R_s_square))
        self.R_ms_sqrt_numerator = 1/self.R_ms_sqrt

        self.R_tot_sqrt = np.sqrt(1-(self.R_m_square+self.R_g_square+self.R_s_square))
        self.R_tot_sqrt_numerator = 1/self.R_tot_sqrt

        if sparse:
            self.R_mT = csr_array(self.R_mT)
            self.R_gT = csr_array(self.R_gT)
            self.R_sT = csr_array(self.R_sT)

        self.specific_add_ons = specific_add_ons
        if specific_add_ons:
            self.unique_combos = unique_combos
            self.n_unique_combos = len(unique_combos)
            self.category_indices = category_indices
```

```
def sim_AIS(
            self,
            N: int,
            random_state: int | None | np.random.RandomState,
            mu_vec_initial: np.ndarray,
            theta_vec_initial: np.ndarray,
            learning_rate_mu: float,
            mu_ind_adapt: np.ndarray,
            q_thres: float,
            eta: float,
            beta: float,
            delta: float,
            B_0: float,
            B_rate: float,
            marginal_add_ons: False,
    ):
        # Set random seed
        if random_state is not None:
```

```
            np.random.seed(random_state)

        # Pre-allocate arrays
        if self.specific_add_ons:
            results_specific = [np.zeros((N, self.n_unique_combos)) for _ in range(len(
                self.K) + 2)]
            weights_specific = [np.zeros((N, self.n_unique_combos)) for _ in range(len(
                self.K) + 2)]

        results = [np.zeros(N) for _ in range(len(self.K) + 2)]
        weights = [np.zeros(N) for _ in range(len(self.K) + 2)]


        # Initialize adaptation
        mu_vec = mu_vec_initial.copy().astype(float)
        theta_vec = theta_vec_initial.copy().astype(float)
        mu_array = np.zeros((N, len(mu_vec[mu_ind_adapt]))) if learning_rate_mu > 0
            else None


        exceed_count = 0
        counts_exceeded_threshold = 0

        # Main loop with progress bar
        for i in tqdm(range(N), desc="Simulation␣Progress"):
            # Draw samples
            Y = multivariate_normal.rvs(mean=mu_vec)
            Z = sp.stats.norm.rvs(loc=theta_vec, size=self.m)

            # Compute the total input X
            rY_m = Y[0] * self.R_m
            rY_g = Y[self.g_0:self.g_1] @ self.R_gT
            rY_s = Y[self.s_0:self.s_1] @ self.R_sT
            X_tot = rY_m + rY_g + rY_s + self.R_tot_sqrt * Z

            # Compute loss and weight
            if self.specific_add_ons:
                L_list = self.l * (X_tot < self.x_d)
                grouped_L = np.zeros(self.n_unique_combos)
                np.add.at(grouped_L, self.category_indices, L_list)
                L = grouped_L.sum()

                log_w = np.zeros(self.K_tot)
                log_w[mu_ind_adapt] = (-mu_vec[mu_ind_adapt] * Y[mu_ind_adapt]
                                       + 0.5 * mu_vec[mu_ind_adapt] ** 2)
                w_grouped = np.array([np.exp(log_w[comb].sum()) for comb in self.
                    unique_combos])
                w = np.exp(log_w.sum())

                weights_specific[4][i] = w_grouped
                results_specific[4][i] = grouped_L
                weights[4][i] = w
                results[4][i] = L
            else:
                L = np.sum(self.l * (X_tot < self.x_d))
                log_w_mu = (-mu_vec[mu_ind_adapt] @ Y[mu_ind_adapt]
                            + 0.5 * (mu_vec[mu_ind_adapt] ** 2).sum())

                w = np.exp(log_w_mu)

                weights[4][i] = w
                results[4][i] = L

                if marginal_add_ons:
                    RHS_m = (self.x_d - rY_m) * self.R_m_sqrt_numerator
                    RHS_mg = (self.x_d - rY_g - rY_m) * self.R_mg_sqrt_numerator
                    RHS_ms = (self.x_d - rY_s - rY_m) * self.R_ms_sqrt_numerator

                    norm_RHS_m = norm.cdf(RHS_m)
                    norm_RHS_mg = norm.cdf(RHS_mg)
                    norm_RHS_ms = norm.cdf(RHS_ms)
```

```
                    X_mz_chunk = rY_m + self.R_m_sqrt * Z

                    results[0][i] = np.sum(self.l * norm_RHS_m)
                    results[1][i] = np.sum(self.l * (X_mz_chunk < self.x_d))
                    results[2][i] = np.sum(self.l * norm_RHS_mg)
                    results[3][i] =  np.sum(self.l * norm_RHS_ms)


            # Adaptation step
            if learning_rate_mu > 0:
                if L > q_thres:

                    w2 = w ** 2
                    gamma =  eta / (beta+delta*exceed_count)
                    Bound = B_0 + B_rate*np.log(counts_exceeded_threshold+1)


                    y_diff = Y[mu_ind_adapt] - mu_vec[mu_ind_adapt]
                    mu_diff = gamma * w2 * y_diff
                    if np.linalg.norm(mu_vec[mu_ind_adapt]+mu_diff) <= Bound:
                        mu_vec[mu_ind_adapt] += mu_diff.astype(mu_vec.dtype)

                    else:
                        counts_exceeded_threshold += 1
                        mu_vec = mu_vec_initial.copy()


                mu_array[i] = mu_vec[mu_ind_adapt]



        # Return results
        if self.specific_add_ons:
            return results, weights, mu_array, results_specific, weights_specific
        return results, weights, mu_array, counts_exceeded_threshold
```

B.1. Additional results

Here we show the specific losses for scenario 2 and 3, with M-IS. The high impact losses (summing up to loss above threshold) are marked blue.

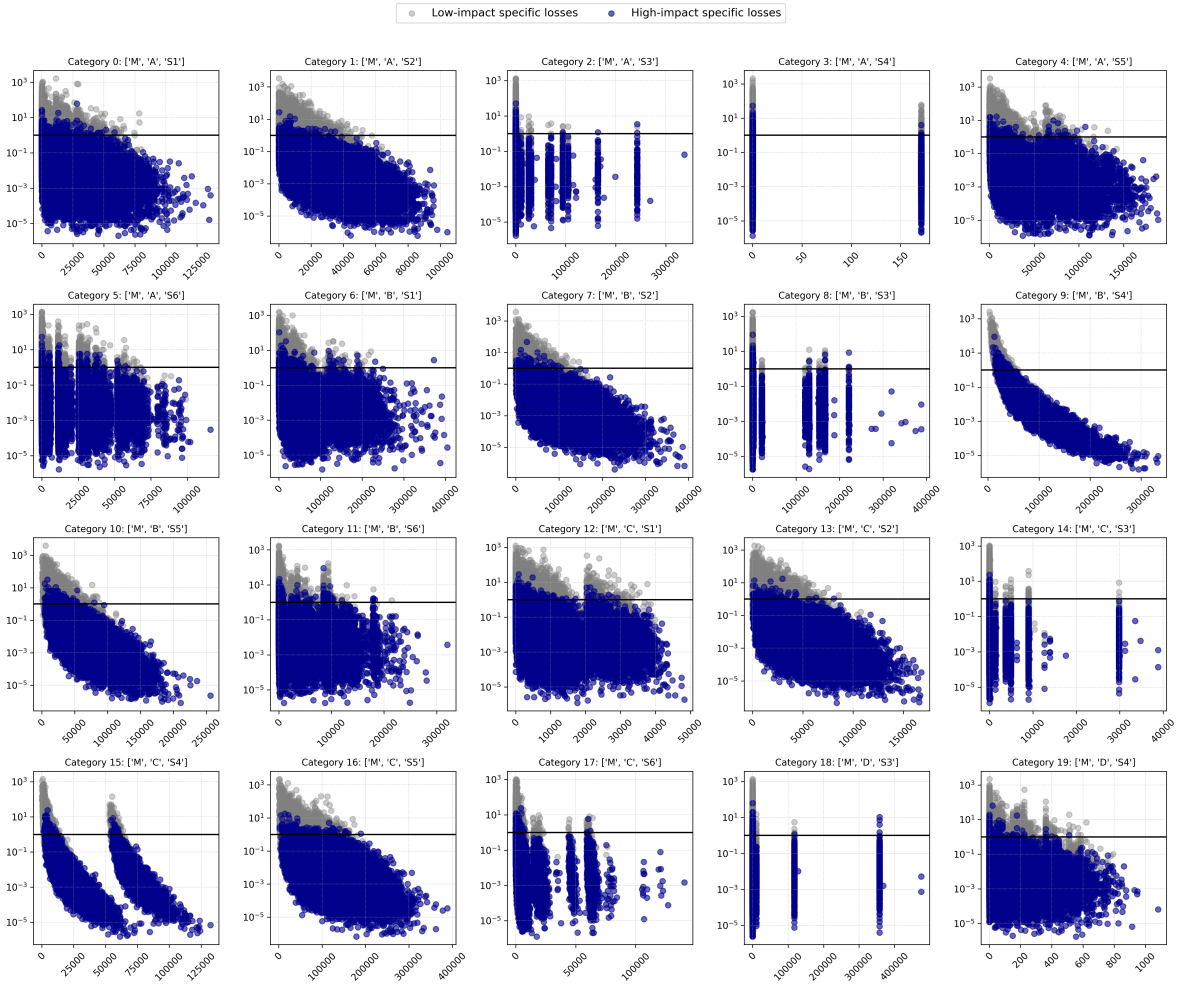


Figure B.1: Scenario 2. Loss vs IS weight, high impact losses marked blue, under M-IS.

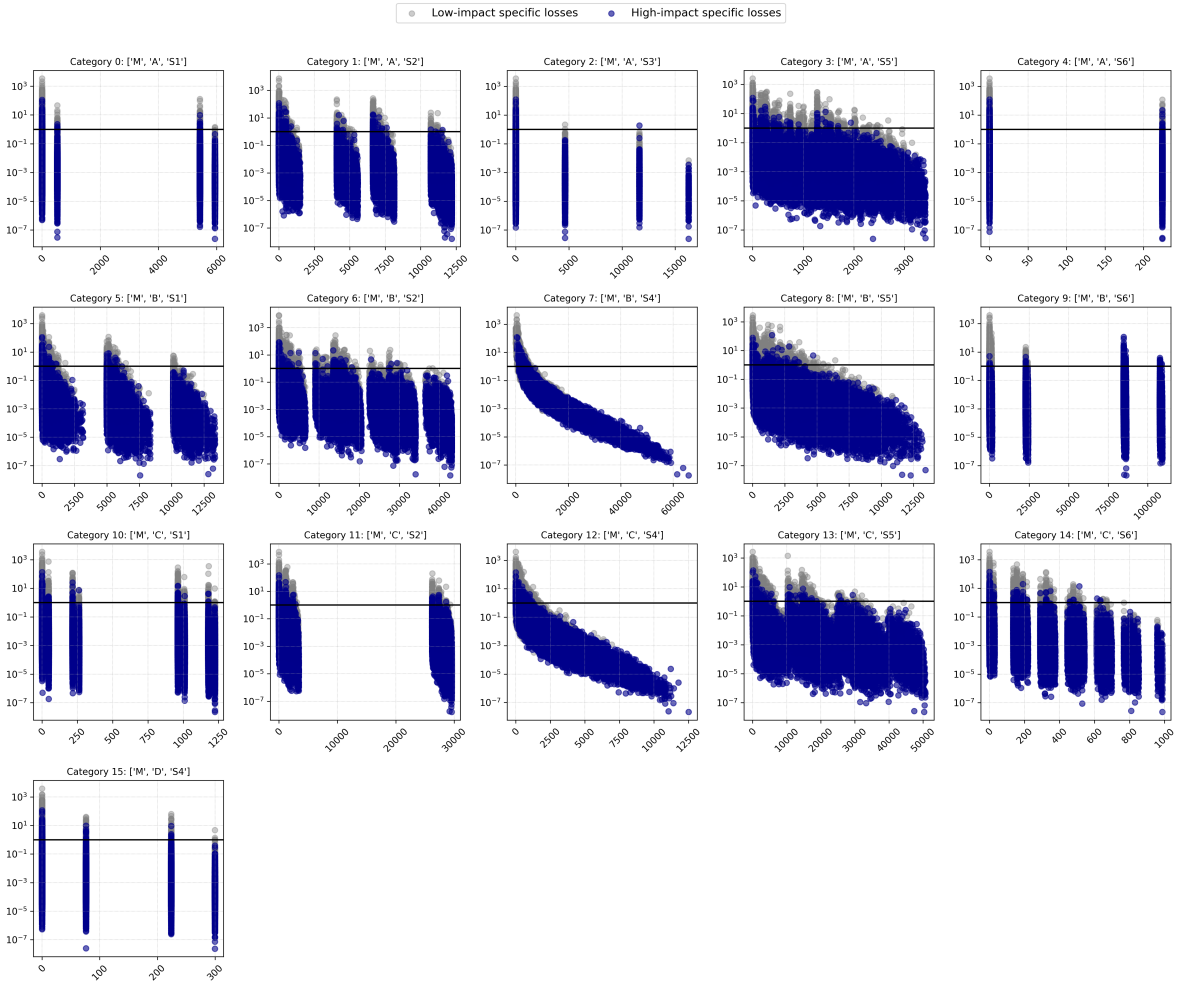


Figure B.2: Scenario 3. Loss vs IS weight, high impact losses marked blue, under M-IS.