Variance Reduction by the COS Method for PFE (Potential Future Exposure)

MSc Thesis

Matt van Kempen

TU Delft

QUANT

Variance Reduction by the COS Method for PFE (Potential Future Exposure)

by

Matt van Kempen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday October 16, 2024 at 13:00.

Student number:	4655583	
Project duration:	January 15, 2024 – October 15, 2024	
Thesis committee:	Dr. F. Fang,	TU Delft, Supervisor
	Dr. X. Shen,	FF Quant Advisory, Supervisor
	Prof. dr. ir. C. Vuik,	TU Delft
	Dr. A.F.F. Derumigny	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.
Cover Image created with Microsoft Copilot Designer

Abstract

Counterparty Credit Risk (CCR) refers to the risk that a counterparty involved in a financial contract will default before the final settlement of the contract, resulting in unrealized financial gains. One risk measure for managing counterparty credit risk is the Potential Future Exposure (PFE). The PFE is defined as the 97.5%-quantile of the exposure distribution. The traditional numerical method for computing the PFE is the Monte Carlo simulation method.

[1] introduced a semi-analytical method based on Fourier-cosine series expansion, shown to produce PFE estimates with at least five times the accuracy of the Monte Carlo simulation in one-tenth of the CPU time, when the trades in the portfolio are driven by 3 risk factors. It was an extention of the COS method [4] initially developed for option pricing. In this thesis, we refer to the method in [1] as the COS-PFE method.

This thesis focuses on utilizing the COS-PFE method to reduce the variance of the Monte Carlo simulation. More specifically, our contribution to the existing literature has three folds:

- We included security financed trades (SFTs), which are trades collateralized by bonds usually, in the COS-PFE framework.
- We developed a new Control Variate method that uses the COS-PFE result as the control variate in the Monte Carlo simulation.
- We developed two new Importance Sampling methods, of which the auxiliary density function for importance sampling is found based on the COS-PFE result.

The control variate method uses the COS method to retrieve the CDF of the portfolio's exposure from which an auxiliary variable, used as the control variate in the Monte Carlo simulation, are sampled using inverse sampling. The first adaptive importance sampling method we developed shifts the joint probability distribution of the driving risk factors, such that more samples around our target quantile are generated. The second adaptive importance sampling algorithm uses the cross-entropy method. To find the probability distribution that minimizes the Kullbeck-Leibler divergence between itself and the zero-variance estimator. The initial guess of the PFE value needed in this procedure is generated by the COS-PFE method. Furthermore, we further improve the latter method by a dimension-reduction approximation in the exact COS-PFE method and by splitting the portfolio into sub-portfolios, each involving fewer risk factors.

These methods are applied to portfolios containing IR and FX derivatives, both with and without collateral in the form of Security Financed Trades. The portfolio's value depends on the domestic short rate, foreign short rate and the exchange rate of the currency pair. The short rates are modelled under the one-factor Hull-White model, the exchange rate follows the geometric Brownian Motion model. The collateral value depends on the domestic short rate and the Z-spread, the latter modelled using geometric Brownian Motion.

The portfolios on which these methods are tested contain 100, 1000 and 10000 derivatives. The control variate method was tested on the test portfolio with 100 derivatives without collateral. The tests concluded that the variance of the PFE values obtained from the control variate method was approximately the same as the variance found using Monte Carlo. For the expected exposure, the control variate method demonstrated the ability to produce a variance reduction given a high enough correlation between the Mark-to-Market value of the portfolio and the control variate.

The adaptive importance sampling method that finds an optimal shift in the driving risk factors' joint distribution successfully reduced the variance of the PFE estimates. Testing on the portfolio with 100 derivatives and collateral found that the variance of the PFE estimates produced by this algorithm

was, on average, 3.5 times lower than the variance found using the straight forward Monte Carlo simulation. However, the algorithm required between 96 and 1773 seconds to produce a PFE estimate.

The adaptive importance sampling method using the cross-entropy approach was tested on all three testing portfolios, both with and without collateral. For portfolios containing 100, 1000 and 10000 derivatives without collateral, the variance of the PFE estimates were on average 35.4, 38.6 and 37.2 times smaller compared to straight forward Monte Carlo simulation, respectively. For the cases with collaterals, the cross-entropy method produced PFE estimates with a variance, on average, 36.3 times lower than the straight forward Monte Carlo simulation. When we switched on the dimension-reduction approximation in the COS-PFE calculations, the impressive performance remains: the variance reduction ratios is about 32.9. Then we further split the portfolio into sub-portfolios, to mimic the real-world situation whereby the exact COS-PFE calculation is suitable for sub-portfolios involving a low number of risk factors. Tests on all three testing portfolios with collateral showed that the variances were on average 31.1, 32.4 and 26.6 times smaller than straight forward Monte Carlo simulation, which is still significant, while the CPU time is greatly reduced. Worth noting that, the variance reduction ratios from the numerical tests match very well our theoretical results on the variance reduction ratio. Translating the significantly reduced variance into accuracy and CPU time. The cross-entropy adaptive importance sampling method combined with COS-PFE produced PFE estimates with much higher accuracy than straight forward Monte Carlo simulation in equal time. We conclude that this method produces more accurate PFE estimations with a significant lower variance than straight forward Monte Carlo simulation within the same CPU time.

Contents

1

Introduction

Counterparty Credit Risk (CCR) refers to the risk that a counterparty involved in a financial contract will default before the final settlement of the contract, resulting in unrealized financial gains. A prime example of poorly managed CCR is the global financial crisis of 2008. During this crisis, the opacity of the over-the-counter (OTC) derivative market, combined with inadequate regulation, led to significant risk mismanagement. The collapse of major institutions like Lehman Brothers and AIG, which were deeply interconnected with other financial entities through their role in the OTC derivative market, triggered widespread counterparty defaults and overall panic in the financial system. This global crisis spurred significant regulatory developments, such as the Dodd-Frank Wall Street Reform and Consumer Protection Act, and Basel III.

One risk measure for managing counterparty credit risk is the Potential Future Exposure (PFE). The PFE is defined as the 97.5%-quantile of the exposure distribution. Currently, the PFE is evaluated using Monte Carlo simulation of the portfolio's underlying risk factors. The Mark-to-Market values driven by different simulated scenarios of the risk factors are calculated. These Mark-to-Market values are then used to calculate exposures, from which the exposure distribution is constructed, allowing for the approximation of the PFE. An illustration of this is given in 1.1.

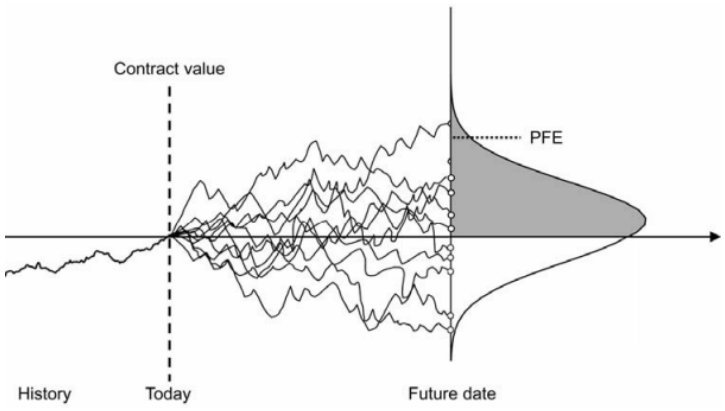

Figure 1.1: An example of the approximation of the PFE [2].

With the continuous growth of the total OTC market, effective regulation and risk management are crucial. Interest rate and foreign exchange derivatives constitute the majority of the OTC market. To illustrate, in the first half of 2023, the total notional amounts of interest rate and foreign exchange derivatives reached $694 trillion, where the total OTC market reached $ 715 trillion [3]. Recent research demonstrates that Fourier expansions can be employed to quickly and precisely calculate the PFE of an uncollateralized portfolio comprising interest rate and FX derivatives [1].

In this thesis, the method found in [1] is combined with the Monte Carlo simulation to reduce the variance of the PFE estimation. The portfolios that will be used for testing contain 100, 1000 and 10000 derivatives. Specifically, these portfolios consist of Forward Rate Agreements, Interest Rate Swaps, FX Forward Contracts and Cross-Currency Swaps. The portfolios are evaluated both with and without collateral. Note, in this thesis the we will focus on non-CSA trades. By "collateralized portfolio" we refer to Security Financed Trades. CSA trades will addressed in subsequent research. a

2

Preliminaries

Major events in the financial sector, such as the global financial crisis, have significantly increased the importance of CCR management. This section provides a literature review, beginning with the importance of CCR management for interest rate derivatives and foreign exchange derivatives, followed by new research in the field of CCR management. This is followed by a review of the variance reduction techniques discussed in this thesis, specifically the control variate method, importance sampling and adaptive importance sampling. The review is concluded with an examination of the applications of importance sampling in finance.

Currently, the Potential Future Exposure is calculated using Monte Carlo simulations of the portfolio's underlying risk factors. From these simulations, an exposure distribution can be formulated from which the Potential Future Exposure can be deduced. Recent research by Fang, Shen and Mast [1] demonstrated the application of the COS method [4] to portfolios containing interest rate derivatives and foreign exchange derivatives with varying amounts of derivatives, as well as general linear derivatives, to efficiently calculate the PFE of the portfolio. The applicability of this method to portfolios containing these type of derivatives is highly useful, as interest rate and foreign exchange derivatives constitute the vast majority of derivatives traded globally. To illustrate, in the last quarter of 2022, the European OTC derivative market's total notional amount was composed of 78% interest rate derivatives and 14% foreign exchange derivatives [5]. The remaining portion consisted of commodities, equity and credit derivatives. The composition of OTC derivatives in the global market can be seen in Figure 2.1.

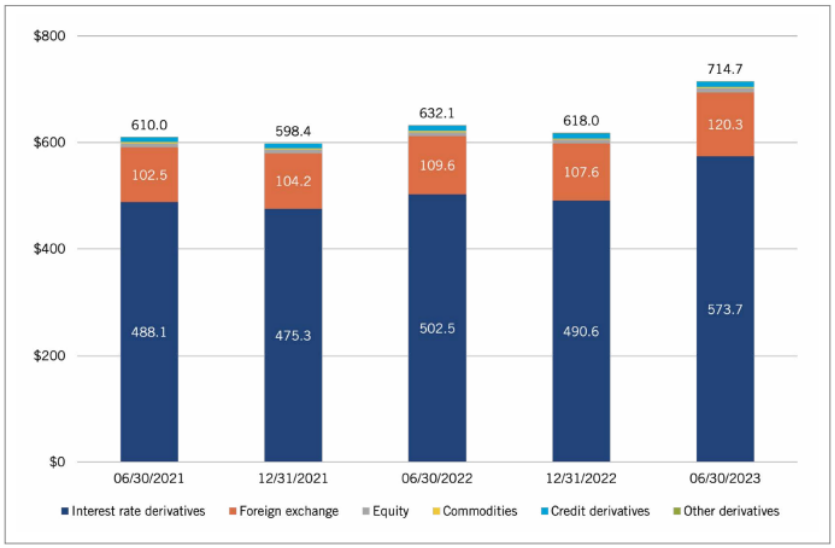

Figure 2.1: Outstanding notional (in US$ trillions) of various OTC derivatives in the global market [6].

Thus, as the vast majority of portfolios consist of interest rate derivatives and foreign exchange derivatives, the method described in [1] can be utilized to rapidly calculate the PFE of most portfolios in practice. The goal of this thesis is to build on this work and integrate it into the original Monte Carlo simulations as tool to reduce it's variance in PFE approximations, as currently no literature can be found on this topic. First, the control variate method will be investigated as a variance reduction technique. Afterwards, the importance sampling method, specifically adaptive importance sampling (AIS), will be discussed.

There is vast literature to be found regarding the control variate method. While the greater part of the literature focuses on the calculation of the expectation, for example [7], a handful focus on quantile estimation. In particular [8], [9] provide an in-depth tutorial on the application of quantile estimation. In literature the control variate has not been found to be applied to Potential Future Exposure calculations. However, in [10] and [11] the control variate method is applied to a Value-at-Risk calculation using a delta-gamma approximation of the portfolio's loss. In the latter, the variance reductions of the control variate method were compared with those of importance sampling and importance sampling using stratification. It was concluded that, although the control variate method reduces the variance compared to the original Monte Carlo, importance sampling and importance sampling with stratification perform better.

The subject of importance sampling was first introduced in [12] where it was used to estimate the probability of nuclear particles penetrating shields. For quantile estimations, importance sampling can be a highly effective technique for reducing variance. In [9] and [7] a detailed description is provided how this is done. Furthermore, [13] and [14] delve deeper into the topic of importance sampling, exploring its variance reduction potential and the adaptive importance sampling method. This method is recursive, utilizing an algorithm to find the most optimal sampler. An early form of adaptive importance sampling is found in [15] where multiple importance sampling (MIS) is introduced. This method combines samples from different distributions to reduce variance. This work highlighted the benefits of optimizing sampling strategies and thus laid the groundwork for adaptive approaches.

Adaptive importance sampling methods can be subdivided in three classes: resampling methods, moment matching methods, and methods that use an independent adaptive process [16]. Firstly, resampling methods eliminate samples with insufficient weight, thereby removing samples that will lead to weight degeneracy [17]. An example of an adaptive importance sampling scheme using resampling is the deterministic mixture population Monte Carlo [18]. Secondly, moment matching methods iteratively fit the moments of a proposed auxiliary density to estimate the moments of the target density. An example of this is given in [19]. The third type of adaptive importance sampling methods are those which make use of independent adaptive processes. These processes vary from method to method. Examples of such methods are found in [20] and [21], where a gradient descent method is applied on the random samples to find the optimal set of parameters for the auxiliary density.

The Cross-Entropy (CE) method, introduced in [22], is another example of the third type of adaptive importance sampling methods. The method is based on minimizing the Kullback-Leibler divergence between the auxiliary density and the zero-variance optimizer. One significant advantage of the CE method is that for some families of distributions the optimizer can be found using an analytical solution. This eliminates the need for an optimization scheme like a gradient descent. Moreover, for distribution families where no analytical solution exists a gradient descent can be used to find the optimizer, as the Kullback-Leibler minimization problem is generally concave [23]. Besides having a wide range of applications, such as combinatorial optimization [24], continuous optimization [25], optimal policy search [26], and multidimensional independent component analysis [27], the method is also applicable to rare-event estimation problems.

The Cross-Entropy method's convergence depends on samples being generated beyond a rare event. Therefore, if insufficient samples are generated beyond this level, the algorithm suffers from a lack of convergence [24]. If this is the case a multi-level Cross-Entropy method is used. Unfortunately, the multi-level CE method fails in specific high-dimensional settings, as the auxiliary density found by the method is not optimal [28]. To counter this problem several algorithms have been proposed. Examples

include the bottleneck screening algorithm [29] and the CE-$m^\star$ and iCE-$m^\star$ algorithms found in [23]. The bottleneck screening algorithm only considers the parameters of importance in the optimization, which decreases the dimensionality. The CE-$m^\star$ and iCE-$m^\star$ methods optimize the covariance matrix in only one direction, which significantly decreases the number of parameters to be optimized, thereby solving the problem. Furthermore, [28] proposes a remedy for this problem by changing the sampling density to be independent of the parameters found using the multi-level Cross Entropy method. In this way finding a non-optimal parameter due to weight degeneracy will not influence the subsequent parameters found. Note that our CE method, when applied to the calculation of the Potential Future Exposure of the testing portfolios, converges in one step given a sufficient number of paths. As a result, the multi-level procedure is not needed. Consequently, the Cross-Entropy method gives an accurate estimator even in high dimensions [28].

In both [30] and [28], the CE method is compared to different adaptive importance sampling methods. The authors conclude that their examples suggest that the importance sampling densities obtained by both methods are at least asymptotically similar. However, because the CE estimators are generally easier to obtain, it is concluded that the CE method is more optimal.

In finance, importance sampling has been extensively researched for applications such as the pricing of options [31] and securities [32], where expected payoffs need to be calculated. Adaptive importance sampling has also been applied to the calculation of the Value-at-Risk and expected shortfall. Examples of this are [11], [33], and [34]. In [11], importance sampling is applied to the delta-gamma approximations of a portfolio's fluctuations in value to reduce the variance of the Value-at-Risk approximation. In [33], a new importance sampling method was applied to a copula-based credit-loss model. Finally, in [34], the authors use an adaptive importance sampling method to forecast the Value-at-Risk and expected shortfall in a Bayesian framework.

2.1. Fundamental definition of stochastic calculus

The following definitions and theorems are used in stochastic calculus and are applied throughout this thesis.

Definition 2.1.1 (Filtration). *A filtration on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a collation* $(\mathcal{F}_t)_{0\le t\le\infty}$ *indexed by* $[0,\infty]$ *of sub-*σ*-fields of* $\mathcal{F}$*, such that* $\mathcal{F}_s \subset \mathcal{F}_t$ *for every* $s \le t \le \infty$.

Definition 2.1.2 (Brownian motion). *A real-valued process* $\{W(t) : t \ge 0\}$ *is called a Brownian motion if:*

1. *Starting at 0:* $W(0) = 0$.
2. *Normality of increments:* $\forall s, t$ *such that* $0 \le s \le t$, $W(t) - W(s) \sim \mathcal{N}(0, t-s)$.
3. *Independent increments: For* $0 \le t_0 < \cdots < t_n$, *the random variables* $Y_i = W(t_i) - W(t_{i-1})$, $i = 1, \ldots, n$ *are independent.*
4. *Continuous trajectories: The map* $t \to W(t)$ *is continuous,*

where the usual filtration is $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$.

Definition 2.1.3 (Martingale). *The process* $\{M_t : t \ge 0\}$ *is a* $(\mathcal{F}_t)_{t\ge0}$ *if*

1. *Adapted:* M_t *is* $\mathcal{F}_t$ *measurable for all* $t \ge 0$.
2. *Integrable:* M_t *is integrable for all* $t \ge 0$.
3. *Martingale property:* $\forall s, t$ *such that* $0 \le s \le t$

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s. \tag{2.1}$$

Before defining a semimartingale, two preliminary definitions must be stated.

Definition 2.1.4 (Local martingale). *A process* M *is a local martingale if there exists an increasing sequence of stopping times* $(\tau_n)_n^\infty$, *where* $\tau_n \to \infty$ *almost surely, such that the stopped process* $(M_{\min(\tau_n,t)})_{t\ge0}$ *is a martingale.*

Definition 2.1.5 (Finite variation). *A process* A *has finite variation if for every* $\omega \in \Omega$, *the path* $A_t(\omega)$ *has finite variation for each finite* $[0,t]$. *Or,*

$$V_{A_t(\omega)}[0,t] = \sup\left\{\sum_{i=1}^{\tau} \left|A_{t_k}(\omega) - A_{t_{k-1}}(\omega)\right| : 0 \le t_0 < \cdots < t_\tau \le t\right\} < \infty. \tag{2.2}$$

Definition 2.1.6 (Semimartingale). *A semimartingale* $S = ()S_t)_{t\ge0}$ *is a càdlàg, adapted process of the form*

$$S_t = S_0 + M_t + A_t \tag{2.3}$$

where $t \ge 0$. *Here* S_0 *is finite and* $\mathcal{F}_0$ *measurable,* $(M_t)_{t\ge0}$ *is a local martingale with* $M_0 = 0$ *and* $(A_t)_T t \ge 0$ *is a with* $A_0 = 0$ *and has finite variation.*

Definition 2.1.7 (Itô integral). *For any square-integrable adapted process* $f(t)$ *with continuous sample paths, we can define the stochastic integral, also known as the Itô integral, by*

$$I(T) = \int_0^T f(s)dW(s) = \lim_{n\to\infty} I_n(T) \tag{2.4}$$

in L^2. *Where* $I_m(T) = \int_0^T f_m(s)dW(s)$ *for some elementary process* $f(t) = \sum_{j=0}^{n-1} \eta_j \left(W(t_{j+1}) - W(t_j)\right)$ *satisfying*

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (f_m(s) - f(s))^2 ds\right] = 0. \tag{2.5}$$

Here η_j *is* $\mathcal{F}_{t_j}$ *measurable for all* $j = 0, \ldots, n-1$ *and square integrable.*

Definition 2.1.8 (Itô's formula). *Let* $f \in C^2(\mathbb{R})$ *and consider a continuous semimartingale* $S = M + A$ *where* M *is a local martingale and* A *is a process of finite variation. Then* $(f(S_t))_{t\ge0}$ *is also a semimartingale, and*

$$f(S_t) = f(S_0) + \int_0^t \frac{\partial f}{\partial s}(S_u)dS_u + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial s^2}(S_u)d[S]_u. \tag{2.6}$$

Here the first and second partial derivatives are with respect to the considered semimartingale, $[S]$ *is the quadratic variation of the stochastic process* $(S_t)^{t\ge0}$.

2.2. Liquid derivatives and their pricing formulas

As previously stated, the test portfolios contain interest rate derivatives and foreign exchange derivatives. This section introduces the four types of derivative products considered and explains how they are priced. The section begins with basic definitions and derivations used in valuation. Following this, the interest rate derivatives, specifically the Forward Rate Agreement and Interest Rate Swap, are discussed. The section will conclude with a discussion on foreign exchange derivatives, specifically the FX Forward Agreement and the Cross-Currency Swap.

2.2.1. Basic definitions

Before discussing the valuation of derivatives, it is necessary to define some concepts. We will start with the definition of the money-market account, or bank account, which is a risk-free investment that grows continuously at a risk-free rate.

Definition 2.2.1 (Money-market account)**.** *Define $B(t)$ as the value of the money-market account at time $t \geq 0$. Assume that $B(0) = 1$ and that the account evolves according to*

$$dB(t) = r_t B(t)dt,\ B(0) = 1, \tag{2.7}$$

where r_t is a positive function of time. From this equation, we find that

$$B(t) = e^{\int_0^t r_s ds}. \tag{2.8}$$

Where r_t is the instantaneous rate, or short rate, with which the bank account grows.

In the definition above, the term 'instantaneous rate' arises from the fact that the bank account grows over the time interval $[t, t + \Delta t)$ by $r_t \Delta t$, as shown in the equation 2.9.

$$\frac{B(t + \Delta t) - B(t)}{B(t)} = r_t \Delta t. \tag{2.9}$$

The zero-coupon bond (ZCB) is a tool used in the valuation of derivatives. A ZCB with maturity T is a contract that represents the present value of one unit of currency to be paid at the bond's maturity.

Definition 2.2.2 (Zero-coupon bond)**.** *A T-maturity zero-coupon bond is a contract that guarantees its holder the payment of one unit of currency at time T, or $P(T, T) = 1$. The contract is characterized by having no intermediate payments. The contract value at time $t < T$ is denoted by $P(t, T)$.*

Next is the simply compounded spot interest rate. An example of this rate is the LIBOR (London Interbank Offered Rate). This rate is an interbank rate that facilitates the exchange of deposits and swap transactions among banks.

Definition 2.2.3 (Simply-compounded spot interest rate)**.** *This rate $L(t, T)$ at time t for the maturity T, is the constant rate at which an investment $P(t, T)$ must grow to produce one unit of currency at maturity T. $L(t, T)$ is calculated using the following expression*

$$L(t, T) = \frac{1 - P(t, T)}{\tau(t, T)P(t, T)}, \tag{2.10}$$

where $\tau(t, T)$ is the time difference $T - t$.

2.2.2. Interest rate derivatives

First, we will discuss the interest rate derivatives contained in the test portfolios. Interest rate derivatives are financial contracts whose value is derived from interest rates. The MtM value of these contracts can be calculated by discounting the future cash flows using the money-market account.

The first interest rate derivative to be discussed is the Forward Rate Agreement (FRA). The FRA is a contract that provides its holder with an interest rate payment for the period from T to S. Here, T is the expiry, the time point a floating rate $L(T, S)$ is agreed upon, and S is the maturity of the contract, where $T < S$. At maturity, a fixed payment based on a fixed rate K will be exchanged against a floating

payment based on the spot rate $L(T,S)$.

From [35] we find that the value of such a contract at time t is

$$V_{FRA}(t,T,S,\tau(T,S),N,K) = N\left[P(t,S)\tau(T,S)K - P(t,T) + P(t,S)\right]. \tag{2.11}$$

To make the Forward Rate Agreement a fair contract at time t it must be true that $V_{FRA}(t,T,S,\tau(T,S),N,K) = 0$. From this, the following definition can be formulated.

Definition 2.2.4 (Simply-compounded forward interest rate)**.** *The simply-compounded forward interest rate at time t for expiry T and maturity S is*

$$F(t;T,S) = \frac{1}{\tau(T,S)}\left[\frac{P(t,T)}{P(t,S)} - 1\right]. \tag{2.12}$$

The value of $F(t;T,S)$ is the value that, when substituted for K, makes the FRA contract a fair contract. This follows from

$$\begin{aligned} V_{FRA}(t,T,S,\tau(T,S),N,K) = N\left[P(t,S)\tau(T,S)K - P(t,T) + P(t,S)\right] = 0,\\ \tau(T,S)K = \frac{P(t,T) - P(t,S)}{P(t,S)},\\ K = \frac{1}{\tau(T,S)}\left[\frac{P(t,T)}{P(t,S)} - 1\right]. \end{aligned}$$

In the case where the maturity S approaches the expiry T the instantaneous forward rate is found.

Definition 2.2.5 (Instantaneous forward interest rate)**.** *The instantaneous forward interest rate at time t for maturity T is found by*

$$\begin{aligned} f(t,T) &= \lim_{S\to T^+} F(t;T,S),\\ &= -\lim_{S\to T^+} \frac{1}{P(t,S)}\frac{P(t,S) - P(t,T)}{S-T},\\ &= -\frac{1}{P(t,T)}\frac{\partial P(t,T)}{\partial T},\\ &= -\frac{\partial \ln(P(t,T))}{\partial T}. \end{aligned} \tag{2.13}$$

Then we also get that

$$P(t,T) = e^{-\int_t^T f(t,u)du}. \tag{2.14}$$

The second interest rate derivative to be discussed is the Interest Rate Swap (IRS). An IRS is a contract in which counterparties exchange fixed-rate payments and floating-rate payments on a set of prespecified dates. At each date in the set $\{T_{\alpha+1},\ldots,T_\beta\}$, a fixed leg $N\tau(T_{i-1},T_i)K$ and a floating leg $N\tau(T_{i-1},T_i)L(T_{i-1},T_i)$ are exchanged, where the floating rate resets at T_{i-1} for each payment at T_i.

Using what is known about the FRA, it can be seen that the IRS can be seen as a portfolio of forward rate agreements with expiry dates T_{i-1} and maturity T_i for $i = \alpha+1,\ldots,\beta$. Using this the following

expression for V_{IRS} can be derived

$$\begin{aligned}
V_{IRS}(t,\mathcal{T},\tau,N,K) &= \sum_{i=\alpha+1}^{\beta} V_{FRA}(t,T_{i-1},T_i,\tau(T_{i-1},T_i),N,K),\\
&= N \sum_{i=\alpha+1}^{\beta} \tau(T_{i-1},T_i)P(t,T_i)\left[K - F(t;T_{i-1},T_i)\right],\\
&= N \sum_{i=\alpha+1}^{\beta} \tau(T_{i-1},T_i)KP(t,T_i),\\
&\quad - P(t,T_i)\tau(T_{i-1},T_i)\left[\frac{1}{\tau(T_{i-1},T_i)}\left(\frac{P(t,T_{i-1})}{P(t,T_i)} - 1\right)\right],\\
&= N \sum_{i=\alpha+1}^{\beta} \tau(T_{i-1},T_i)KP(t,T_i) - P(t,T_{i-1}) + P(t,T_i),\\
&= N\left[P(t,T_\beta) - P(t,T_\alpha) + \sum_{i=\alpha+1}^{\beta} \tau(T_{i-1},T_i)KP(t,T_i)\right]. \qquad (2.15)
\end{aligned}$$

2.2.3. Foreign exchange derivatives

A foreign exchange (FX) derivative is a financial contract whose value is derived from the exchange rate between two currencies. The exchange rate between two currencies at time t as $X(t)$. In this thesis, $X(t)$ can be viewed as the amount of domestic currency received for one unit of foreign currency. In the test portfolios the chosen currencies were USD and JPY. Similarly to the interest rate derivatives, the Mark-to-Market value of FX derivatives is calculated by discounting cash flows. This can be done in two ways: First, discount the foreign currency cash flows using the foreign money-market account, then convert to domestic currency using the exchange rate. Alternatively, convert the foreign currency cash flows to domestic currency first, then discount them using the domestic money-market account. Due to the absence of arbitrage, either approach can be used.

The first FX derivative discussed is the FX forward contract. This contract is similar to an FRA, but instead of exchanging a fixed interest rate for a floating one, a specified amount of domestic currency is exchanged for a specified amount of foreign currency at an agreed-upon exchange rate. Its value is calculated using,

$$V_{FX}(t,T,N^f,N^d) = N^f P^f(t,T)X(t) - N^d P^d(t,T). \qquad (2.16)$$

Here $t \leq T$, N^f and N^d are the amount of foreign and domestic currency exchanged at T, $X(t)$ is the exchange spot rate at time t, and $P^f(t,T)$ and $P^d(t,T)$ are the values of the foreign and domestic ZCB at time t with maturity T.

Similar to the simply-compounded forward interest rate of the FRA defined in Definition 2.2.4, it is also possible to determine an FX forward rate that makes the FX forward contract fair.

Definition 2.2.6 (FX forward rate)**.** *The FX forward rate $X_F(t,T)$ is the exchange rate that makes the FX forward rate contract fair. It can be calculated using*

$$X_F(t,T) = X(t)\frac{P^d(t,T)}{P^f(t,T)}. \qquad (2.17)$$

The expression for the FX forward rate is obtained by setting $V_{FX}(t,T) = 0$. Then, by using the FX rate to rewrite $N^d = N^f X(t)$ it can be derived that from

$$N^f P^f(t,T)X_F(t,T) - N^d P^d(t,T) = 0,$$

it follows that

$$X_F(t,T) = \frac{N^d P^d(t,T)}{N^f P^f(t,T)},$$
$$X_F(t,T) = \frac{N^f X(t) P^d(t,T)}{N^f P^f(t,T)},$$
$$X_F(t,T) = X(t)\frac{P^d(t,T)}{P^f(t,T)}.$$

The second type of foreign exchange derivative to be discussed is the Cross-Currency Swap (XCS). In this contract, the counterparties exchange an amount of foreign currency for an amount of domestic currency, with the exchanged amounts based on either a fixed or floating rate. When a fixed rate is exchanged for a floating rate, the XCS can be viewed as an IRS. However, the FX rate introduces an additional risk factor. The value of an XCS, where an amount of foreign currency with a fixed rate is swapped for an amount of domestic currency with a floating rate over a set of payment dates $\{T_{\alpha+1}, \ldots, T_\beta\}$ is

$$\begin{aligned} V_{XCS}(t, \mathcal{T}, N^f, N^d, K) = &\sum_{i=\alpha+1}^{\beta} N^f X(t) K \tau(T_{i-1}, T_i) P^f(t, T_i), \\ &- N^d L^d(t; T_{i-1}, T_i) \tau(T_{i-1}, T_i) P^d(t, T_i), \\ =&N^d \Bigg[P^d(t, T_\beta) - P^d(t, T_\alpha), \\ &+ X(0) \sum_{i=\alpha+1}^{\beta} X(t) K \tau(T_{i-1}, T_i) P^f(t, T_i) \Bigg]. \end{aligned} \tag{2.18}$$

Here, N^f and N^d represent the notional amounts in foreign and domestic currency, $X(t)$ is the FX rate at time t, K is the fixed rate, τ is the year fraction, and P^f and P^d are the foreign and domestic ZCBs.

2.3. Models for IR and FX risk factors

The test portfolios depend on three risk factors: the domestic short rate, the foreign short rate, and the exchange rate. This section discusses the models used to simulate these risk factors. Firstly, the Gaussian one-factor model used to generate the short rates is discussed. The model's solution will be given as well as an equation for valuing the zero-coupon bond. Next, the Geometric Brownian Motion model used to simulate the exchange rate is discussed. Additionally, its solution will be provided.

2.3.1. The IR Model

The Gaussian one-factor model (G1++) model used to model the short rate $r(t)$ using the following expression,

$$r(t) = x(t) + \beta(t). \tag{2.19}$$

Here, $\beta(t)$ is a deterministic function that is used to calibrate the model to the market and $x(t)$ is the shifted short rate which follows the stochastic differential equation (SDE),

$$\begin{aligned} dx(t) &= -ax(t)dt + \sigma dW(t), \\ x(0) &= 0. \end{aligned} \tag{2.20}$$

Here, a is the constant mean reversion coefficient, σ is the volatility, and $dW(t)$ is the incremental Wiener process.

Proposition 1. *In the G1++ model, it is derived that when $\beta(t)$ is calibrated to the market it follows that,*

$$\beta(t) = f^M(0,t) + \frac{\sigma^2}{2a^2}\left(1 - e^{-at}\right)^2. \tag{2.21}$$

The proof of Proposition 1 can be found in Proof A.1 in the Appendix.

Taking the integral of both sides of Equation 2.20 yields the solution

$$x(t) = e^{-a(t-s)}x(s) + \sigma \int_s^t e^{-a(t-u)}dW(u). \tag{2.22}$$

From this, we easily find the solution to Equation 2.19, namely,

$$r(t) = e^{-a(t-s)}x(s) + \sigma \int_s^t e^{-a(t-s)}dW(u) + \beta(t). \tag{2.23}$$

Then, for the domestic and foreign short rates, the solution 2.22 yields

$$x_d(t) = e^{-a_d t}x_d(0) + \sigma_d \int_0^t e^{-a_d(t-s)}dW_d(s), \tag{2.24}$$

$$x_f(t) = e^{-a_f t}x_f(0) + \sigma_f \int_0^t e^{-a_f(t-s)}dW_f(s). \tag{2.25}$$

2.3.2. The affine term structure

The shifted short rate in Equation 2.20 is an affine function because we can write the SDE as

$$dx(t) = \mu(x(t))dt + \Sigma(x(t))dW(t),$$

where

$$\begin{aligned} \mu(x(t)) &= -ax(t), \\ \Sigma(x(t)) &= \sigma. \end{aligned}$$

From [36], it is known that due to this affine term structure, we can write

$$P(t,T) = A(t,T)e^{-B(t,T)x(t)}. \tag{2.26}$$

This brings us to Proposition 2, which is proven in the Appendix under Proof A.2.

Proposition 2. *In the G1++ model, the ZCB price is formulated by the following equation*

$$P(t,T) = A(t,T)e^{-B(t,T)x(t)}, \tag{2.27}$$

where

$$A(t,T) = \frac{P^M(0,T)}{P^M(0,t)} e^{\frac{1}{2}(V(t,T)-V(0,T)+V(0,t))}, \tag{2.28}$$

$$B(t,T) = \frac{1-e^{-a(T-t)}}{a}. \tag{2.29}$$

Where

$$V(t,T) = \frac{\sigma^2}{a^2}\left(T-t-2\frac{1-e^{-a(T-t)}}{a}+\frac{1-e^{-2a(T-t)}}{2a}\right). \tag{2.30}$$

2.3.3. The FX Model

The FX rate, $X(t)$, is modelled using the Geometric Brownian Motion (GBM) model. The model is characterized by the following stochastic differential equation

$$dX(t) = \mu_X X(t)dt + \sigma_X X(t)dW(t). \tag{2.31}$$

Here, μ_X and σ_X are the drift and volatility of the exchange rate. From applying Itô's Lemma on $g(X(t)) = \log(X(t))$ it can be seen that

$$\begin{aligned} dg(X(t)) =& \frac{\partial g(X(t))}{\partial t}dt + \frac{\partial g(X(t))}{\partial X(t)}dX(t) + \frac{1}{2}\frac{\partial^2 g(X(t))}{\partial X(t)^2}dX(t)dX(t), \\ =& \frac{1}{X(t)}\left(\mu_X X(t)dt + \sigma_X X(t)dW_t^{\mathbb{P}}\right) - \frac{1}{2}\frac{1}{X(t)^2}\left(\sigma_X^2 X(t)^2 dt\right), \\ =& \left(\mu_X - \frac{1}{2}\sigma_X^2\right)dt + \sigma_X dW(t). \end{aligned}$$

Then taking the integral on both sides yields the solution of Equation 2.31,

$$\log(X(t)) = \log(X(0)) + \left(\mu_X - \frac{1}{2}\sigma_X^2\right)t + \sigma_X W(t). \tag{2.32}$$

2.4. Entropy and information theory

Information theory is a field focused on the quantification, accumulation, and exchange of information. Information theory was first introduced by Claude Shannon in his 1948 paper [37], where he defined information as a measurable quantity, which meets several axioms.

1. An event that occurs with probability 1 is perfectly unsurprising and yields no information.
2. The lower the probability of an event, the more surprising it is and the more information it yields.
3. For two independent events the total amount of information is the sum of the self-informations of the individual events.

Self-information is defined as the measure that quantifies the amount of information a variable holds. The self-information, or suprisal, is formally defined as,

Definition 2.4.1 (Self-information). *Let X be a continuous random variable that takes values in the set $\mathcal{X}$ and is distributed according to $f : \mathcal{X} \to \mathbb{R}$ with probability density function $f(x) = \mathbb{P}(X = x)$. The self-information of an event with probability $f(x)$ is*

$$I_X(x) = \log_b \left(\frac{1}{f(x)} \right). \tag{2.33}$$

The base b corresponds to different units of information. For example, if $b = 2$ the unit is the shannon, or bit. When $b = e$, the unit is the natural unit of information nat.

From the definition of self-information, it is evident that the information a random variable possesses is based on its probability. As the probability of an outcome decreases, the value $1/f(x)$ increases, and so does the self-information. Conversely, as the probability of an outcome increases, the value of $1/f(x)$ and the self-information decrease. Thus, the less probable an outcome, the more information it contains. In practice, a more interesting value is the entropy of a random variable [38]. For a discrete random variable the entropy is defined as follows.

Definition 2.4.2 (Entropy). *Let X be a discrete random variable taking values in the set $\mathcal{X}$ with probability mass function $f(x)$. The entropy $H(X)$ is*

$$H(X) = \mathbb{E}_f[I_X] = -\sum_{x \in \mathcal{X}} f(x) \log_b(f(x)). \tag{2.34}$$

Entropy, as defined in equation 2.34, measures the expected amount of information provided by the random variable to the observer. It can thus be seen as the average Shannon information weighted by the probability of all possible realizations of the random variable [39].

For a continuous random variable the entropy is defined as the differential entropy.

Definition 2.4.3 (Differential entropy). *Let X be a continuous random variable taking values in $\mathcal{X}$ with probability density function $f : \mathcal{X} \to \mathbb{R}$. The entropy $h(X)$ is*

$$h(X) = \mathbb{E}_f[I_X] = -\int_{\mathcal{X}} f(x) \log_b(f(x)) dx. \tag{2.35}$$

The concept of cross-entropy finds it's roots in the Kraft-McMillan theorem originally stated in [40].

Theorem 2.4.1 (Kraft-McMillan Theorem). *Let $l_1, \ldots, l_n$ be lengths of codewords in a code with n codewords using an alphabet of size D. Then a prefix-free code with these codeword lengths exists if and only if*

$$\sum_{i=1}^{n} D^{-l_i} \leq 1. \tag{2.36}$$

A prefix-free code is a code in which no codeword is a prefix of any other codeword. For example, encoding 'A' as 1 and 'B' as 11 is not a prefix-free code because receiving 111 can be decoded into various sequences containing 'A's and 'B's. The theorem implies that if $\mathcal{X}$ is the set of all possible messages that

can be encoded, an implicit probability density $g : \mathcal{X} \to \mathbb{R}$ can be constructed. The density g is defined as,

$$g(x_i) = D^{-l_i}. \tag{2.37}$$

Here, $x_i \in \mathcal{X}$ is a value that can be encoded with a codeword of length l_i using an alphabet containing D symbols. The length of a codeword corresponds with the frequency of the message. As the more frequent a message occurs, the shorter its codeword length, and thus the higher the probability. For example, using binary digits to encode the messages gives us $D = 2$, as we have either a 0 or 1. If a codeword has a length of 1, the probability of this codeword occurring is $2^{-1} = 0.5$. Note that this probability distribution is implicit as it only reflects the relative frequency of each codeword.

From equation 2.37 it is found that $l_i = \log_D(g(x_i))$. From this last expression, the concept of cross-entropy follows. It is the expected value of the lengths of the codewords used to encode messages from distribution f using an approximating distribution g.

$$\mathbb{E}_f[l] = -\mathbb{E}_f[\log_D(g(x))] = -\sum_{\mathcal{X}} f(x)\log_D(g(x)) = H(f, g).$$

Formally, cross-entropy is defined in the following way.

Definition 2.4.4 (Cross-entropy)**.** *The cross-entropy of a discrete distribution g relative to the discrete distribution f over the set of events $\mathcal{X}$ which follows the distribution g is*

$$H(f, g) = -\sum_{x \in \mathcal{X}} f(x)\log_b(g(x)). \tag{2.38}$$

Similarly, in the case that g and f are continuous probability distributions the cross-entropy is calculated by

$$H(f, g) = -\int_{\mathcal{X}} f(x)\log_b(g(x)). \tag{2.39}$$

In terms of codeword length, cross-entropy can be viewed as the expected extra number of bits needed to encode the events using distribution g instead of the optimal distribution f. It reflects how well the distribution g models the true distribution f. If g approximates f well, the cross-entropy will be close to the entropy of f.

The concept of cross-entropy is directly used in the Kullback-Leibler divergence. The Kullback-Leibler divergence measures the amount of information lost when using probability density g to approximate f. It was first established in [41]. This divergence is closely related to cross-entropy and differential entropy, as it represents the difference of the two.

Definition 2.4.5 (Kullbeck-Leibler divergence)**.** *Let X be a continuous random variable with probability density function $f : \mathcal{X} \to \mathbb{R}$. Then for probability density function $g(x)$ the Kullbeck-Leibler divergence is defined by,*

$$D(f, g) = H(f, g) - H(X) = \mathbb{E}_f\left[\log_b\left(\frac{f(X)}{g(X)}\right)\right] = \int f(x)\log_b\left(\frac{f(x)}{g(x)}\right)dx. \tag{2.40}$$

Then from equation 2.40 it can be seen that $D(f, g) \geq 0$, and $D(f, g) = 0$ if and only if $f = g$ almost everywhere. The base used in the logarithm of the Kullbeck-Leibler divergence can vary. For calculations in statistical interference and information theory the base e is mostly used. This is because it simplifies many calculations, particularly when dealing with distributions like those of the exponential family. Therefore, in this thesis the base e will also be used.

The convergence of g to f as the Kullbeck-Leibler divergence goes to zero can be proven using the following theorem.

Theorem 2.4.2 (Pinsker's inequality)**.** *If f and g are two probability distributions on $\mathcal{X}$, then*

$$\|f - g\|_{TV} \leq \sqrt{\frac{1}{2}D(f, g)}, \tag{2.41}$$

where $\|f - g\|_{TV}$ *is the total variation distance which is defined as*

$$\|f - g\|_{TV} = \frac{1}{2} \|f - g\|_1 . \tag{2.42}$$

Thus, if $D(f, g)$ goes to 0 then the total variation distance $\|f - g\|_{TV}$ also goes to 0, this implies convergence in probability.

3

Counterparty credit risk and its quantification

This section will cover the fundamental definitions involved in quantifying counterparty credit risk and Potential Future Exposure. A more thorough insight into these definitions can be found in [2]. Counterparty credit risk is quantified using the Mark-to-Market value (MtM) and exposure of a portfolio. The Mark-to-Market value is defined in the following way.

Definition 3.0.1 (Mark-to-Market)**.** *The Mark-to-Market indicates the potential loss or gain that could occur today in relation to a counterparty. At the portfolio level, the MtM value is calculated by summing the present values of all payments to be received from the counterparty and subtracting the sum of the present values of all payments to be made to the counterparty. Therefore, the MtM can be either a positive or negative number.*

The MtM value of a portfolio can be either positive or negative. In the case of a negative MtM value, the counterparty is still owed money. In the case of a positive MtM value, the counterparty still owes money. Thus, only when the MtM value of a portfolio is positive is the institution is subjected to CCR, as in the case of a counterparty default the cash flow is not received. The MtM of the unrealized cash flow is formally defined as the exposure of a portfolio.

Definition 3.0.2 (Exposure)**.** *The exposure of a contract can be defined as*

$$E(t) = \max\left(V(X(t)), 0\right) \tag{3.1}$$

where $V(X(t))$ is the MtM value of the contract under the risk factors at time t denoted by $X(t)$.

The CCR of a portfolio can be reduced in multiple ways. Two of these methods are posting collateral and using netting. Firstly, when using collateral, the institution can retain the collateral in the event of a default. The use of collateral will be discussed later in this thesis. Secondly, the exposure of a portfolio can be reduced by using netting. This method allows the counterparty to offset the MtM value of one trade with that of another. For example, in the case of two trades, A and B, with corresponding MtM values 10 and -5, the exposure of the trades without netting would be the sum of the exposure values of the trades, which is 10. With netting, the exposure value of trades A and B is the exposure value of the sum of the MtM values, or $10 - 5 = 5$. The exposure value using netting is defined as follows.

Definition 3.0.3 (Netting)**.** *Netting is the act of offsetting the MtM values of the contracts with a particular counterparty. The exposure using netting n different contracts then becomes*

$$E_{netted}(t) = \max\left(\sum_{i=1}^{n} V_i(X(t)), 0\right). \tag{3.2}$$

Here $V_i(X(t))$ is the MtM value of contract i, where $i = 1, \ldots, n$, under the risk factors at time t denoted by $X(t)$.

To manage the CCR of a portfolio, institutions are interested in metrics involving future exposure. One such metric is the expected exposure (EE).

Definition 3.0.4 (Expected exposure)**.** *The expected exposure depending on risk factor $X(t)$ under risk measure at time t is defined as*

$$EE(t) = \mathbb{E}\left[E(t)\right]. \tag{3.3}$$

A commonly used metric is the Potential Future Exposure.

Definition 3.0.5 (Potential Future Exposure)**.** *The Potential Future Exposure at time t is defined as the worst-case gain for a certain level α, and is calculated using the following expression*

$$PFE_\alpha(t) = \inf\{x \in \mathbb{R} : \mathbb{P}(E(t) > x) \leq 1 - \alpha\}. \tag{3.4}$$

Here $\alpha \in (0, 1)$ is the confidence level, and $E(t)$ is the exposure at time t. In other words, the $PFE_\alpha(t)$ is the smallest number x such that the probability of the exposure at time t being bigger than x is less than $1 - \alpha$.

In this thesis, the PFE is defined as the 97.5%-quantile of the exposure distribution.

3.1. Monte Carlo

Monte Carlo methods are computational techniques used to estimate the value of a function or model through random sampling. It was first introduced by physicist Stanislaw Ulam in March 1947. While considering the win rate of a game of Solitaire, and attempting to solve the problem using combinatorial calculations, he thought of playing games and recording the win frequency [42]. Together with John von Neumann, Robert Richtmyer and Nick Metropolis the method was developed further [43]. The method's foundation lies in the law of large numbers which states that the average result of a number of i.i.d. random variables converges to the true expectation as the number of independent samples increases. The Monte Carlo method is applied countless times across a wide variety of fields. An example of an application in finance is calculating the expected payoff of an option. Using a model the dynamics of the underlying asset is modelled and its value is simulated over a given period. After this, the payoff of the option is calculated for each simulated value of the underlying asset. From these payoffs the average payoff can be calculated.

The Monte Carlo method for calculating the expected value of a random variable X is

$$\hat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Here $\hat{X}_n$ is the unbiased Monte Carlo estimator for $\mathbb{E}[X]$, where n is the number of simulations or paths, and X_i is the i.i.d. result of the random simulation of X at simulation i. This estimator is an unbiased estimator since

$$\mathbb{E}\left[\hat{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X] = \mathbb{E}[X].$$

As stated, the Monte Carlo method is rooted in the law of large numbers that is formally formulated in the following way.

Theorem 3.1.1 (Weak law of large numbers)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X] = \mu < \infty$. Then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\hat{X}_n - \mu\right| \geq \epsilon\right] = 0.$$

From this theorem it can indeed be concluded that as the number of samples increases the approximation $\hat{X}_n$ of $\mathbb{E}[X]$ becomes more accurate.

A second theorem that it very useful for evaluating the accuracy of the Monte Carlo method is the central limit theorem.

Theorem 3.1.2 (Central limit theorem)**.** *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu < \infty$ and variance $0 < var[X_i] = \sigma_X^2 < \infty$. Then,*

$$\frac{\hat{X}_n - \mu}{\frac{\sigma_X}{\sqrt{n}}} \to^d \mathcal{N}(0,1),$$

where $\to^d$ implies convergence in distribution.

Using the central limit theorem it can be seen that the error of the sample mean $\hat{X}_n$ and the true mean $\mathbb{E}[X]$ is distributed as

$$\hat{X}_n - \mathbb{E}[X] \to^d \mathcal{N}\left(0, \frac{\sigma_X^2}{n}\right). \tag{3.5}$$

From this distribution it can be seen that the error of the Monte Carlo estimator scales with a factor $\frac{1}{\sqrt{n}}$. This means that to achieve an error that is ten times smaller the amount of simulations n must be increased by a factor 100.

Approximating the PFE using the Monte Carlo method begins by simulating the portfolio's underlying risk factors. Using these risk factors, the MtM value and exposure of the portfolio can be calculated. Repeating this process n times results in a range of exposure values for different risk factor values. Using this range the exposure distribution, or cumulative distribution function (CDF) of the exposure, of the portfolio can be constructed from which the PFE can be calculated. The PFE is the 97.5%-quantile of the exposure distribution. Thus the Monte Carlo method will be used to estimate the theoretical exposure distribution $F(x)$ from which the 97.5%-quantile can be approximated.

The general Monte Carlo cumulative distribution estimator is defined as

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i \leq y\}, \tag{3.6}$$

where $\mathbf{1}\{X_i \leq y\}$ is the indicator function. Similarly to the estimator $\hat{X}_n$ seen prior, the estimator $\hat{F}_n(x)$ is also an unbiased estimator of $F(x)$ since

$$\mathbb{E}\left[\hat{F}_n(y)\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}\{X_i \leq y\}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\mathbf{1}\{X \leq y\}\right] = \mathbb{P}\left(X \leq y\right) = F(y).$$

The CDF estimator used for the calculation of the PFE is defined as

$$\hat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\max\left(V\left(\hat{\mathbf{X}}_{(i)}(t)\right), 0\right) \leq y\right\}.$$

Here, $\hat{\mathbf{X}}_{(i)}$ denotes the i-th simulated sample of a vector of risk factors in a Monte Carlo simulation. An estimator for the PFE is obtained using the inverse of the Monte Carlo CDF estimator. The quantile estimation $\hat{q}_\alpha$ is then defined as

$$\hat{q}_\alpha = \inf\{x : \hat{F}_n(y) > \alpha\}. \tag{3.7}$$

Similarly to the Monte Carlo estimator $\hat{X}_n$ the convergence of $\hat{F}_n(x)$ goes at a rate of $\frac{1}{\sqrt{n}}$.

The Monte Carlo method is easily implemented and very intuitive. Unfortunately, the convergence rate of the error is very slow. Consequently, significant effort has been invested in researching variance reduction techniques. These methods aim at reducing the variance of an estimator without requiring additional simulations. Some examples of variance reduction techniques are antithetic variates, the control variate method, and importance sampling. The latter two will be discussed in more detail in future sections.

In this thesis the paths of the risk factors are generated using direct sampling. From Equations 3.21 up to and including 3.23 and 4.4 it can be seen that the risk factors are distributed in the following way,

$$\begin{aligned}
x_d(t) &\sim \mathcal{N}\left(x_d(0)e^{-a_d t}, \left(\sigma_d \sigma_{Z_d}\right)^2\right), \\
x_f(t) &\sim \mathcal{N}\left(x_f(0)e^{-a_f t}, \left(\sigma_f \sigma_{Z_f}\right)^2\right), \\
\log(X(t)) &\sim \mathcal{N}\left(\log(X(0)) + \left(\mu_X - \frac{1}{2}\sigma_X^2\right)t, \left(\sigma_X \sigma_{Z_X}\right)^2\right), \\
\log(z(t)) &\sim \mathcal{N}\left(\log(z(0)) - \frac{1}{2}\sigma_z^2 t, (\sigma_z \sigma_{Z_z})^2\right).
\end{aligned}$$

Using their individual distributions and correlation matrix, a mean vector and covariance matrix can be constructed to generate the risk factors using a multivariate normal distribution.

For the uncollateralized portfolio the multivariate normal distribution $\mathcal{N}(\mu(t), \Sigma(t))$ used for sampling the risk factors has parameters,

$$\mu(t) = \begin{bmatrix} x_d(0)e^{-a_d t} \\ x_f(0)e^{-a_f t} \\ \log(X(0)) + (\mu_X - \frac{1}{2}\sigma_X^2)t \end{bmatrix}$$

$$\Sigma(t) = V \cdot V^\top \odot \begin{bmatrix} 1 & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_{x_f}) & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_X) \\ \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_{x_d}) & 1 & \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_X) \\ \text{Corr}(\hat{Z}_X, \hat{Z}_{x_d}) & \text{Corr}(\hat{Z}_X, \hat{Z}_{x_f}) & 1 \end{bmatrix}(t),$$

where $\odot$ signifies element-wise multiplication, and

$$V = \begin{bmatrix} \sigma_d \sigma_{Z_d} \\ \sigma_f \sigma_{Z_f} \\ \sigma_X \sigma_{Z_X} \end{bmatrix}.$$

For the Security Financed Trades the risk factors are generated using a multivariate normal distribution $\mathcal{N}\left(\mu(t), \Sigma(t)\right)$ where

$$\mu(t) = \begin{bmatrix} x_d(0)e^{-a_d t} \\ x_f(0)e^{-a_f t} \\ \log(X(0)) + (\mu_X - \frac{1}{2}\sigma_X^2)t \\ \log(z(0)) - \frac{1}{2}\sigma_z^2 t \end{bmatrix}$$

$$\Sigma(t) = V \cdot V^\top \odot \begin{bmatrix} 1 & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_{x_f}) & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_X) & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_{x_d}) & 1 & \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_X) & \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_X, \hat{Z}_{x_d}) & \text{Corr}(\hat{Z}_X, \hat{Z}_{x_f}) & 1 & \text{Corr}(\hat{Z}_X, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_z, \hat{Z}_{x_d}) & \text{Corr}(\hat{Z}_z, \hat{Z}_{x_f}) & \text{Corr}(\hat{Z}_z, \hat{Z}_X) & 1 \end{bmatrix}(t),$$

where $\odot$ signifies element-wise multiplication, and

$$V = \begin{bmatrix} \sigma_d \sigma_{Z_d} \\ \sigma_f \sigma_{Z_f} \\ \sigma_X \sigma_{Z_X} \\ \sigma_z \sigma_{Z_z} \end{bmatrix}.$$

Thus, at time t a sample $\begin{bmatrix} x_d(t) & x_f(t) & \log(X(t)) \end{bmatrix}$ or $\begin{bmatrix} x_d(t) & x_f(t) & \log(X(t)) & \log(z(t)) \end{bmatrix}$ is generated using $\mathcal{N}(\mu(t), \Sigma(t))$ and their respective parameters.

3.2. COS method

The COS method [4] finds its foundation in the Fourier pair

$$\varphi(\omega) = \int_{\mathbb{R}} e^{i\omega x} f(x) dx, \tag{3.8}$$

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} \varphi(\omega) d\omega. \tag{3.9}$$

Where Equation 3.9 can be solved using its Fourier-cosine expansion. For a function supported on $[a, b] \in \mathbb{R}$ this cosine expansion is given by

$$f(x) = {\sum_{k=0}^{\infty}}' A_k \cos\left(k\pi \frac{x-a}{b-a}\right), \tag{3.10}$$

where

$$A_k = \frac{2}{b-a} \int_a^b f(x) \cos\left(k\pi \frac{x-a}{b-a}\right) dx. \tag{3.11}$$

The accent on the sum in Equation 3.10 indicates that the first term is halved.

As can be seen by comparing Equations 3.9 and 3.11 the integration range has been truncated from $[-\infty, \infty]$ to $[a, b]$. Performing this truncation introduces a truncation error. One of the characteristics of the Fourier transform integrands, found in Equation 3.9, is that it decays to zero at $-\infty$ and ∞. Due to this, we are given the room to choose an integration range $[a, b]$ such that truncating the bounds does not result in a reduction of accuracy. Choosing such an integration range yields

$$\varphi_1(\omega) = \int_a^b e^{i\omega x} f(x) dx \approx \int_{\mathbb{R}} e^{i\omega x} f(x) dx = \varphi(\omega). \tag{3.12}$$

In [1] the truncation range $[a, b]$ is given by

$$[a, b] = \left[\mu_E - L\sigma_E, \mu_E + L\sigma_E\right] \qquad \text{with} \qquad L = 8, \tag{3.13}$$

where, for the PFE calculation, μ_E is the first moment of the exposure, and σ_E is the exposure's standard deviation.

By combining Equation 3.12 and the fact that $e^{ix} = \cos(x) + i\sin(x)$, it can be found that

$$A_k \approx F_k = \frac{2}{b-a} \mathrm{Re}\left\{\varphi\left(\frac{k\pi}{b-a}\right) \cdot \exp\left(-i\frac{ka\pi}{b-a}\right)\right\}. \tag{3.14}$$

Then, by filling in F_k and truncating the infinite sum in equation 3.10 to the first $N-1$ terms, it can be seen that

$$f(x) = {\sum_{k=0}^{N-1}}' F_k \cos\left(k\pi \frac{x-a}{b-a}\right). \tag{3.15}$$

Using the characteristic function $\varphi_X(\omega)$ of a random variable X, the COS method allows for the recovery of its probability density function $f_X(x)$. Consequently, the cumulative density function $F_X(x)$ can be obtained by integration,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u) du \\ &\approx \int_a^x {\sum_{k=0}^{N-1}}' F_k \cos\left(k\pi \frac{u-a}{b-a}\right) du \\ &= {\sum_{k=0}^{N-1}}' F_k \int_a^x \cos\left(k\pi \frac{u-a}{b-a}\right) du \\ &= \frac{1}{2}(x-a)F_0 + \sum_{k=1}^{\infty} F_k \frac{b-a}{k\pi} \sin\left(k\pi \frac{x-a}{b-a}\right). \end{aligned} \tag{3.16}$$

3.2.1. Numerical Integration

To find the COS coefficients F_k in Equation 3.15, it is required to find the characteristic function, which requires to perform an integration. For this integration the Clenshaw-Curtis quadrature will be used. This quadrature makes use of a change of variables $x = \cos\theta$. For a function $f(x)$ this results in

$$\int_{-1}^{1} f(x)dx = \int_{0}^{\pi} f(\cos(\theta))\sin(\theta)d\theta. \tag{3.17}$$

Using the cosine expansion of $f(\cos\theta)$, the integral can be rewritten in the following way,

$$\int_{0}^{\pi} f(\cos(\theta))\sin(\theta)d\theta = a_0 + \sum_{i=1}^{\infty} \frac{2a_{2k}}{1-(2k)^2} = \sum_{i=0}^{\infty} a_{2k}w_{2k}. \tag{3.18}$$

In the last equation of Equation 3.18, $w_{2k} = 1$ for $k = 0$ and $w_{2k} = \frac{2}{1-(2k)^2}$ for $k > 0$. Furthermore, a_2k can be calculated using the type I discrete cosine transform

$$a_k \approx \frac{2}{N}\left[\frac{f(1)}{2} + \frac{f(-1)}{2}(-1)^k + \sum_{n=1}^{N-1} f\left(\cos\left(\frac{n\pi}{N}\right)\right)\cos\left(\frac{nk\pi}{N}\right)\right], \tag{3.19}$$

where N is the number of quadrature points used.

It can be seen that by using the change of variables $x = \cos(\theta)$, the characteristic function used in the COS method in Equation 3.12 can be rewritten as

$$\begin{aligned}\varphi\left(\frac{k\pi}{b-a}\right) &= \int_{l}^{u} \exp\left(i\frac{k\pi}{b-a}x\right) f(x)dx \\ &= \frac{u-l}{2}\int_{0}^{\pi} g(\cos(\theta))\sin(\theta)d\theta\end{aligned}$$

where

$$g(\cos(\theta)) = \exp\left(i\frac{k\pi}{b-a}\left(l + \frac{u-l}{2}(\cos(\theta)+1)\right)\right) f\left(l + \frac{u-l}{2}(\cos(\theta)+1)\right). \tag{3.20}$$

Sequentially, this characteristic function is used in the COS method to retrieve the cumulative density function.

3.2.2. COS method for PFE calculations

This section provides a brief summary of how the COS method described above can be applied to the calculation of the Potential Future Exposure. The full description of this method can be found in [1].

The COS method was applied to calculate the PFE of a portfolio containing interest rate derivatives and foreign exchange derivatives. The value of these derivatives depend on the domestic short rate, foreign short rate, and the exchange rate between the domestic and foreign currencies. Equations 2.24, 2.25, and 2.32 show the solutions of the dynamics of these risk factors to be

$$\begin{aligned}x_d(t) &= x_d(0)e^{-a_d t} + \sigma_d \int_0^t e^{-a_d(t-s)}dW_d(s), \\ x_f(t) &= x_f(0)e^{-a_f t} + \sigma_f \int_0^t e^{-a_f(t-s)}dW_f(s), \\ \log(X(t)) &= \log(X(0)) + \left(\mu - \frac{1}{2}\sigma_X^2\right)t + \sigma_X W_X(t).\end{aligned}$$

Here W_d, W_f, and W_X are standard Brownian motions as the modelling is done under the real world measure.

By using

$$\int_0^t e^{-a_d(t-s)} dW_d(s) = \sigma_{Z_d} \hat{Z}_d,$$
$$\int_0^t e^{-a_f(t-s)} dW_f(s) = \sigma_{Z_f} \hat{Z}_f,$$
$$W_X(t) = \sigma_{Z_X} \hat{Z}_X(t),$$

where $\hat{Z}_\cdot \sim \mathcal{N}(0,1)$ the solutions can be rewritten as,

$$x_d(t) = x_d(0)e^{-a_d t} + \sigma_d \sigma_{Z_d} \hat{Z}_d(t), \tag{3.21}$$
$$x_f(t) = x_f(0)e^{-a_f t} + \sigma_f \sigma_{Z_f} \hat{Z}_f(t), \tag{3.22}$$
$$\log(X(t)) = \log(X(0)) + \left(\mu_X - \frac{1}{2}\sigma_X^2\right) t + \sigma_X \sigma_{Z_X} \hat{Z}_X(t). \tag{3.23}$$

Where, due to

$$\int_0^t e^{-a(t-s)} dW(s) \sim \mathcal{N}\left(0, \frac{1 - e^{-2at}}{2a}\right),$$

and $W_X(t) \sim \mathcal{N}(0,t)$ it can be seen that

$$\sigma_{Z_d} = \sqrt{\frac{1 - e^{-2a_d t}}{2a_d}},$$
$$\sigma_{Z_f} = \sqrt{\frac{1 - e^{-2a_f t}}{2a_f}},$$
$$\sigma_{Z_X} = \sqrt{t}.$$

The correlation matrix $\Sigma(t)$ at time t is,

$$\Sigma(t) = \begin{bmatrix} 1 & \text{Corr}(\hat{Z}_d, \hat{Z}_f) & \text{Corr}(\hat{Z}_d, \hat{Z}_X) \\ \text{Corr}(\hat{Z}_f, \hat{Z}_d) & 1 & \text{Corr}(\hat{Z}_f, \hat{Z}_X) \\ \text{Corr}(\hat{Z}_X, \hat{Z}_d) & \text{Corr}(\hat{Z}_X, \hat{Z}_f) & 1 \end{bmatrix}(t), \tag{3.24}$$

As a correlation matrix is SPD by definition, a Cholesky decomposition $L(t)$ at time t can be constructed. Using $L(t)$, the risk factors $\hat{Z}_d$, $\hat{Z}_f$, and $\hat{Z}_X$ are modeled by computing $\begin{bmatrix} \hat{Z}_d & \hat{Z}_f & \hat{Z}_X \end{bmatrix} = L(t) \cdot \begin{bmatrix} \tilde{Z}_d & \tilde{Z}_f & \tilde{Z}_X \end{bmatrix}^\top$ where $\tilde{Z}_d$, $\tilde{Z}_f$, and $\tilde{Z}_X$ are independent standard normal random variables.

Using these risk factors, the portfolio can be priced and its exposure $E_t(\mathbf{X})$ can be calculated. Sequentially, the characteristic function of the exposure can be calculated via

$$\varphi(\omega) = \iiint_{\mathbb{R}^3} e^{i\omega E_t(\tilde{z}_d, \tilde{z}_f, \tilde{z}_X)} f(\tilde{z}_d) f(\tilde{z}_f) f(\tilde{z}_X) d\tilde{z}_d d\tilde{z}_f d\tilde{z}_X. \tag{3.25}$$

Using the Clenshaw-Curtis quadrature described above, Equation 3.25 can be evaluated. Then, using the characteristic function, the COS method can be applied to retrieve the CDF of the exposure from which the PFE of the portfolio can be calculated.

4

Our contribution 1: Adding collaterals to the COS-PFE framework

In this chapter, we add collaterals to the earlier work in [1] to the COS-PFE framework. As stated in Section 3, requiring collateral is a very effective means to reduce the CCR of a netting set. Such an agreement requires that the party for which a netting set has a negative MtM to provide a collateral (with good liquidity and credit rating) with a higher or equal MtM value to the other party to mitigate the CCR faced by the other party. In this way, the party holding the collateral can retain it if the counterparty fails to meet payment obligations. Collateral arrangements can be bilateral, requiring both parties to provide collateral in response to a negative MtM value. Additionally, it can also be one-sided where only one party has a collateral obligation, which typically happens when the other party is of a much higher credit rating.

In a fully collateralized position, the exposure E driven by $\mathbf{X}$ at time t follows,

$$E(t) = \max\left(V(\mathbf{X}(t)) - C_t, 0\right) = 0. \tag{4.1}$$

Here, $V(\mathbf{X}(t))$ is the MtM value of the position driven by risk factor values $\mathbf{X}$ at time t, and C_t is the value of the collateral at time t.

Unfortunately, the exposure cannot always be fully collateralized. Generally, there are three main sources of partial-collateralization: threshold amount (under which collateral transfer is not required), minimum transfer amount (the minimal amount of collateral transfer), and margin period of risk (the time period from the last reception of collateral from a defaulting counterparty until all trades with this counterparty are closed out and the resulting market risk is re-hedged). In this work, we do not consider the impact of the threshold amount and minimal transfer amount for simplicity. The methods developed in this thesis are still applicable in the presence of margin call threshold and the minimum transfer amount. The margin period of risk is also out of the scope, as it is associated to margined trades under CSA, which we briefly discuss below and is out of scope of this thesis.

Collaterals can be posted in two ways in practice: direct via the Credit Support Annex (CSA) agreement and indirectly via the Security Financed Trades (SFT).

Firstly, in a bilateral CSA agreement both parties are required to post collateral to each other whenever the MtM value of the netting set portfolio, from the perspective of the party, exceeds a certain threshold. These thresholds can vary for each party depending on the counterparty's creditworthiness, and the party's risk appetite. For example, if the counterparty is rated from BBB- to AAA according to the S&P, the threshold is $1,000,000. If the rating is BB+ or below, the threshold can drop to zero [2]. If the party is in the money and the MtM or the relevant netting set seen from the perspetive of the party is beyond the threshold, the party will send a margin call to the counterparty. The collateral value called by this party equals the MtM price of the netting set observed on the date of the margin call. That is,

$$C_t = V(X(\tau))$$

and thus

$$E(t) = \max\left(V(X(t)) - V(X(\tau))\right),$$

where $\tau = t - I$ with I being the number of days between two margin calls. The same applies in reverse. The CSA trades are not covered.

In the case of SFTs, the counterparty posts a security or collection of securities with a value that falls within a range around the MtM value of the trade at initialization. Examples of such securities include equities, cash, bonds or ETFs. The value of the collateral depends on its liquidity, risk and credit quality. A haircut for non-cash collaterals is usually applied. This haircut adjusts the value of the collateral to account for the uncertainty in the collateral value by the time of liquidation, due to market movement and/or volatility in FX rate. In this way, posting riskier securities as collaterals may result in more collateral being needed relative to the MtM value of the derivative trades. In this thesis, we assume the security that is posted is a single bond and it is posted at the initialization of the portfolio. The extension of our methods to using more products as collaterals is straight forward. Additionally, we assume no haircut is applied to the value of the bond. Note that it is trivial to extend our methods to including haircuts.

4.1. Adding the Z-spread

As stated, it is assumed that the counterparty posts a bond that fully collateralizes the portfolio's MtM value at the initialization, $t = 0$, of the portfolio. The bond is posted in the domestic currency and is valued using the so-called Z-spread model

Definition 4.1.1 (Z-spread)**.** *A bond's Z-spread measures the spread along the entire risk-free yield curve. With this, it is a measure of the extra compensation, on top of the risk-free rate, that is rewarded for taking the risk of holding the bond.*

The Z-spread of a bond can be calculated by finding the value z such that the discounted cashflows from this bond equalizes its market price, i.e.

$$V_{\text{bond}}(t) = \sum_{i=1}^{n} c_{t_i} P(r_{t_i} + z, t_i, T), \tag{4.2}$$

where c_{t_i} is the expected cashflow at time t_i, and P is the discount function under r_{t_i} which is the risk-free rate at t_i. Here, $i = 1, \ldots, n$ such that $t < t_i \leq t_n = T$ where T is the maturity of the bond.

The Z-spread is incorporated into the COS method as a new risk factor. Its dynamic is modelled by a driftless Geometric Brownian Motion which has the following solution,

$$\log(z(t)) = \log(z(0)) - \frac{1}{2}\sigma_z^2 t + \sigma_z W_z(t). \tag{4.3}$$

Similarly as done for the original dynamics in Equations 3.21, the above expression can be rewritten as

$$\log(z(t)) = \log(z(0)) - \frac{1}{2}\sigma_z^2 t + \sigma_z \sqrt{t}\hat{Z}_z(t), \tag{4.4}$$

following $W_z(t) \sim \mathcal{N}(0, t)$. Using Equation 4.4 the correlation matrix can be expanded to include the Z-spread. The correlation matrix now reads

$$\rho(t) = \begin{bmatrix} 1 & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_{x_f}) & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_X) & \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_{x_d}) & 1 & \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_X) & \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_X, \hat{Z}_{x_d}) & \text{Corr}(\hat{Z}_X, \hat{Z}_{x_f}) & 1 & \text{Corr}(\hat{Z}_X, \hat{Z}_z) \\ \text{Corr}(\hat{Z}_z, \hat{Z}_{x_d}) & \text{Corr}(\hat{Z}_z, \hat{Z}_{x_f}) & \text{Corr}(\hat{Z}_z, \hat{Z}_X) & 1 \end{bmatrix}(t). \tag{4.5}$$

Here, it was found that

$$\text{Corr}(\hat{Z}_z, \hat{Z}_{x_d})(t) = \text{Corr}(\hat{Z}_{x_d}, \hat{Z}_z)(t) = \frac{\frac{\rho_{zd}}{a_d}\left(1 - e^{-a_d t}\right)}{\sqrt{\frac{1}{2a_d}\left(1 - e^{-2a_d t}\right)}\sqrt{t}},$$

$$\text{Corr}(\hat{Z}_z, \hat{Z}_{x_f})(t) = \text{Corr}(\hat{Z}_{x_f}, \hat{Z}_z)(t) = \frac{\frac{\rho_{zf}}{a_f}\left(1 - e^{-a_f t}\right)}{\sqrt{\frac{1}{2a_f}\left(1 - e^{-2a_f t}\right)}\sqrt{t}},$$

$$\text{Corr}(\hat{Z}_z, \hat{Z}_X)(t) = \text{Corr}(\hat{Z}_X, \hat{Z}_z)(t) = \frac{\rho_{zX}}{t}.$$

As $\rho(t)$ is a positive semi-definite matrix a Cholesky decomposition can be applied resulting in

$$L(t) = \begin{bmatrix} L_{1,1} & 0 & 0 & 0 \\ L_{2,1} & L_{2,2} & 0 & 0 \\ L_{3,1} & L_{3,2} & L_{3,3} & 0 \\ L_{4,1} & L_{4,2} & L_{4,3} & L_{4,4} \end{bmatrix}(t), \tag{4.6}$$

where

$$L_{j,j}(t) = \sqrt{\rho(t)_{j,j} - \sum_{k=1}^{j-1} L_{j,k}(t)^2},$$

$$L_{i,j}(t) = \frac{\left(\rho(t)_{i,j} - \sum_{k=1}^{j-1} L_{i,k}(t)L_{j,k}(t)\right)}{L_{j,j}(t)}.$$

Then $\rho(t) = L^\top(t)L(t)$. Using $\mathbf{Z} = L \cdot \mathbf{Z}$, similarly done as in Section 3.2.2, but now including the standard normal random variable $\tilde{Z}_z$, it follows that

$$z(t) = z(0)e^{-\frac{1}{2}\sigma_z^2 t + \sigma_z \sigma_{\hat{Z}_z}\left(L_{4,1}(t)\tilde{z}_{x_d}(t) + L_{4,2}(t)\tilde{z}_{x_f}(t) + L_{4,3}(t)\tilde{z}_X(t) + L_{4,4}\tilde{z}_z(t)\right)}. \tag{4.7}$$

After this, the characteristic function of the collateralized exposure can be calculated using

$$\begin{aligned}\varphi(\omega) = &\iiiint_{\mathbb{R}^4}\Big(e^{i\omega\cdot\max(V_{\text{portfolio}} - V_{\text{bond}}, 0)(\tilde{z}_{x_d}, \tilde{z}_{x_f}, \tilde{z}_X, \tilde{z}_z)} \\ &f_{\tilde{Z}_{x_d}, \tilde{Z}_{x_f}, \tilde{Z}_X, \tilde{Z}_z}(\tilde{z}_{x_d}, \tilde{z}_{x_f}, \tilde{z}_X, \tilde{z}_z)\Big) d\tilde{z}_{x_d} d\tilde{z}_{x_f} d\tilde{z}_X d\tilde{z}_z, \qquad (4.8) \\ = &\iiiint_{\mathbb{R}^4}\Big(e^{i\omega\cdot\max(V_{\text{portfolio}} - V_{\text{bond}}, 0)(\tilde{z}_{x_d}, \tilde{z}_{x_f}, \tilde{z}_X, \tilde{z}_z)} \\ &f_{\tilde{Z}_{x_d}}(\tilde{z}_{x_d}) f_{\tilde{Z}_{x_f}}(\tilde{z}_{x_f}) f_{\tilde{Z}_X}(\tilde{z}_X) f_{\tilde{Z}_z}(\tilde{z}_z)\Big) d\tilde{z}_{x_d} d\tilde{z}_{x_f} d\tilde{z}_X d\tilde{z}_z. \qquad (4.9)\end{aligned}$$

This characteristic function can be used to recover the CDF of the collateralized exposure of the portfolio, from which the PFE can be calculated.

In this thesis, the Z-spread is independent of the processes of x_d, x_f and X. From this, it follows that $\rho_{zd} = \rho_{zf} = \rho_{zX} = 0$.

4.2. Bond pricing via moment matching

By adding the Z-spread in Equation 4.2, a new expression for $P(t,T)$ can be found by calculating

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T (r(s) + z(s))ds}\middle|\mathcal{F}_t\right]. \tag{4.10}$$

By filling in

$$r(t) = x(t) + \beta(t), \tag{4.11}$$

into Equation 4.10 yields

$$P(t,T) = e^{-\int_t^T \beta(s)ds}\mathbb{E}^{\mathbb{Q}}\left[e^{-\left(\int_t^T x(s)ds+\int_t^T z(s)ds\right)}\middle|\mathcal{F}_t\right]. \tag{4.12}$$

As previously seen,

$$\int_t^T x(s)ds = x(t)\frac{1-e^{-a(T-t)}}{a} + \frac{\sigma_x}{a}\int_t^T (1-e^{-a(T-u)})dW(u). \tag{4.13}$$

From this, it was seen that

$$\int_t^T x(s)ds \sim \mathcal{N}\left(x(t)\frac{1-e^{-a(T-t)}}{a}, V(t,T)\right),$$

where

$$V(t,T) = \frac{\sigma_x^2}{a^2}\left(T-t-\frac{2\left(1-e^{-a(T-t)}\right)}{a}+\frac{1-e^{-2a(T-t)}}{2a}\right).$$

From Equation 4.3 it follows that,

$$z(t) = z(s)e^{-\frac{1}{2}\sigma_z^2(t-s)+\sigma_z(W(t)-W(s))}. \tag{4.14}$$

Taking the integral of both sides yields

$$\int_t^T z(s)ds = \int_t^T z(t)e^{-\frac{1}{2}\sigma_z^2(s-t)+\sigma_z(W(s)-W(t))}ds. \tag{4.15}$$

While it is known that $z(t)$ follows the log-normal distribution, the distribution of $\int_t^T z(s)ds$ is not known analytically. Because of this it is not possible to use the moment generating function to evaluate the bond pricing formula in Equation 4.12. To solve this problem moment-matching will be used. This method approximates the unknown distribution by matching its sample moments to those of a parametric probability distribution via tuning the parameters. Fortunately, the paper [44] provides a recursive equation for the analytical solution for the moments of the time average of a Geometric Brownian Motion. In this way, instead of the sample moments, these analytical solutions can be used to find the moments. Subsequently, the time-averaged Geometric Brownian Motion can be used for approximating the moments of $\int_t^T z(s)ds$.

In the paper the authors define the time average $A(T)$ as

$$A(T) = \frac{1}{T}\int_0^T S(t)dt, \tag{4.16}$$

where $S(t)$ is the Geometric Brownian Motion

$$S(t) = e^{\left(r-\frac{\sigma^2}{2}\right)t+\sigma B(t)}. \tag{4.17}$$

Here, $B(t)$ is a Brownian motion. It can be easily seen that if $r = 0$, then

$$\begin{aligned}\int_0^T z(s)ds &= \int_0^T z(0)e^{-\frac{\sigma_z^2}{2}s+\sigma_z W(s)}ds,\\ &= z(0)T\cdot\frac{1}{T}\int_0^T e^{-\frac{\sigma_z^2}{2}s+\sigma_z W(s)}ds,\\ &= z(0)T\cdot A(T).\end{aligned}$$

Similarly, it can be seen that

$$\begin{aligned}\int_t^T z(s)ds &= z(t)(T-t)\cdot\frac{1}{T-t}\int_t^T e^{-\frac{\sigma_z^2}{2}s+\sigma_z W(s)}ds,\\ &= z(t)(T-t)\cdot A(T-t).\end{aligned}$$

Theorem 4.2.1 stated below, initially provided in [44], is used to calculate the moments of $A(T)$.

Theorem 4.2.1. *Let*

$$b_k = kr + \frac{\sigma^2}{2} \cdot k(k-1), \quad k = 0, 1, \ldots \tag{4.18}$$

then

$$\mathbb{E}\left[A(T)^m\right] = m! \exp\left[b_0 T, b_1 T, \ldots, b_m T\right], \quad m \geq 0. \tag{4.19}$$

Furthermore, the convention $f[a_0, \ldots, a_n]$ is defined in [44] as

Theorem 4.2.2.

$$f\left[a_0, \ldots, a_n\right] = \frac{f\left[a_1, \ldots, a_n\right] - f\left[a_0, \ldots, a_{n-1}\right]}{a_n - a_0}, \tag{4.20}$$

for any distinct complex numbers $a_0, \ldots, a_n$, which follows from the relation

$$f[a_0, a_1] = \frac{f(a_1) - f(a_0)}{a_1 - a_0} = \int_0^1 f'((1-t)a_0 + ta_1)dt. \tag{4.21}$$

Using this theorem, the first three moments can be derived, as shown below, and are then used for moment matching.

For the first moment

$$\mathbb{E}\left[\int_t^T z(s)ds\right] = z(t)(T-t)\mathbb{E}[A(T-t)]. \tag{4.22}$$

From Theorem 4.2.1 it follows that $b_0 = 0$ and $b_1 = 0$. Additionally, if $f(x) = e^x$ then $f'(x) = e^x$. From this it can be seen that

$$\begin{aligned}\mathbb{E}[A(T-t)] &= 1! \exp[b_0(T-t), b_1(T-t)], \\ &= \exp[0, 0], \\ &= \int_0^1 e^{(1-t)\cdot 0 + t\cdot 0} dt = 1.\end{aligned}$$

Filling this into Equation 4.22 it follows that

$$\mathbb{E}\left[\int_t^T z(s)ds\right] = z(t)(T-t).$$

Similarly, calculating the second moment shows that

$$\mathbb{E}\left[\left(\int_t^T z(s)ds\right)^2\right] = z(t)^2(T-t)^2\mathbb{E}[(A(T-t))^2]. \tag{4.23}$$

Then as $b_2 = 2 \cdot 0 + \frac{\sigma_z^2}{2} \cdot 2 \cdot (2-1) = \sigma_z^2$ it can be seen that

$$\mathbb{E}\left[(A(T-t))^2\right] = 2! \exp[0, 0, \sigma_z^2(T-t)]$$

where

$$\exp[0, 0, \sigma_z^2(T-t)] = \frac{\exp[0, \sigma_z^2(T-t)] - \exp[0, 0]}{\sigma_z^2(T-t)}.$$

Then as

$$\exp[0, \sigma_z^2(T-t)] = \frac{e^{\sigma_z^2(T-t)} - 1}{\sigma_z^2(T-t)}$$

and $\exp[0, 0]$ is known it is found that

$$\exp[0, 0, \sigma_z^2(T-t)] = \frac{\frac{e^{\sigma_z^2(T-t)}-1}{\sigma_z^2(T-t)} - 1}{\sigma_z^2(T-t)} = \frac{e^{\sigma_z^2(T-t)} - 1}{\sigma_z^4(T-t)^2} - \frac{1}{\sigma_z^2(T-t)}.$$

It follows from Equation 4.23 that

$$\begin{aligned}\mathbb{E}\left[\left(\int_t^T z(s)ds\right)^2\right] &= z(t)^2(T-t)^2\mathbb{E}\left[(A(T-t))^2\right], \\ &= 2z(t)^2(T-t)^2\left(\frac{e^{\sigma_z^2(T-t)}-1}{\sigma_z^4(T-t)^2}-\frac{1}{\sigma_z^2(T-t)}\right), \\ &= 2z(t)^2\left(\frac{e^{\sigma_z^2(T-t)}-1}{\sigma_z^4}-\frac{T-t}{\sigma_z^2}\right) \end{aligned} \tag{4.24}$$

Finally, the third moment shows that

$$\mathbb{E}\left[\left(\int_t^T z(s)ds\right)^3\right] = z(t)^3(T-t)^3\mathbb{E}\left[A(T-t)^3\right],$$

where, using $b_3 = \frac{\sigma^2}{2}\cdot 3\cdot(3-1) = 3\sigma_z^2$, it follows that

$$\mathbb{E}\left[A(T-t)^3\right] = 3!\exp\left[0,0,\sigma_z^2(T-t),b_3(T-t)\right].$$

From

$$\exp\left[0,0,\sigma^2(T-t),3\sigma^3(T-t)\right] = \frac{\exp\left[0,\sigma^2(T-t),3\sigma^3(T-t)\right]-\exp\left[0,0,\sigma^2(T-t)\right]}{3\sigma^2(T-t)},$$

and working out that

$$\begin{aligned}\exp\left[0,\sigma^2(T-t),3\sigma^3(T-t)\right] &= \frac{\exp\left[\sigma^2(T-t),3\sigma^3(T-t)\right]-\exp\left[0,\sigma^2(T-t)\right]}{3\sigma^2(T-t)}, \\ &= \frac{e^{3\sigma^2(T-t)}-3e^{\sigma^2(T-t)}+2}{6\sigma^4(T-t)^2}, \\ \exp\left[0,0,\sigma^2(T-t)\right] &= \frac{\exp\left[0,\sigma^2(T-t)\right]-\exp\left[0,0\right]}{\sigma^2(T-t)}, \\ &= \frac{e^{\sigma^2(T-t)}-1-\sigma^2(T-t)}{\sigma^4(T-t)^2}, \end{aligned}$$

it is found that

$$\exp\left[0,0,\sigma^2(T-t),3\sigma^3(T-t)\right] = \frac{6\sigma^2(T-t)+e^{3\sigma^2(T-t)}-9e^{\sigma^2(t-t)}+8}{18\sigma^6(T-t)^3}.$$

It then follows that

$$\begin{aligned}\mathbb{E}\left[\left(\int_t^T z(s)ds\right)^3\right] &= 6z(t)^3(T-t)^3\mathbb{E}\left[A(T-t)^3\right] \\ &= 6z(t)^3(T-t)^3\left(\frac{6\sigma^2(T-t)+e^{3\sigma^2(T-t)}-9e^{\sigma^2(t-t)}+8}{18\sigma^6(T-t)^3}\right), \\ &= z(t)^3\left(\frac{6\sigma^2(T-t)+e^{3\sigma^2(T-t)}-9e^{\sigma^2(t-t)}+8}{3\sigma^6}\right). \end{aligned} \tag{4.25}$$

As Monte Carlo simulations of the integral show skewness in the resulting distribution density, the skew-normal, Beta, and Gamma distributions are selected as candidates and are tested in the approach. The Monte Carlo simulations showed a skewness greater than 1. While the skewness parameter of the

skew-normal distribution has a theoretical maximum of 1, the skewness parameters of the Beta and Gamma distributions can accommodate skewness in excess of 1. For these distributions, the analytical solutions for the moments which we derived earlier, i.e. the moments of the time-integration of the Brownian Motion, are used to back-out the parameters of these three parametric distributions. Recall that the first moments are defined as follows,

$$\begin{aligned}
\mu &= \mathbb{E}\left[\int_0^T z(s)ds\right], \\
\sigma &= \sqrt{\mathbb{E}\left[\left(\int_0^T z(s)ds\right)^2\right] - \mu^2}, \\
\gamma &= \mathbb{E}\left[\left(\frac{\int_0^T z(s)ds - \mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}\left[\left(\int_0^T z(s)ds\right)^3\right] - 3\mu\sigma^2 - \mu^3}{\sigma^3}.
\end{aligned}$$

Fortunately, the moments of the skew-normal distribution $\mathcal{SN}(\xi, \omega, \alpha)$ have closed-form expressions using the moments, i.e.

$$|\delta| = \sqrt{\frac{\pi|\gamma_1|^{2/3}}{2\left(|\gamma_1|^{2/3} + \left(\frac{4-\pi}{2}\right)^{2/3}\right)}}, \tag{4.26}$$

$$\alpha = \frac{\delta}{\sqrt{1-\delta^2}}, \tag{4.27}$$

$$\omega = \frac{\sigma}{\sqrt{1-\frac{2\delta^2}{\pi}}}, \tag{4.28}$$

$$\xi = \mu - \omega\delta\sqrt{\frac{2}{\pi}}. \tag{4.29}$$

In Equation 4.26 the sign of δ is put equal to the sign of γ_1, where

$$\gamma_1 = \min\left(0.99, \gamma\right).$$

In the last equation, γ_1 is bounded above by the maximal theoretical skewness which is accepted by the skew-normal distribution.

To match the moments of the Beta distribution the values of α and β need to be found such that

$$\begin{aligned}
\mu &= \frac{\alpha}{\alpha+\beta}, \\
\sigma^2 &= \frac{\alpha\beta}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)}, \\
\gamma &= \frac{2\left(\beta-\alpha\right)\sqrt{\alpha+\beta+1}}{\left(\alpha+\beta+2\right)\sqrt{\alpha\beta}}.
\end{aligned}$$

In the last equation, γ is not bounded from above as the Beta distribution allows for larger values of γ. The values of α and β were found using a root finding algorithm.

For the Gamma distribution the values of k and θ need to be matched such that

$$\begin{aligned}
\mu &= k\theta, \\
\sigma^2 &= k\theta^2, \\
\gamma &= \frac{2}{\sqrt{k}}.
\end{aligned}$$

Here, γ is again unbounded for the same reason as for the Beta distribution.

To assess the accuracy in matching the moments, the theoretical CDFs of the three parametric distributions are compared with the empirical CDF of the time-integration of Brownian Motion. This empirical CDF was generated using 500000 path using Monte Carlo. The comparison can be found in Figure 4.1.

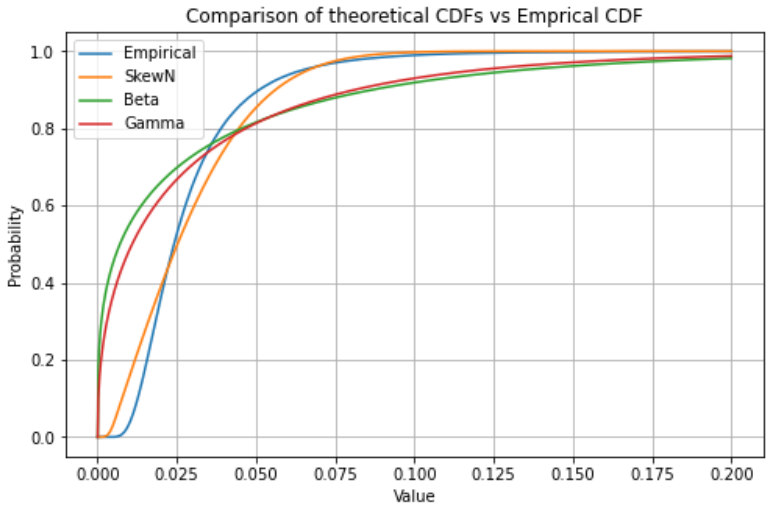

Figure 4.1: The CDFs of the three parametric distributions compared to the empirical CDF of the target distribution, after matching the first three moments.

Figure 4.1 demonstrates that the skew-normal distribution gives the best fit for approximating the probability distribution of $\int_t^T z(s)ds$. The moment generating function of the skew-normal distribution is therefore utilized in calculating the value of the bond. That is, the moment generating function of $\int_t^T z(s)ds$ is approximated by

$$M_{\int_t^T z(s)ds}(t) \approx 2\Phi(\omega\delta t)e^{\xi t+\frac{\omega^2 t^2}{2}},$$

where $\Phi(\cdot)$ is the standard normal CDF. It then follows that

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T z(s)ds}\right] = 2\Phi(\omega\delta\cdot(-1))e^{\xi\cdot(-1)+\frac{\omega^2\cdot(-1)^2}{2}} = 2\Phi(-\omega\delta)e^{-\xi+\frac{\omega^2}{2}}$$

By the independence of the processes $x(t)$ and $z(t)$, the bond price becomes

$$\begin{aligned}
P(t,T) &= e^{-\int_t^T \beta(s)ds}\mathbb{E}^{\mathbb{Q}}\left[e^{-\left(\int_t^T x(s)ds+\int_t^T z(s)ds\right)}\middle|\mathcal{F}_t\right] \\
&= e^{-\int_t^T \beta(s)ds}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T x(s)ds}\middle|\mathcal{F}_t\right]\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T z(s)ds}\middle|\mathcal{F}_t\right], \\
&= e^{-\int_t^T \beta(s)ds}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T x(s)ds}\right]\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T z(s)ds}\right]
\end{aligned}$$

By recognizing the expression of the zero-coupon bond it can be written that

$$\begin{aligned}
P(t,T) =& e^{-\int_t^T \beta(s)ds}\cdot e^{-\frac{x(t)}{a}\left(1-e^{-a(T-t)}\right)+\frac{\sigma_x^2}{2a^2}\left(T-t-\frac{2\left(1-e^{-a(T-t)}\right)}{a}+\frac{1-e^{-a(T-t)}}{2a}\right)} \\
&\cdot 2\Phi(-\omega\delta)e^{-\xi+\frac{\omega_t^2}{2}}, \\
=& 2\Phi(-\omega\delta)\frac{P^M(0,T)}{P^M(0,t)}e^{-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}(V(t,T)-V(0,T)+V(0,t))-\xi+\frac{\omega^2}{2}}.
\end{aligned}$$

4.3. The COS method including collateral

To demonstrate the effectiveness of the COS method for a portfolio with collateral, the method was applied to the two testing portfolios containing 100 and 1000 derivatives, respectively. In these tests, the bonds used as the collateral were chosen to have a notional of \$15000 and a coupon rate of 4%, with the maturities coinciding with the longest maturity found in the corresponding portfolios. The payment dates of these bonds coincide with the time points at which the PFE of the portfolios is calculated. Furthermore, as stated before, the bonds are assumed to fully collateralize the portfolios at $t = 0$. Therefore, the initial Z-spread, $z(0)$, is calibrated, via a root finding algorithm, to equalize the bonds' prices to the initial values of the corresponding portfolios, i.e.

$$\max\left(V_{\text{portfolio}}(x_d(0), x_f(0), X(0), z(0)), 0\right) = V_{\text{bond}}(x_d(0), z(0)). \tag{4.30}$$

4.3.1. Results from the exact COS-PFE method

Using 64 expansion terms for the COS method and 40 quadrature points, the PFE at each time point was calculated and compared with the Monte Carlo method using 5000000 paths. The comparison is given below in Figure 4.2. A good match is observed for both testing cases. Note that we refer to the straight forward extend of the original COS-PFE method to including the bond collateral as the "COS-PFE-exact" method, as no further approximation is added to the original COS-PFE method.

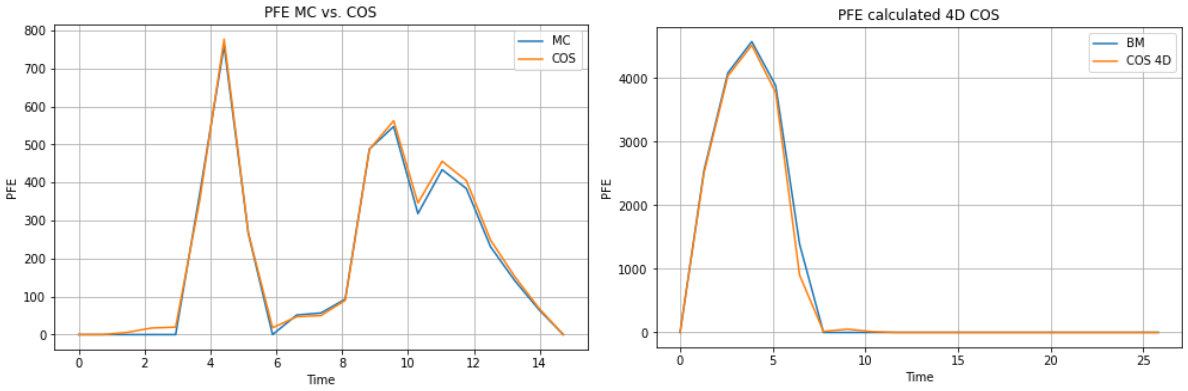

Figure 4.2: Comparison of the PFE estimated using straight forward Monte Carlo simulation with 5000000 paths and the COS-PFE-exact for a test portfolios containing 100 (left) and 1000 derivatives (right).

4.3.2. Dimension reduction via splitting the portfolio

As seen in Equation 4.9, applying the COS-PFE method to four risk factors requires four-dimensional integration. To calculate this integral with reasonable accuracy, a sufficient number of quadrature points must be used. Unfortunately, this requirement increases CPU time.

Our idea to speed up this process is to reduce the dimensionality by seperating the collateral from the rest of the portfolio. We then approximate the Ch.f. of the total portfolio by combining the standalone Ch.f. of the collateral bonds alone and the standalone Ch.f. of the rest of the portfolio. In this way, the four-dimensional integral is broken down in a three-dimensional and a two-dimensional problem as follows

$$\varphi_{\text{portfolio}}(\omega) = \iiint_{\mathbb{R}^3} e^{i\omega V_{\text{portfolio}}(\tilde{z}_{x_d}, \tilde{z}_{x_f}, \tilde{z}_X)} f_{\tilde{Z}_{x_d}} f_{\tilde{Z}_f} f_{\tilde{Z}_X}(\tilde{z}_{x_d}, \tilde{z}_{x_f}, \tilde{z}_X) d\tilde{z}_{x_d} d\tilde{z}_{x_f} d\tilde{z}_X, \tag{4.31}$$

$$\varphi_{\text{bond}}(\omega) = \iint_{\mathbb{R}^2} e^{i\omega V_{\text{portfolio}}(\tilde{z}_{x_d}, \tilde{z}_z)} f_{\tilde{Z}_{x_d}} f_{\tilde{Z}_z}(\tilde{z}_{x_d}, \tilde{z}_z,) d\tilde{z}_{x_d} d\tilde{z}_z. \tag{4.32}$$

From these separate characteristic functions, the parameters of their distribution can be calculated and combined to find the exposure distribution as follows. Using the Clenshaw-Curtis quadrature we calculate $\mathbb{E}\left[V_{portfolio}(t)\right]$, $\mathbb{E}\left[V_{bond}(t)\right]$, $\text{var}\left[V_{portfolio}(t)\right]$, $\text{var}\left[V_{bond}(t)\right]$, and $\text{cov}\left[V_{portfolio}(t), V_{bond}(t)\right]$. With these parameters, the parameters of the exposure distribution can be found by the following

relations:

$$\mathbb{E}[V(t)] = \mathbb{E}\left[V_{portfolio}(t)\right] - \mathbb{E}[V_{bond}(t)],$$
$$\text{var}[V(t)] = \text{var}\left[V_{portfolio}(t)\right] + \text{var}[V_{bond}(t)] - 2\text{cov}\left[V_{portfolio}(t), V_{bond}(t)\right].$$

Here, $V(t)$ is the value of the collateralized portfolio. This method we refer to as the COS-PFE-split method.

Figure 4.3 presents the comparison of the PFE approximations using this split approximation, assuming the exposure is normally distributed, and without the split approximation, which we refer to as the COS-PFE-exact method.

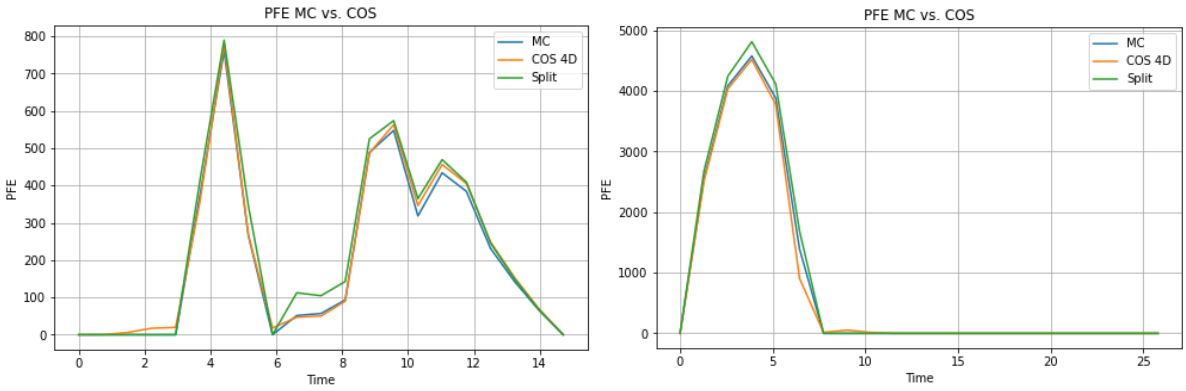

Figure 4.3: Comparison of the PFE among Monte Carlo with 5000000 paths, COS-PFE-exact method and the COS-PFE-split method for a test portfolios containing 100 (left) and 1000 derivatives (right).

In Figure 4.3 it can be seen that by splitting up the four-dimensional integration some accuracy is lost. However, it significantly decreases the CPU time. To demonstrate this, the CPU times using the COS-PFE-exact and -split method were tested five times for each PFE calculation and averaged. These averaged CPU times can be seen in Figure 4.4. Averaged out over all time points, the COS-PFE-exact method took 4.77 and 4.72 seconds for the two portfolios, respectively, while the COS-PFE-split method took only 0.07 and 0.14 seconds. This performance improvement is significant.

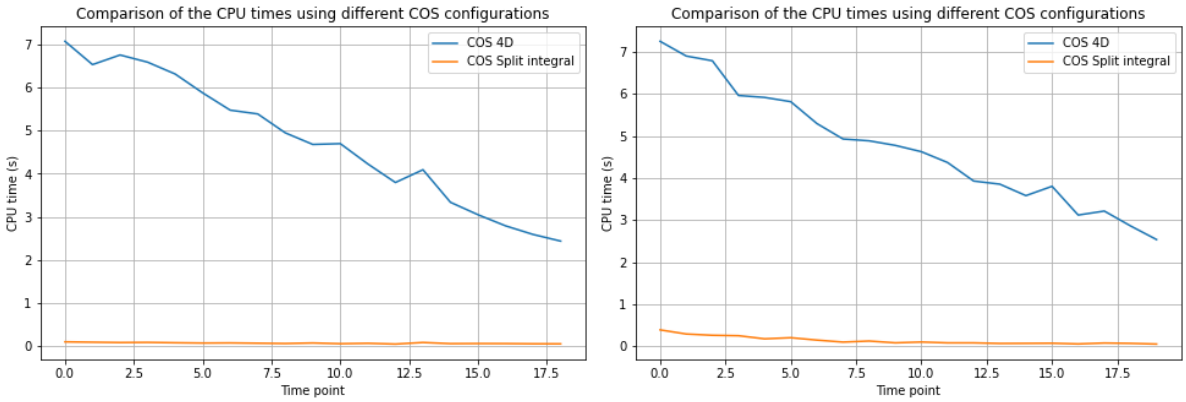

Figure 4.4: Comparison of the CPU time used for computing the PFE between the COS-PFE-exact method and the COS-PFE-split method, for a test portfolios containing 100 (left) and 1000 derivatives (right).

This method will be referred to as the "COS-PFE-split" method.

4.3.3. Splitting the portfolio further for real-world situation

The COS-PFE-exact method suffers from the curse of dimensionality, as it involves multi-dimensional numerical integration. We have demonstrated in the previous subsection that the dimensionality can

be reduced via dividing the portfolio into sub-portfolios. Here in this subsection, we apply this idea further, to mimic the real-world situation that a portfolio can involve up to 100 risk factors, which is not feasible for COS-PFE-exact method without splitting it into sub-portfolios.

This split is based on derivative types. In this subsection the portfolios are split up in two sub-portfolios based on the derivative type. One sub-portfolio contains all interest rate derivatives, while the second contains all foreign exchange derivatives. The idea behind this split is to mimic the different trading desks and their corresponding derivatives on a trading floor. Each desk can apply the COS method to their part of the portfolio to find the PFE of the sub-portfolio. From these approximations, the probability density of the whole portfolio can be estimated, allowing for an approximation of the PFE of the entire portfolio.

For a combination of portfolios, [45] suggests that the Value-at-Risk of the whole portfolio can be calculated using

$$VaR_{portfolio}(t) = \sqrt{VaR_{IR}^2(t) + VaR_{FX}^2(t) + 2 \cdot \rho_{IR,FX} \cdot VaR_{IR}(t) \cdot VaR_{FX}(t)}.$$

Here, $\rho_{IR,FX}$ is the correlation between the portfolio containing all interest rate derivatives and the portfolio containing all foreign exchange derivatives. In the case of the PFE calculation, the $VaR_{\cdot}(t)$ is substituted by the $PFE_{\cdot}(t)$ for the portfolios in the following way

$$PFE_{portfolio}(t) = \sqrt{PFE_{IR}^2(t) + PFE_{FX}^2(t) + 2 \cdot \rho_{IR,FX} \cdot PFE_{IR}(t) \cdot PFE_{FX}(t)}. \tag{4.33}$$

The correlation between the two can be quickly calculated using the Clenshaw-Curtis quadrature. By using $\text{cov}[V_{IR}(t), V_{FX}(t)] = \mathbb{E}[(V_{IR}(t) - \mathbb{E}[V_{IR}(t)])(V_{FX}(t) - \mathbb{E}[V_{FX}(t)])]$, where the first moments can be calculated using the Clenshaw-Curtis quadrature.

Unfortunately, as seen in Figure 4.5, this decomposition is not effective for getting a good approximation of the PFE.

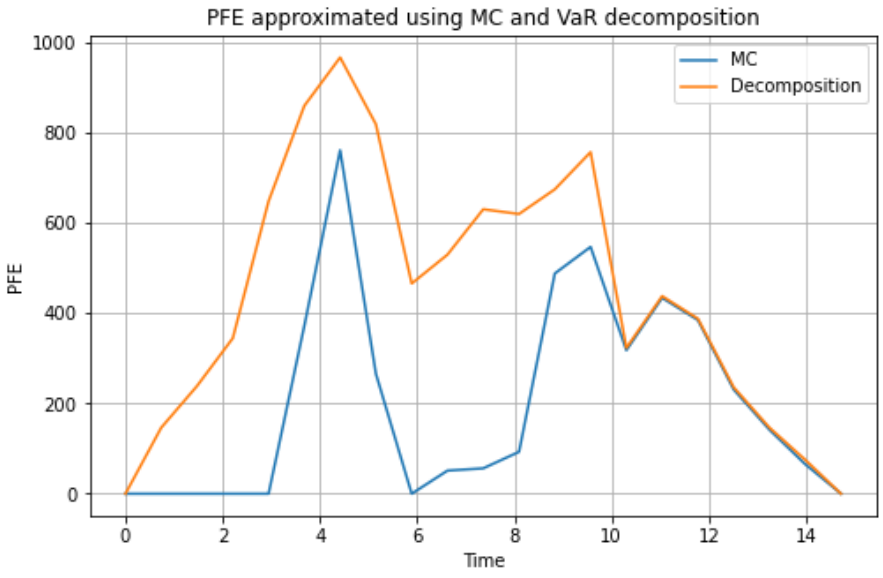

Figure 4.5: The PFE approximations of the Monte Carlo method compared with the PFE found using the PFE decomposition for the portfolio containing 100 derivatives without collateral.

As shown in the last figure, the PFE of the entire portfolio using the decomposition is estimated to be significantly larger than that of approximated using the Monte Carlo method. This is because the decomposition cannot fully capture the influence of the negative PFE. As an example, if the 0.975-quantiles of the MtM values V_{IR} and V_{FX} is 100 and -100, then the PFE of both parts are 100 and

0. The decomposition PFE then becomes $\sqrt{(100)^2 + (0)^2 + 2 \cdot \rho_{IR,FX} \cdot (100) \cdot (0)} = 100$, in this case the negative MtM values of V_{FX} are not taken into account, thus resulting in a much higher PFE using the decomposition compared to the real portfolio.

The method used for splitting up the portfolio relies on the Clenshaw-Curtis quadrature to calculate the distribution parameters of each part, and is essentially the same as the COS-PFE-split method. These parameters are then used to construct a CDF from which the PFE can be found.

For the portfolio without collateral,

$$\begin{aligned}\mathbb{E}\left[V_{portfolio}(t)\right] &= \mathbb{E}\left[V_{FX}(t)\right] + \mathbb{E}\left[V_{FX}(t)\right], \\ \text{var}\left[V_{portfolio}(t)\right] &= \text{var}\left[V_{IR}(t)\right] + \text{var}\left[V_{FX}(t)\right] + 2\text{cov}\left[V_{IR}(t), V_{FX}(t)\right].\end{aligned}$$

For the portfolio with collateral,

$$\begin{aligned}\mathbb{E}\left[V_{portfolio}(t)\right] =& \mathbb{E}\left[V_{FX}(t)\right] + \mathbb{E}\left[V_{FX}(t)\right] - \mathbb{E}\left[V_{bond}(t)\right], \\ \text{var}\left[V_{portfolio}(t)\right] =& \text{var}\left[V_{IR}(t)\right] + \text{var}\left[V_{FX}(t)\right] + \text{var}\left[V_{bond}(t)\right] \\ &+ 2\text{cov}\left[V_{IR}(t), V_{FX}(t)\right] - 2\text{cov}\left[V_{bond}(t), V_{IR}(t)\right] - 2\text{cov}\left[V_{bond}(t), V_{FX}(t)\right].\end{aligned}$$

From Figure 4.6 it can be seen that this method is far more effective compared to the PFE decomposition in Equation 4.33. The following figures were made making with the assumption that the CDF made using the parameters above follows a normal distribution.

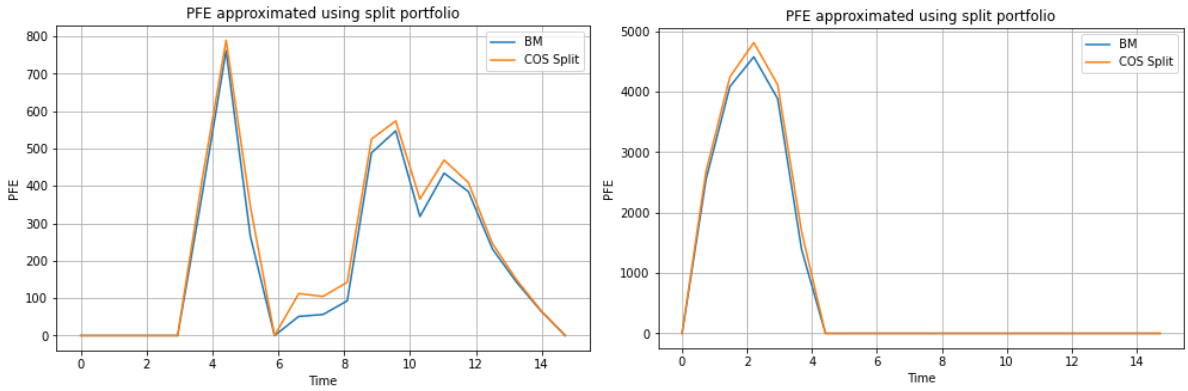

Figure 4.6: The PFE approximations using the split portfolio compared to the benchmark for the portfolio containing 100 derivatives without collateral (left) and the portfolio containing 1000 derivatives with collateral (right).

Unfortunately, the PFE approximations using the sub-portfolios are less accurate than those of the COS-PFE-exact method. However, as will be shown in the results section, the accuracy of the PFE approximation is still sufficient to yield a significant variance reduction in importance sampling.

5

Our contribution 2: Using COS as the control variate for Monte Carlo simulation

The Control Variate method is a variance reduction technique that makes use of the high correlation value between an auxiliary random variable with known properties (such as the mean value) and the variable of interest. This auxiliary random variable is referred to as the control variate.

Define X as the random variable for which the expected value needs to be approximated. If Z is used as the control variate, a new random variable $\tilde{X}$ can defined as

$$\tilde{X} = X + c \cdot (Z - \mathbb{E}[Z]). \tag{5.1}$$

$\tilde{X}$ is unbiased as

$$\begin{aligned} \mathbb{E}[\tilde{X}] &= \mathbb{E}\left[X + c \cdot (Z - \mathbb{E}[Z])\right], \\ &= \mathbb{E}[X] + c \cdot (\mathbb{E}[Z] - \mathbb{E}\left[\mathbb{E}[Z]\right]), \\ &= \mathbb{E}[X]. \end{aligned}$$

The variance of this new random variable is

$$\text{var}[\tilde{X}] = \text{var}[X] + c^2 \cdot \text{var}[Z] + 2c \cdot \text{cov}\left[X, Z\right] = \text{var}[X] + c \cdot \text{cov}\left[X, Z\right]. \tag{5.2}$$

One can see that the variance of $\tilde{X}$ can be made smaller than that of X by tuning the value of c. Differentiating Equation 5.2 with respect to c and solving it results in Equation 5.3 which minimizes the variance of $\tilde{X}$.

$$c = -\frac{\text{cov}[X, Z]}{\text{var}[Z]}. \tag{5.3}$$

From this equation it can be seen that the higher the correlation between X and Z, the lower the variance of $\tilde{X}$.

5.1. Definitions

Recall that the PFE, is defined as the 97.5% quantile of the exposure distribution. This requires one to first obtain the exposure distribution $F(y)$. Also recall that the exposure is defined as $Y(\mathbf{X}) = \max(V(\mathbf{X}), 0)$, where $V(\mathbf{X})$ is the MtM-value of the portfolio driven by the risk factors $\mathbf{X}$. Then we have the corresponding Monte Carlo CDF estimator

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) \leq y\right\}. \tag{5.4}$$

The Control Variate CDF estimator reads:

$$\hat{F}_{n,CV}(y) := \hat{F}_n(y) + c_x \cdot (\hat{g}_n - \mathbb{E}[g(Z)]), \tag{5.5}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) \leq y\right\} + c_x \cdot \left(g(\hat{Z}_i) - \mathbb{E}\left[g(Z)\right]\right). \tag{5.6}$$

In the above equation $\hat{Z}_i$ are the samples of the control variate, $g : \mathbb{R} \to \mathbb{R}$, $\hat{g}_n = \frac{1}{n}\sum_{i=1}^{n} g(\hat{Z}_i)$, and $\hat{F}_n(y)$ is as in Equation 3.6. Again, $\hat{\mathbf{X}}_{(i)}$ denotes the i-th simulated sample of a vector of risk factors in a Monte Carlo simulation.

Similarly to how c was found in Equation 5.3 it can be found that

$$\begin{aligned} c_x &= -\frac{\text{cov}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}, g(Z)\right]}{\text{var}[g(Z)]}, \\ &= -\frac{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) \leq y\right\} g(\hat{Z}_i) - \hat{F}_n(y)\hat{g}_n}{\frac{1}{n}\sum_{i=1}^{n}(g(\hat{Z}_i) - \hat{g}_n)^2}. \end{aligned} \tag{5.7}$$

In [9] it was shown that, by combining Equations 5.6 and 5.7, the control variate estimator $\hat{F}_{n,CV}(y)$ can be rewritten as

$$\hat{F}_{n,CV}(y) = \sum_{i=1}^{n} W_i \cdot \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) \leq y\right\}, \tag{5.8}$$

where, it can be shown that by defining the function g as $g(x) = \mathbf{1}\{x \leq q'_\alpha\}$, where q'_α is the α-quantile of the known distribution of the control variate Z. It follows that $\mathbb{E}[g(Z)] = \alpha$, and $g_n = \frac{N_0}{n}$, where $N_0 = \sum_{i=1}^{n} \mathbf{1}\{\hat{Z}_i \leq q'_\alpha\}$. Sequentially, it follows that

$$W_i = \mathbf{1}\{\hat{Z}_i \leq q'_\alpha\} \cdot \frac{\alpha}{N_0} + \mathbf{1}\{\hat{Z}_i > q'_\alpha\} \cdot \frac{1-\alpha}{n - N_0}. \tag{5.9}$$

5.2. A control variate using the COS-PFE method

The control variate needs to be highly correlated with the exposure value and needs to have known statistical properties, if not the distribution.

The key insight we have here is that the exposure distribution obtained by the COS-PFE method can be used to construct an effective control variate. More specifically, the control variate has a marginal distribution obtained by the COS-PFE method. A high correlation between this control variate and the exposure in the Monte Carlo simulation is ensured via an intermediate variable, which we assume to follow a drift-less GBM for simplicity.

This intermediate variable we denote as $C(t)$ and following our assumption we have

$$dC(t) = C(t)dW(t). \tag{5.10}$$

Applying Itô's lemma to the SDE yield the solution,

$$\log(C(t)) = \log(C(0)) - \frac{1}{2}dt + dW(t), \tag{5.11}$$

$$= \log(C(0)) - \frac{1}{2}dt + \sqrt{t}\hat{Z}_C(t). \tag{5.12}$$

To make sure the control variate is highly correlated with the portfolio exposure, we force it to be highly correlated with the risk factors. The extended correlation matrix of all risk factors including the Gaussian variable from the control variate looks as follows:

$$\Sigma = \begin{bmatrix} 1 & \text{Cor}(\hat{Z}_d, \hat{Z}_f) & \text{Cor}(\hat{Z}_d, \hat{Z}_X) & \text{Cor}(\hat{Z}_d, \hat{Z}_C) \\ \text{Cor}(\hat{Z}_f, \hat{Z}_d) & 1 & \text{Cor}(\hat{Z}_f, \hat{Z}_X) & \text{Cor}(\hat{Z}_f, \hat{Z}_C) \\ \text{Cor}(\hat{Z}_X, \hat{Z}_d) & \text{Cor}(\hat{Z}_X, \hat{Z}_f) & 1 & \text{Cor}(\hat{Z}_X, \hat{Z}_C) \\ \text{Cor}(\hat{Z}_C, \hat{Z}_d) & \text{Cor}(\hat{Z}_C, \hat{Z}_f) & \text{Cor}(\hat{Z}_C, \hat{Z}_X) & 1 \end{bmatrix}, \tag{5.13}$$

where

$$\text{Cor}(\hat{Z}_d, \hat{Z}_C) = \text{Cor}(\hat{Z}_C, \hat{Z}_d) = \frac{\frac{\rho_{dC}}{a_d}(1 - e^{-a_d t})}{\sqrt{\frac{1}{2a_d}(1 - e^{-2a_d t})}\sqrt{t}}, \tag{5.14}$$

$$\text{Cor}(\hat{Z}_f, \hat{Z}_C) = \text{Cor}(\hat{Z}_C, \hat{Z}_f) = \frac{\frac{\rho_{fC}}{a_f}(1 - e^{-a_f t})}{\sqrt{\frac{1}{2a_f}(1 - e^{-2a_f t})}\sqrt{t}}, \tag{5.15}$$

$$\text{Cor}(\hat{Z}_X, \hat{Z}_C) = \text{Cor}(\hat{Z}_C, \hat{Z}_X) = \frac{\rho_{XC}}{\sqrt{t}\sqrt{t}} = \frac{\rho_{XC}}{t}. \tag{5.16}$$

To have a positive correlation between $Y(\mathbf{X})$ and Z it must be true that the correlations between $Y(\mathbf{X})$ and the risk factors are comparable to the correlations between Z and the risk factors. For example, if a decrease in a risk factor causes a decrease in $Y(\mathbf{X})$, then for $Y(\mathbf{X})$ and Z to have positive correlation it must also be true that Z decreases in proportion to $Y(\mathbf{X})$.

Each time the risk factors are generated and then the exposure values are calculated. From these values, the correlation values between each risk factor and the exposure value are calculated and set as ρ_{dC}, ρ_{fC} and ρ_{XC}. Using these correlation coefficients the paths of $C(t)$ can be generated.

Inverse sampling is then applied to translate the values generated by $C(t)$ to samples of control variable. As the distribution of $C(t)$ is known, its CDF, $F_C(x)$, can be used to find the probabilities of the values of $\hat{C}_1(t), \ldots, \hat{C}_n(t)$. Next, the COS method can be used to retrieve the CDF of the exposure of the portfolio, $F_{COS}(x)$. Using $F_{COS}(x)$ the probabilities found by $F_C(x)$ can be translated into the auxiliary random variable Z by

$$\hat{Z}_i(t) = F^{-1}_{COS}(p) \qquad \text{where} \qquad p = F_C(\hat{C}_i(t)). \tag{5.17}$$

Doing this for n simulations, $\hat{C}_1(t), \ldots, \hat{C}_n(t)$ are translated into the samples of the control variate, $\hat{Z}_1(t), \ldots, \hat{Z}_n(t)$. Additionally, the COS method is used to estimate the PFE of the portfolio which is used as the $q'_{0.975}$-quantile in Equation 5.9. Finally, by using the variables $\hat{Z}_1(t), \ldots, \hat{Z}_n(t)$, the quantile $q'_{0.975}$, and $g(x) = \mathbf{1}\{x \leq q'_{0.975}\}$ the estimator $\hat{F}_{n,CV}$, of Equation 5.8, can be constructed and applied on the PFE calculation.

5.3. Theoretical variance reduction

The CDF estimators $\hat{F}_n(y)$ and $\hat{F}_{n,CV}(y)$ obey the central limit theorem. For the estimator $\hat{F}_n(y)$ it is seen that for the theoretical CDF $F(yx)$,

$$\sqrt{n}\left(\hat{F}_n(y) - F(y)\right) \to \mathcal{N}(0, \sigma^2), \tag{5.18}$$

as $n \to \infty$, where $\sigma^2 = F(y)(1 - F(y))$.

Doing the same for $\hat{F}_{n,CV}(y)$ results in

$$\sqrt{n}\left(\hat{F}_{n,CV}(y) - F(y)\right) \to \mathcal{N}(0, \sigma^2_{CV}), \tag{5.19}$$

as $n \to \infty$. Here, the variance σ^2_{CV} is defined as

$$\begin{aligned}\sigma^2_{CV} &= F(y)(1 - F(y)) - \frac{\text{cov}[\mathbf{1}\left\{Y(\mathbf{X}) \leq y\right\}, \mathbf{1}\{Z \leq q'_\alpha\}]^2}{\text{var}[\mathbf{1}\{Z \leq q'_\alpha\}]}, \\ &= F(y)(1 - F(y))(1 - \rho^2).\end{aligned}$$

From this last equation it can be seen that the theoretical variance reduction is fully dependent on ρ, or the correlation between $\mathbf{1}\{\max(V(\mathbf{X}(t)), 0) \leq q_\alpha\}$ and $\mathbf{1}\{Z \leq q'_\alpha\}$. From this it can be derived that the theoretical variance reduction is

$$\frac{\sigma^2}{\sigma^2_{CV}} = \frac{F(y)(1 - F(y))}{F(y)(1 - F(y))(1 - \rho^2)} = (1 - \rho^2)^{-1}. \tag{5.20}$$

6

Our contribution 3: Using COS for importance sampling in Monte Carlo simulation

Importance sampling (IS) is a variance reduction technique that relies on a change in the underlying probability distribution that is driving a stochastic process. IS allows for more samples to be generated in and around the area of interest. For this reason, it can be very effective for estimating extreme quantiles [9].

Let us assume that we have a function $Y : \mathbb{R}^m \to \mathbb{R}$, a random vector $\mathbf{X} \in \mathbb{R}^m$ and a set of parameters θ_0 with joint probability density function $p_{\theta_0}(\mathbf{x})$ and joint CDF $P_{\theta_0}(\mathbf{x})$. Then if $p_\theta(\mathbf{x})$ is another joint probability density function with joint CDF $P_\theta(\mathbf{x})$ such that P_{θ_0} is absolutely continuous with respect to P_θ. Where absolutely continuous is defined as follows.

Definition 6.0.1 (Absolutely continuous)**.** *If μ and ψ are two measures on a σ-algebra $\mathcal{B}$ of subsets of X, then ψ is absolutely continuous with respect to μ if $\psi(A) = 0$ for any $A \in \mathcal{B}$ such that $\mu(A) = 0$.*

Here, the absolute continuity of the probability measure is required for there to exist a Radon-Nikodym derivative that can be used to generate samples from a more favorable probability distribution. Then, a change of measure can be applied in the following way,

$$\begin{aligned}\mathbb{E}_{\theta_0}\left[Y(\mathbf{X})\right] &= \int_{\mathbb{R}^m} Y(\mathbf{x}) p_{\theta_0}(\mathbf{x}) d\mathbf{x}, \\ &= \int_{\mathbb{R}^m} Y(\mathbf{x}) p_\theta(\mathbf{x}) \frac{p_{\theta_0}(\mathbf{x})}{p_\theta(\mathbf{x})} d\mathbf{x}, \\ &= \mathbb{E}_\theta\left[Y(\mathbf{X}) \frac{p_{\theta_0}(\mathbf{X})}{p_\theta(\mathbf{X})}\right], \end{aligned} \tag{6.1}$$

where $\mathbb{E}_{\theta_0}$ and $\mathbb{E}_\theta$ are the expectations under P_{θ_0} and P_θ. For notational purposes, define

$$w_{\mathbf{X}}(\theta) = \frac{p_{\theta_0}(\mathbf{X})}{p_\theta(\mathbf{X})}. \tag{6.2}$$

The variance of $Y(\mathbf{X}) w_{\mathbf{X}}(\theta)$ under the new probability density P_θ then is

$$\mathrm{var}_\theta\left[Y(\mathbf{X}) w_{\mathbf{X}}(\theta)\right] = \mathbb{E}_\theta\left[\left(Y(\mathbf{X}) w_{\mathbf{X}}(\theta)\right)^2\right] - \left(\mathbb{E}_\theta\left[Y(\mathbf{X}) w_{\mathbf{X}}(\theta)\right]\right)^2. \tag{6.3}$$

Equation 6.1 shows that as $\mathbb{E}_\theta\left[Y(\mathbf{X}) w_{\mathbf{X}}(\theta)\right] = \mathbb{E}_{\theta_0}\left[Y(\mathbf{X})\right]$ a variance reduction is achieved when

$$\mathbb{E}_\theta\left[\left(Y(\mathbf{X}) w_{\mathbf{X}}(\theta)\right)^2\right] < \mathbb{E}_{\theta_0}\left[Y(\mathbf{X})^2\right]. \tag{6.4}$$

In the case of the PFE estimation, a right quantile estimation is done. To achieve this, IS will be used to construct an estimator of the exposure distribution of the portfolio at time t. To demonstrate how the same change of measure as in Equation 6.1 can be applied to the right tail distribution we define our random variable as $Y(\mathbf{X}) = \max(V(\mathbf{X}), 0)$. Here, $V(\mathbf{X})$ is the MtM-value of the portfolio at time t driven by risk factors $\mathbf{X} \in \mathbb{R}^m$. This yields,

$$\begin{aligned} 1 - F(y) &= P_{\theta_0}\left[Y(\mathbf{X}) > y\right], \\ &= \mathbb{E}_{\theta_0}\left[\mathbf{1}\{Y(\mathbf{X}) > y\}\right], \\ &= \int_{\mathbb{R}^m} \mathbf{1}\{Y(\mathbf{x}) > y\} p_{\theta_0}(\mathbf{x}) d\mathbf{x}, \\ &= \int_{\mathbb{R}^m} \mathbf{1}\{Y(\mathbf{x}) > y\} p_{\theta}(\mathbf{x}) w_{\mathbf{x}}(\theta) d\mathbf{x}, \\ &= \mathbb{E}_{\theta}\left[\mathbf{1}\{Y(\mathbf{X}) > y\} w_{\mathbf{X}}(\theta)\right]. \end{aligned} \tag{6.5}$$

Using this transformation the IS estimator of $F(y)$ becomes

$$\hat{F}_{n,\text{IS}}(y) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) > y\right\} w_{\hat{\mathbf{X}}_{(i)}}(\theta), \tag{6.6}$$

with corresponding α-quantile estimator $\hat{q}_\alpha = \hat{F}^{-1}_{n,\text{IS}}(\alpha)$. Similarly as before, $\hat{\mathbf{X}}_{(i)}$ denotes the i-th simulated sample of a vector of risk factors in a Monte Carlo simulation.

The goal of IS is to reduce the variance of the estimator of $\hat{q}_\alpha$, by reducing the variance of $\hat{F}_{n,IS}(q_\alpha)$. The variance of the IS estimator at a level y is

$$\text{var}_\theta\left[\hat{F}_{n,IS}(y)\right] = \frac{\mathbb{E}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) > y\} w^2_{\mathbf{X}}(\theta)\right] - F(y)^2}{n}. \tag{6.7}$$

Here, the last expression is found by seeing that

$$\begin{aligned} \mathbb{E}_\theta\left[\hat{F}_{n,IS}(y)\right] &= 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\theta\left[\mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) > y\right\} w_{\hat{\mathbf{X}}_{(i)}}(\theta)\right], \\ &= 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left[\mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) > y\right\}\right], \\ &= 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left[\mathbf{1}\{Y(\mathbf{X}) > y\}\right], \\ &= F(y). \end{aligned}$$

Equation 6.7 indicates that the variance can be reduced vai a carefully chosen the auxiliary density $p_\theta(\mathbf{x})$. On the one hand, a well-chosen auxiliary density can greatly reduce the variance of the estimator. On the other hand, choosing a poor auxiliary density can result in only a very few useful samples being generated. For the samples that are not useful their value $w_{\mathbf{X}}(\theta)$ is usually very small. From this it can be seen that by choosing an auxiliary density that produces a small amount of useful samples a lot of samples are not evaluated in the final estimator as $w_{\mathbf{X}}(\theta) \approx 0$. This results in a final estimator which has a higher variance and a possible bias. This concept is known as weight degeneracy [14].

In theory, there exists an optimal auxiliary density that eliminates the variance. From Equation 6.7 it can be seen that this is the case when the auxiliary density solves

$$\mathbb{E}_{\theta_{\text{opt}}}\left[\mathbf{1}\{Y(\mathbf{X}) > y\} w^2_{\mathbf{X}}(\theta_{\text{opt}})\right] = F^2(y). \tag{6.8}$$

The probability density $p_{\theta_{\text{opt}}}$ that solves the equation above can be expressed as

$$\mathbf{1}\{Y(\mathbf{X}) > y\} \left(\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta_{\text{opt}}}(\mathbf{X})} \right)^2 = F^2(y),$$
$$p_{\theta_{\text{opt}}}(\mathbf{X}) = \frac{\mathbf{1}\{Y(\mathbf{X} > y\} \cdot p_{\theta_0}(\mathbf{X})}{F(y)},$$

and is referred to as the zero-variance estimator.

Unfortunately, this density is just a theoretical one, as for quantile estimation $y = q_\alpha$ the quantile q_α is an unknown value.

It is very hard to find a suitable density before using Importance Sampling in the Monte Carlo simulations. To counter this problem adaptive importance sampling (AIS) was introduced in [15]. AIS constructs the auxiliary density using an algorithm to find an approximation of the optimal density. In the rest of this section two algorithms will be introduced. The first algorithm is the shift-AIS-COS algorithm, which finds a shift that shifts the joint distribution of the risk factors towards the target quantile. In this way the risk factors are sampled much more often (than in the straight forward Monte Carlo simulation) in our area of interest. The second algorithm is the CE-AIS-COS method which aims to achieve the same goal but finds the best "manipulation" to the original density function using the Cross-Entropy (CE) method.

6.1. Adaptive Importance Sampling using the optimal shift

Our aim of applying IS is to generate risk factor scenarios that exhibit a smaller variance of the PFE calculation. As stated in [22], to reduce the variance, the likelihood ratio must be small on the set $\{\max(V(\mathbf{X}), 0) > \text{PFE}\}$. To achieve this the IS density must be chosen such that $p_\theta(\mathbf{x})$ is large on this set. The risk factors are generated by taking random samples of a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. The aim of the algorithm introduced in this section is to find a shift δ such that the risk factors generated by $\mathcal{N}(\boldsymbol{\mu}+\boldsymbol{\delta}, \Sigma)$ result in a larger amount of exposure values being generated around the PFE.

In [20], the authors introduce a new quantile estimator with the use of adaptive importance sampling. This estimator approximates the non-unique quantiles for a general distribution using an iterative algorithm. In the article, the authors introduce Lemma 6.1.1 where $p_\theta(x)$ is defined on a measurable space $(\mathcal{X}, \mathcal{F})$, with some reference measure λ, and a countably generated σ-field $\mathcal{F}$.

Lemma 6.1.1. *Let $(\theta_n)_{n\geq 0}$ be a sequence of parameters and $X_n \sim p_{\theta_{n-1}} d\lambda$. Define $\mathcal{F}_n = \sigma(\theta_0, \dots, \theta_n, X_1, \dots, X_n)$. Then for $f \in \mathcal{L}_1(\theta_0)$,*

$$M_n = \sum_{i=1}^{n} (w_{X_i}(\theta_{i-1}) f(X_i) - \mathbb{E}_{\theta_0}[f(X)]) \tag{6.9}$$

is a martingale with respect to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$.

Here,

$$\mathcal{L}_1(\theta) = \{f : \mathcal{X} \to \mathbb{R} : f \text{ is } \mathcal{F}\text{-measurable}, ||f||^1_{\theta,1} = \mathbb{E}_\theta \left[|f(X)| \right] < \infty\}, \tag{6.10}$$

and the sequence $(\theta_n)_{n\geq 0}$ estimates the optimal solution θ^*.

In the case of the PFE calculation $f\left(\hat{\mathbf{X}}_{(i)}\right) = \mathbf{1}\left\{\max\left(V\left(\hat{\mathbf{X}}_{(i)}\right), 0\right) > PFE\right\}$ from which it is clear to see that $f \in \mathcal{L}_1(\theta_0)$, where $\theta_0 = \boldsymbol{\mu}$. From this Lemma 6.1.1 follows.

With Lemmma 6.1.1 in mind a reference to [46] can be made. In this article, Arouna combines the strong law of large numbers and the central limit theorem with classical martingale convergence to construct Theorem 6.1.1. Let $\boldsymbol{\Theta}$ be the parameter space $\{p_\theta(\mathbf{x}) \mid \theta \in \Theta\}$ is the set of all possible distributions. Because $\hat{\mathbf{X}}_{(i)} \sim N(\boldsymbol{\mu}, \Sigma)$ it follows that the parameter space is defined as $\Theta = \{\theta \in \mathbb{R}^m\}$ where m is the number of risk factors.

Theorem 6.1.1. *Let θ_n, X_n, and $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ be as in Lemma 6.1.1. Assume that $\theta_n \to \theta^* \in \boldsymbol{\Theta}$ converges almost surely and that there exists $a > 1$ such that for all $\theta \in \boldsymbol{\Theta}$*

$$\mathbb{E}_\theta \left[|w_X(\theta) f(X)|^{2a} \right] < \infty, \tag{6.11}$$

the function $g : \theta \to \mathbb{E}_\theta \left[|w_X(\theta) f(X)|^{2a} \right]$ is continuous in θ^*, and

$$\mathbb{E}\left[g(\theta_n) \right] < \infty \qquad \forall n \geq 0. \tag{6.12}$$

Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_{X_i}(\theta_{i-1}) f(X_i) = \mathbb{E}_{\theta_0}[f(X)] \qquad a.s. \tag{6.13}$$

and

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} w_{X_i}(\theta_{i-1}) f(X_i) - \mathbb{E}_{\theta_0}[f(X)] \right) \to^d N\left(0, \sigma_f^2(\theta^*)\right), \tag{6.14}$$

where $\to^d$ denotes convergence in distribution.

In [20] Theorem 6.1.1 was used to conclude that $\frac{1}{n}\sum_{i=1}^{n} w_i \to 1$ almost surely, where $w_i = w_{X_i}(\theta_{i-1})$.

As the PFE is a right quantile by definition, the empirical CDF can be defined in a way that emphasizes the right tail of the distribution,

$$F^r_{n,w,v}(y) = 1 - \frac{1}{v(n)} \sum_{i=1}^{n} w_i \mathbf{1}\{Y_i > y\}. \tag{6.15}$$

Here $v : \mathbb{N} \to \mathbb{R}^+$ is a normalization function.

To prove the convergence, Egloff and Leippold first made an assumption.

Assumption 1. *$(\mathcal{K}_j)_{j\in\mathbb{N}}$ is a compact exhaustion of the parameter space Θ. The sequence $(\theta_n)_{n\geq 0}$ satisfies*

$$\theta_n \to \theta^* \in \Theta \qquad a.s. \tag{6.16}$$

For any $\rho \in (0,1)$, there exists a constant $C(\rho)$ such that

$$\mathbb{P}\left(\sup_{n\geq 1} \kappa_n \geq j \right) \leq C(\rho)\rho^j, \tag{6.17}$$

where κ_n is the counter of the active truncation set of $(\theta_n)_{n\geq 0}$ such that $\theta_j \in \mathcal{K}_{\kappa_n}$ for all $j \leq n$. For some $p^* > 1$, there exists $W \in \mathcal{L}_{p^*}(\theta_0)$ such that for any compact set $\mathcal{K} \subset \Theta$,

$$\mathbf{1}\{\theta \in \mathcal{K}\} w_x(\theta) \leq C_{p^*}(diam(\mathcal{K}))W(x), \tag{6.18}$$

where $C_{p^*}(diam(\mathcal{K}))$ is a constant only depending on p^* and the diameter of $\mathcal{K}$. The compact covering $\Theta = \bigcup_{j=1}^{\infty} \mathcal{K}_j$ where $\mathcal{K}_j \subset int(\mathcal{K}_{j+1})$ is selected such that

$$C_{p^*}(diam(\mathcal{K}_j)) \leq e^{k_{p^*} + j m_{p^*}} \tag{6.19}$$

for some positive constants k_{p^*}, m_{p^*}.

These assumptions imply that firstly, every θ_n remains in Θ. Secondly, the continuity of moments as a function of θ. Thirdly, there is an upper bound to the growth of the set $\mathcal{K}_j$. Using these assumptions the Egloff and Leippold formulated the following theorem.

Theorem 6.1.2. *Assume that the distribution function $F(y) = \mathbb{P}(Y \leq y)$ is strictly increasing at q_α. Under Assumption 1,*

$$q^r_{n,w,v}(\alpha) \to q_\alpha \qquad a.s. \quad as \quad n \to \infty \tag{6.20}$$

for $v(n) = n$.

Sketch of the proof. To prove the claim the authors proved that

$$\mathbb{P}(q_{n,w,v}(\alpha) \leq q_\alpha - \delta \;\; i.o.) = \mathbb{P}(q_{n,w,v}(\alpha) > q_\alpha + \delta \;\; i.o.) = 0. \tag{6.21}$$

The authors show that the sets

$$A_n^r(\delta) = \Big\{ \sum_i (w_i \mathbf{1}\{Y_i \leq q_\alpha - \delta\} - F(q_\alpha - \delta)) \tag{6.22}$$

$$\geq v(n)\alpha \sum_i w_i - nF(q_\alpha - \delta) \Big\}, \tag{6.23}$$

and

$$B_n^r(\delta) = \{ q_{n,w,v}^r(\alpha) > q_\alpha + \delta \}, \tag{6.24}$$

$$= \Big\{ \sum_i (w_i \mathbf{1}\{Y_i > q_\alpha + \delta\} - F(q_\alpha + \delta))$$

$$< \alpha \sum_i w_i - nF(q_\alpha + \delta) \Big\} \tag{6.25}$$

satisfy $A_n^r \cap W_n(\eta) \subset A_n^{r,LIL}(\delta, \eta)$ and $B_n^r \cap W_n(\eta) \subset B_n^{r,LIL}(\delta, \eta)$ where

$$W_n(\eta) = \Big\{ \Big| \sum_i (w_i - 1) \Big| \leq (1+\eta)\phi(nv_\alpha) \Big\}, \tag{6.26}$$

$$A_n^{r,LIL}(\delta, \eta) = \Big\{ \sum_i (w_i \mathbf{1}\{Y_i > q_\alpha - \delta\} - (1 - F(q_\alpha - \delta)))$$

$$\leq -(1+\eta)\phi(nv_{q_\alpha - \delta}) \Big\}, \tag{6.27}$$

$$B_n^{r,LIL}(\delta, \eta) = \Big\{ \sum_i (w_i \mathbf{1}\{Y_i > q_\alpha + \delta\} - (1 - F(q_\alpha + \delta))) \tag{6.28}$$

$$\geq (1+\eta)\phi(nv_{q_\alpha + \delta}) \Big\} \tag{6.29}$$

Where $\phi(t) = \sqrt{2t \log\left(\log(t)\right)}$. From there the authors show that

$$\mathbb{P}(A_n^{r,LIL}(\delta, \eta) \;\; i.o.) = \mathbb{P}(B_n^{r,LIL}(\delta, \eta) \;\; i.o.) = 0 \tag{6.30}$$

from which the claim follows. The complete proof is provided in [20].

Furthermore, the authors proved Theorem 6.1.3.

Theorem 6.1.3. *Suppose the conditions in Assumption 1 hold. If there exists $\eta > 0$, $k > 0$, and $0 < \gamma < \frac{1}{2}$ such that*

$$n + \frac{1+\eta}{1-\alpha}\sqrt{2nv_\alpha \log_e\left(\log_e(nv_\alpha)\right)} \leq v(n) \leq n + kn^{\frac{1}{2}+\gamma}, \tag{6.31}$$

then

$$q_{n,w,v}^r(\alpha) \to q_\alpha \quad a.s. \quad as \quad n \to \infty. \tag{6.32}$$

Here $v_\alpha = \sigma^2_{\mathbf{1}\{(q_\alpha,\infty)\}\circ E_t}(\theta^*)$

Sketch of the proof. The proof of this theorem is along the lines of the proof of Theorem 6.1.2. By using Equation 6.15 and the fact that if F is a right continuous increasing function we have that $F(x) \geq \alpha$ if

and only if $F^{\leftarrow}(\alpha) \leq x$ it can be seen that

$$\begin{aligned} A_n^r(\delta) =& \Big\{ \sum_i (w_i \mathbf{1}\{Y_i > q_\alpha - \delta\} - (1 - F(q_\alpha - \delta))) \\ &\leq v(n)(1-\alpha) - n(1 - F(q_\alpha - \delta)) \Big\}, \end{aligned} \tag{6.33}$$

$$\begin{aligned} B_n^r =& \Big\{ \sum_i (w_i \mathbf{1}\{Y_i > q_\alpha\} - (1 - F(q_\alpha))) \\ &> v(n)(1-\alpha) - n(1 - F(q_\alpha)) \Big\}. \end{aligned} \tag{6.34}$$

Then again two sets $A_n^{r,LIL}(\delta, \eta)$ and $B_n^{r,LIL}(\eta)$ are constructed that contain $A_n^r(\delta)$ and B_n^r. Then in a similar fashion to the previous sketch of proof, it is shown that

$$\mathbb{P}(A_n^{r,LIL}(\delta, \eta) \ \ i.o.) = \mathbb{P}(B_n^{r,LIL}(\eta) \ \ i.o.) = 0 \tag{6.35}$$

from which the statement of the theorem follows. Again, the full proof can be found in [20].

From Theorem 6.1.2 it is known that for $q_{n,w,v}^r(\alpha)$ to converge to q_α it must hold that $v(n) = n$. Additionally, the bounds in Equation 6.31 of Theorem 6.1.3 imply that $v(n)$ is closest to n if $\sqrt{2nv_\alpha \log_e\left(\log_e(nv_\alpha)\right)}$ is minimized, which is the case if v_α is minimized. From this, we have

$$\theta^* = \arg\min_\theta \ \text{var}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > q_\alpha\} w_{\mathbf{X}}(\theta)\right]. \tag{6.36}$$

This is equivalent to finding the θ that minimizes the second moment

$$\theta^* = \arg\min_\theta \ \mathbb{E}_\theta\left[|\mathbf{1}\{\max(V(\mathbf{X}), 0) > q_\alpha\} w_{\mathbf{X}}(\theta)|^2\right] \tag{6.37}$$

which is the minimization suggested in [47].

In [48], Glasserman et al. evaluated an algorithm for estimating the Value-at-Risk for credit portfolios. In the paper, the authors use a delta-gamma approximation to approximate the portfolio's loss. From this approximation, an upper bound for the second moment is derived. This upper bound is then minimized by varying θ such that the first moment of the auxiliary distribution is equal to the quantile that needs to be estimated.

Combining the results from these two papers, a minimization problem can be defined in our case, which examines all combinations of shifts and finds the one using which the mean of the shifted exposure distribution is equal to the PFE and which simultaneously minimizes the second moment of the shifted exposure distribution. Stated as an optimization problem it reads

$$\begin{aligned} \min_\theta \quad & \text{var}\left[\mathbf{1}\{\max(V(\mathbf{X}(t)), 0) > q_\alpha\} w_{\mathbf{X}}(\theta)\right] \\ \text{subjected to} \quad & |\mathbb{E}_\theta\left[E(\mathbf{X}(t))\right] - PFE| \leq \beta. \end{aligned} \tag{6.38}$$

An optimizer is used which minimizes the loss function $\text{var}\left[\mathbf{1}\{\max(V(\mathbf{X}(t)), 0) > q_\alpha\} w_{\mathbf{X}}(\theta)\right]$ under the constraint $|\mathbb{E}_\theta\left[E(\mathbf{X}(t))\right] - PFE| \leq \beta$. Here, $\beta \in \mathbb{R}^+$ defines an area around the PFE in which an optimal θ can be sought for.

Note that the PFE in 6.38 is the quick estimation from the COS-PFE method.

To summerize, this minimization problem is solved in several steps.

First, generate a low number of paths of risk factors and calculate the corresponding MtM values of the portfolio. From these MtM values the correlation between the MtM values and each of the risk factors can be calculated. This quick estimation on the correlation values suggests shifting which

risk factor would result in the most efficient push to portfolios exposures. This step is very necessary as, randomly pushing a risk factor that is negatively correlated with the MtM value of the portfolio would only result in generating MtM values further away from it's PFE.

Secondly, the COS method is used to calculate a "good enough" approximation of the PFE and the EE of the portfolio.

Finally, for the risk factor that has the highest correlation with the MtM values of the portfolio, the Bisection method is used to find the shift value that makes the expected exposure of the portfolio, under the new probability distribution, coincide with the PFE. In this way a new distribution of the risk factors is found which causes the new exposure distribution to have the original PFE as the new expected value. This shift is then used as the initial value in the minimization problem.

The complete algorithm can be summarized as follows.

Algorithm 1 The shift-AIS-COS algorithm.

Initialize:

Generate the paths of x_d, x_f, X and z.

1. Calculate the MtM values for each of the paths.
2. Find the correlation between the MtM valeus and the risk factors from a quick MC simulation using a low number of paths.
3. Use the COS method to find a quick estimation of the expected exposure and PFE of the portfolio, from which a relatively big error can be tolerated.
4. Use the Bisection method on the risk factor having the highest correlation with the portfolio exposure found in step 2, to find the first initial guess of the shift value for 6.38.
5. Solve the minimization problem in 6.38.
6. Use the new probability distribution to sample $\hat{\mathbf{X}}_1, \dots, \hat{\mathbf{X}}_n$.
7. Calculate $\hat{F}_{n,IS}(y) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{Y\left(\hat{\mathbf{X}}_{(i)}\right) > y\right\} w_{\hat{\mathbf{X}}_{(i)}}(\theta)$.
8. Find the PFE estimate by computing $\hat{F}^{-1}_{n,IS}(\alpha)$.

6.2. Adaptive Importance Sampling using the Cross-Entropy method

The Cross-Entropy method is an adaptive importance sampling method which is highly efficient as for multivariate-normal distributions, which our auxiliary density is assumed to be, the optimal distribution has analytical solutions for the parameters. This makes gradient descent methods obsolete. In this section the adaptive importance sampling method using the Cross-Entropy algorithm (CE-AIS) will be discussed. First, the general multi-level CE algorithm will be analyzed. It is followed by the CE-AIS-COS method that will be used for the PFE approximation. After this, the consistency of the parameters and convergence of the algorithm will be discussed. The section is concluded by investigating the theoretical variance reduction of the method we propose here.

6.2.1. The algorithm

The CE-AIS method is based on the minimization of the Kullback-Leibler divergence [23][49]. The goal of the algorithm is to find an auxiliary sampling density $g_{\mu,\Sigma}$ that minimizes this divergence between itself and the zero-variance density $g^\star$. Define $\theta_0 = (\mu_0, \Sigma_0)$ as the set of parameters used in the original probability density and $\theta = (\mu, \Sigma) = \{(\mu, \Sigma) : \mu \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}\}$ as the set of possible parameters for the auxiliary density. Furthermore, define the optimal parameters as

$$\theta^\star = \left(\mu^\star, \Sigma^\star\right) = \underset{\mu\in\mathbb{R}^m, \Sigma\in\mathcal{M}^+_{m\times m}}{\arg\min} \; D\left(g_{\mu,\Sigma}, g^\star\right).$$

Where m is the number of the risk factors, and $\mathcal{M}^{+}_{m\times m}$ is the set of all symmetric, positive-definite matrices in $\mathbb{R}^{m\times m}$. From the definition of the Kullback-Leibler divergence it can be seen that

$$D(g^{\star}, g_{\mu,\Sigma}) = \mathbb{E}_{g^{\star}}\left[\log\left(\frac{g^{\star}(\mathbf{X})}{g_{\mu,\Sigma}(\mathbf{X})}\right)\right] = \mathbb{E}_{g^{\star}}\left[\log(g^{\star}(\mathbf{X})\right] - \mathbb{E}_{g^{\star}}\left[\log(g_{\mu,\Sigma}(\mathbf{X})\right], \tag{6.39}$$

where, in our testing framework, $\mathbf{X} = [x_d, x_f, \log(X), \log(z)]$. Recall, that the logarithms found in this section have base e. To find an auxiliary sampling density $g_{\mu,\Sigma}$ that minimizes the information lost when approximating the optimal auxiliary sampling distribution $g^{\star}$, the Kullback-Leibler divergence $D(g^{\star}, g_{\mu,\Sigma})$ needs to be minimized. From Equation 6.39 it can be seen that as $\mathbb{E}_{g^{\star}}\left[\log(g^{\star}(\mathbf{X}))\right]$ is a constant, $D(g^{\star}, g_{\mu,\Sigma})$ is minimized by maximizing $\mathbb{E}_{g^{\star}}\left[\log(g_{\mu,\Sigma}(\mathbf{X})\right]$.

As already known, the optimal density, or zero-variance estimator, defined for the quantile q_{α} is

$$g^{\star}(\mathbf{X}) = \frac{\mathbf{1}\{\max(V(\mathbf{X}), 0) > q_{\alpha}\} \cdot p_{\theta_0}(\mathbf{X})}{1-\alpha}.$$

Here, p_{θ_0} is the original probability density. It then follows that

$$\begin{aligned}
(\boldsymbol{\mu}^{\star}, \Sigma^{\star}) &= \underset{\mu\in\mathbb{R}^m, \Sigma\in\mathcal{M}^{+}_{m\times m}}{\arg\max} \; \mathbb{E}_{g^{\star}}\left[\log(g_{\mu,\Sigma}(\mathbf{X}))\right], \\
&= \underset{\mu\in\mathbb{R}^m, \Sigma\in\mathcal{M}^{+}_{m\times m}}{\arg\max} \int_{\mathbb{R}^m} \log\left(g_{\mu,\Sigma}(\mathbf{x})\right) g^{\star}(\mathbf{x}) d\mathbf{x}, \\
&= \underset{\mu\in\mathbb{R}^m, \Sigma\in\mathcal{M}^{+}_{m\times m}}{\arg\max} \int_{\mathbb{R}^m} \log\left(g_{\mu,\Sigma}(\mathbf{x})\right) \cdot \frac{\mathbf{1}\{\max(V(\mathbf{x}), 0) > q_{\alpha}\} \cdot p_{\theta_0}(\mathbf{x})}{1-\alpha} d\mathbf{x}, \\
&= \underset{\mu\in\mathbb{R}^m, \Sigma\in\mathcal{M}^{+}_{m\times m}}{\arg\max} \; \mathbb{E}_{p_{\theta_0}}\left[\log(g_{\mu,\Sigma}(\mathbf{X})) \cdot \mathbf{1}\left\{\max(V(\mathbf{X}), 0) > q_{\alpha}\right\}\right].
\end{aligned} \tag{6.40}$$

Equation 6.40 is known as the cross-entropy problem. For the exponential family, of which the multivariate normal distribution is part, analytical solutions exists. For other families of probability densities, a gradient method can be used to find the optimal parameter, since the problem is generally concave and differentiable [50][51]. In the case of the multivariate normal distribution the analytical solutions exist:

$$\boldsymbol{\mu}^{\star} = \mathbb{E}_{p_{\theta_0}}\left[\mathbf{X} | \max(V(\mathbf{X}), 0) > q_{\alpha}\right], \tag{6.41}$$

$$\Sigma^{\star} = \mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\boldsymbol{\mu}^{\star})(\mathbf{X}-\boldsymbol{\mu}^{\star})^{\top} | \max(V(\mathbf{X}), 0) > q_{\alpha}\right]. \tag{6.42}$$

The full derivations of Equations 6.41 and 6.42 can be found in the Appendix. Additionally, in Appendix A.3 it is shown that $\boldsymbol{\mu}^{\star}$ and $\Sigma^{\star}$ give the global maximum of Equation 6.40. Using the Monte Carlo method, these parameters can be approximated by

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{n} w_i \hat{\mathbf{X}}_{(i)},$$

$$\hat{\Sigma} = \sum_{i=1}^{n} w_i (\hat{\mathbf{X}}_{(i)} - \hat{\boldsymbol{\mu}})(\hat{\mathbf{X}}_{(i)} - \hat{\boldsymbol{\mu}})^{\top},$$

where

$$w_i = \frac{w_i'}{\sum_{i=1}^{n} w_i'},$$

$$w_i' = \mathbf{1}\left\{\max\left(V(\hat{\mathbf{X}}_{(i)}), 0\right) > q_{\alpha}\right\} \frac{p_{\theta_0}(\hat{\mathbf{X}}_{(i)})}{g_{\mu,\Sigma}(\hat{\mathbf{X}}_{(i)})}.$$

The general multi-level cross-entropy method for finding the optimal importance sampling density to approximate the α-quantile of the exposure distribution is summarized in the following algorithm.

Algorithm 2 The multi-layer CE-AIS algorithm applied to PFE estimations

Initialize:

Set the iteration $k = 0$, $g_{\mu_0,\Sigma_0} = p_{\theta_0}$ where g_{μ_k,Σ_k} is the auxiliary density used in iteration k. Furthermore, set the number of paths n and the maximum number of iterations k_{max}. Lastly, set a low quantile ρ.

1. Draw a samples of $\hat{\mathbf{X}}_1^{(k)}, \ldots, \hat{\mathbf{X}}_n^{(k)}$ from $g_{\hat{\mu}_k,\hat{\Sigma}_k}$ where

if collateral is used **then**

$$\hat{\mathbf{X}}_{(i)}^{(k)} = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; \log\left(\hat{X}_i^{(k)}\right) \;\; \log\left(\hat{z}_i^{(k)}\right)\right].$$

else

$$\hat{\mathbf{X}}_{(i)}^{(k)} = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; \log\left(\hat{X}_i^{(k)}\right)\right].$$

2. Apply the domain transformation

if collateral is used **then**

$$\hat{\mathbf{Y}}_{(i)}^{(k)} = h\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right) = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; e^{\log\left(\hat{X}_i^{(k)}\right)} \;\; e^{\log\left(\hat{z}_i^{(k)}\right)}\right],$$

else

$$\hat{\mathbf{Y}}_{(i)}^{(k)} = h\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right) = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; e^{\log\left(\hat{X}_i^{(k)}\right)}\right].$$

3. Find the $\hat{q}_\alpha$ quantile estimate of the weighted samples

$$\max\left(V\left(\hat{\mathbf{Y}}_{(1)}^{(k)}\right), 0\right), \ldots, \max\left(V\left(\hat{\mathbf{Y}}_{(n)}^{(k)}\right), 0\right).$$

4. Compute the ρ-quantile γ_t^k of the samples

$$\max\left(V\left(\hat{\mathbf{Y}}_{(1)}^{(k)}\right), 0\right), \ldots, \max\left(V\left(\hat{\mathbf{Y}}_{(n)}^{(k)}\right), 0\right)$$

and set $\gamma = \min\left(\gamma_t^k, \hat{q}_\alpha\right)$.

5. Calculate

$$w_i = \frac{w_i'}{\sum_{i=1}^n w_i'},$$

$$w_i' = \mathbf{1}\left\{\max\left(V\left(\hat{\mathbf{Y}}_{(i)}^{(k)}\right), 0\right) > \gamma\right\} \frac{p_{\theta_0}\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right)}{g_{\hat{\mu}_k,\hat{\Sigma}_k}\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right)}.$$

6. Calculate

$$\hat{\boldsymbol{\mu}}_{k+1} = \sum_{i=1}^n w_i \hat{\mathbf{X}}_{(i)}^{(k)},$$

$$\hat{\Sigma}_{k+1} = \sum_{i=1}^n w_i \left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)\left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)^\top.$$

if $k = k_{max}$ **then**

calculate the final PFE estimate using $g_{\hat{\mu}_{k_{max}},\hat{\Sigma}_{k_{max}}}$,

else

return to step 1 with $k = k + 1$.

The algorithm that is used for the PFE calculation has two major advantages to the original multi-level CE method.

Firstly, the original multi-level CE algorithm is based on the assumption that the rare event q_α has a probability 10^{-5} or lower of being exceeded [49]. As the probability of generating samples above q_α, using p_{θ_0}, is very small the algorithm has a very small probability of converging. To aid the convergence, the multi-level procedure in step 3 is introduced. In this step, a low quantile γ_t^k of the unweighted samples is computed. Using γ_t^k the value of γ can be found by taking the minimum of the low quantile γ_t^k and the empirical quantile $\hat{q}_\alpha$. In this way, at each step the parameters $\hat{\boldsymbol{\mu}}_{k+1}$ and $\hat{\Sigma}_{k+1}$ are calculated using the expectation conditional on $\max(V(\mathbf{X}), 0) > \gamma$. As γ_t^k is increasing in each iteration the parameters $\hat{\boldsymbol{\mu}}_{k+1}$ and $\hat{\Sigma}_{k+1}$ also change accordingly, this causes γ_t^k to increase. This process is repeated until γ_t^k overtakes $\hat{q}_\alpha$. After this happens $\gamma = \hat{q}_\alpha$ and the final parameters are calculated.

To illustrate the shift of these distributions an example of the multi-level CE algorithm is given in Figure 6.1. In this figure the optimal sampling distribution for a high quantile of the standard normal distribution was found. Fortunately, since the PFE is the 0.975-quantile a sufficient number of paths will generate exposure values above the PFE level. Therefore, finding the quantile γ_t^k of the unweighted samples is not necessary and the multi-level step can be skipped.

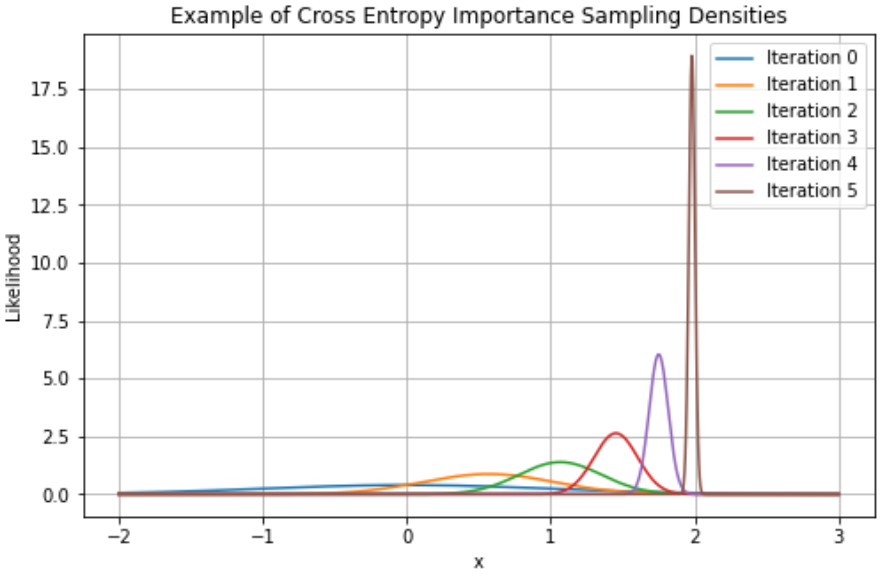

Figure 6.1: An example of the intermediate results from multi-level CE algorithm, which finds the auxiliary distribution for importance sampling to capture a rare event of the standard normal distribution.

Secondly, the biggest advantage is that the COS method can be used to very quickly estimate the PFE of the portfolio. Due to this it is not needed to calculate the empirical quantile of the weighted samples as is done in step 1 of the multi-level CE algorithm.

In the original paper the authors give the option to use smoothing when calculating the new parameters. In this way, an $\alpha \in [0, 1]$ can be chosen such that

$$\hat{\boldsymbol{\mu}}_{k+1} = \alpha \cdot \sum_{i=1}^{n} w_i \hat{\mathbf{X}}_{(i)}^{(k)} + (1 - \alpha) \cdot \hat{\boldsymbol{\mu}}_k,$$

$$\hat{\Sigma}_{k+1} = \alpha \cdot \sum_{i=1}^{n} w_i \left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)\left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)^\top + (1 - \alpha) \cdot \hat{\Sigma}_k.$$

These expressions for $\hat{\mu}_{k+1}$ and $\hat{\Sigma}_{k+1}$ can replace those in step 6 of Algorithm 2. Adding this smoothing parameter α can prevent the convergence to wrong solutions as the parameters $\hat{\mu}_{k+1}$ and $\hat{\Sigma}_{k+1}$ are less influenced by potential outliers. The original paper does not use this smoothing as it did not improve the results. Both dynamic smoothing, where the smoothing parameter changes at each iteration, or static smoothing, using a constant value for α, can be used.

Our CE algorithm for PFE calculations does not use smoothing, as it does not improve the results. However, an extra step is added at the end of the original CE algorithm. This extra step involves averaging the means and covariance matrices found throughout the iterations of the algorithm, instead of only using the mean and covariance matrix found in the last iteration, as is done in the original CE algorithm. The reason for this is that, with a sufficient number of paths, the algorithm converges in one step because the PFE is the 97.5%-quantile, which is not very extreme. In the tests, using the mean instead of the parameters from the last iteration not only improved the accuracy of the results but also allowed the algorithm to achieve this with significantly fewer paths. This is due to the fact that averaging makes the algorithm less vulnerable for outliers being simulated. Additionally, the need for fewer paths sped up the calculation of the new mean and covariance matrix considerably.

The complete algorithm is as follows:

Algorithm 3 The CE-AIS-COS algorithm for PFE estimation.

Initialize:

Set PFE as the PFE value found by the COS method. Set $k = 0$ and $g_{\hat{\boldsymbol{\mu}}_0,\hat{\Sigma}_0} = p_{\theta_0}$ where $g_{\hat{\boldsymbol{\mu}}_k,\hat{\Sigma}_k}$ is the auxiliary density at iteration k. Then, set the number of paths n and maximum number of iterations k_{max}.

1. Draw a sample of $\hat{\mathbf{X}}_{(1)}^{(k)}, \ldots, \hat{\mathbf{X}}_{(n)}^{(k)}$ from $g_{\hat{\boldsymbol{\mu}}_k,\hat{\Sigma}_k}$

if collateral is used **then**

$$\hat{\mathbf{X}}_{(i)}^{(k)} = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; \log\left(\hat{X}_i^{(k)}\right) \;\; \log\left(\hat{z}_i^{(k)}\right)\right],$$

else

$$\hat{\mathbf{X}}_{(i)}^{(k)} = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; \log\left(\hat{X}_i^{(k)}\right)\right].$$

2. Apply the domain transformation

if collateral is used **then**

$$\hat{\mathbf{Y}}_{(i)}^{(k)} = h\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right) = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; e^{\log\left(\hat{X}_i^{(k)}\right)} \;\; e^{\log\left(\hat{z}_i^{(k)}\right)}\right],$$

else

$$\hat{\mathbf{Y}}_{(i)}^{(k)} = h\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right) = \left[\hat{x}_{d,i}^{(k)} \;\; \hat{x}_{f,i}^{(k)} \;\; e^{\log\left(\hat{X}_i^{(k)}\right)}\right].$$

3. Calculate

$$w_i = \frac{w_i'}{\sum_{i=1}^n w_i'},$$

$$w_i' = \mathbf{1}\left\{\max\left(V\left(\hat{\mathbf{Y}}_{(i)}^{(k)}\right), 0\right) > \text{PFE}\right\} \frac{p_{\theta_0}\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right)}{g_{\hat{\boldsymbol{\mu}}_k,\hat{\Sigma}_k}\left(\hat{\mathbf{X}}_{(i)}^{(k)}\right)}.$$

4. Calculate

$$\hat{\boldsymbol{\mu}}_{k+1} = \sum_{i=1}^n w_i \hat{\mathbf{X}}_{(i)}^{(k)},$$

$$\hat{\Sigma}_{k+1} = \sum_{i=1}^n w_i \left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)\left(\hat{\mathbf{X}}_{(i)}^{(k)} - \hat{\boldsymbol{\mu}}_{k+1}\right)^\top.$$

if $k = k_{max}$ **then** calculate the final PFE estimate using $g_{\hat{\boldsymbol{\mu}}_{k_{max}},\hat{\Sigma}_{k_{max}}}$, where

$$\hat{\boldsymbol{\mu}}^\star = \frac{1}{k_{max}} \sum_{i=1}^{k_{max}} \hat{\boldsymbol{\mu}}_i,$$

$$\hat{\Sigma}^\star = \frac{1}{k_{max}} \sum_{i=1}^{k_{max}} \hat{\Sigma}_i.$$

else

return to step 1 with $k = k + 1$.

6.2.2. Consistency and convergence

It is highly desirable for the parameters found by the CE-AIS-COS method to converge to the optimal parameters, which are equal to Equations 6.41 and 6.42. If that is the case, it can be said that the sequence of estimators is asymptotically consistent. In this section we will demonstrate that the estimators produced by the CE-AIS-COS algorithm are indeed asymptotically consistent. First, an introduction to M-estimators and consistency will be provided. Following this, a theorem proving the consistency of the parameters will be presented.

In [47] the M-estimator is defined in the following way.

Definition 6.2.1 (M-estimator). *Let $m_\theta : \mathcal{X} \to \mathbb{R}$ be a known function, and let*

$$\theta \mapsto M_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m_\theta\left(\hat{\mathbf{X}}_{(i)}\right). \tag{6.43}$$

An estimator maximizing $M_n(\theta)$ over a set Θ is called an M-estimator.

The CE-AIS-COS algorithm tries to find $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ such that,

$$\theta^\star = (\boldsymbol{\mu}^\star, \Sigma^\star) = \underset{\boldsymbol{\mu}\in\mathbb{R}^m, \Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max} \; \mathbb{E}_{p_{\theta_0}}\left[\log(g_{\boldsymbol{\mu},\Sigma}(\mathbf{X}))\cdot \mathbf{1}\{\max(V(\mathbf{X}),0) > q_\alpha\}\right], \tag{6.44}$$

from where it was found that the analytical solutions for $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ are

$$\boldsymbol{\mu}^\star = \mathbb{E}_{p_{\theta_0}}\left[\mathbf{X} | \max(V(\mathbf{X}),0) > q_\alpha\right], \tag{6.45}$$

$$\Sigma^\star = \mathbb{E}_{p_{\theta_0}}\left[\left(\mathbf{X}-\boldsymbol{\mu}^\star\right)\left(\mathbf{X}-\boldsymbol{\mu}^\star\right)^\top | \max(V(\mathbf{X}),0) > q_\alpha\right]. \tag{6.46}$$

For a certain random sample of $\mathbf{X}$ with size n, Equation 6.44 is approximated by

$$M_n(\boldsymbol{\mu}, \Sigma) = \frac{1}{n}\sum_{i=1}^{n} \log(g_{\boldsymbol{\mu},\Sigma}(\hat{\mathbf{X}}_{(i)}))\cdot \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}),0) > q_\alpha\}.$$

By defining

$$(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) = \underset{\boldsymbol{\mu}\in\mathbb{R}^m, \Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max} \; \frac{1}{n}\sum_{i=1}^{n} \log(g_{\boldsymbol{\mu},\Sigma}(\hat{\mathbf{X}}_{(i)}))\cdot \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}),0) > q_\alpha\},$$

it can be concluded that the estimators $(\boldsymbol{\mu}^\star, \Sigma^\star)$ are asymptotically consistent if the estimators $(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n)$ converge in probability to $(\boldsymbol{\mu}^\star, \Sigma^\star)$. Theorem 6.2.1, found in [47], will be used to prove the consistency of the parameters produced by the CE-AIS-COS algorithm.

Theorem 6.2.1. *Let M_n be random functions and let M be a fixed function of θ such that for every $\epsilon > 0$ it satisfies*

1. $\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \to^{\mathbb{P}} 0$,
2. $\sup_{\theta: d(\theta,\theta^\star)\geq\epsilon} M(\theta) < M(\theta^\star)$,
3. $M_n(\hat{\theta}_n) \geq M_n(\theta^\star) - o_{\mathbb{P}}(1)$.

Then any sequence of estimators $\hat{\theta}_n$ converges in probability to $\theta^\star$.

In the previous theorem $o_{\mathbb{P}}(1)$ is a sequence that converges to zero in probability as n goes to ∞. To prove the asymptotical consistency of the CE-AIS-COS estimators Lemma 6.2.1, found in [52], is used.

Lemma 6.2.1. *If the data are i.i.d., Θ is compact, $f(\mathbf{X},\theta)$ is continuous at each $\theta\in\Theta$ with probability one, and there is $d(\mathbf{X})$ with $\|f(\mathbf{X},\theta)\|_2 \leq d(\mathbf{X})$ for all $\theta\in\Theta$ and $\mathbb{E}_{p_{\theta_0}}[d(\mathbf{X})] < \infty$, then $\mathbb{E}_{p_{\theta_0}}[f(\mathbf{X},\theta)]$ is continuous and*

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n} f(\hat{\mathbf{X}}_{(i)},\theta) - \mathbf{E}_{p_{\theta_0}}[f(\mathbf{X},\theta)]\right\|_2 \to^{\mathbb{P}} 0.$$

Proof. First, define

$$f(\mathbf{X}, \boldsymbol{\mu}, \Sigma) = \mathbf{1}\{\max(V(\mathbf{X}), 0) > q_\alpha\} \cdot \log(g_{\mu,\Sigma}(\mathbf{X})).$$

First, we will show that $f(\mathbf{X}, \boldsymbol{\mu}, \Sigma)$ is a continuous function for every $(\boldsymbol{\mu}, \Sigma) \in \Theta$. In the case that $\max(V(\mathbf{X}), 0) > q_\alpha$ we get that $f(\mathbf{X}, \boldsymbol{\mu}, \Sigma) = \log(g_{\mu,\Sigma}(\mathbf{X}))$. Since the probability density function is non-zero for each $(\boldsymbol{\mu}, \Sigma) \in \Theta$, where $\Theta = \{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}\}$, the logarithm of $g_{\mu,\Sigma}(\mathbf{X})$ is continuous.

Secondly, in the case that $\max(V(\mathbf{X}), 0) \leq q_\alpha$ we get that $f(\mathbf{X}, \boldsymbol{\mu}, \Sigma) = 0$. And thus, again, $f(\mathbf{X}, \boldsymbol{\mu}, \Sigma)$ is continuous for all $(\boldsymbol{\mu}, \Sigma) \in \Theta$.

Next it is shown that

$$\left\|f(\mathbf{X}, \boldsymbol{\mu}, \Sigma)\right\|_2 \leq \left\|\log(g_{\mu,\Sigma}(\mathbf{X}))\right\|_2 \leq \left\|\log(g_{\mu,\Sigma}(\mathbf{X}))\right\|_1 = |\log(g_{\mu,\Sigma}(\mathbf{X}))| = d(\mathbf{X}),$$

for all $(\boldsymbol{\mu}, \Sigma) \in \Theta$. Then it follows, assuming g is normally distributed, that:

$$\begin{aligned}
\mathbb{E}_{p_{\theta_0}}\left[\left|\log(g_{\mu,\Sigma}(\mathbf{X}))\right|\right] &= \int_{\mathbb{R}^m} |\log(g_{\mu,\Sigma}(\mathbf{x}))| \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&= \int_{\mathbb{R}^m} \left|-\frac{k}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right| \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&\leq \int_{\mathbb{R}^m} \left|-\frac{k}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|)\right| + \left|\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right| \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&= \int_{\mathbb{R}^m} |C| + \left|\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right| \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Let $K \geq |C|$. Then it is found that

$$\begin{aligned}
\mathbb{E}_{p_{\theta_0}}\left[\left|\log(g_{\mu,\Sigma}(\mathbf{x}))\right|\right] &= \int_{\mathbb{R}^m} |C| + \left|\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right| \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&\leq \int_{\mathbb{R}^m} \left(K + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&= K\int_{\mathbb{R}^m} p_{\theta_0}(\mathbf{x})d\mathbf{x} + \frac{1}{2}\int_{\mathbb{R}^m} (\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

Whenever g is normally distributed, as is the case in our application, the squared Mahalabonis distance defined by $(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})$ is non-negative and has a chi-squared distribution with m degrees of freedom. Thus as $(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2_m$, then

$$\int_{\mathbb{R}^m} (\mathbf{x}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x} = \mathbb{E}_{p_{\theta_0}}[(\mathbf{X}-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})] = m.$$

Secondly, as $\int_{\mathbb{R}^m} p_{\theta_0}(\mathbf{x})d\mathbf{x} = 1$ it is given that

$$\mathbb{E}_{p_{\theta_0}}\left[\left|\log(g_{\mu,\Sigma}(\mathbf{X}))\right|\right] = K + \frac{m}{2} < \infty.$$

□

Using Theorem 6.2.1 and Lemma 6.2.1 the following Lemma 6.2.2 can be formulated and proven.

Lemma 6.2.2. *The estimators obtained by the CE-AIS-COS algorithm are asymptotically consistent.*

Proof. Define

$$
\begin{aligned}
M_n(\boldsymbol{\mu}, \Sigma) &= \frac{1}{n} \sum_{i=1}^{n} \log(g_{\boldsymbol{\mu},\Sigma}(\hat{\mathbf{X}}_{(i)})) \cdot \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}), 0) > q_\alpha\}, \\
M(\boldsymbol{\mu}, \Sigma) &= \mathbb{E}_{p_{\theta_0}} \left[\log(g_{\boldsymbol{\mu},\Sigma}(\mathbf{X})) \cdot \mathbf{1}\{\max(V(\mathbf{X}), 0) > q_\alpha\} \right], \\
(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) &= \underset{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}}{\arg\max} \; M_n(\boldsymbol{\mu}, \Sigma), \\
(\boldsymbol{\mu}^\star, \Sigma^\star) &= \underset{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}}{\arg\max} \; M(\boldsymbol{\mu}, \Sigma), \\
\Theta &= \{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}\}
\end{aligned}
$$

Now, it is proven that the requirements of Theorem 6.2.1 are met.

1. First it needs to be shown that

$$\sup_{(\boldsymbol{\mu},\Sigma)\in\Theta} |M_n(\boldsymbol{\mu}, \Sigma) - M(\boldsymbol{\mu}, \Sigma)| \to^{\mathbb{P}} 0.$$

   For this, the Uniform Law of Large Numbers of Lemma 6.2.1 will be used, which we copy below:

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} f(\hat{\mathbf{X}}_{(i)}, \theta) - \mathbf{E}_{p_{\theta_0}}[f(\mathbf{X}, \theta)] \right\|_2 \to^{\mathbb{P}} 0.$$

   where

$$f(\hat{\mathbf{X}}_{(i)}, \boldsymbol{\mu}, \Sigma) = \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}), 0) > q_\alpha\} \cdot \log(g_{\boldsymbol{\mu},\Sigma}(\hat{\mathbf{X}}_{(i)})).$$

   Then, from the proof of the lemma, the statement follows.

2. Next, it needs to be shown that

$$\sup_{(\boldsymbol{\mu},\Sigma): d((\boldsymbol{\mu},\Sigma),(\boldsymbol{\mu}^\star,\Sigma^\star))\geq\epsilon} M(\boldsymbol{\mu}, \Sigma) < M(\boldsymbol{\mu}^\star, \Sigma^\star).$$

   As shown, $M(\boldsymbol{\mu}, \Sigma)$ is maximized by 6.45 and 6.46. Then for all $(\boldsymbol{\mu}, \Sigma) \neq (\boldsymbol{\mu}^\star, \Sigma^\star)$, $M(\boldsymbol{\mu}^\star, \Sigma^\star) - M(\boldsymbol{\mu}, \Sigma) > 0$. Thus, if $\epsilon > 0$, then for all $(\boldsymbol{\mu}, \Sigma)$ such that $d((\boldsymbol{\mu}, \Sigma), (\boldsymbol{\mu}^\star, \Sigma^\star)) \geq \epsilon$ it is true that $(\boldsymbol{\mu}, \Sigma) \neq (\boldsymbol{\mu}^\star, \Sigma^\star)$, and thus

$$\sup_{(\boldsymbol{\mu},\Sigma): d((\boldsymbol{\mu},\Sigma),(\boldsymbol{\mu}^\star,\Sigma^\star))\geq\epsilon} M(\boldsymbol{\mu}^\star, \Sigma^\star) - M(\boldsymbol{\mu}, \Sigma) > 0. \tag{6.47}$$

3. Lastly, we need to show that

$$M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) \geq M_n(\boldsymbol{\mu}^\star, \Sigma^\star) - o_{\mathbb{P}}(1).$$

   If $(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n)$ maximizes $M_n(\boldsymbol{\mu}, \Sigma)$ then it can be seen that as

$$M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) \geq \sup_{(\boldsymbol{\mu},\Sigma)\in\Theta} M_n(\boldsymbol{\mu}, \Sigma) - o_{\mathbb{P}}(1),$$

   it can be concluded that

$$M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) \geq M_n(\boldsymbol{\mu}^\star, \Sigma^\star) - o_{\mathbb{P}}(1).$$

Then by combining $M_n(\boldsymbol{\mu}^\star, \Sigma^\star) \to^{\mathbb{P}} M(\boldsymbol{\mu}^\star, \Sigma^\star)$ with the third condition, it can be seen that

$$M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) \geq M(\boldsymbol{\mu}^\star, \Sigma^\star) - o_{\mathbb{P}}(1).$$

It then follows that

$$
\begin{aligned}
M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) &\geq M(\boldsymbol{\mu}^\star, \Sigma^\star) - o_{\mathbb{P}}(1), \\
M_n(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) - M(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n) + o_{\mathbb{P}}(1) &\geq M(\boldsymbol{\mu}^\star, \Sigma^\star) - M(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n), \\
\sup_{\boldsymbol{\mu},\Sigma} \left| M_n(\boldsymbol{\mu}, \Sigma) - M(\boldsymbol{\mu}, \Sigma) \right| + o_{\mathbb{P}}(1) &\geq M(\boldsymbol{\mu}^\star, \Sigma^\star) - M(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n).
\end{aligned}
$$

As the first condition shows that $\sup_{\mu,\Sigma} \left|M_n(\mu,\Sigma) - M(\mu,\Sigma)\right| \to 0$, and as $o_{\mathbb{P}} \to 0$ it follows that $\sup_{\mu,\Sigma} \left|M_n(\mu,\Sigma) - M(\mu,\Sigma)\right| + o_{\mathbb{P}} \to 0$.

From the second condition of the proof it follows that that for every $\epsilon > 0$ there exists a number $\zeta > 0$ such that for every (μ,Σ) where $d((\mu,\Sigma),(\mu^\star,\Sigma^\star)) \geq \epsilon$ it can be seen that $M(\mu,\Sigma) < M(\mu^\star,\Sigma^\star) - \zeta$. Thus, the event $\{M(\hat{\mu},\hat{\Sigma}) < M(\mu^\star,\Sigma^\star) - \zeta\}$ contains $\{d((\mu,\Sigma),(\mu^\star,\Sigma^\star)) \geq \epsilon\}$. Therefore, by having shown that

$$M(\mu^\star,\Sigma^\star) - M(\hat{\mu},\hat{\Sigma}) \to 0,$$

it follows that the probability of $\{M(\hat{\mu},\hat{\Sigma}) < M(\mu^\star,\Sigma^\star) - \zeta\}$ goes to zero which implies that the probability of $\{d((\mu,\Sigma),(\mu^\star,\Sigma^\star)) \geq \epsilon\}$ goes to zero. Thus as $\mathbb{P}\left(\{d((\mu,\Sigma),(\mu^\star,\Sigma^\star)) \geq \epsilon\}\right) \to 0$ it can be concluded that $(\mu,\Sigma) \to (\mu^\star,\Sigma^\star)$ in probability. □

From Lemma 6.2.2 it can be seen that the algorithm produces consistent estimators $\hat{\mu}_n$ and $\hat{\Sigma}_n$. The final estimators $\mu^\star$ and $\Sigma^\star$ are then computed by taking the average of the estimators found at each iteration of the algorithm. This step is justified by the Law of Large Numbers which states that the sample mean of independent identically distributed values, which in this case are the estimators $\hat{\mu}_1,\ldots,\hat{\mu}_{k_{max}}$ and $\hat{\Sigma}_1,\ldots,\hat{\Sigma}_{k_{max}}$, converges to the true mean which is their analytical solution.

In [53] and [54] the convergence of the CE algorithm is proven. In the papers Proposition 3 is formulated which provides the condition needed for the CE algorithm to converge in a finite number of iterations. For the proof of Proposition 3 the authors in [54] introduce a modified multi-layer CE algorithm where the low quantile ρ_t, where t is the iteration, is adaptive, as choosing an acceptable ρ a prior can be difficult. Notice, that by choosing $\rho_t = \rho$ the algorithm is equivalent to the original multi-layer CE algorithm. This multi-layer CE algorithm is summarized below:

Algorithm 4 The multi-layer CE algorithm with adaptive low quantile ρ_t used to prove convergence, provided in [54].

Initialize:

Set $t = 1$, $\rho_0 = \rho$, $\mathbf{v}_0 = \mathbf{u}$.

1. Compute

$$Q(\mathbf{v},\mathbf{v}_{t-1},\rho_{t-1}) = \mathbb{E}_{\mathbf{v}_{t-1}}\left[\mathbf{1}\left\{\mathbb{M}(\mathbf{Z}) \geq \min\left(x,\gamma(\mathbf{v}_{t-1},\rho_{t-1})\right)\right\} \cdot \mathbf{W} \cdot \log(f(\mathbf{Z},\mathbf{v}))\right].$$

Here, $\mathbf{W}$ is the likelihood ratio of $\mathbf{u}$ and $\mathbf{v}_{t-1}$.

2. Compute

$$\mathbf{v}_t \in \arg\max_{\mathbf{v}\in V} Q(\mathbf{v},\mathbf{v}_{t-1},\rho_{t-1}).$$

if $\gamma(\mathbf{v}_{t-1},\rho_{t-1}) \geq x$ **then** stop.

else

Move to the next step.

3. Let ρ_t be such that $\gamma(\mathbf{v}_t,\rho_t) \geq \min\left(x,\gamma(\mathbf{v}_{t-1},\rho_{t-1}) + \delta\right)$, where δ is a positive constant.

Let $t = t + 1$ and go to step 1.

In this algorithm $\mathbf{v}_t$ is the auxiliary sampling density found in iteration t, u is the original sampling density, $\mathbb{M}$ is the sample performance under the random vector $\mathbf{Z}$, and $\gamma(\mathbf{v}_t,\rho_t)$ is the $(1-\rho_t)$-quantile of $\mathbb{M}(\mathbf{Z})$ under the probability density $\mathbf{v}_t$.

Proposition 3. *Let $S^\star$ be the set containing the $\mathbf{v}$'s that maximize Equation 6.40, then if there exists a set V such that $V \cap S^\star \neq \emptyset$ and $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) \geq x) > 0$ for all $\mathbf{v} \in V$, then the multi-level CE algorithm converges with probability 1 to a solution of 6.40 after a finite number of iterations.*

Proof. Let t be an iteration of the algorithm. Define $\gamma(\mathbf{v}_t,\rho)$ as an arbitrary $(1-\rho)$-quantile of $\mathbb{M}(\mathbf{Z})$ under the probability density $\mathbf{v}_t$.

Define $\rho_x^\star = \mathbb{P}_{\mathbf{v}_t}(\mathbb{M}(\mathbf{Z}) \geq x)$. Then, from the condition of the proposition it holds that, as $\mathbb{P}_{\mathbf{v}_t}(\mathbb{M}(\mathbf{Z}) \geq x) > 0$ since $\mathbf{v}_t \in V$ it must be true that $\rho_x^\star > 0$. Then, define an arbitrary $\rho^\star \in (0, \rho_x^\star)$. It follows that,

$$\begin{aligned} \mathbb{P}_{\mathbf{v}_t}\left(\mathbb{M}(\mathbf{Z}) \geq \gamma(\mathbf{v}_t, \rho^\star)\right) &= \rho^\star, \\ \mathbb{P}_{\mathbf{v}_t}\left(\mathbb{M}(\mathbf{Z}) \leq \gamma(\mathbf{v}_t, \rho^\star)\right) &= 1 - \rho^\star > 1 - \rho_x^\star. \end{aligned} \tag{6.48}$$

In the last equation we used the fact that $\rho^\star \in (0, \rho_x^\star)$ implies that $\rho^\star < \rho_x^\star$.

Now, suppose that $\gamma(\mathbf{v}_t, \rho^\star) < x$, then

$$\mathbb{P}_{\mathbf{v}_t}\left(\mathbb{M}(\mathbf{Z}) \leq \gamma(\mathbf{v}_t, \rho^\star)\right) \leq \mathbb{P}_{\mathbf{v}_t}(\mathbb{M}(\mathbf{Z}) < x) = 1 - \rho_x^\star.$$

This last statement contradicts Equation 6.48. From this it follows that $\gamma(\mathbf{v}_t, \rho^\star) \geq x$. This causes the algorithm to reach step 4 and stop. If this happens at $t = T$ it follows that $\gamma\left(\mathbf{v}_{T-1}, \rho_{T-1}\right) \geq x$. Consequently, in step 2 of the algorithm we find,

$$\mathbf{v}_T \in \arg\max_{\mathbf{v} \in V} \mathbb{E}_{\mathbf{v}}\left[\mathbf{1}\{\mathbb{M}(\mathbf{Z}) \geq x\} \cdot \mathbf{W} \cdot \log(f(\mathbf{Z}, \mathbf{v}))\right].$$

Here, $\mathbf{v}_T$ is the CE-optimal solution which proves Proposition 3. □

In practice the expectation and quantile of the last algorithm are approximated using the sample mean and the sample quantile. This transforms the calculation in step 3 of the algorithm into

$$\mathbf{v}_t \in \arg\max_{\mathbf{v} \in V} \frac{1}{N} \sum_{i=1}^{N} \left[\mathbf{1}\left\{\mathbb{M}(\hat{\mathbf{Z}}_{(i)}) \geq \min\left(x, \hat{\gamma}(\mathbf{v}_t, \rho_{t-1})\right)\right\} \cdot \mathbf{W} \cdot \log\left(f(\hat{\mathbf{Z}}_{(i)}, \mathbf{v})\right)\right], \tag{6.49}$$

where $\hat{\gamma}_n(\mathbf{v}_t, \rho_t)$ is the sample quantile of $\mathbb{M}(\mathbf{Z})$. In Proposition 4, found in [54], it is shown that the conditions needed for convergence, as shown in Proposition 3, are equal to those using the sample mean and sample quantile.

Proposition 4. *Suppose that $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) \geq x) > 0$ for all $\mathbf{v} \in V$. Let $\hat{\mathbf{Z}}_{(1)}, \hat{\mathbf{Z}}_{(2)}, \ldots$ be i.i.d. with common density $f(\mathbf{Z}, \mathbf{v})$. Then, there exists $\rho_x > 0$ and a random $N_x > 0$ such that, with probability one, $\hat{\gamma}_N(\mathbf{Z}, \rho) \geq x$ for all $\rho \in (0, \rho_x)$ and all $N \geq N_x$. Moreover, the probability that $\hat{\gamma}_N(\mathbf{Z}, \rho) \geq x$ for a given N goes to one exponentially fast with N.*

Proof. In the proof found in [54] the proof is split up for distributions where $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) > x) > 0$ and $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) = x) > 0$. As the auxiliary density used in the CE-AIS-COS algorithm for PFE calculations is assumed to be normally distributed, this proof is restricted to the case where $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) > x) > 0$. For auxiliary densities following finitely supported probability distributions the full proof can be found in [54].

First it will be proven that the sample quantile $\hat{\gamma}$ goes to the theoretical quantile γ as n goes to ∞. After this it is proven that $\hat{\gamma}_N(\mathbf{Z}, \rho) \geq x$ for all $\rho \in (0, \rho_x)$ and $N \geq N_x$.

A $(1-\rho)$-quantile of a random variable Y can be expressed as the optimal solution of the minimization problem

$$\min_{\theta} \mathbb{E}_h(Y, \theta),$$

where

$$h(Y, \theta) = (1-\rho)(Y - \theta) \cdot \mathbf{1}\{\theta \leq Y\} + \rho(Y - \theta) \cdot \mathbf{1}\{\theta \geq Y\}.$$

Using the subdifferential of $\mathbb{E}_h(Y, \theta)$, which is defined as the set of subderivatives, with respect to θ it can be seen that

$$\partial_\theta \mathbb{E}_h(Y, \theta) = \left[\rho - \mathbb{P}(Y \geq \theta), -(1-\rho) + \mathbb{P}(Y \leq \theta)\right].$$

To minimize $\mathbb{E}_h(Y, \theta)$ a θ must be found such that

$$\begin{aligned} \rho - \mathbb{P}(Y \geq \theta) &= 0, \\ -(1-\rho) + \mathbb{P}(Y \leq \theta) &= 0. \end{aligned}$$

From this it follows that $\mathbb{P}(Y \geq \theta) = \rho$ and $\mathbb{P}(Y \leq \theta) = 1 - \rho$.

Similarly, it can be shown that $\hat{\theta}$ is the sample quantile of Y as it can be shown to be the solution to,

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} h(Y_i, \theta).$$

Then, the second order subdifferential yields

$$\partial_\theta^2 \mathbb{E}_h(Y, \theta) = \left[\partial_\theta \left(\rho - \mathbb{P}(Y \geq \theta)\right), \partial_\theta \left(-(1-\rho) + \mathbb{P}(Y \leq \theta)\right)\right].$$

where

$$\begin{aligned} \partial_\theta \left(\rho - \mathbb{P}(Y \geq \theta)\right) &= \partial_\theta \left(\rho - (1 - \mathbb{P}(Y < \theta))\right), \\ &= \partial_\theta \rho - \partial_\theta \left(1 - \mathbb{P}(Y < \theta)\right), \\ &= f_Y(\theta) > 0, \end{aligned}$$

and

$$\begin{aligned} \partial_\theta \left(-(1-\rho) + \mathbb{P}(Y \leq \theta)\right) &= \partial_\theta \left(-(1-\rho) + \mathbb{P}(Y \leq \theta)\right), \\ &= \partial_\theta(-(1-\rho)) + \partial_\theta \mathbb{P}(Y \leq \theta), \\ &= f_Y(\theta) > 0. \end{aligned}$$

Due to this, it can be concluded that $\mathbb{E}_h(Y, \theta)$ is convex in θ. From this it follows that the sample quantile $\hat{\theta}$ goes to θ with probability 1 as n goes to ∞.

Next, it will be proven that $\hat{\gamma}_N(\mathbf{Z}, \rho) > x$ for all $\rho \in (0, \rho_x)$ and $N \geq N_x$.

Let $\hat{\mathbf{Z}}_{(1)}, \ldots, \hat{\mathbf{Z}}_{(N)}$ be i.i.d. samples, and $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) > x) > 0$ for all $\mathbf{v} \in V$. Then using the same arguments as in the proof of Proposition 3, it can be argued that for any $\rho \in (0, \rho_x)$, where $\rho_x = \mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) > x)$, we have that $\gamma(\mathbf{v}, \rho) \geq x$. Then by using the argument proven earlier it can be stated that as $\hat{\gamma}$ goes to γ with probability 1, for N large enough, that $\hat{\gamma}(\mathbf{v}, \rho^\star) > x$. Besides this the probability of $\hat{\gamma}(\mathbf{v}, \rho^\star) > x$ for a given N goes to one exponentially fast [55]. □

The CE-AIS-COS algorithm used in this thesis is not multi-layered. However, the prior proposition can still be used to prove the convergence of our algorithm. The reason for this is that our algorithm is equivalent to a multi-layer CE algorithm in which $\gamma = \min\left(q_{0.975}, \hat{\gamma}\right)$ where $\hat{\gamma}$ is the sample 0.975-quantile. Since the original probability density is used in the first iteration, provided that our PFE approximation $q_{0.975}$ calculated using the COS method is accurate enough, we have that $q_{0.975} \approx \hat{\gamma}$. In this way, in the first iteration, $\gamma = q_{0.975} = \hat{\gamma}$. From the second iteration onward $\gamma = q_{0.975} \leq \hat{\gamma}$. Then, as the CE-AIS-COS method developed in this section can be seen as a multi-layer CE algorithm Lemma 5 can be proven.

Proposition 5. *The CE-AIS-COS algorithm converges with probability 1 to the solution of 6.40 after a finite number of iterations.*

Proof. To prove this lemma it must be shown that $\mathbb{P}_{\mathbf{v}}(\mathbb{M}(\mathbf{Z}) \geq x) > 0$ for all $\mathbf{v} \in V$, where in our case $\mathbf{v}$ is $g_{\mu,\Sigma}$, V is $\{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}\}$, $\mathbb{M}(\mathbf{Z})$ is $\max(V(\mathbf{X}), 0)$, and x is q_α. Thus, to prove is that for all $(\boldsymbol{\mu}, \Sigma) \in \{(\boldsymbol{\mu}, \Sigma) : \boldsymbol{\mu} \in \mathbb{R}^m, \Sigma \in \mathcal{M}^+_{m\times m}\}$ it must be true that $\mathbb{P}_{g_{\mu,\Sigma}}(\max(V(\mathbf{X}), 0) \geq q_\alpha) > 0$. As $g_{\mu,\Sigma}$ follows the normal distribution, which has infinite tails for all values of $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\Sigma \in \mathcal{M}^+_{m\times m}$, is can be concluded that for all values of q_α it is true that $\mathbb{P}_{g_{\mu,\Sigma}}(\max(V(\mathbf{X}), 0) \geq q_\alpha) > 0$. Thus, using proposition 3 the lemma has been proven. □

6.3. Theoretical variance reduction

In this section the theoretical variance reduction will be discussed. The CDF estimator using the Monte Carlo method and the importance sampling are

$$\hat{F}_n(y) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}), 0) > y\},$$
$$\hat{F}_{n,IS}(y) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\max(V(\hat{\mathbf{X}}_{(i)}), 0) > y\} \cdot \frac{p_{\theta_0}(\hat{\mathbf{X}}_{(i)})}{p_{\theta}(\hat{\mathbf{X}}_{(i)})}.$$

As shown, the variances of these estimators are

$$\begin{aligned}
\text{var}_{p_{\theta_0}}\left[\hat{F}_n(y)\right] =& \frac{1}{n}\left(\mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\right] - \mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\right]^2\right), \\
\text{var}_{p_{\theta}}\left[\hat{F}_{n,IS}(y)\right] =& \frac{1}{n}\left(\mathbb{E}_{p_{\theta}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\left(\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}\right)^2\right]\right. \\
& \left. - \mathbb{E}_{p_{\theta}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}\right]^2\right), \\
=& \frac{1}{n}\left(\mathbb{E}_{p_{\theta}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\left(\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}\right)^2\right]\right. \\
& \left. - \mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\right]^2\right).
\end{aligned}$$

From this it can be seen that the variance reduction is dependent on the second moments. To be more precise, it can be shown that the variance reduction is dependent on the ratio $\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}$ since,

$$\begin{aligned}
\mathbb{E}_{p_{\theta}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\left(\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}\right)^2\right] &= \int_{\mathbb{R}^m} \mathbf{1}\{\max(V(\mathbf{x}, 0)) > y\} \cdot \left(\frac{p_{\theta_0}(\mathbf{x})}{p_{\theta}(\mathbf{x})}\right)^2 \cdot p_{\theta}(\mathbf{x})d\mathbf{x}, \\
&= \int_{\mathbb{R}^m} \mathbf{1}\{\max(V(\mathbf{x}, 0)) > y\} \cdot \frac{p_{\theta_0}(\mathbf{x})}{p_{\theta}(\mathbf{x})} \cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}, \\
&= \mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{\max(V(\mathbf{X}), 0) > y\}\frac{p_{\theta_0}(\mathbf{X})}{p_{\theta}(\mathbf{X})}\right].
\end{aligned}$$

As a next step, we use the central limit theorem to derive the influence of the likelihood ratio on the theoretical variance reduction.

The quantile estimate $\hat{q}_\alpha$ obeys the central limit theorem [9]. Suppose y is the quantile that solves $F(y) = p$. Define the indicator functions $\mathbf{1}\{Y(\mathbf{X}) \leq y\}$ as a new random variable with $\mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}\right] = p$, where $p = \mathbb{P}\left[Y(\mathbf{X}) \leq y\right]$. From this construction, we have that $\text{var}_{p_{\theta 0}}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}\right] = p(1-p)$. Based on the central limit theorem it holds that

$$\sqrt{n}\left(\hat{F}_n(y) - p\right) \rightarrow^d \mathcal{N}\left(0, p(1-p)\right) \tag{6.50}$$

as n goes to ∞.

Similarly as done for $\hat{F}_n$ above, it follows from the central limit theorem that the estimator $\hat{F}_{n,IS}$ also converges in distribution to a normal distribution. By seeing that $\mathbb{E}_{p_\theta}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}w_{\mathbf{X}}(\theta)\right] = \mathbb{E}_{p_{\theta_0}}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}\right] = p$, it follows that

$$\sqrt{n}\left(\hat{F}_{n,IS}(y) - p\right) \rightarrow \mathcal{N}\left(0, \text{var}_{p_\theta}\left[\mathbf{1}\{Y(\mathbf{X}) \leq y\}w_{\mathbf{X}}(\theta)\right]\right).$$

In [56], it is found that the estimator of the sample quantile $\hat{q}_\alpha = \hat{F}_{\text{IS}}^{-1}(p)$ also converges in distribution following a central limit theorem

$$\sqrt{n}\left(\hat{q}_\alpha - q_\alpha\right) \to \mathcal{N}\left(0, \tau^2\right) \tag{6.51}$$

as n goes to ∞. Here, q_α is the true $\alpha-$quantile, and τ^2 is the asymptotic variances of the estimator, defined by

$$\tau^2 = \frac{p(1-p)}{f^2(q_\alpha)}, \tag{6.52}$$

where $f(q_\alpha)$ is the derivative of the CDF F at the α-quantile q_α, and is assumed to satisfy $f(q_\alpha) > 0$.

The proof of the central limit theorem for the quantile estimator can be found in [56] and starts by defining the Berry-Esséen theorem, which we reformulate below as Theorem 6.3.1 retrieved from [57]. In this theorem F denotes the CDF which possesses left- or right-hand derivatives at q_α, denoted by $F'(q_\alpha-)$ or $F'(q_\alpha+)$, and Φ is the CDF of the standard normal distribution function.

Theorem 6.3.1 (Berry-Esséen). *Let $(X_j)_{j\geq 1}$ be independent identically distributed. Suppose that $\mathbb{E}\left[|X_j|^3\right] < \infty$. Let $G_n(x) = \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right)$ where $\mu = \mathbb{E}[X_j]$ and $\sigma^2 = \sigma^2_{X_j} < \infty$. Let $\Phi(x) = \mathbb{P}(Y \leq x)$, then*

$$\sup_x |G_n(x) - \Phi(x)| \leq C \cdot \frac{\mathbb{E}\left[|X_1|^3\right]}{\sigma^3\sqrt{n}}, \tag{6.53}$$

where $S_n = \sum_{j=1}^n X_j$ and C is a constant.

The central limit theorem for the quantile estimate $\hat{q}_\alpha$ is stated in Theorem 6.3.2.

Theorem 6.3.2. *Let $0 < p < 1$. Suppose that F is continuous at q_p.*

1. *If there exists $F'(q_p-) > 0$, then for $t < 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\sqrt{n}(\hat{q}_p - q_p)}{\sqrt{p(1-p)}/F'(q_p-)} \leq t\right] = \Phi(t).$$

2. *If there exists $F'(q_p+) > 0$, then for $t > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{\sqrt{n}(\hat{q}_p - q_p)}{\sqrt{p(1-p)}/F'(q_p+)} \leq t\right] = \Phi(t).$$

3. *In any case,*

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}(\hat{q}_p - q_p \leq 0)\right] = \Phi(0) = \frac{1}{2}.$$

Proof. Fix t, and let $A > 0$ be a normalizing constant. Then define

$$G_n(t) = \mathbb{P}\left[\frac{\sqrt{n}(\hat{q}_p - q_p)}{A} \leq t\right].$$

Using the fact that $F(x) \geq t$ if and only if $x \geq F^{-1}(t)$, it can be seen that

$$\begin{aligned}
G_n(t) &= \mathbb{P}\left[\hat{q}_p \leq q_p + \frac{tA}{\sqrt{n}}\right], \\
&= \mathbb{P}\left[F_n(\hat{q}_p) \leq F_n\left(q_p + \frac{tA}{\sqrt{n}}\right)\right], \\
&= \mathbb{P}\left[p \leq F_n\left(q_p + \frac{tA}{\sqrt{n}}\right)\right], \\
&= \mathbb{P}\left[np \leq Z(n, F\left(q_p + \frac{tA}{\sqrt{n}}\right))\right].
\end{aligned}$$

Here, $Z(n,\beta) \sim \text{Binom}(n,\beta)$ with $\beta = F\left(q_p + \frac{tA}{\sqrt{n}}\right)$. Then we define the standardized binomial distribution as:

$$Z^\star(n,\beta) = \frac{Z(n,\beta) - n\beta}{\sqrt{n\beta(1-\beta)}}. \tag{6.54}$$

Using the previous equations we yield that

$$\begin{aligned} G_n(t) &= \mathbb{P}\left[np \le Z(n,\beta)\right], \\ &= \mathbb{P}\left[\frac{np - n\beta}{\sqrt{n\beta(1-\beta)}} \le \frac{Z(n,\beta) - n\beta}{\sqrt{n\beta(1-\beta)}}\right], \\ &= \mathbb{P}\left[Z^\star(n,\beta) \ge -a(n,t)\right]. \end{aligned}$$

The Berry-Esséen Theorem is applied to yield that

$$\sup_{-\infty \le x \le \infty} |\mathbb{P}\left[Z^\star(n,\beta) < x\right] - \Phi(x)| \le C \cdot \frac{\mathbb{E}|Z(1,\beta) - \beta|^3}{\text{var}[Z(1,\beta)]^{\frac{3}{2}}\sqrt{n}}.$$

Then we fill in $\text{var}[Z(1,\beta)] = \beta(1-\beta)$, and $\mathbb{E}|Z(1,\beta) - \beta|^3 = \beta(1-\beta)((1-\beta)^2 - \beta^2)$, to simplify the expression to:

$$\sup_{-\infty \le x \le \infty} |\mathbb{P}\left[Z^\star(n,\beta) < x\right] - \Phi(x)| \le C \cdot \frac{(1-\beta)^2 - \beta^2}{\sqrt{n\beta(1-\beta)}}.$$

Next, it can be written that

$$\begin{aligned} \Phi(t) - G_n(t) &= \Phi(t) - \mathbb{P}\left[Z^\star(n,\beta) \ge -a(n,t)\right], \\ &= \Phi(t) - (1 - \mathbb{P}\left[Z^\star(n,\beta) < -a(n,t)\right]), \\ &= \mathbb{P}\left[Z^\star(n,\beta) < -a(n,t)\right] - (1 - \Phi(t)), \\ &= \mathbb{P}\left[Z^\star(n,\beta) < -a(n,t)\right] - (\Phi(-a(n,t)) + \Phi(a(n,t)) - \Phi(t)), \end{aligned}$$

where the last equation follows from:

$$\begin{aligned} \Phi(a(n,t)) &= \int_{-\infty}^{a(n,t)} \phi(x)dx = \int_{-\infty}^{\infty} \phi(x)dx - \int_{a(n,t)}^{\infty} \phi(x)dx \\ &= 1 - \int_{a(n,t)}^{\infty} \phi(x)dx = 1 - \int_{-\infty}^{-a(n,t)} \phi(x)dx = 1 - \Phi(-a(n,t)). \end{aligned}$$

We then again apply the Berry-Esséen Theorem to have

$$\begin{aligned} |G_n(t) - \Phi(t)| &= |\mathbb{P}\left[Z^\star(n,\beta) < -a(n,t)\right] - (\Phi(-a(n,t)) + \Phi(a(n,t)) - \Phi(t))|, \\ &\le |\mathbb{P}\left[Z^\star(n,\beta) < -a(n,t)\right] - \Phi(-a(n,t))| + |\Phi(t) - \Phi(a(n,t))|, \\ &\le C \cdot \frac{(1-\beta)^2 - \beta^2}{\sqrt{n\beta(1-\beta)}} + |\Phi(t) - \Phi(a(n,t))|. \end{aligned} \tag{6.55}$$

As $n \to \infty$ it is found that $q_p + \frac{tA}{\sqrt{n}}$ goes to q_p, from which it follows that $\beta = F\left(q_p + \frac{tA}{\sqrt{n}}\right)$ goes to $F(q_p) = p$. Consequently, it can be concluded that $\frac{(1-\beta)^2-\beta^2}{\sqrt{n\beta(1-\beta)}} \to 0$, as $n \to \infty$.

Then to show that $G_n(t) \to \Phi(t)$, it must be shown that $a(n,t) \to t$ as seen from Equation 6.55. As $a(n,t)$ is defined as,

$$a(n,t) = \frac{\sqrt{n}(\beta - p)}{\sqrt{\beta(1-\beta)}},$$

it follows that,

$$\frac{\sqrt{n}(\beta - p)}{\sqrt{\beta(1-\beta)}} = \frac{\sqrt{n}\left(F\left(q_p + \frac{tA}{\sqrt{n}}\right) - p\right)}{\sqrt{\beta(1-\beta)}},$$
$$= \frac{tA}{\sqrt{\beta(1-\beta)}} \cdot \frac{F\left(q_p + \frac{tA}{\sqrt{n}}\right) - F(q_p)}{\frac{tA}{\sqrt{n}}}.$$

Here, in the last term of the above equation, the expression for the derivative can be recognized. That is, as $n \to \infty$, if $t > 0$

$$a(n,t) \to \frac{tA}{\sqrt{p(1-p)}} \cdot F'(q_p+).$$

And if $t < 0$

$$a(n,t) \to \frac{tA}{\sqrt{p(1-p)}} \cdot F'(q_p-).$$

Thus to get $a(n,t) \to t$ it must hold that, if $t > 0$

$$A = \frac{\sqrt{p(1-p)}}{F'(q_p+)}.$$

Similarly, if $t < 0$

$$A = \frac{\sqrt{p(1-p)}}{F'(q_p-)}.$$

This proves the theorem. □

If F is differentiable at q_p, then $F'(q_p+) = F'(q_p-) = F'(q_p)$. In this case, as demonstrated in [56], Theorem 6.3.2 leads to Corrolary 6.3.2.1.

Corollary 6.3.2.1. *Let $0 < p < 1$. If F is differentiable at q_p and $F'(q_p) > 0$, then*

$$\sqrt{n}(\hat{q}_p - q_p) \sim \mathcal{N}\left(0, \frac{p(1-p)}{F'(q_p)^2}\right) \tag{6.56}$$

From Corollary 6.3.2.1 it can be seen that theoretically the quantile estimate found using $\hat{F}_n$ is normally distributed, i.e.

$$\hat{q}_p \sim \mathcal{N}\left(q_p, \frac{p(1-p)}{n \cdot F'(q_p)^2}\right). \tag{6.57}$$

However, for the quantile estimate produced using the importance sampling estimator $\hat{F}_{n,IS}$, this variance changes. Using the same steps as in the proof of Theorem 6.3.2 it can be seen that, by standardizing $Z(n,\beta)$, as done in Equation 6.54, the expression $Z^\star(n,\beta)$ becomes

$$Z^\star(n,\beta) = \frac{Z(n,\beta) - n\beta}{\sqrt{n \cdot \text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \le y\} w_{\mathbf{X}}(\theta)\right]}}.$$

Similar to Equation 6.55, it can be shown that $G_n(t) \to \Phi(t)$. Finally, by repeating the last step for $a(n,t)$ we have that for $t > 0$ and $t < 0$

$$A = \frac{\sqrt{n \cdot \text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \le y\} w_{\mathbf{X}(\theta)}\right]}}{F'(q_p+)}$$

or

$$A = \frac{\sqrt{n \cdot \text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \le y\} w_{\mathbf{X}}(\theta)\right]}}{F'(q_p-)}$$

respectively. As F is differentiable at q_p and $F'(q_p) > 0$, it is found that the importance sampling quantile estimate also follows a normal distribution,

$$\hat{q}_{p,IS} \sim \mathcal{N}\left(q_p, \frac{\text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \leq q_p\} w_{\mathbf{X}}(\theta)\right]}{n \cdot F'(q_p)^2}\right). \tag{6.58}$$

Now the asymptotical variance reduction can be found by calculating:

$$\begin{aligned}\frac{\text{var}[\hat{q}_{p,MC}]}{\text{var}[\hat{q}_{p,IS}]} &= \frac{\frac{p(1-p)}{n \cdot F'(q_p)^2}}{\frac{\text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \leq q_p\} w_{\mathbf{X}}(\theta)\right]}{n \cdot F'(q_p)^2}}, \\ &= \frac{p(1-p)}{\text{var}_\theta\left[\mathbf{1}\{Y(\mathbf{X}) \leq q_p\} w_{\mathbf{X}}(\theta)\right]}.\end{aligned}$$

To get an idea of how much variance reduction can be achieved according to the above theoretical analysis, we compute the theoretical variance reduction ratio for a hypothetical portfolio containing 100 derivatives using Monte Carlo simulation. First it was estimated for a portfolio without collateral and then with collateral. For both cases, the theoretical variance reduction was given for the time points at which the straight forward MC variance is non-zero. For these portfolios the PFE was estimated for 20 equidistant time points starting at the initialization of the portfolio and ending at the maturity of the portfolio. In Tables 6.1 and 6.2 the t refers to these time points.

For the uncollateralized portfolio, the theoretical variance reductions are

t	5	6	7	9	10	11	12
$\sigma^2_{PFE,MC}/\sigma^2_{PFE,CE-AIS}$	41.14	40.89	40.77	42.50	41.86	40.63	39.76
t	13	14	15	16	17	18	19
$\sigma^2_{PFE,MC}/\sigma^2_{PFE,CE-AIS}$	40.57	40.27	39.46	40.03	45.17	43.04	43.64

Table 6.1: The theoretical asymptotic variance reductions for a uncollateralized portfolio containing 100 derivatives.

This table indicates that the variance reduction of our CE-AIS method is significant. The same is observed for the collateralized portfolio:

t	5	6	7	9	10	11	12
$\sigma^2_{PFE,MC}/\sigma^2_{PFE,CE-AIS}$	41.66	41.52	41.99	41.31	41.42	41.88	42.03
t	13	14	15	16	17	18	19
$\sigma^2_{PFE,MC}/\sigma^2_{PFE,CE-AIS}$	42.24	41.59	42.10	41.14	42.21	42.25	41.71

Table 6.2: The theoretical asymptotic variance reductions for a collateralized portfolio containing 100 derivatives.

7 Results

In this section, the results of the control variate method, adaptive importance sampling using the optimal shift say shift-AIS-COS, and the CE-AIS-COS method will be discussed.

The sample values of the domestic and foreign short rate x_d, x_f, exchange rate X, and for the collateralized portfolio, Z-spread z were directly drawn from their dynamics,

$$x_d(t) = x_d(0)e^{-a_d t} + \sigma_d \int_0^t e^{-a_d(t-s)} dW_d(s,) \tag{7.1}$$

$$x_f(t) = x_f(0)e^{-a_f t} + \sigma_f \int_0^t e^{-a_f(t-s)} dW_f(s), \tag{7.2}$$

$$\log(X(t)) = \log(X(0)) + \left(\mu_X - \frac{\sigma_X^2}{2}\right) t + \sigma_X W_X(t), \tag{7.3}$$

$$\log(z(t)) = \log(z(0)) - \frac{\sigma_z^2}{2} t + \sigma_z W_z(t). \tag{7.4}$$

Here, $a_d = 0.01$, $\sigma_d = 0.007$, $a_f = 0.05$, $\sigma_f = 0.012$, $\mu_X = 0.008$, $\sigma_X = 0.02$, $X(0) = \frac{1}{105}$, $\sigma_z = 0.010359$ of which the latter was obtained from data found online. The value of $z(0)$ is obtained by equalizing the bond's model price to the MtM value of the portfolio at $t = 0$. The correlations between the risk factors were set as $\rho_{df} = 0.25$, $\rho_{dX} = \rho_{fX} = -0.15$, $\rho_{dz} = \rho_{fz} = \rho_{Xz} = 0$.

The number of paths tested were 100, 250, 500, 2500, 7500, and 25000. In the CE-AIS-COS algorithm 2500 paths were used with $k_{max} = 5$.

For the uncollateralized portfolios, the benchmark PFE was calculated using the COS method with 150 expansion terms and 130 quadrature points, similar as in the original paper [1]. For the collateralized portfolio, the benchmark was calculated using the Monte Carlo method with 5000000 paths.

In each of the methods the PFE of the test portfolios without collateral are calculated using 64 expansion terms for the COS method and 40 quadrature points for the Clenshaw-Curtis quadrature. The PFE of the test portfolios with collateral are calculated using 32 expansion terms for the COS method and 30 quadrature points for the Clenshaw-Curtis quadrature. Additionally, all CPU times measured are the total time of the PFE calculation.

7.1. Results of using COS as the control variate

The method of using COS as the control variate was tested using a hypothetical portfolio containing 100 derivatives without collateral. Table 7.1 shows the variance reductions averaged out over all non-zero variances at all time points per number of paths.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE,MC}}}{\sigma^2_{\text{PFE,CV}}} \approx$	0.91	0.95	0.97	1.2	1.2	1.2

Table 7.1: The average variance reduction using COS as the control variate.

On average, using the COS as control variate approximately reduced the PFE estimator's variance by 1.09 times, compared to straight forward Monte Carlo simulation.

For the expected exposure the variance reduction was much more effective for some time points as can be seen from Figure 7.1. Table 7.2 shows the variance reduction for each time point using 25000 paths.

Time points	1	2	3	4	5	6	7	8	9	10
$\frac{\sigma^2_{\text{MC}}}{\sigma^2_{\text{CV}}} \approx$	1.00	5.80	3.21	1.68	1.34	1.17	1.15	1.02	1.00	1.09
Time points	11	12	13	14	15	16	17	18	19	
$\frac{\sigma^2_{\text{MC}}}{\sigma^2_{\text{CV}}} \approx$	1.17	1.25	1.09	1.02	1.90	2.41	2.32	1.93	1.93	

Table 7.2: The average variance reduction using COS as the control variate on the calculation of the expected exposure for a portfolio with 100 derivatives without collateral.

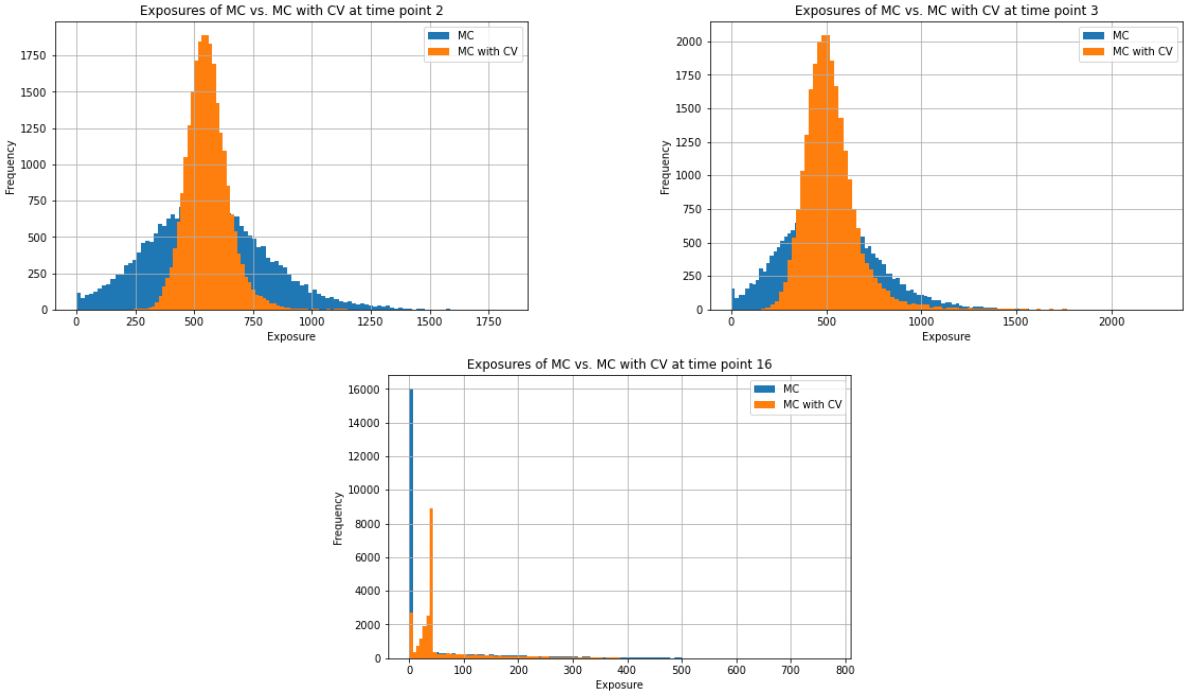

Figure 7.1: The comparison of the exposures for time points $t = 2, 3$ and 16 generated using COS as control variate and straight forward MC simulation.

The correlation values over time between the auxiliary variable and the MtM value of the portfolio are plotted in Figure 7.2, and those between the indicators $\mathbf{1}\{\max(E_t(\mathbf{X}), 0) > q_{0.975}\}$ and $\mathbf{1}\{Z > q'_{0.975}\}$ can be found in Figure 7.3. These correlation values show that, as expected, as the correlation values between the control variate and the MtM value of the portfolio are higher than those between the indicators, using COS can help to reduce the variance for EE estimations yields.

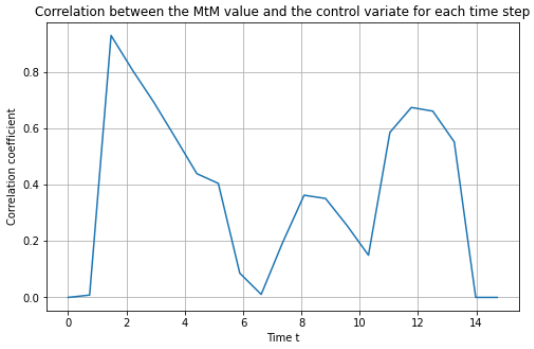

Figure 7.2: The correlation values between the auxiliary variable and the MtM value of a portfolio containing 100 derivatives without collateral.

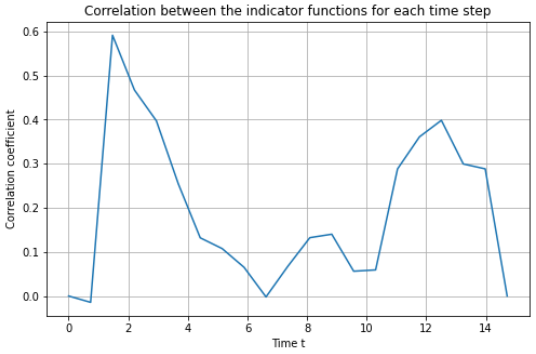

Figure 7.3: The correlation values between the indicators $\mathbf{1}\{\max(E_t(\mathbf{X}), 0) > q_{0.975}\}$ and $\mathbf{1}\{Z > q'_{0.975}\}$.

7.2. Results of Adaptive Importance Sampling using the optimal shift

In this section the variance reduction is tested for the collateralized portfolio containing 100 derivatives using the adaptive importance sampling method we developed. It uses the COS-PFE result to find the optimal shift to be applied to the original joint density function of the risk factors.

The algorithm was ran 100 times from which the variance of the PFE estimations was measured and compared to that from the straight forward Monte Carlo simulation. The reduction in variance, for an increasing number of paths, are presented in Table 7.3.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, SHIFT}}} \approx$	2.7	3.0	2.6	2.0	3.2	3.1

Table 7.3: The variance ratio between a straight forward MC simulation (numerator) and the shift-AIS-COS method (denominator), in the calculation of the PFE of a collateralized portfolio containing 100 derivatives.

Averaged out over all paths and all non-zero variances, the variance of the shift-AIS-COS method is approximately less than one-third of that of the straight forward Monte Carlo simulation.

For time points 2, 4, and 8, the convergence of the PFE estimations, as well as the associated 95% confidence interval, is shown in Figure 7.4

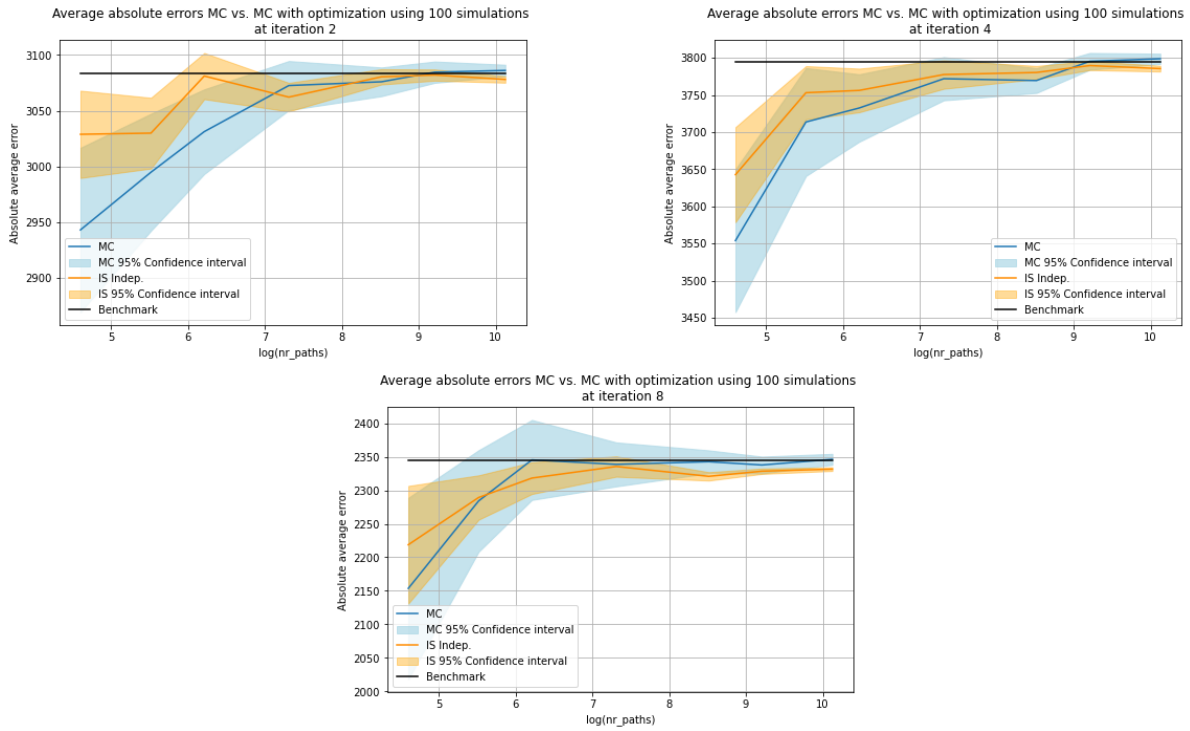

Figure 7.4: The convergence of the PFE of a collateralized portfolio with 100 derivatives at time points $t = 2, 4$ and 8 using the shift-AIS-COS method developed in Section 6.1 with the initial guess of the PFE from the split COS-PFE method.

In these plots the black line represents the benchmark PFE value, the solid blue line and solid orange line represents the PFE estimate averaged over the 100 PFE estimations, the blue and orange shaded areas depict the 95% confidence interval.

Although the variance reduction is effective, the optimal shift algorithm can take a very long time to compute the auxiliary density. For example, using 25000 paths the CPU time for all PFE estimations were between 96 and 1773 seconds.

7.3. Results of the Cross-Entrophy Adaptive Importance Sampling

In this section the results of the CE-AIS-COS method developed in Section 6.2 are presented and discussed. First, the results for the test portfolios without collateral are presented. Next, we compare the results of two variants of the CE-AIS-COS method for a test portfolio containing 100 derivatives with collateral: one using the exact COS-PFE method to provide the initial guess of the PFE value, which we refer to as the "CE-AIS-COS-exact" method; and the other that splits the collateral from the rest of the portfolio to approximate the initial guess of the PFE value, which we refer to as the "CE-AIS-COS-split" method. Note that CE-AIS-COS-split variant van be very useful in practise due to the fast calculation speed. Finally, to demonstrate this, we present the results of the CE-AIS-COS-split method applied to all test portfolios, whereby the portfolio splitting method is applied further, to illustrate the real-world application of this method. The results are based on 250 repeating runs of the PFE calculation.

7.3.1. Without collateral: results from the CE-AIS-exact method

The portfolio containing 100 derivatives

Table 7.4 presents In this table the averaged ratio between the variance of the straight forward Monte Carlo method (the numerator) and the CE-AIS-COS-exact simulation (the denominator) for different number of simulation paths. These variances are averaged over all non-zero variances.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}} \approx$	23.0	31.8	35.6	40.8	41.6	42.5

Table 7.4: The variance reductions of the CE-AIS-COS-exact method in PFE calculations for a uncollateralized portfolio containing 100 derivatives.

Averaged over all number of paths and all non-zero variances, the variance of the CE-AIS-COS-exact method is approximately 35.4 times lower than that of the straight forward Monte Carlo simulation.

Worth noting that these variance-reduction ratios match the theoretical reduction rations that we calculated in Table 6.1.

This significant variance reduction can also be seen in Figure 7.5, where the convergence to the benchmark PFE for the time steps $t = 5, 10$ and 15 are plotted.

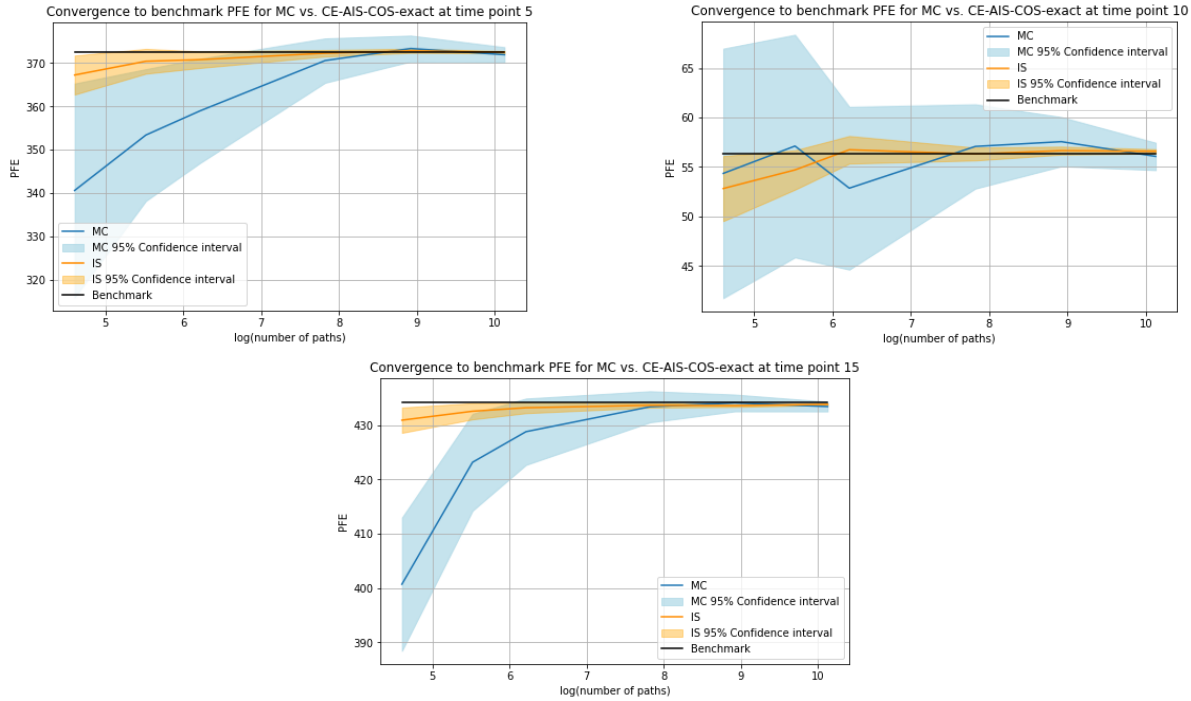

Figure 7.5: The convergence of the PFE, using straight forward Monte Carlo method and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 100 derivatives without collateral.

For the same time points, Figure 7.6 shows the comparison between the averaged absolute error and the CPU time for the CE-AIS-COS-exact algorithm and the straight forward Monte Carlo simulation.

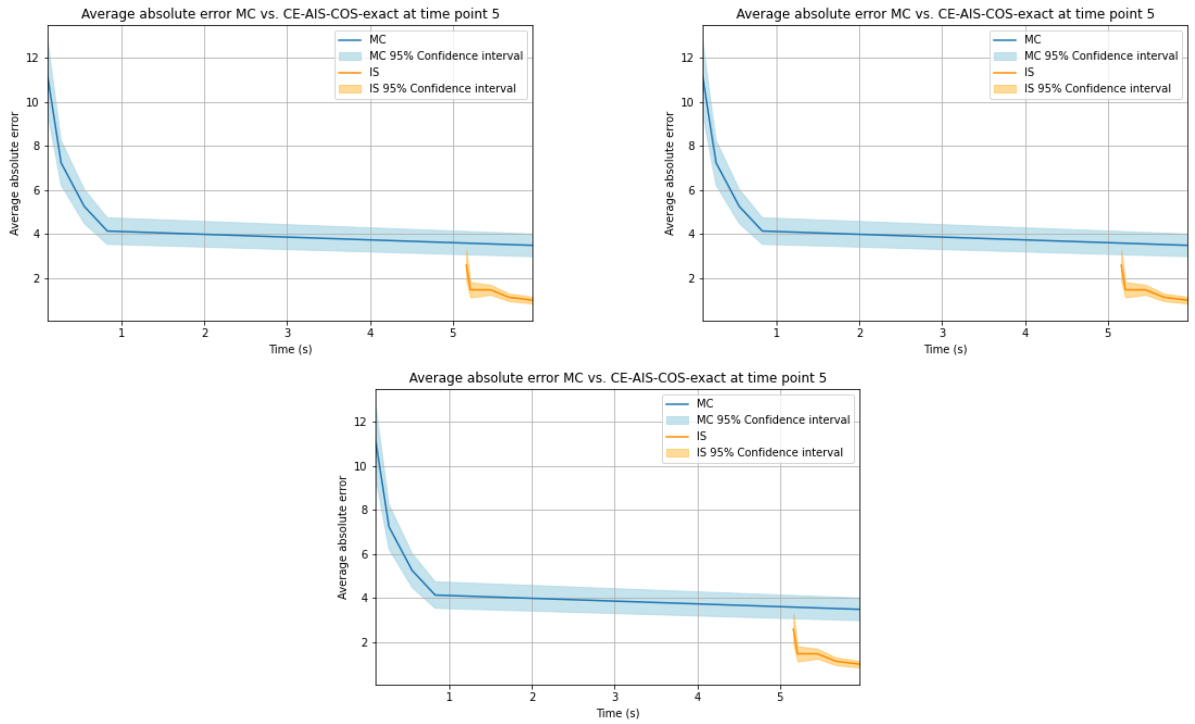

Figure 7.6: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 100 derivatives without collateral.

The portfolio containing 1000 derivatives

To test the applicability of the method to real-world portfolios, we increased the number of derivatives to 1000 and repeated the same tests. Table 7.5 again shows the significant variance reduction.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}} \approx$	30.6	36.2	38.7	41.6	42.1	42.5

Table 7.5: The variance reductions of the CE-AIS-COS-exact method in PFE calculations for a uncollateralized portfolio containing 1000 derivatives.

Averaged out over all paths and all non-zero variances, the variance of the CE-AIS-COS-exact algorithm is approximately 38.6 times lower compared to the straight forward Monte Carlo simulation.

The impressive increase in performance using the CE-AIS-COS-exact method is again demonstrated in Figure 7.7.

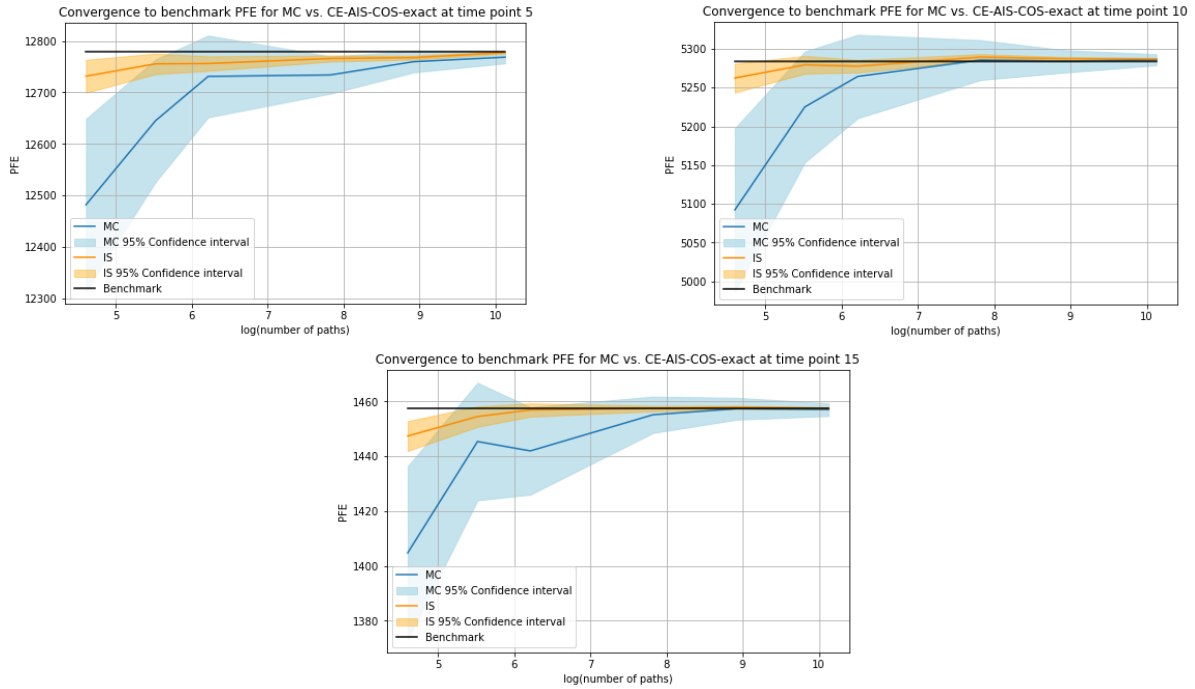

Figure 7.7: The convergence of the PFE, using the straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 1000 derivatives without collateral.

The average absolute errors and total CPU times of the two methods are compared. This comparison, for time points $t = 5, 10$ and 15, the comparison can be found in Figure 7.8.

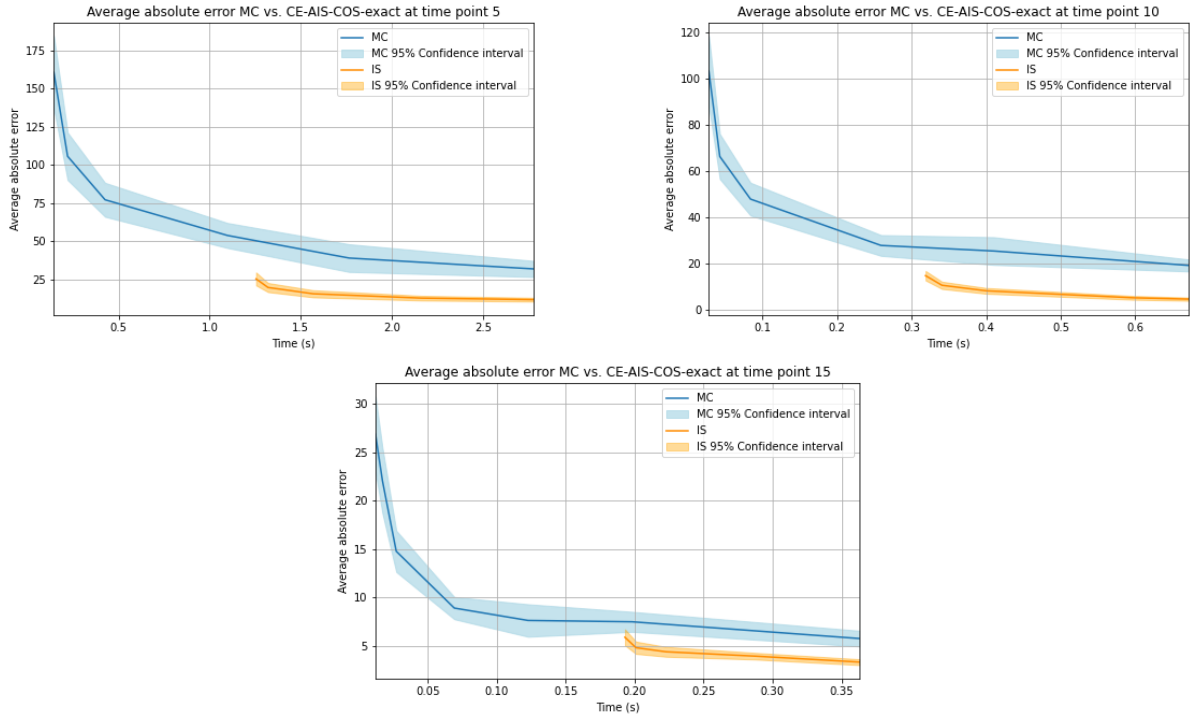

Figure 7.8: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 1000 derivatives without collateral.

The portfolio containing 10000 derivatives

To further test the applicability and performance of our methods, we repeated the same tests on a really large portfolio containing 10000 derivatives without collateral. Table 7.6 evidences the same, significant variance reduction as for smaller portfolios.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}} \approx$	28.8	35.5	37.8	43.4	40.7	36.9

Table 7.6: The variance reductions of the uncollateralized portfolio containing 10000 derivatives.

Averaged over all number of paths, and all non-zero the variance of the CE-AIS-COS-exact method is approximately 37.2 times lower than that of the straight forward Monte Carlo simulation.

As before, Figure 7.9 was made to showcase the convergence to the benchmark PFE using both methods.

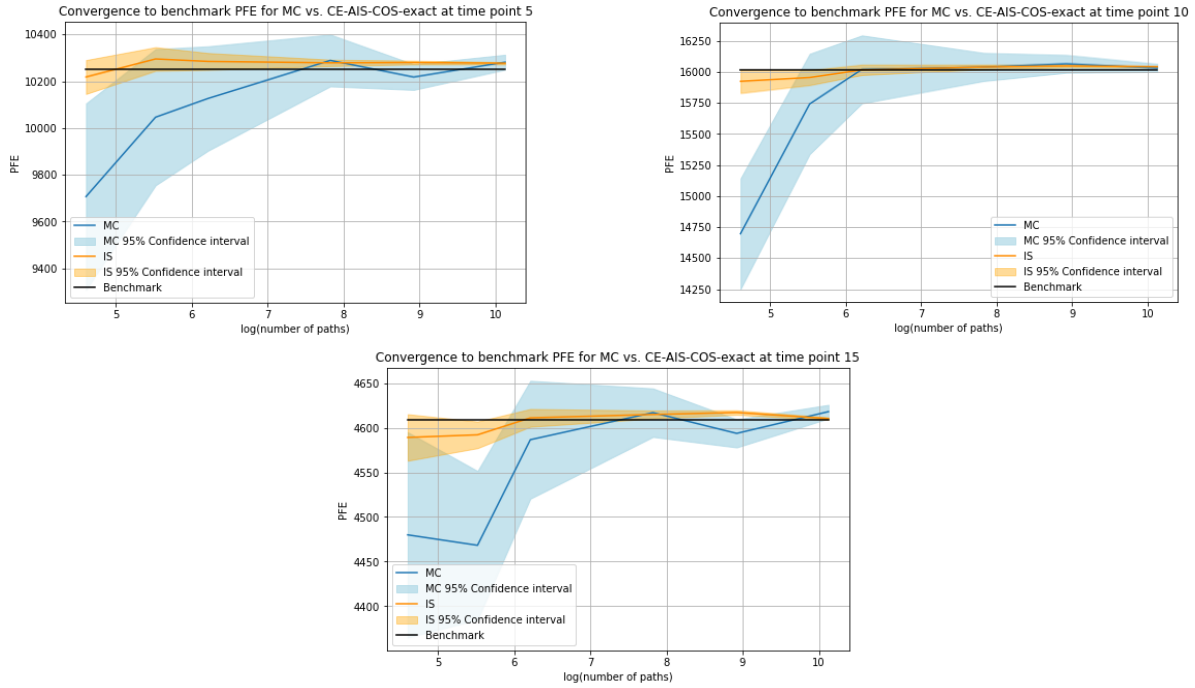

Figure 7.9: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 10000 derivatives without collateral.

The averaged absolute error and CPU time needed to obtain the approximations for the two methods can be seen in Figure 7.10.

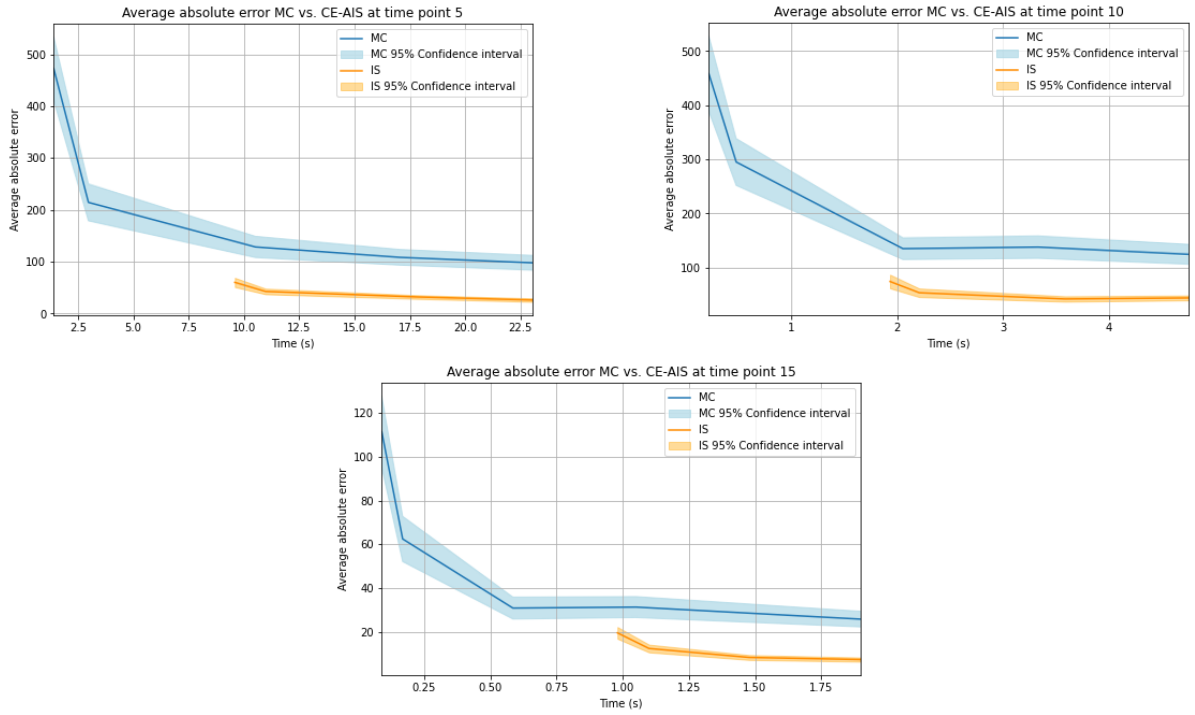

Figure 7.10: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10$ and 15 for a portfolio containing 10000 derivatives without collateral.

An illustration of the CE-AIS-COS applied to the PFE calculation

To check how the CE-AIS-COS method alters the original distribution of the underlying risk factors, Figures 7.11 and 7.12 were generated. Figure 7.11 shows how the CE-AIS-COS method changes the short rates samples after each iteration. The resulting shift in the exposure distribution can be seen in Figure 7.12.

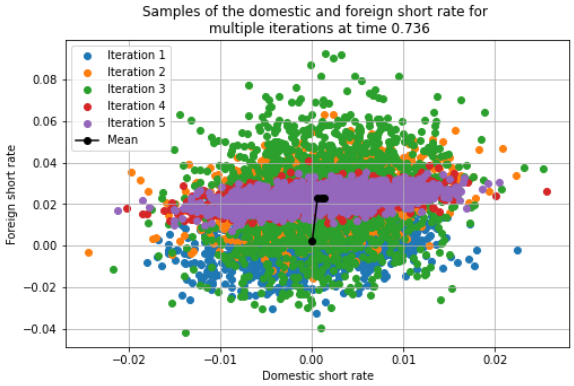

Figure 7.11: Random samples of the domestic and foreign short rate.

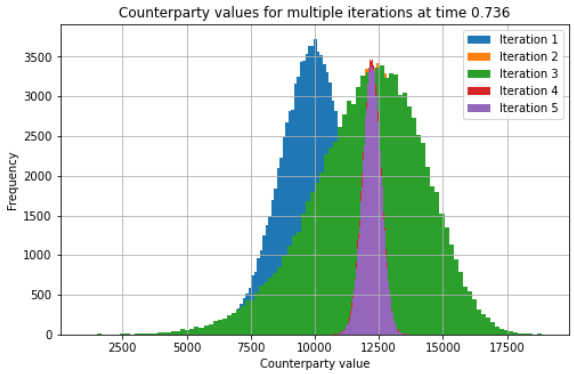

Figure 7.12: The shift in the exposure distribution.

7.3.2. With collateral: results from the CE-AIS-exact method

In this section the variance reduction and convergence of the CE-AIS-COS-exact method are tested on a collateralized portfolio. Table 7.7 summarizes the variance reduction compared to the straight forward Monte Carlo method for a test portfolio containing 100 derivatives.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}} \approx$	26.2	33.1	37.4	39.9	41.5	39.9

Table 7.7: The variance reduction for a collateralized portfolio containing 100 derivatives using the CE-AIS-COS-exact method.

The same, significant reduction in variance is observed as in the case without collateral.

Similar to the case without collateral, these variance-reduction ratios match the theoretical reduction rations that we calculated in Table 6.2.

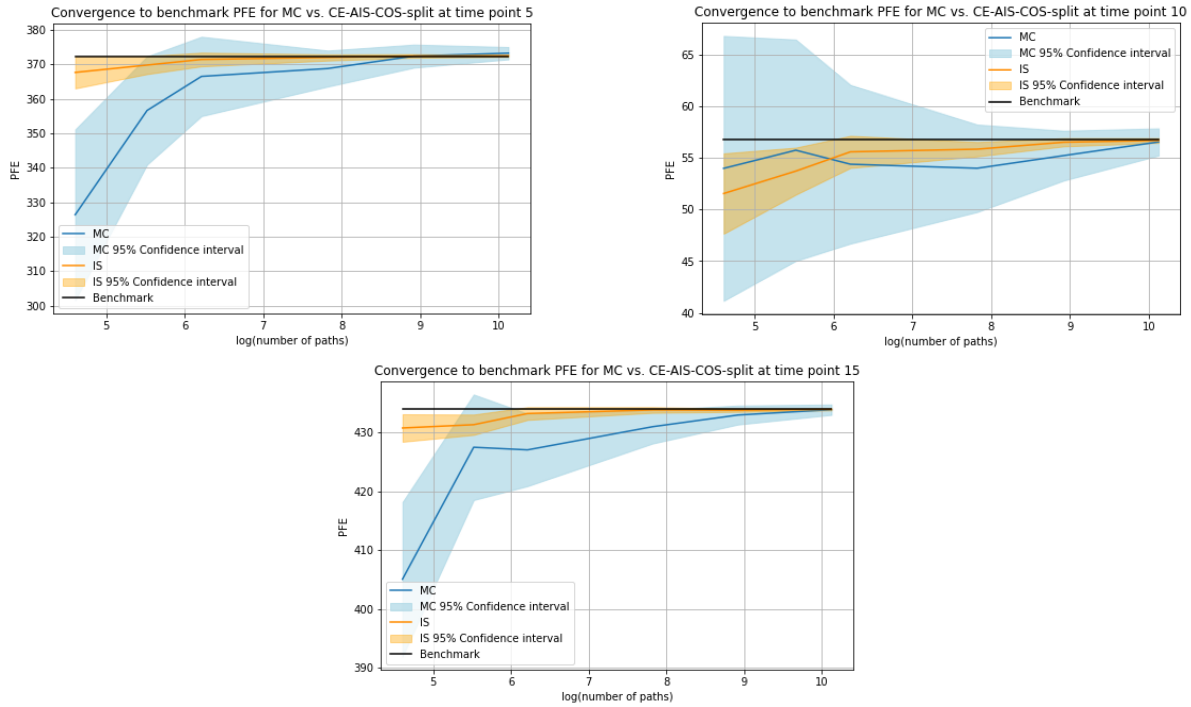

Figure 7.13: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-exact method, at time points $t = 5, 10, 15$ for a portfolio containing 100 derivatives with collateral.

For the same time points the average absolute error and CPU times are compared between the two methods. These can be seen in Figure 7.14.

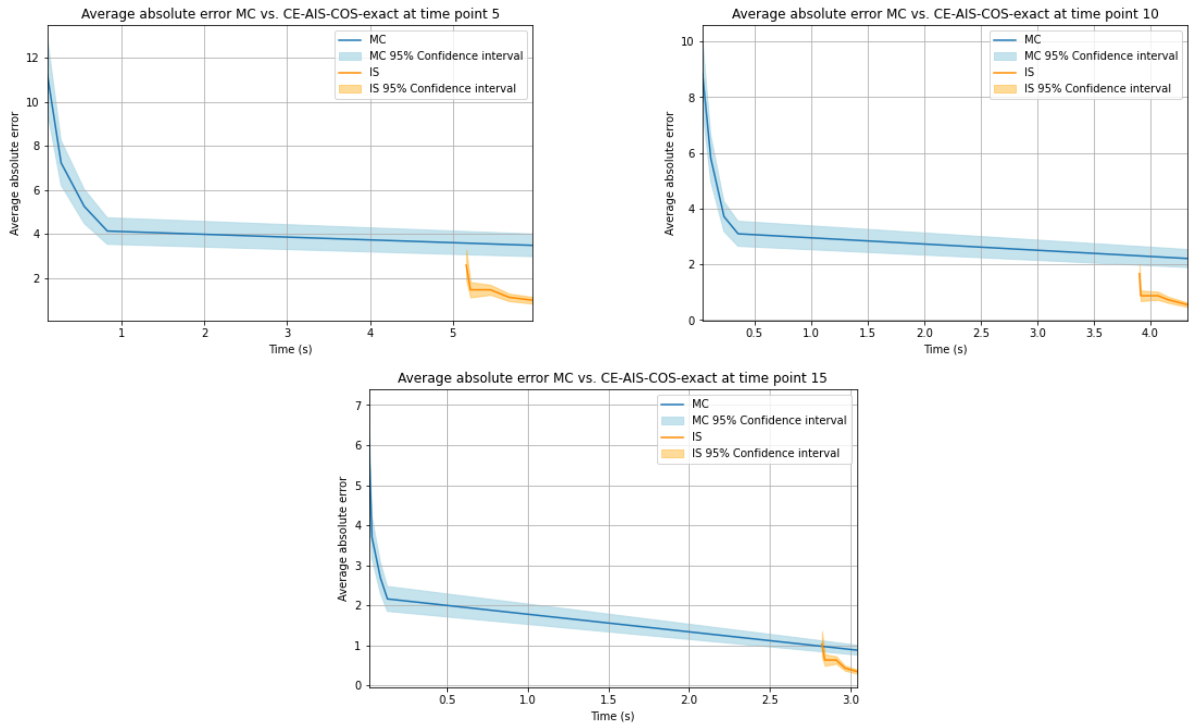

Figure 7.14: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-split, at time points $t = 5, 10$ and 15 for a portfolio containing 100 derivatives with collateral.

7.3.3. With collateral: results from the CE-AIS-COS-split method

After running the simulations Table 7.8 was constructed showing the variance reductions.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}}$	20.1	25.8	31.5	34.1	38.4	35.1

Table 7.8: The average variance reduction of a collateralized portfolio containing 100 derivatives with collateral using the CE-AIS-COS-split method.

The variance averaged over all non-zero variances and all number of paths 32.9 times lower compared to using the straight forward Monte Carlo simulation.

For time points $t = 5, 10$ and 15 the convergence of the PFE values and the confidence interval can be seen in Figure 7.15.

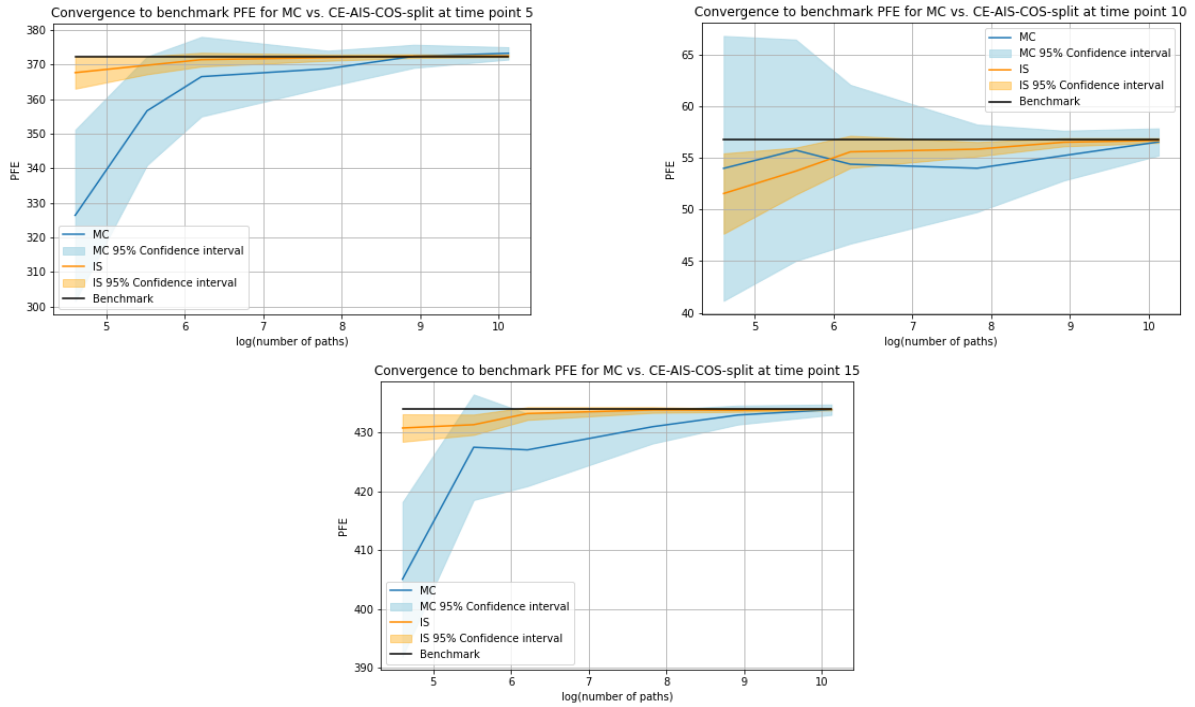

Figure 7.15: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method, at time points $t = 5, 10, 15$ for a portfolio containing 100 derivatives with collateral.

For the same time points the average absolute error and CPU times are compared between the two methods. These can be seen in Figure 7.16.

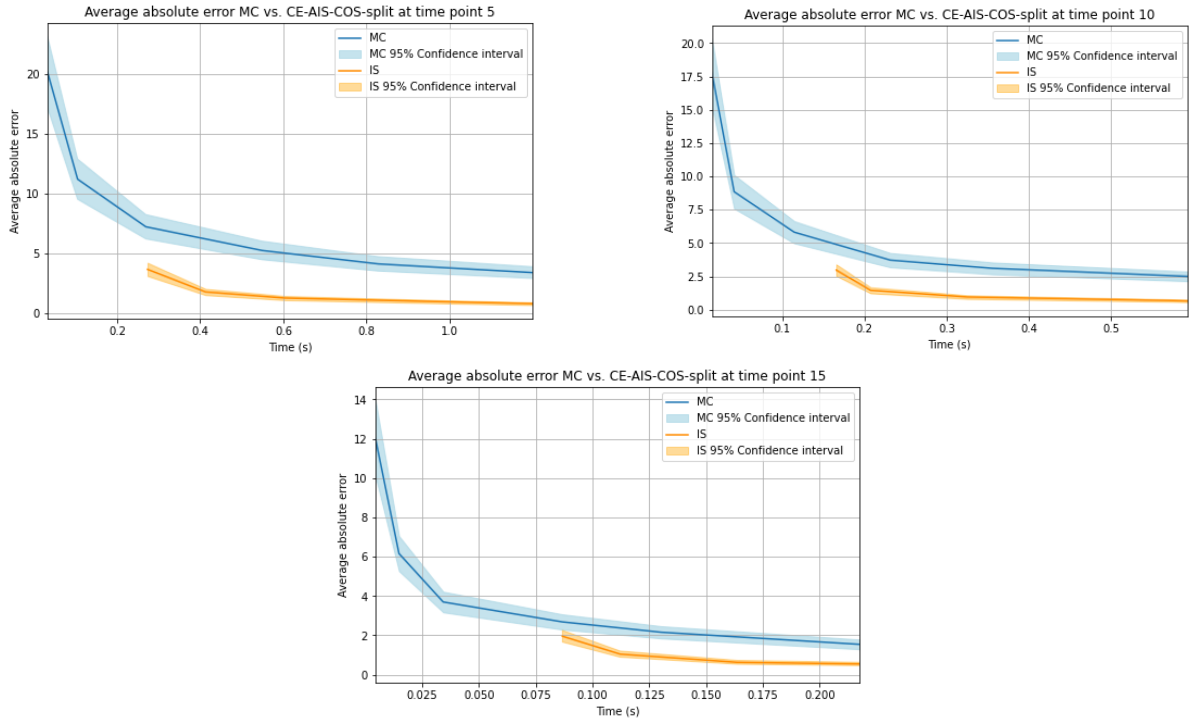

Figure 7.16: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-split, at time points $t = 5, 10$ and 15 for a portfolio containing 100 derivatives with collateral.

7.3.4. With collateral: results from the CE-AIS-COS-split using more sub-portfolios

In this section, the variance reduction for collateralized portfolios containing 100, 1000 and 10000 derivatives is show. Now, the PFE approximation is made using the COS-PFE-split method and applied on a further division of the portfolio.

The portfolio containing 100 derivatives

After running the simulations Table 7.9 could be constructed showing the variance reductions.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}}$	20.3	26.0	31.6	36.4	39.3	36.2

Table 7.9: The averaged variance reduction of a collateralized portfolio containing 100 derivatives with collateral using the CE-AIS-COS-split method applied on the sub-portfolios.

The variance averaged over all non-zero variances and all number of paths is 31.1 times lower compared to using the straight forward Monte Carlo simulation.

For three time points the convergence of the average PFE values and the confidence interval can be seen in Figure 7.17.

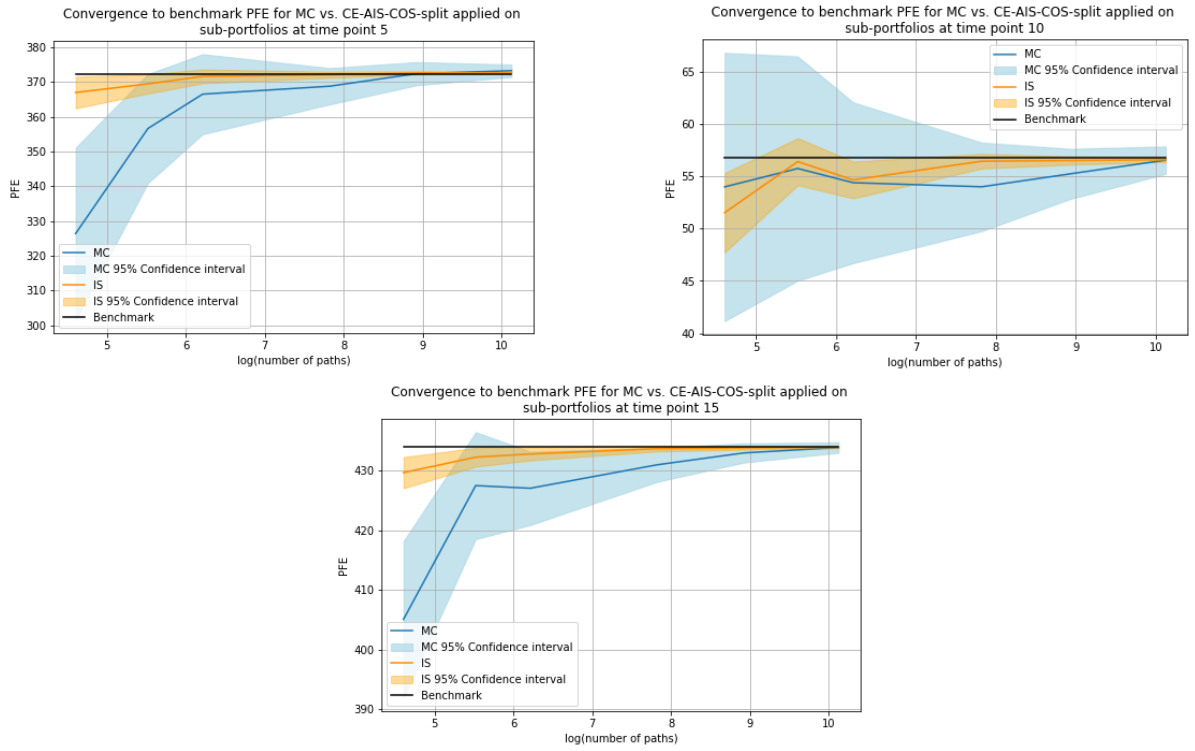

Figure 7.17: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 5, 10, 15$ for a portfolio containing 100 derivatives with collateral.

The comparison of the average absolute error and CPU time for three time points are showcased in Figure 7.18.

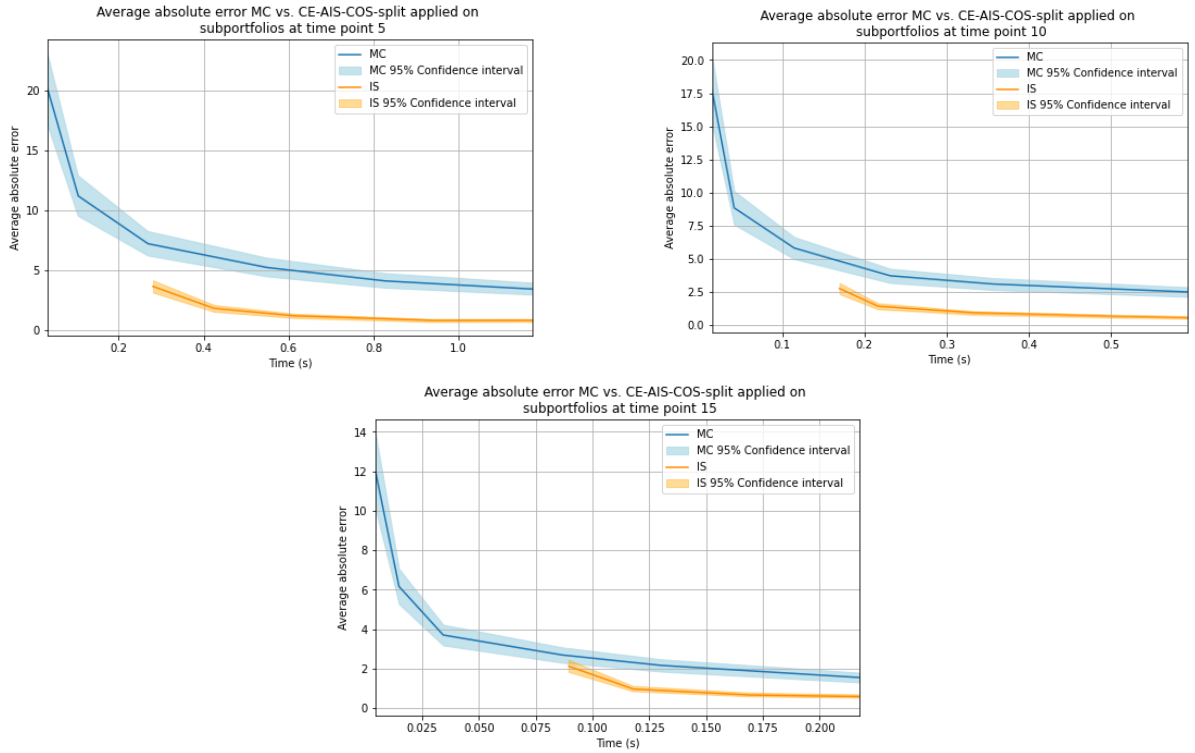

Figure 7.18: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 5, 10$ and 15 for a portfolio containing 100 derivatives with collateral.

The portfolio containing 1000 derivatives

For the test portfolio containing 1000 derivatives the variance reduction and averaged variances can be found in Table 7.10.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}}$	17.9	31.4	30.7	34.6	37.7	35.5

Table 7.10: The averaged variance reduction of a collateralized portfolio containing 1000 derivatives with collateral using the CE-AIS-COS-split method applied on the sub-portfolios.

Averaging out the variance reduction over all number of paths and all non-zero variances the CE-AIS method has a variance 32.4 times lower than that of the straight forward Monte Carlo simulation.

To illustrate the increase in performance using the CE-AIS method the convergence and confidence interval of the PFE can be seen in Figure 7.19.

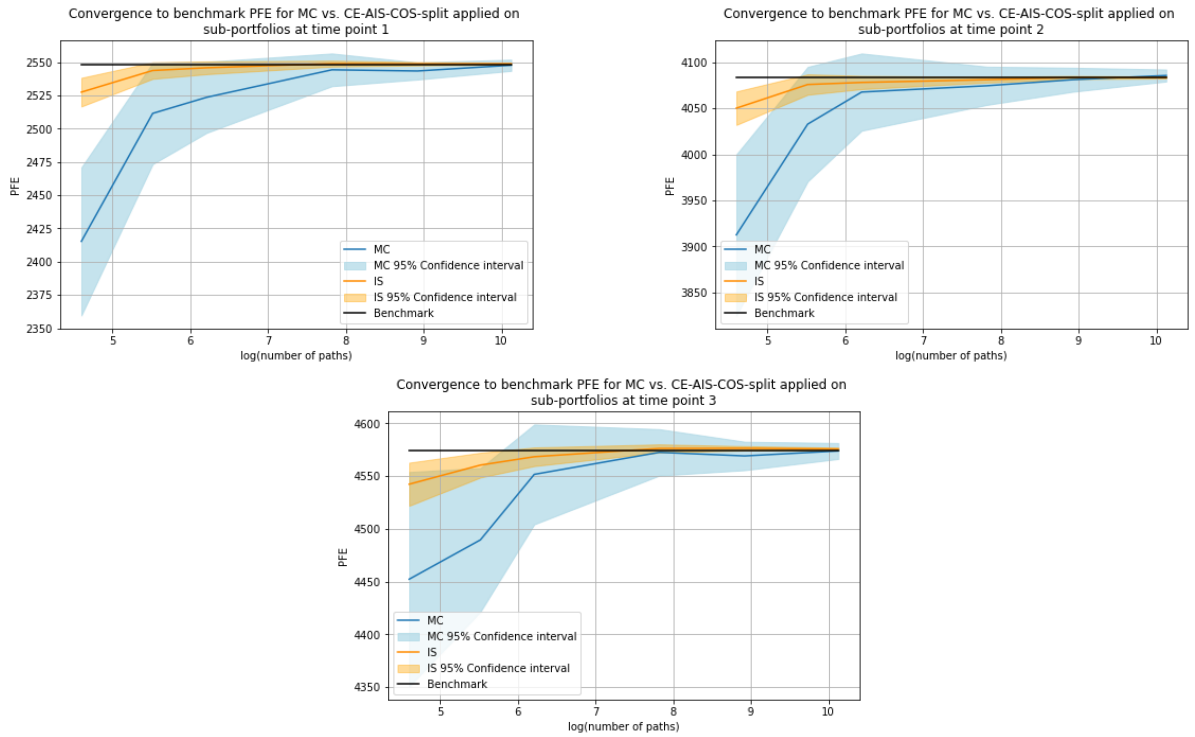

Figure 7.19: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 1, 2, 3$ for a portfolio containing 1000 derivatives with collateral.

The comparison of the average absolute error and CPU time for the three time points are shown in Figure 7.20.

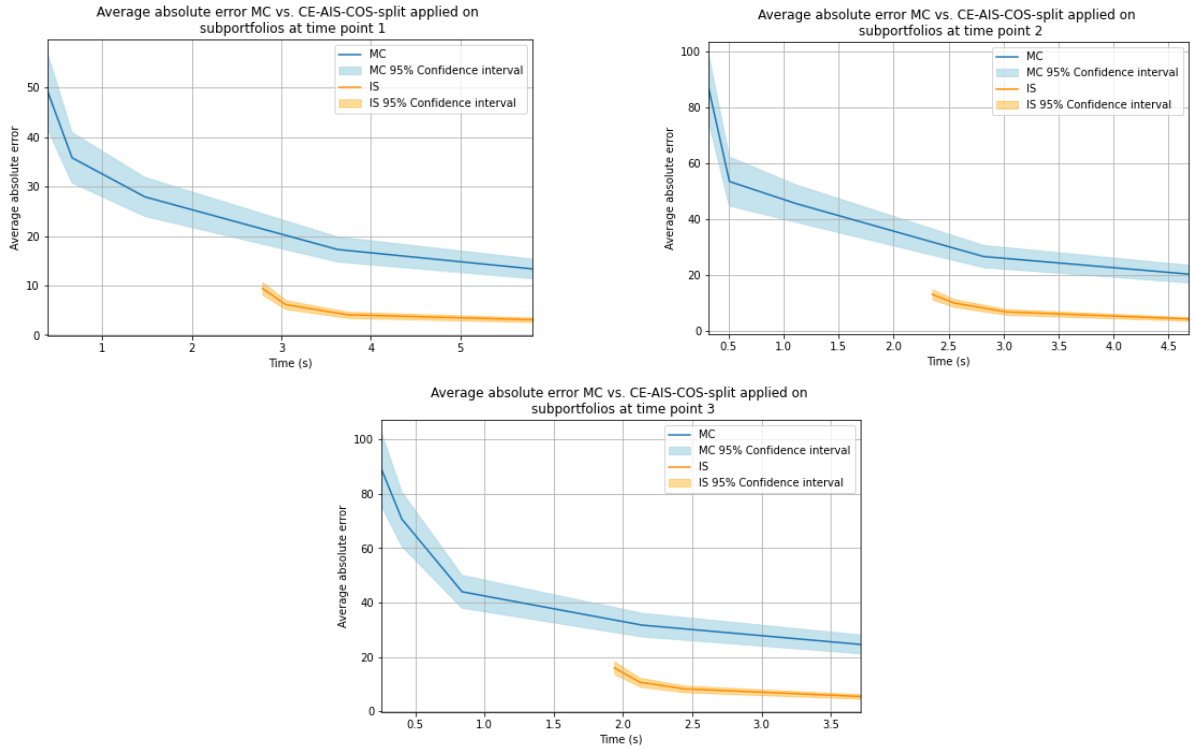

Figure 7.20: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 1, 2$ and 3 for a portfolio containing 1000 derivatives with collateral.

The portfolio containing 10000 derivatives

For the test portfolio containing 10000 derivatives the variance reduction and averaged variances can be found in Table 7.11.

Number of paths	100	250	500	2500	7500	25000
$\frac{\sigma^2_{\text{PFE, MC}}}{\sigma^2_{\text{PFE, CE-AIS}}}$	20.5	23.4	27.3	29.4	26.8	32.1

Table 7.11: The averaged variance reduction of a collateralized portfolio containing 10000 derivatives with collateral using the CE-AIS-COS-split method applied on the sub-portfolios.

Averaging out the variance reduction over all number of paths, and all non-zero variances the CE-AIS method has a variance 26.6 times lower than that of the straight forward Monte Carlo simulation.

To illustrate the increase in performance using the CE-AIS method the convergence and confidence interval of the PFE can be seen in Figure 7.21.

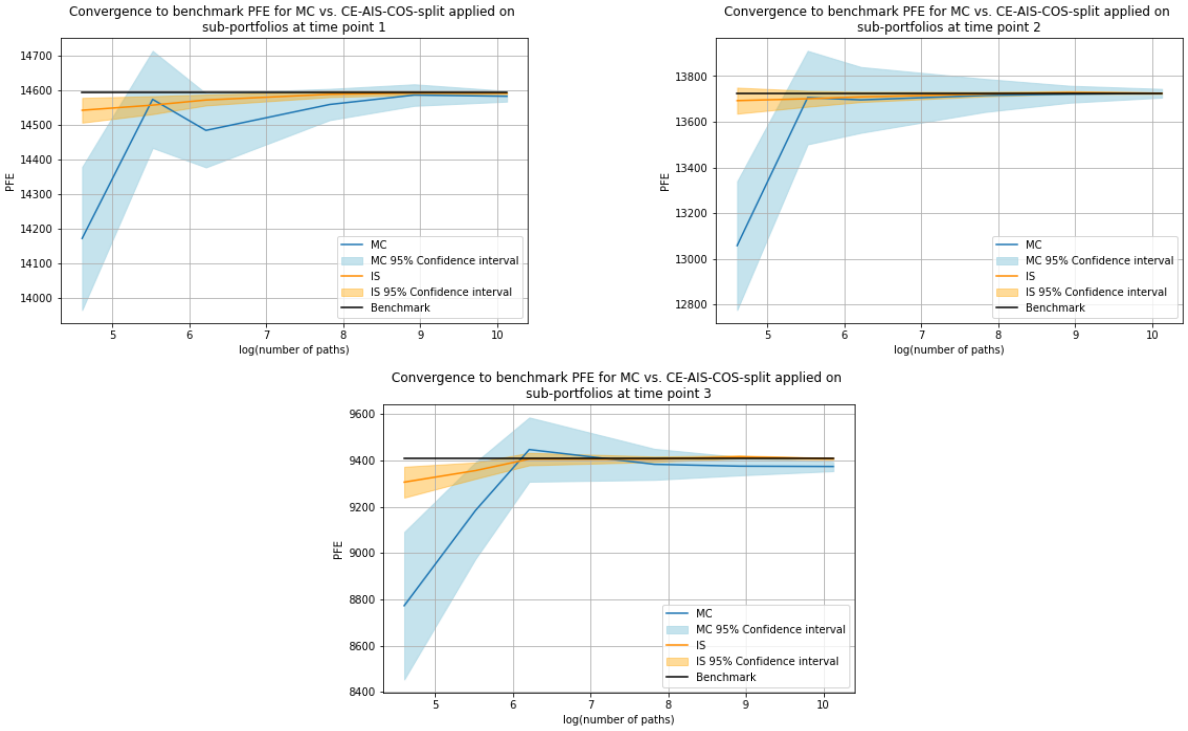

Figure 7.21: The convergence of the PFE, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 1, 2, 3$ for a portfolio containing 10000 derivatives with collateral.

For three time points the comparison of the average absolute error and CPU time can be seen in Figure 7.22.

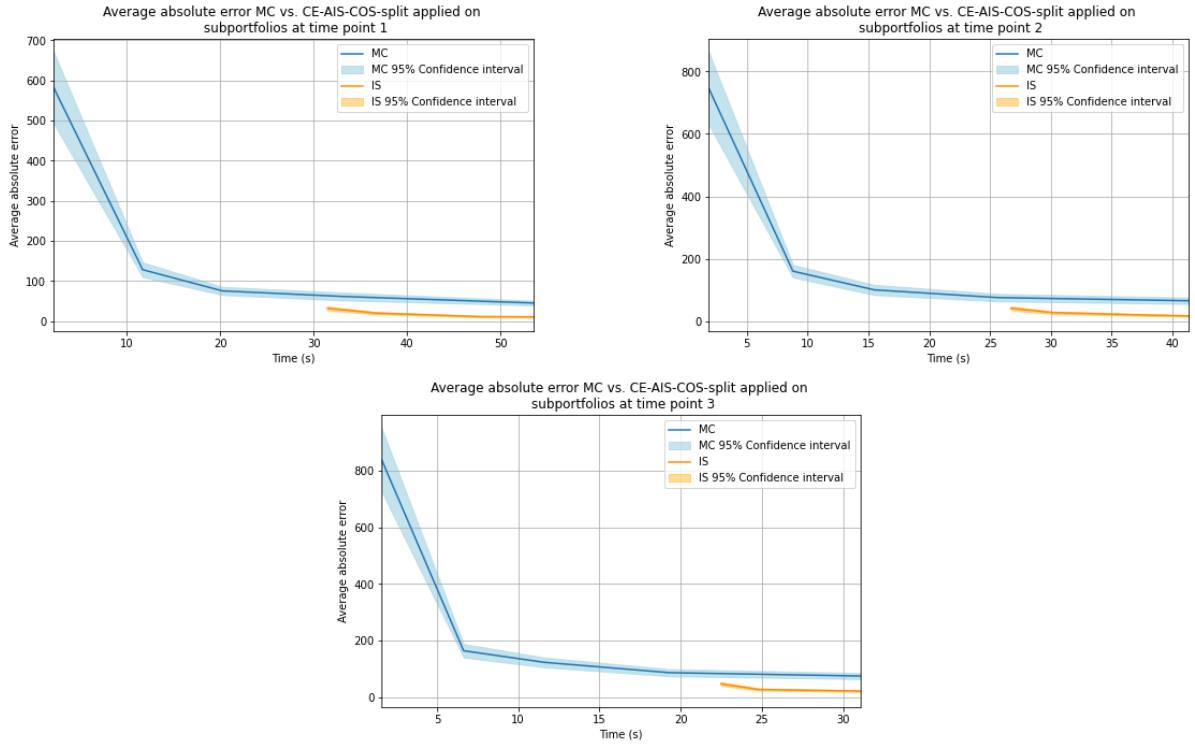

Figure 7.22: The averaged absolute error and CPU time, using straight forward Monte Carlo simulation and the CE-AIS-COS-split method applied on the sub-portfolios, at time points $t = 1, 2$ and 3 for a portfolio containing 10000 derivatives with collateral.

7.3.5. Impact of the accuracy in COS step on the variance reduction of CE-AIS-COS method

To understand how the CE-AIS-COS algorithm's performance depends on the accuracy of initial guess of the PFE value using the COS-PFE method, the variance reduction for different levels of accuracy is investigated, ranging from 0% to 120% of the real PFE. For example, the 0% means that the PFE approximation is 0 for each time point, and the 80% means that the PFE approximation from COS is 80% of the reference PFE at that time step. The algorithm was run 100 times for 1000000 paths at each time step. For a portfolio containing 100 derivatives with and without collateral, the variance reduction ratio and the accuracy in the COS step are plotted in Figure 7.23.

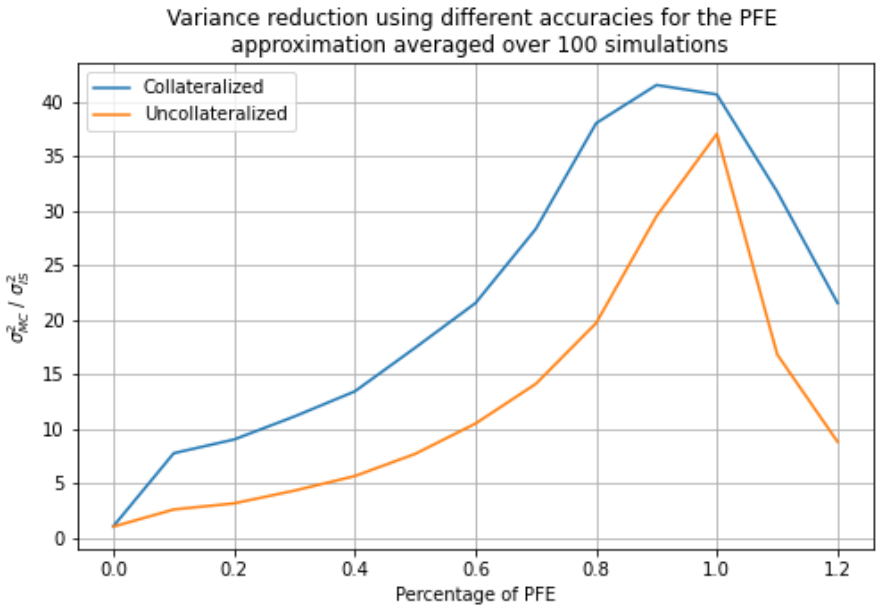

Figure 7.23: The variance reduction of the uncollateralized and collateralized portfolio containing 100 derivatives with varying accuracies of PFE approximations.

Not surprising, it can be seen that the variance reduction seems to increase exponentially from both sides of 100% accuracy of the COS-estimated PFE. This strongly suggests the great applicability of the COS method, in practice: a very fast calculation using the COS-PFE method with a "good enough" accuracy would already yield 20 times up to 40 times reduction in the variance of the Monte Carlo simulation.

8

Conclusions and discussions

In this thesis we combined the COS-PFE method, the extended COS method presented in [1], and the Monte Carlo simulation to reduce the variance of the PFE calculation of a portfolio. We researched on Control Variate method and Importance Sampling method, both being well studied variance reduction techniques in literature.

The first method that we developed was a Control Variate method with the control variate being yielded from the COS-PFE method. This method, however, was tested to be not effecitive in reducing the variance of the PFE. One potential reason could be that the correlation between the indicator function used in the Monte Carlo CDF estimator and the control variate is not high enough. Testing results of a portfolio containing 100 derivatives without collateral suggests that, the variance using the control variate method is approximately the same as that of straight forward Monte Carlo simulation. Averaged over all non-zero variances the variance reduction was approximately 1.09. Further testing showed, however, it is very successful in reducing the variance of the expected exposures.

Secondly, adaptive importance sampling was combined with the COS-PFE method as such: via an iterative procedure, a shift is searched for the risk factor with the highest correlation to the MtM value of the portfolio, such that the resulting expected exposure based on the shifted distribution coincides with the PFE obtained from the COS-PFE method, while the variance is minimized. Extensive testing results indicate that the variance of the PFEs from this method were less than one third of that from the straight forward Monte Carlo simulation. While being successful in reducing the variances, the computation time of this method is not superior. The total CPU time of the algorithm, using 25000 paths, was between 96 and 1773 seconds.

Then the majority research time of this thesis was spent on the development of an adaptive importance sampling method based on cross-entropy. This method finds a probability distribution that minimized the Kullbeck-Leibler divergence between itself and the theoretical zero-variance probability density. The COS-PFE method is used to provide the initial guess of the PFE value. Therefore, we name it the CE-AIS-COS method. The CE-AIS-COS method was tested for three portfolios both with and without collateral. For the uncollateralized portfolios containing 100, 1000 and 10000 derivatives the variance of the PFE using the CE-AIS-COS method was approximately 35.4, 38.6 and 37.2 times lower than that of the straight forward Monte Carlo simulation, which is significant and perfectly align with our theoretical results of the variance reduction. The same impressive performance was observed for portfolios with collateral. To reduce the CPU time, we applied a dimension-reduction approximation in the original COS-PFE method. This variant we call CE-AIS-COS-split method. Using the same test portfolio with collaterals, the CE-AIS-COS-split method produced PFE estimates with an average variance 32.9 times lower than the straight forward Monte Carlo simulation, while the CPU time is much lower, too. When we further split the portfolio into sub-portfolios, to mimic the real-world situation that the COS-PFE is only suitable for portfolios involving a few risk factors. That is, we use stand-alone COS-PFE estimates to approximate the portfolio level PFE, which is fed to the adaptive importance sampling method as the initial guess. Results demonstrated the variance

reduction ratios are slightly less than before but are still significant: on average 31.1, 32.4 and 26.6 for portfolios having 100, 1000 and 10000 derivatives, respectively.It can be concluded that the CE-AIS-COS method has the potential to be applied on real-world portfolios, to produce PFE estimations with a lower average absolute error than the straight forward Monte Carlo simulation using the same CPU time.

While this study have has provided an intuitive and easy to implement method for reducing the variance of PFE calculations some limitations may suggest avenues for future research. First is that the Z-spread dynamic is assumed to be independent of the processes of the other risk factors, which simplifies the bond pricing formula. Future research could implement a correlation between the Z-spread and the other processes. Additionally, this work only concerns SFTs. Future work could be done to extend the method to CSA agreements.

Bibliography

[1] F. Fang, G. Mast, and X. Shen, "Fast calculation of counterparty credit exposures and associated sensitivities using fourier series expansion," 2023.

[2] J. Gregory, *Counterparty Credit Risk*. Wiley, 2010.

[3] B. for International Settlements, "Otc derivatives statistics at end-june 2023," 2023.

[4] F. Fang and C. W. Oosterlee, "A novel pricing method for european options based on fourier-cosine series expansions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, 2009.

[5] E. Securities and M. Authority, *Eu derivatives markets 2023*, Online, accessed August 15, 2024, 2023. [Online]. Available: https://www.esma.europa.eu/sites/default/files/2023-12/ESMA50-524821-2930%5C_EU%5C_Derivatives%5C_Markets%5C_2023.pdf.

[6] ISDA, "Key trends in the size and composition of otc derivatives markets in the first half of 2023," 2023.

[7] P. Glasserman, *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

[8] C. Cannamela, J. Garnier, and B. Iooss, "Controlled stratification for quantile estimation," *The Annals of Applied Statistics*, vol. 2, no. 4, pp. 1554–1580, 2008.

[9] H. Dong and M. K. Nakayama, "A tutorial on quantile estimation via monte carlo," in *Monte Carlo and Quasi-Monte Carlo Methods*, Springer, 2020, pp. 3–30.

[10] J. Cardenas, E. Fruchard, J. Picron, *et al.*, "Monte carlo within a day," *Risk*, vol. 12, pp. 55–59, 1999.

[11] P. Glasserman, P. Heidelberger, and P. Shahbuddin, "Importance sampling and stratification for value-at-risk," 2000.

[12] H. Kahn, "Random sampling (monte carlo) techniques in neutron attenuation problems – i," *Nucleonics*, vol. 6, no. 5, 1950.

[13] J. A. Bucklew, *Introduction to Rare Event Simulation*. Springer, 2004.

[14] G. Casella and C. Robert, *Monte Carlo Statistical Methods*. Springer, 2004.

[15] L. Guibas and E. Veach, "Optimally combining sampling techniques for monte carlo rendering," 1995.

[16] M. Buggalo, P. Djuric, V. Elvira, *et al.*, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 60–79, 2017.

[17] M. Bolic, P. Djuric, and T. Li, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 70–86, 2015.

[18] M. Bugallo, V. Elvira, D. Luengo, and L. Martino, "Improving population monte carlo: Alternative weighting and resampling schemes," *Signal Processing*, vol. 131, no. 12, pp. 77–91, 2017.

[19] P. Bückner, T. Paananen, J. Piironen, and A. Vehtari, "Imp;icitly adaptive importance sampling," *Statistics and Computing*, vol. 31, no. 16, 2021.

[20] D. Egloff and M. Leippold, "Quantile estimation with adaptive importance sampling," *The Annals of Statistics*, vol. 38, no. 2, pp. 1244–1278, 2010.

[21] S. He, G. Jiang, H. Lam, and M. C. Fu, "Adaptive importance sampling for efficient stochastic root finding and quantile estimation," 2023.

[22] R. Rubinstein, "Optimization of computer simulation models with rare events," *European Journal of Operational Research*, vol. 99, no. 1, pp. 89–112, 1997.

[23] M. E. Masri, J. Morio, and F. Simatos, "Improvement of the cross-entropy method in high dimension for failure probability estimation through a one-dimensional projection without gradient estimation," *Reliability Engineering and System Safety*, vol. 216, 2021.

[24] Z. Botev, D. Kroese, P. L'Ecuyer, and R. Rubinstein, "The cross-entropy method for optimization," *Methodology and Computing in Applied Probability*, vol. 8, pp. 383–407, 206.

[25] D. Kroese, S. Porotsky, and R. Rubinstein, "The cross-entropy method for continuous multi-extremal optimization," *Handbook of Statistics*, vol. 31, pp. 35–59, 2013.

[26] R. Babuska, L. Busoniu, D. Ernst, and B. d. Schutter, *Reinforcement Learning and Dynamic Programming Using Function Approximators*. CRC Press, 2010.

[27] A. Lörincz, B. Póczos, and B. Szabó, "Cross-entropy optimization for independent process analysis," *Independent Component Analysis and Blind Signal Separation*, vol. 3889, pp. 909–916, 2006.

[28] J. Chan and D. Kroese, "Improved cross-entropy method for estimation," *Statistics and Computing*, vol. 22, pp. 1031–1040, 2011.

[29] R. Rubinstein, "How to deal with the curse of dimensionality of likelihood ratios in monte carlo simulations," *Stochastic Models*, vol. 25, no. 4, pp. 547–568, 2009.

[30] J. Chan, P. W. Glynn, and D. Kroese, "A comparison of cross-entropy and variance minization strategies," *Journal of Applied Probability*, vol. 48, pp. 183–194, 2010.

[31] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Asymptotically optimal importance sampling and stratification for pricing path-dependent options," *Mathematical finance*, vol. 9, no. 2, pp. 117–152, 1999.

[32] M. Fu and Y. Su, "Optimal importance sampling in securities pricing," *Journal of Computational Finance*, vol. 5, no. 4, pp. 27–50, 2002.

[33] T. Reitan and K. Aas, "A new robust importance-sampling method for measuring value-at-risk and expected shortfall allocations for credit portfolios," *The Journal of Credit Risk*, vol. 6, no. 4, p. 113, 2010.

[34] L. Hoogerheide and H. K. van Dijk, "Bayesian forecasting of value at risk and expected shortfall using adaptive importance sampling," *International Journal of Forecasting*, vol. 26, no. 2, pp. 231–247, 2010.

[35] D. Brigo and F. Mercurio, *Interest Rate Models - Theory and Practice*. Springer, 2006.

[36] D. Duffie and R. Kan, "A yield-factor model of interest rates," *Mathematical Finance*, vol. 6, no. 4, pp. 379–406, 1996.

[37] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[38] J. Stone, *Information Theory: A Tutorial Introduction*. 2019.

[39] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[40] B. McMillan, "Two inequalities implied by unique decipherability," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 115–116, 1956.

[41] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[42] R. Eckhardt, "Stan ulam, john von neumann, and the monte carlo method," *Los Alamos Science*, vol. Special Issue, no. 15, pp. 131–137, 1987.

[43] O. Summerscales, "Hitting the jackpot: The birth of the monte carlo method," 2023. (visited on 08/30/2024).

[44] B. Baxter and R. Brummelhuis, "Functionals of exponential brownian motion and divided differences," *Journal of Computational and Applied Mathematics*, vol. 236, no. 4, pp. 424–433, 2010.

[45] JPMorgan, "Risk metrics tm - technical document," 1995.

[46] B. Arouna, "Adaptive monte carlo method, a variance reduction technique," *Monte Carlo Methods Appl.*, vol. 10, no. 1, pp. 1–24, 2004.

[47] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 1998.

[48] P. Glasserman, P. Heidelberger, and P. Shahabuddin, "Variance reduction techniques for estimating value-at-risk," *Management Science*, vol. 46, no. 10, pp. 1349–1364, 2000.

[49] P. de Boer, D. Kroese, S. Mannor, and R. Rubinstein, "A tutorial on the cross-entropy method," *Annals of Operations Research*, vol. 134, pp. 19–67, 2005.

[50] D. Kroese and R. Rubinstein, *The Cross-Entropy Method*. Springer, 2004.

[51] D. Kroese and R. Rubinstein, *Simulation and the Monte Carlo method. Wiley series in probability and statistics.* Wiley, 2017.

[52] W. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, vol. 4, pp. 2111–2245, 1994.

[53] T. Homem-de-Mello and R. Rubinstein, "Estimation of rare event probabilities using cross-entropy," in *Proceedings of the Winter Simulation Conference*, vol. 1, 2002, pp. 310–319.

[54] T. Homem-de-Mello and R. Rubinstein, "Rare event estimation for static models via cross-entropy and importance sampling," 2002.

[55] Y. Kaniovski, A. J. King, and R. Wets, "Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems," *Annals of Operations Research*, vol. 56, pp. 189–208, 1995.

[56] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.

[57] J. Jacob and P. Protter, *Probability Essentials, Second Edition*. Springer, 2004.

A

Appendix

A.1. Proof of Proposition 1

Proof. From the G1++ model we have

$$r(t) = x(t) + \beta(t)$$

With

$$\begin{aligned} dx(t) &= -ax(t)dt + \sigma dW(t) \\ x(0) &= 0 \end{aligned}$$

We get that

$$\begin{aligned} d(e^{at}x(t)) &= ae^{at}x(t)dt + e^{at}dx(t) \\ &= ae^{at}x(t)dt + e^{at}[-ax(t)dt + \sigma dW(t)] \\ &= e^{at}\sigma dW(t) \end{aligned}$$

Then we find that for $t \geq s$

$$\begin{aligned} \int_s^t d(e^{au}x(u)) &= \int_s^t e^{au}\sigma dW(u) \\ e^{at}x(t) - e^{as}x(s) &= \sigma \int_s^t e^{au}dW(u) \end{aligned}$$

Then we find that $x(t) = e^{-a(t-s)}x(s) + \sigma \int_s^t e^{-a(t-u)}dW(u)$, and filling into $r(t)$ we get that

$$r(t) = e^{-a(t-s)}x(s) + \sigma \int_s^t e^{-a(t-s)}dW(u) + \beta(t)$$

We find that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[\,r(t)|\,\mathcal{F}_s] &= \mathbb{E}^{\mathbb{Q}}\left[e^{-a(t-s)}x(s) + \sigma\int_s^t e^{-a(t-s)}dW(u) + \beta(t)\middle|\mathcal{F}_s\right] \\
&= e^{-a(t-s)}x(s) + \beta(t)\mathbb{E}^{\mathbb{Q}}\left[\sigma\int_s^t e^{-a(t-u)}dW(u)\middle|\mathcal{F}_s\right] \\
&= e^{-a(t-s)}x(s) + \beta(t) \\
\mathrm{var}^{\mathbb{Q}}[\,r(t)|\,\mathcal{F}_s] &= \mathrm{var}^{\mathbb{Q}}\left[e^{-a(t-s)}x(s) + \sigma\int_s^t e^{-a(t-u)}dW(u) + \beta(t)\middle|\mathcal{F}_s\right] \\
&= \mathrm{var}^{\mathbb{Q}}\left[\sigma\int_s^t e^{-a(t-u)}dW(u)\middle|\mathcal{F}_s\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\left(\sigma\int_s^t e^{-a(t-u)}dW(u)\right)^2\middle|\mathcal{F}_s\right] \\
&=^{\star} \mathbb{E}^{\mathbb{Q}}\left[\sigma^2\int_s^t e^{-2a(t-u)}du\middle|\mathcal{F}_s\right] \\
&= \frac{\sigma^2}{2a}\left[1 - e^{-2a(t-s)}\right]
\end{aligned}$$

$\star$ By Itô's isometry.

We know that

$$\begin{aligned}
P(0,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r(s)ds}\middle|\mathcal{F}_s\right] &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T x(s)ds - \int_0^T \beta(s)ds}\middle|\mathcal{F}_s\right] \\
&= e^{-\int_0^T \beta(s)ds}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T x(s)ds}\middle|\mathcal{F}_s\right]
\end{aligned}$$

We know that since $x(t)$ is normally distributed $\int x(t)dt$ is normally distributed. We find that

$$\begin{aligned}
x(t) &= e^{-a(t-0)}x(0) + \int_0^t e^{-a(t-u)}dW(u) \\
&= \int_0^t e^{-a(t-u)}dW(u)
\end{aligned}$$

Using Fubini's Theorem for stochastic integrals and substitution we find that

$$\begin{aligned}
\int_0^T x(t)dt &= \int_0^T\int_0^t \sigma e^{-a(t-u)}dW(u)dt \\
&= \frac{\sigma}{a}\int_0^t \left(1 - e^{-a(T-u)}\right)dW(u)
\end{aligned}$$

We see that $dW(u)$ is normally distributed with mean zero. The variance is given by

$$\begin{aligned}
V(0,T) &= \mathrm{var}^{\mathbb{Q}}\left[\int_0^T x(t)dt\middle|\mathcal{F}_0\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\left(\frac{\sigma}{a}\int_0^T \left(1 - e^{-a(T-u)}\right)dW(u)\right)^2\middle|\mathcal{F}_0\right] \\
&=^{\star} \mathbb{E}^{\mathbb{Q}}\left[\frac{\sigma^2}{a^2}\int_0^T \left(1 - e^{-a(T-u)}\right)^2 du\middle|\mathcal{F}_0\right] \\
&= \frac{\sigma^2}{a^2}\int_0^T \left(1 - e^{-a(T-u)}\right)^2 du \\
&= \frac{\sigma^2}{a^2}\left(T - 2\frac{1 - e^{-aT}}{a} + \frac{1 - e^{-aT}}{2a}\right)
\end{aligned}$$

⋆ By Itô's isometry.

Now filling in what we know yields

$$\begin{aligned} P(0,T) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T x(t)dt - \int_0^T \beta(t)dt} \middle| \mathcal{F}_0 \right] \\ &= e^{-\int_0^T \beta(t)dt} \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T x(t)dt} \middle| \mathcal{F}_0 \right] \\ &=^{\star} e^{-\int_0^T \beta(t)dt} e^{\frac{1}{2}V(0,T)} \end{aligned}$$

⋆ By the moment generating function of the normal distribution $\mathbb{E}^{\mathbb{Q}}\left[e^{tX}\right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ where X is normally distributed with mean μ and variance σ^2.

We know that if the model is calibrated to the market $P(0,T) = P^M(0,T) = e^{-\int_0^T f_M(0,s)ds}$. From this we find that

$$\begin{aligned} e^{-\int_0^T f^M(0,s)ds} &= e^{-\int_0^T \beta(s)ds + \frac{1}{2}V(0,T)} \\ \int_0^T \beta(s)ds &= \int_0^T f^M(0,s)ds + \frac{1}{2}V(0,T) \end{aligned}$$

After differentiation, we get that

$$\beta(t) = f^M(0,t) + \frac{\sigma^2}{2a^2}(1 - e^{-at})^2$$

□

A.2. Proof of Proposition 2

Proof. In Proof A.1 we found that

$$x(t) = x(s)e^{-a(t-s)} + \sigma \int_s^t e^{-a(t-u)} dW(u)$$

We also have that

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s)ds} \middle| \mathcal{F}_t \right] = e^{-\int_t^T \beta(s)ds} \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T x(s)ds} \middle| \mathcal{F}_t \right]$$

Again, since we know that $x(t)$ is normally distributed, we know that $\int_t^T x(s)ds$ is also normally distributed. After finding the mean and variance of $\int_t^T x(s)ds$ we can use the moment generating function to compute the last expectation.

$$\begin{aligned} \int_t^T x(s)ds &= \int_t^T e^{-a(s-t)} x(t)ds + \sigma \int_t^T \int_t^s e^{-a(s-u)} dW(u)ds \\ &= x(t)\frac{1 - e^{-a(T-t)}}{a} + \sigma \int_t^T \int_t^s e^{-a(s-u)} dW(u)ds \\ &= x(t)\frac{1 - e^{-a(T-t)}}{a} + \frac{\sigma}{a} \int_t^T \left(1 - e^{-a(T-u)}\right) dW(u) \end{aligned}$$

From this we find that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[\int_t^T x(s)ds\middle|\mathcal{F}_s\right] &= x(t)\frac{1-e^{-a(T-t)}}{a}\\
\text{var}^{\mathbb{Q}}\left[\int_t^T x(s)ds\middle|\mathcal{F}_s\right] &= \text{var}^{\mathbb{Q}}\left[x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{\sigma}{a}\int_t^T\left(1-e^{-a(T-u)}\right)dW(u)\middle|\mathcal{F}_s\right]\\
&= \text{var}^{\mathbb{Q}}\left[\frac{\sigma}{a}\int_t^T\left(1-e^{-a(T-u)}\right)dW(u)\middle|\mathcal{F}_s\right]\\
&=^{\star}\frac{\sigma^2}{a^2}\mathbb{E}^{\mathbb{Q}}\left[\left(\int_t^T\left(1-e^{-a(T-u)}\right)dW(u)\right)^2\right]\\
&=^{\star\star}\frac{\sigma^2}{a^2}\mathbb{E}^{\mathbb{Q}}\left[\int_t^T\left(1-e^{-a(T-u)}\right)^2du\right]\\
&= \frac{\sigma^2}{a^2}\left(T-t-2\frac{1-e^{-a(T-t)}}{a}+\frac{1-e^{-2a(T-t)}}{2a}\right)\\
&= V(t,T)
\end{aligned}$$

$\star$ by using that $\mathbb{E}^{\mathbb{Q}}\left[\int_t^T\left(1-e^{-a(T-u)}\right)dW(u)\right]=0$.
$\star\star$ by Itô's isometry.

When we fill in what we know we find

$$\begin{aligned}
P(t,T) &= \exp\left\{-\int_t^T\beta(s)ds\right\}\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T x(s)ds}\middle|\mathcal{F}_t\right]\\
&=^{\star}\exp\left\{-\int_t^T\beta(s)ds-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}V(t,T)\right\}\\
&= \exp\left\{-\int_0^T\beta(s)ds+\int_0^t\beta(s)ds-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}V(t,T)\right\}\\
&= \exp\left\{-\int_0^T f^M(0,s)ds-\frac{1}{2}V(0,T)+\int_0^t f^M(0,s)ds+\frac{1}{2}V(0,t)\right\}\\
&\quad\exp\left\{-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}V(t,T)\right\}\\
&= \frac{\exp\left\{\int_0^t f^M(0,s)ds\right\}}{\exp\left\{\int_0^T f^M(0,s)ds\right\}}\exp\left\{-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}\left(V(0,t)+V(t,T)-V(0,T)\right)\right\}\\
&= \frac{P^M(0,t)}{P^M(0,T)}\exp\left\{-x(t)\frac{1-e^{-a(T-t)}}{a}+\frac{1}{2}\left(V(0,t)+V(t,T)-V(0,T)\right)\right\}\\
&= A(t,T)e^{-B(t,T)x(t)}
\end{aligned}$$

$\star$ by using the moment generating function of the normal distribution since we know that $\int_t^T x(s)ds\middle|\mathcal{F}_t\sim\mathcal{N}\left(x(t)\frac{1-e^{-a(T-t)}}{a},V(t,T)\right)$. □

A.2.1. The pricing functions

When looking at the valuation of the MtM of a XCS it can be seen that the value involves the foreign and domestic ZCB. The values of these ZCBs are calculated using the foreign and domestic short rate. In order to be able to combine these short rates a change of measure must be performed, this gives rise to the following proposition.

Proposition 6. *The foreign short-rate process $x_f(t)$ under the domestic risk-neutral measure $\mathbb{Q}^d$ follows the stochastic differential equation,*

$$dx_f(t) = [-a_f x_f(t) + \sigma_f \sigma_X \rho_{fX}]dt + \sigma_f dW_f^{\mathbb{Q}^d}. \tag{A.1}$$

Proof. This proof starts by finding a change of measure that makes it possible to move from the foreign risk-neutral measure to the domestic risk-neutral measure. After we have found such a change of measure it becomes possible to use the Brownian motion defined on the foreign risk-neutral measure, combined with Girsanov's theorem, to find an expression in terms of a Brownian motion defined on the domestic risk-neutral measure.

We have that

$$dx_f(t) = -a_f x_f(t)dt + \sigma_f dW_f^{\mathbb{Q}^f}$$

We know that we can use the foreign and domestic money market accounts, $B^f(t)$ and $B^d(t)$, as a numeraire. Under these numeraires, we know that the payoff of any traded asset $V^f(t)$ in the foreign market and $V^d(t)$ in the domestic market relative to their corresponding numeraire is a martingale. That is,

$$\frac{V^f(t)}{B^f(t)} = \mathbb{E}^{\mathbb{Q}^f}\left[\frac{V^f(T)}{B^f(T)}\middle|\mathcal{F}_t\right]$$
$$\frac{V^d(t)}{B^f(t)} = \mathbb{E}^{\mathbb{Q}^d}\left[\frac{V^d(T)}{B^d(T)}\middle|\mathcal{F}_t\right]$$

$X(t)$ is defined as the exchange rate. It indicates the amount of domestic currency is received per unit of foreign currency. From this, we can write that

$$\begin{aligned}
V^d(t) &= X(t)V^f(t) \\
&= X(t)B^f(t)\mathbb{E}^{\mathbb{Q}^f}\left[\frac{V^f(T)}{B^f(T)}\middle|\mathcal{F}_t\right] \\
&= X(t)B^f(t)\mathbb{E}^{\mathbb{Q}^f}\left[\frac{V^f(T)}{B^f(T)}\frac{X(T)}{X(T)}\middle|\mathcal{F}_t\right] \\
&= B^d(t)\mathbb{E}^{\mathbb{Q}^d}\left[\frac{X(T)V^f(T)}{B^d(T)}\middle|\mathcal{F}_t\right]
\end{aligned}$$

By using this and putting $t = 0$ we can see that

$$X(0)V^f(0) = B^d(0)\mathbb{E}^{\mathbb{Q}^d}\left[\frac{X(T)V^f(T)}{B^d(T)}\middle|\mathcal{F}_0\right]$$
$$X(0)B^f(0)\mathbb{E}^{\mathbb{Q}^f}\left[\frac{V^f(T)}{B^f(T)}\middle|\mathcal{F}_0\right] = B^d(0)\mathbb{E}^{\mathbb{Q}^d}\left[\frac{X(T)V^f(T)}{B^d(T)}\middle|\mathcal{F}_0\right]$$
$$\mathbb{E}^{\mathbb{Q}^f}\left[\frac{X(0)V^f(T)}{B^f(T)}\right] = \mathbb{E}^{\mathbb{Q}^d}\left[\frac{X(T)V^f(T)}{B^d(T)}\right]$$

Where in the last step we used that $B^f(0) = B^d(0) = 1$, and the expecation is independent of $\mathcal{F}_0$.

To be able to move from the expectation with measure $\mathbb{Q}^f$ to the expectation with measure $\mathbb{Q}^d$

we need a change of measure. From the equality above we can write that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}^f}\left[\frac{X(0)V^f(T)}{B^f(T)}\right] &= \int \frac{X(0)V^f(T)}{B^f(T)} d\mathbb{Q}^f \\
&= \int \frac{X(0)V^f(T)}{B^f(T)} \frac{d\mathbb{Q}^f}{d\mathbb{Q}^d} d\mathbb{Q}^d \\
&= \mathbb{E}^{\mathbb{Q}^d}\left[\frac{d\mathbb{Q}^f}{d\mathbb{Q}^d}\frac{X(0)V^f(T)}{B^f(T)}\right] \\
&= \mathbb{E}^{\mathbb{Q}^d}\left[\frac{X(T)V^f(T)}{B^d(T)}\right]
\end{aligned}$$

From this we see that the change of measure must have the form

$$\frac{d\mathbb{Q}^f}{d\mathbb{Q}^d} = \frac{B^f(T)X(T)}{X(0)B^d(T)}$$

To be able to find the expression of the Brownian motion under the domestic risk-neutral measure we must find the expression of the stochastic exponential. This expression is equal to the expression of the change of measure. To find this expression we must fill in the equations for $B^f(T)$, $B^d(T)$, $X(T)$, and $X(0)$ above.

We know that

$$\begin{aligned}
B^f(t) &= e^{\int_0^t r_f(s)ds} \\
B^d(t) &= e^{\int_0^t r_d(s)ds}
\end{aligned}$$

Now we only need to find the expression for $X(t)$ to be able to find the expression of the change of measure.

We know that the $X(t)$ follows the following dynamic

$$dX(t) = \left(r_d(t) - r_f(t)\right) X(t)dt + \sigma_X X(t) dW_X^{\mathbb{Q}^d}$$

Taking $g(X(t)) = \ln(X(t))$ and using Itô's Lemma we get that

$$\begin{aligned}
\partial \ln(X(t)) &= \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial X(t)}dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial X(t)^2}dX(t)dX(t) \\
&= \frac{1}{X(t)}dX(t) - \frac{1}{2}\frac{1}{X(t)^2}dX(t)dX(t) \\
&= \frac{1}{X(t)}\left(\left(r_d(t) - r_f(t)\right)X(t)dt + \sigma_X X(t) dW_X^{\mathbb{Q}^d}\right) \\
&- \frac{1}{2}\frac{1}{X(t)^2}\left(\left(r_d(t) - r_f(t)\right)X(t)dt + \sigma_X X(t) dW_X^{\mathbb{Q}^d}\right)^2 \\
&= (r_d(t) - r_f(t) - \frac{1}{2}\sigma_X^2)dt + \sigma_X dW_X^{\mathbb{Q}^d}
\end{aligned}$$

Taking the integral on both sides results in

$$\begin{aligned}
\ln(X(t)) &= \ln(X_0) + \int_0^t \left(r_d(s) - r_f(s) - \frac{1}{2}\sigma_X^2\right)ds + \int_0^t \sigma_X dW_X^{\mathbb{Q}^d} \\
X(t) &= X(0)\exp\left\{\int_0^t \left(r_d(s) - r_f(s) - \frac{1}{2}\sigma_X^2\right)ds + \int_0^t \sigma_X dW_X^{\mathbb{Q}^d}\right\}
\end{aligned}$$

Filling this into the change of measure results in the expression for the stochastic exponential $\mathcal{E}(L)_t$

$$\begin{aligned} Z_t = \frac{d\mathbb{Q}^f}{d\mathbb{Q}^d} &= \frac{B^f(T)X(T)}{X(0)B^d(T)} \\ &= \exp\left\{\int_0^T \sigma_X dW_X^{\mathbb{Q}^d} - \frac{1}{2}\int_0^T \sigma_X^2 dt\right\} \\ &= \exp\left\{L_t - \frac{1}{2}\langle L\rangle_t\right\} \\ &= \mathcal{E}(L)_t \end{aligned}$$

From these last expressions, we find that $L_t = \int_0^T \sigma_X dW_X^{\mathbb{Q}^d}$.
Then by Girsanov's theorem for Brownian motion, we find that since $W_f^{\mathbb{Q}^d}$ is a Brownian motion under $\mathbb{Q}^d$, then $W_f^{\mathbb{Q}^f}$ is a Brownian motion in $\mathbb{Q}^f$

$$\begin{aligned} W_f^{\mathbb{Q}^f} &= W_f^{\mathbb{Q}^d} + \langle W_f^{\mathbb{Q}^d}, \int_0^T \sigma_X dW_X^{\mathbb{Q}^d}\rangle \\ &= W_f^{\mathbb{Q}^d} + \int_0^T \sigma_X d\langle W_f^{\mathbb{Q}^d}, W_X^{\mathbb{Q}^d}\rangle \\ &= W_f^{\mathbb{Q}^d} + \int_0^T \sigma_X \rho_{fX} dt \end{aligned}$$

From this we find that

$$dW_f^{\mathbb{Q}^f} = dW_f^{\mathbb{Q}^d} + \sigma_X \rho_{fX} dt$$

Then filling in the expression for $dW_f^{\mathbb{Q}^f}$ in the dynamics of $x_f(t)$ yields

$$dx_f(t) = (-a_f x_f(t) + \sigma_f \sigma_X \rho_{fX})dt + \sigma_f dW_f^{\mathbb{Q}^d}$$

□

Using Itô's Lemma on equation A.1 and taking the integral on both sides yields the solution

$$x_f(t) = x_f(0)e^{-a_f t} + \frac{\sigma_f \sigma_X \rho_{fX}}{a_f}(1 - e^{-a_f t}) + \sigma_f \int_0^t e^{-a_f(t-s)} dW_f(s). \tag{A.2}$$

As can be seen in equation 2.31, the dynamics of the FX rate is modelled using the real world measure $\mathbb{P}$. To be able to combine them with the shifted short rates that is modelled under the risk-neutral measure of the domestic market $\mathbb{Q}^d$ we need to change the measure of the FX rate by performing a change of measure. This gives room to the following proposition.

Proposition 7. *Under the risk-neutral measure of the domestic measure $\mathbb{Q}^d$ we have that*

$$dX(t) = (r_d(t) - r_f(t))dt + \sigma_X dW^{\mathbb{Q}}(t) \tag{A.3}$$

Where $r_d(t)$ and $r_f(t)$ are the domestic and foreign short rates at time t, and σ_X is the volatility of the FX rate.

Proof. We have that by choosing the numeraire to be the domestic money-market account $X(t)\frac{B^f(t)}{B^d(t)}$ must be a martingale under the domestic risk-neutral measure. Since we have that $B^f(t) = \exp\left\{-\int_0^t r_f(s)ds\right\}$ and $B^d(t) = \exp\left\{-\int_0^t r_d(s)ds\right\}$ we know that we can write $B^f(t)/B^d(t) = \exp\left\{\int_0^t (r_d - r_f)(s)ds\right\}$.

Using this we find that

$$\begin{aligned} d\left(X(t)\frac{B^f(t)}{B^d(t)}\right) &= d(X(t))\frac{B^f(t)}{B^d(t)} + X(t)d\left(\frac{B^f(t)}{B^d(t)}\right), \\ &= \left(\mu X(t)dt + \sigma_X X(t)dW^{\mathbb{P}}\right)\frac{B^f(t)}{B^d(t)} - (r_d(t) - r_f(t))X(t)\frac{B^f(t)}{B^d(t)}dt, \\ &= \left((\mu - (r_d(t) - r_f(t))dt + \sigma_X dW^{\mathbb{P}}\right)X(t)\frac{B^f(t)}{B^d(t)}. \end{aligned}$$

In order to be a martingale there mustn't be drift. This is only the case if $\mu = r_d(t) - r_f(t)$. Thus, we find that under the risk-neutral domestic measure $\mathbb{Q}^d$,

$$dX(t) = (r_d(t) - r_f(t))X(t)dt + \sigma_X dW^{\mathbb{Q}^d}(t).$$

□

A.3. Proof of parameters of CE-AIS being global maximizers

$$\begin{aligned} &\underset{\mu\in\mathbb{R}^m,\Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max}\ \mathbb{E}_{p_{\theta_0}}\left[\log(g_{\mu,\Sigma}(\mathbf{X}))\cdot\mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right], \\ =&\underset{\mu\in\mathbb{R}^m,\Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max}\int_{-\infty}^{\infty}\log(g_{\mu,\Sigma}(\mathbf{X}))\cdot\mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\ =&\underset{\mu\in\mathbb{R}^m,\Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max}\frac{\mathbb{E}_{p_{\theta_0}}\left[\log\left(\frac{\exp\{-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\}}{\sqrt{(2\pi)^m\det(\Sigma)}}\right)\middle|\mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right]}{\mathbb{P}_{p_{\theta_0}}\left[\max(E_t(\mathbf{X}),0) > q_\alpha\right]}, \\ =&\underset{\mu\in\mathbb{R}^m,\Sigma\in\mathcal{M}^+_{m\times m}}{\arg\max}\ \mathbb{E}_{p_{\theta_0}}\left[\log\left(\frac{\exp\{-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\}}{\sqrt{(2\pi)^m\det(\Sigma)}}\right)\middle|\mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right]. \end{aligned} \tag{A.4}$$

From the latter expression it can be seen that the following expression must be maximized,

$$\mathbb{E}_{p_{\theta_0}}\left[-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu) - \log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)\middle|\mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right].$$

The derivatives of this expression can be taken with respect to μ and Σ to find the values that maximizes

A.4. Using $\Omega = \{\max(E_t(\mathbf{X}), 0) > q_\alpha\}$, it can be seen that

$$
\begin{aligned}
&\frac{\partial}{\partial \mu}\mathbb{E}_{p_{\theta_0}}\left[-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu) - \log\left(\sqrt{(2\pi)^m \det(\Sigma)}\right)\middle| \mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right], \\
=&\frac{\partial}{\partial \mu}\left(\int_\Omega \left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right) p_{\theta_0}(\mathbf{X})d\mathbf{X} - \int_\Omega \left(-\log\left(\sqrt{(2\pi)^m \det(\Sigma)}\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}\right)\right), \\
=&\int_\Omega \frac{\partial}{\partial \mu}\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right) p_{\theta_0}(\mathbf{X})d\mathbf{X} - \int_\Omega \frac{\partial}{\partial \mu}\left(-\log\left(\sqrt{(2\pi)^m \det(\Sigma)}\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}\right), \\
=&\int_\Omega \frac{\partial}{\partial \mu}\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
=&-\frac{1}{2}\int_\Omega \frac{\partial}{\partial \mu}\left(\mathbf{X}^\top\Sigma^{-1}\mathbf{X} - \mathbf{X}^\top\Sigma^{-1}\mu - \mu^\top\Sigma^{-1}\mathbf{X} + \mu^\top\Sigma^{-1}\mu\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
=&-\frac{1}{2}\int_\Omega \left(\frac{\partial}{\partial \mu}\left(\mathbf{X}^\top\Sigma^{-1}\mathbf{X}\right) - \frac{\partial}{\partial \mu}\left(\mathbf{X}^\top\Sigma^{-1}\mu\right) - \frac{\partial}{\partial \mu}\left(\mu^\top\Sigma^{-1}\mathbf{X}\right) + \frac{\partial}{\partial \mu}\left(\mu^\top\Sigma^{-1}\mu\right)\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
=&\int_\Omega \left(\frac{1}{2}\Sigma^{-1}\mathbf{X} + \frac{1}{2}\mathbf{X}\Sigma^{-1} - \Sigma^{-1}\mu\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
=&\frac{1}{2}\Sigma^{-1}\int_\Omega \mathbf{X}p_{\theta_0}(\mathbf{X})d\mathbf{X} + \frac{1}{2}\int_\Omega \mathbf{X}p_{\theta_0}(\mathbf{X})d\mathbf{X}\Sigma^{-1} - \Sigma^{-1}\int_\Omega \mu p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
=&\frac{1}{2}\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right] + \frac{1}{2}\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right]\Sigma^{-1} \\
&-\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mu\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right], \\
=&\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right] - \Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mu\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right]
\end{aligned}
$$

Solving the last expression it follows that

$$
\begin{aligned}
&\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right] - \Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mu\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right] = 0, \\
&\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right] - \Sigma^{-1}\mu = 0
\end{aligned}
$$

From which it follows that

$$
\mu = \mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right]. \tag{A.5}
$$

To verify that this μ will indeed result in finding the maximum of the equation A.4 the second order derivative will be computed.

$$
\begin{aligned}
&\frac{\partial^2}{\partial \mu^2}\mathbb{E}_{p_{\theta_0}}\left[-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu) - \log\left(\sqrt{(2\pi)^m \det(\Sigma)}\right)\middle| \mathbf{1}\{\max(E_t(\mathbf{X}),0) > q_\alpha\}\right], \\
&= \frac{\partial}{\partial \mu}\left(\int_\Omega \left(-\frac{1}{2}\Sigma^{-1}\mathbf{X} - \frac{1}{2}\mathbf{X}\Sigma^{-1} - \Sigma^{-1}\mu\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}\right), \\
&= \int_\Omega \frac{\partial}{\partial \mu}\left(-\frac{1}{2}\Sigma^{-1}\mathbf{X} - \frac{1}{2}\mathbf{X}\Sigma^{-1} - \Sigma^{-1}\mu\right) p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
&= \int_\Omega -\Sigma^{-1} p_{\theta_0}(\mathbf{X})d\mathbf{X}, \\
&= -\Sigma^{-1}\int_\Omega p_{\theta_0}(\mathbf{X})d\mathbf{X}.
\end{aligned}
$$

Because $\int_\Omega p_{\theta_0}(\mathbf{X})d\mathbf{X}$ is a positive constant Σ^{-1} is still an SPD matrix as the covariance matrix Σ is SPD by definition. From this it follows that $-\Sigma^{-1}\int_\Omega p_{\theta_0}(\mathbf{X})d\mathbf{X}$ is negative definite. Now it can be concluded that since the Hessian matrix, or second derivative, is negative definite for every μ the expectation is concave, therefore the solution found gives the global maximum. Thus, the value of μ in A.5 indeed maximizes equation A.4.

Next, the same will be done for Σ.

$$
\begin{aligned}
&\frac{\partial}{\partial \Sigma}\mathbb{E}_{p_{\theta_0}}\left[-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)-\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)\middle|\mathbf{1}\{\max(E_t(\mathbf{X}),0)>q_\alpha\}\right],\\
=&\frac{\partial}{\partial \Sigma}\left(\int_\Omega\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right)d\mathbf{X}-\int_\Omega\left(\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)p_{\theta_0}(\mathbf{X})d\mathbf{X}\right)\right),\\
=&\int_\Omega\frac{\partial}{\partial \Sigma}\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right)d\mathbf{X}-\int_\Omega\frac{\partial}{\partial \Sigma}\left(\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)p_{\theta_0}(\mathbf{X})d\mathbf{X}\right),\\
=&\int_\Omega\left(-\frac{1}{2}\left(-\Sigma^{-\top}\mathbf{X}^\top\mathbf{X}\Sigma^{-\top}+\Sigma^{-\top}\mathbf{X}^\top\mu\Sigma^{-\top}+\Sigma^{-\top}\mu^\top\mathbf{X}\Sigma^{-\top}-\Sigma^{-\top}\mu^\top\mu\Sigma^{-\top}\right)p_{\theta_0}(\mathbf{X}d\mathbf{X})\right)\\
&-\int_\Omega\frac{\partial}{\partial \Sigma}\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)p_{\theta_0}(\mathbf{X})d\mathbf{X},\\
=&\frac{1}{2}\Sigma^{-\top}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-\top}\\
&-\mathbb{E}_{p_{\theta_0}}\left[\frac{\sqrt{(2\pi)^k\det(\Sigma)}\frac{\partial}{\partial\Sigma}\det(\Sigma)}{2\sqrt{(2\pi)^k\det(\Sigma)}}\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right],\\
=&\frac{1}{2}\Sigma^{-\top}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-\top}-\mathbb{E}_{p_{\theta_0}}\left[\frac{1}{2}\Sigma^{-1}\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right].
\end{aligned}
$$

Solving the last expression shows,

$$
\begin{aligned}
&\frac{1}{2}\Sigma^{-\top}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-\top}-\frac{1}{2}\Sigma^{-1}=0,\\
&\Sigma^{-1}=\Sigma^{-\top}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-\top}
\end{aligned}
$$

From which, by using the fact that $\Sigma^{-\top}=\Sigma^{-1}$, it follows that

$$
\begin{aligned}
\Sigma\Sigma^{-1}\Sigma&=\Sigma\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-1}\Sigma,\\
\Sigma&=\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right].
\end{aligned}
\tag{A.6}
$$

Again, it is verified that this value gives the maximum by looking at the second order derivative.

$$
\begin{aligned}
&\frac{\partial^2}{\partial\Sigma^2}\mathbb{E}_{p_{\theta_0}}\left[-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)-\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)\middle|\mathbf{1}\{\max(E_t(\mathbf{X}),0)>q_\alpha\}\right],\\
=&\frac{\partial^2}{\partial\Sigma^2}\left(\int_\Omega\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right)d\mathbf{X}-\int_\Omega\left(\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)p_{\theta_0}(\mathbf{X})d\mathbf{X}\right)\right),\\
=&\int_\Omega\frac{\partial^2}{\partial\Sigma^2}\left(-\frac{1}{2}(\mathbf{X}-\mu)^\top\Sigma^{-1}(\mathbf{X}-\mu)\right)d\mathbf{X}-\int_\Omega\frac{\partial^2}{\partial\Sigma^2}\left(\log\left(\sqrt{(2\pi)^k\det(\Sigma)}\right)p_{\theta_0}(\mathbf{X})d\mathbf{X}\right),\\
=&\frac{1}{2}\frac{\partial}{\partial\Sigma}\left(\Sigma^{-1}\right)\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-1}\\
&+\frac{1}{2}\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\frac{\partial}{\partial\Sigma}\left(\Sigma^{-1}\right)-\frac{1}{2}\frac{\partial}{\partial\Sigma}\Sigma^{-1},\\
=&-\frac{1}{2}\Sigma^{-2}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-1}\\
&-\frac{1}{2}\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0)>q_\alpha\right]\Sigma^{-2}+\frac{1}{2}\Sigma^{-2}.
\end{aligned}
$$

To find the expression for $\frac{\partial}{\partial\Sigma}\Sigma^{-1}$ it was used that, $\partial I=\partial(\Sigma\Sigma^{-1})=\partial\Sigma\Sigma^{-1}+\Sigma\partial\Sigma^{-1}$ from this it can be seen that as $\partial I=0$ it follows that $\partial\Sigma^{-1}/\partial\Sigma=-\Sigma^{-1}\Sigma^{-1}=-\Sigma^{-2}$.

Then by using that $\Sigma = \mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right]$ it can be seen that

$$\begin{aligned}
&-\frac{1}{2}\Sigma^{-2}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right]\Sigma^{-1}\\
&-\frac{1}{2}\Sigma^{-1}\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right]\Sigma^{-2}+\frac{1}{2}\Sigma^{-2},\\
=&-\frac{1}{2}\Sigma^{-2}\Sigma\Sigma^{-1}\\
&-\frac{1}{2}\Sigma^{-1}\Sigma\Sigma^{-2}+\frac{1}{2}\Sigma^{-2},\\
=&-\Sigma^{-2}+\frac{1}{2}\Sigma^{-2},\\
=&-\frac{1}{2}\Sigma^{-2}.
\end{aligned}$$

Then as the square of the inverse of an SPD matrix is again SPD it can be concluded that $-\frac{1}{2}\Sigma^{-2}$ is negative definite. Thus, as the Hessian is negative definite for all Σ it follows that the expectation is concave for Σ. From this it follows that the value of Σ in A.6 is the global maximizer of the problem. Thus the parameters that globally maximize A.4 are

$$\begin{aligned}
\mu^\star &= \mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(E_t(\mathbf{X}),0) > q_\alpha\right],\\
\Sigma^\star &= \mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\mu)^\top(\mathbf{X}-\mu)\middle|\max(E_t(\mathbf{X}),0) > q_\alpha\right].
\end{aligned}$$

A.4. Using the analytical solution to find the CE-AIS-COS parameters

The results presented in the previous sections were from the CE-AIS-COS variants, where the parameters $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ are approximated using Monte Carlo. Instead of using Monte Carlo, the Clenshaw-Curtis quadrature can be used to find the values of $\boldsymbol{\mu}^\star$ and $\Sigma^\star$. Doing this could make the iterative loop inside the CE-AIS-COS method obsolete. Below we describe how to apply the Clenshaw-Curtis quadrature in finding those parameters.

We start from the definitions of those parameters to estimate:

$$\begin{aligned}
\mu^\star &= \mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}|\max(V(\mathbf{X}),0) > q_\alpha\right],\\
&= \frac{\mathbb{E}_{p_{\theta_0}}\left[\mathbf{X}\cap\{\max(V(\mathbf{X}),0) > q_\alpha\}\right]}{\mathbb{P}_{p_{\theta_0}}\left[\max(V(\mathbf{X}),0) > q_\alpha\}\right]},\\
&= \frac{\int_{\mathbb{R}^m}\mathbf{x}\cdot\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}\cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}}{\int_{\mathbb{R}^m}\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}p_{\theta_0}(\mathbf{x})d\mathbf{x}},
\end{aligned}$$

and

$$\begin{aligned}
\Sigma^\star &= \frac{\mathbb{E}_{p_{\theta_0}}\left[(\mathbf{X}-\boldsymbol{\mu}^\star)(\mathbf{X}-\boldsymbol{\mu}^\star)\cap\{\max(V(\mathbf{X}),0) > q_\alpha\}\right]}{\mathbb{P}_{p_{\theta_0}}\left[\max(V(\mathbf{X}),0) > q_\alpha\}\right]},\\
&= \frac{\int_{\mathbb{R}^m}(\mathbf{x}-\boldsymbol{\mu}^\star)(\mathbf{x}-\boldsymbol{\mu}^\star)\cdot\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}\cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}}{\int_{\mathbb{R}^m}\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}p_{\theta_0}(\mathbf{x})d\mathbf{x}},
\end{aligned}$$

Then the integrals $\int_{\mathbb{R}^m}(\mathbf{x}-\boldsymbol{\mu}^\star)(\mathbf{x}-\boldsymbol{\mu}^\star)\cdot\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}\cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}$ and $\int_{\mathbb{R}^m}(\mathbf{x}-\boldsymbol{\mu}^\star)(\mathbf{x}-\boldsymbol{\mu}^\star)\cdot\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}\cdot p_{\theta_0}(\mathbf{x})d\mathbf{x}$ are calculated using the Clenshaw-Curtis quadrature. In the above fractions we define

$$p = \mathbb{P}_{p_{\theta_0}}\left[\mathbf{1}\{\max(V(\mathbf{X}),0) > q_\alpha\}\right] = \int_{\mathbb{R}^m}\mathbf{1}\{\max(V(\mathbf{x}),0) > q_\alpha\}p_{\theta_0}(\mathbf{x})d\mathbf{x}.$$

As q_α is the 0.975-quantile it is already known that $p = 0.025$. In this way, if the quadrature's accuracy is sufficient the values $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ can be calculated and used in the CE-AIS-COS-exact. Table A.1 shows the

values p calculated using the Clenshaw-Curtis quadrature with 64 expansion terms and 30 quadrature points. These values were calculated using a test portfolio containing 100 derivatives without collateral on even time points.

t	2	4	6	8	10	12	14	16	18
p	0.02494	0.02602	0.02466	0.02567	0.02415	0.02353	0.03821	0.03829	0.03825

Table A.1: The values of $\hat{p}$ on even time points for a uncollateralized portfolio containing 100 derivatives.

As can be seen in the table, the analytical solutions p are very close to the actual 0.025. Sequentially, by using these values as an indication of the accuracy, it was found that using the analytical solution for the values $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ in the CE-AIS-COS method gave good results for the PFE approximations.

For each of the time points in Table A.1 the PFE was calculated 100 times using 25000 paths. The variance of these PFE values, calculated using the analytical solutions, were compared to the variances of the PFE approximations using the CE-AIS-COS method. From this comparison it was found that for approximately half of the time points the variances of the PFE calculated this approach were approximately the same as the CE-AIS-COS method we developed in Section 6.2. For other time points the variances of the CE-AIS method using the Monte Carlo approach were two to six times lower than those found using the analytical solution.

Next, the same check was done for a collateralized portfolio. Table A.2 shows the values of p calculated using 32 expansion terms and 30 quadrature points for the Clenshaw-Curtis quadrature.

t	2	4	6	8	10	12	14	16	18
p	0.0	0.01742	0.01708	0.01147	0.02114	0.01404	0.03819	0.03789	0.01426

Table A.2: The values of p on even time points for a collateralized portfolio using 100 derivatives.

It can be seen that the values of p for the collateralized portfolio are much further away from the real value of p, compared to the uncollateralized case. These values indicate that the errors from the integrals in the numerators of Equation **??** and **??** become dominating. This was verified by comparing the variance of 100 PFE calculations using the analytical solution to that using the CE-AIS-COS method as described in Section 6.2.

These results suggest that using the parameters found by the analytical solution gave a much higher variance than that of the CE-AIS-COS method from Section 6.2. In some cases the variance was more than 200 times higher. To find a better approximation of the solutions of $\boldsymbol{\mu}^\star$ and $\Sigma^\star$ more expansion terms and quadrature points have to be used. It would result in a higher CPU time, thus not efficient.