EFFICIENT MONTE CARLO ESTIMATION OF CREDIT CONCENTRATION RISK

Matteo Barbagli, Frédéric Vrins

LIDAM Discussion Paper LFIN
2025 / 03

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications

Efficient Monte Carlo estimation of credit concentration risk

Matteo Barbagli[a,*], Frédéric Vrins[a,b]

[a]*LFIN/LIDAM, UCLouvain, 1348 Louvain-la-Neuve, Belgium*
[b]*HEC Montréal, Department of Decision Sciences, H3T 2A7 Montréal, Canada*

Abstract

In this paper we address the explicit exclusion of credit concentration risk from the Pillar 1 minimum capital requirements formulas of the Basel framework. Leveraging on a well established Gaussian multi-factor model, we introduce a novel control variate estimator of value-at-risk (VaR), suitable for measuring sector concentration risk under the Pillar 2 guidelines. This estimator integrates the precision of Monte Carlo simulations with the speed and simplicity of the Large Pool approximation, aiming for a more efficient quantile estimation tool. We conduct numerical experiments in a two systematic factor setup to test the validity of our methodology, achieving consistent variance reduction compared to the benchmark Monte Carlo estimator. Our results are robust across various pool parameters and increasing number of Monte Carlo simulations.

Keywords: Credit risk; Factor model; Control variate; Value-at-risk; Basel regulation

JEL classification: G21: G28; G32

*Corresponding author at: Université catholique de Louvain, Place des Doyens 1, Boite L2.01.01, Office b 208, 1348 Louvain-la-Neuve, Belgium. Tel.: +32 (0)10 478 449.

Email addresses: `matteo.barbagli@uclouvain.be` (Matteo Barbagli), `frederic.vrins@uclouvain.be` (Frédéric Vrins)

1. Introduction

Concentration risk is defined as the risk arising from the concentration of a single exposure or group of exposures with the potential to threaten the bank's solvency or ability to maintain its core operations (see paragraph 30.20 of BCBS, 2023g). Within a bank's assets, concentration risk can take many forms, arising from obligors concentration both in the banking book (credit portfolio) or in the trading book (those assets held for trading). Additionally, concentration risk can arise from credit risk mitigation techniques, such as exposures in similar collateral instruments or closely related protection sellers, or stem from concentration of exposures in a specific currency. It can result from a lack of diversification in terms of obligors and/or counterparties, a concentration of exposures in too few sectors, countries, or currencies. As lending remains the primary activity of most banks, credit concentration risk (hereafter referred to as concentration risk) is recognised as the most material type of concentration risk (BCBS, 2023g).

The Basel Committee on Banking Supervision (BCBS) distinguishes two sorts of credit concentration risk. The first type, called *name concentration*, features a credit pool where the total exposure at default (EAD) is mainly driven by a set of exposures to a single counterparty, or to a set of exposures to economically interdependent counterparties. This phenomenon is usually occurring in banks holding a small banking book, featuring a relatively small number of exposures with large EADs due to their size or specialisation degree. Nevertheless, this has historically been the case also for global systemically important banks (G-SIBs), who usually display a more diversified portfolio (BCBS, 2021). The second type, called *sector concentration*, refers to a credit pool featuring a significant share of the total EAD related to exposures in the same geographic area or economic sector.

To mitigate credit risk, the Basel Committee requires banks and financial institutions to comply with a series of minimum capital requirements (MCR) and thus set aside an amount of economic capital, i.e., an amount of capital needed to absorb credit losses at a given confidence level (see Gordy, 2003 and BCBS, 2009). Since the release of the final version of Basel II in 2006, the majority of G-SIBs compute the MCR by relying on the internal ratings-based (IRB) approach (BCBS, 2006a). The latter, detailed in the Pillar 1 section of the Accords, defines the MCR as the capital needed to absorb the unexpected loss (UL) at a 99.9% confidence level, that is, the difference between the 99.9% value-at-risk (VaR) and the expected value of the portfolio losses (see BCBS, 2009, p. 9 and Section 2 of BCBS, 2005). In the IRB approach, the VaR is computed under a specific version of the asymptotic single risk factor (ASRF) credit portfolio model developed by Gordy (2003).[1]

The ASRF model relies on two key assumptions: (i) the portfolio is infinitely granular, so that when the portfolio size grows to infinity each individual exposure at default (EAD) accounts for a smaller share of the total portfolio EAD, hence idiosyncratic risk is perfectly

[1] Any loan, bond or debt instrument subject to credit risk in the banking book (BCBS, 2000).

diversified away and (ii) there is a single systematic risk factor driving the dependency between the individual obligor losses L_i. The practical benefit of those assumptions is twofold: first, a closed-form expression is obtained for the MCR; second, the model exhibits the so-called *portfolio-invariance* property asymptotically, which means that the total portfolio VaR is the sum of individual marginal contributions that solely depend on the risk-profile of each borrower. The price to pay, however, is that the ASRF model fails to capture concentration risk. For instance, *name concentration* implies imperfect granularity as the default likelihood of the portfolio would then mainly be driven by the idiosyncratic risk of a single counterparty or group of interconnected obligors. Similarly, *sector concentration* relates to imperfect diversification of the credit pool across economic sectors or regions. In contrast with idiosyncratic risk which can be mitigated by increasing the size of the pool, the specific risks associated with the various sectors or regions remain, calling for a dependency structure governed by multiple systematic factors. Consequently, computing the MCR according to the ASRF model underlying the IRB approach triggers the risk to overestimate the diversification level achieved by the credit portfolio, hence, to underestimate the unexpected loss and the associated capital requirements (BCBS, 2006b, p. 5).

The importance of credit concentration risk is well acknowledged in the Basel documents (see paragraphs 31.6 and 32.7 of Pillar 2 regulations of BCBS, 2023g): (i) 'Risk concentrations are arguably the single most important cause of major problems in banks', and (ii) 'Because lending is the primary activity of most banks, credit risk concentrations are often the most material risk concentrations within a bank'. Nonetheless, the Committee decided that the IRB formulas, meant to deal with credit risk on the first place, would explicitly not address this credit risk subcategory: 'Such concentrations are not addressed in the Pillar 1 capital charge for credit risk' (BCBS, 2023g, §32.8). As highlighted by a research task force mandated by the Committee in 2006, this choice was mainly guided by a willingness to preserve simplicity of the capital computations, specifically to guarantee its most appealing regulatory feature: 'The desire for portfolio invariance, however, makes recognition of institution-specific diversification effects within the framework difficult' (BCBS, 2006b). Using data on corporate exposures, this research group unveiled that amongst the two potential types of credit concentration risk, sector concentration is likely to represent a larger marginal contribution to economic capital than name concentration for a typical commercial bank with a medium to large size banking book.

The Committee's desired *simplicity*, *comparability* and *risk-sensitivity* of the regulatory capital computations (BCBS, 2014) still relegates concentration risk to be dealt under the scope of Pillar 2. Hence, differently from Pillar 1 requirements, there is no specific credit risk model or formulas assigned to quantify this risk dimension in terms of economic capital. National competent authorities must conduct an individual assessment of capital adequacy to concentration risk according to the guidelines found in paragraphs 32.6 till 32.12 of BCBS (2023g). Given the recent world COVID-19 pandemic, the restrictions on international trade and linked negative economic outlook, there is little willingness from regulators to modify the IRB formulas. Nevertheless, a dedicated multi-factor credit risk model is well suited as a stress testing or back testing tool for national regulators and supervisors, completing the general guidelines found in Pillar 2: ‘Banks should employ a number of techniques, as appropriate, to measure risk concentrations. [...] and the use of integrated stress testing and economic capital models’ (BCBS, 2023g, §30.26). In order to avoid pitfalls in the design of a framework for measuring sector concentration risk in the banking book, we design our extended credit VaR model around the one underlying the IRB approach. Hence, we comply with the Committees guidelines prescribing that any such tools should be consistent and plausible with the underlying economic capital model (BCBS, 2018).

Our starting point is seminal work of Pykhtin (2004), which proposes a general multi-factor extension fitting the Gaussian one-factor model of the Basel ASRF. Most importantly, the author introduces an analytical approximation of the multi-factor VaR based on a linear component built upon a comparable ASRF model and two non-linear adjustment terms. The first one reflects the existence of a multi-factor structure, i.e., an unbalanced distribution of the risk across sectors (buckets). The second one reflects, in such a multi-factor structure, the difference between the VaR in a finite *vs.* infinite pool. Düllmann and Masschelein (2007) explore a simplified version of Pykhtin (2004)'s model, where the risk parameters (PD, EAD, LGD and ρ controlling the default dependence) are only sector dependent, resulting in the correlation matrix between the obligors having a block structure. We propose a model where similarly to Düllmann and Masschelein (2007) the correlation matrix has a block structure, but where the obligors are free to feature their own PD, LGD and EAD. Naturally, the quality of Pykhtin (2004)'s approximation varies depending on the characteristics of the credit pool at hand. Even if the latter works generally better for a heterogeneous than an homogeneous pool, the performance is observed to decrease in three cases: (i) when the credit portfolio is made of sub-pools that may represent different industrial sectors or geographic regions (buckets), (ii) when the inter-sector (inter-buckets) correlations are low, and (iii) when the total portfolio EAD is not equally partitioned across the obligors.

From the above, it is clear that there are setups in which a bank still needs to have an idea of the gap between the approximate VaR and the benchmark Monte Carlo (MC) estimate. This is especially true for smaller and more specialised banks applying to national competent authorities for approval to use the IRB approach. Nevertheless, the latter results in computationally intensive exercises that require considerable amount of resources to implement. Therefore, our objective is to address this issue by proposing a control variate estimator blending the power of Pykhtin (2004)'s analytical approximation with the precision of the MC estimator. The goal is reduce the variance of the benchmark MC estimator, reducing the computation time while getting a precise VaR estimate.

In this paper we make the following contributions to the literature on economic capital models. First, following Pykhtin (2004)'s idea, we propose an alternative route to find the factor loadings of the comparable one-factor model. We achieve the latter by exploiting properties derived from the Gaussian conditional distribution as found in Vrins (2018). The latter circumvents the need to specify intermediate factor models featuring interdependent idiosyncratic terms. Second, we propose a methodology to build a control variate estimator of the multi-factor VaR, leveraging on the comparable ASRF quantile. Third, we propose a methodology to reduce the bias of the proposed control variate estimator using the granularity adjustment formula of Gordy and Lutkebohmert (2013). Fourth, through a series of numerical experiments, we benchmark the performance of our control variate estimator against a series of alternative quantiles. Our control variate estimator delivers a consistent variance reduction and increased accuracy compared to the benchmark MC estimator.

The above contributes to the efforts of strengthening methodologies used to stress-test capital requirements under the supervisory review process featured in Pillar 2: 'A bank should ensure that it has sufficient capital to meet the Pillar 1 requirements and the results (where a deficiency has been indicated) of the credit risk stress test performed as part of the Pillar 1 internal ratings-based (IRB) minimum requirements' (BCBS, 2023g, 32.1). At last, it is important to note that the Basel Committee is not new to the use of multi-factor models with latent systematic random variables for measuring credit risk (Laurent et al., 2016). Following the completion of the fundamental review of the trading book in 2019 (BCBS, 2019) a two factor model is prescribed for measuring credit risk in the trading book under internal models approach (IMA)(BCBS, 2023e). Precisely, for credit and equity positions a default risk charge (DRC), computed under a VaR framework, is set to cover against: '[...] the risk of direct loss due to an obligor's default, as well as the potential for indirect losses that may arise from a default event' (BCBS, 2023f, §33.18-33.39).

This paper is structured as follows. Section 2 presents the rationale behind the MCR found in the Basel regulations, introduces the ASRF model and details its link with the formulas to compute those same MCR under the IRB approach. Section 3 presents the considered Gaussian multi-factor structure to capture credit concentration risk in the banking book. Section 4 presents Pykhtin (2004)'s methodology to find an analytical approximation of the multi-factor VaR. In Section 5, we introduce our variance reduction technique based on a control variate estimator. The empirical analysis on two synthetic credit portfolios is conducted in Section 6. Section 7 concludes.

2. Capital requirements in Basel regulation

Under the IRB approach, capital requirements on the banking book are defined through a one-year VaR paradigm. For a given credit portfolio, it is defined as the amount of capital a financial institution needs to absorb the unexpected losses (UL) on the portfolio over a one-year horizon, at confidence level α:

$$K_\alpha := \mathbb{Q}_\alpha[L] - \mathbb{E}[L] \,. \tag{1}$$

In this expression, L stands for the random variable representing the total loss on the portfolio, $\mathbb{E}[L]$ is the expected loss (EL) on the pool and $\mathbb{Q}_\alpha[L] = \mathrm{VaR}_\alpha(L)$ defines the portfolio loss quantile, that is, the VaR, at confidence level α.

The computation of the EL is relatively straightforward due to the linearity of the expectation operator, it is just the sum of the EL on each constituent of the portfolio. The computation of the quantile, however, requires to assume a dependence structure between the individual losses. The formulas associated with the IRB approach rely on an ASRF model postulating a one-factor Gaussian copula governing the dependence of the defaults. Since Basel II (BCBS, 2006a), the latter idea has remained at the core of the RWA formulas found under the IRB approach, translating (1) into specific MCR for a series of credit portfolio asset classes defined in the Accords (BCBS, 2017, 2023a,d).

2.1. Asymptotic portfolio loss model

In the asymptotic portfolio loss model proposed by Gordy (2003), we consider a portfolio made of I credit exposures where we assume that exposures to each obligor i have been aggregated, so that there is a single exposure per obligor.[2] We note L_i the loss resulting from the default of obligor i and w_i the weight of exposure i, defined as the EAD of exposure i relative to the total EAD of the portfolio. The default indicator of obligor i is B_i. In case of default, the loss severity on exposure i is modelled as a random variable Λ_i with support in $[0, 1]$. Hence, the credit loss on the portfolio relative to the total EAD is modelled as

$$L = \sum_{i=1}^{I} w_i L_i \quad \text{where} \quad L_i := \Lambda_i \times B_i \,. \tag{2}$$

In this expression, $B_i \sim Ber(p_i)$ is a Bernoulli random variable whose parameter p_i corresponds to the obligors' unconditional default probability and Λ_i is the corresponding LGD of exposure i.

The expected loss on the pool only depends on the expectations of the individual losses,

$$\mathbb{E}[L] = \sum_{i=1}^{I} w_i \, \mathbb{E}[L_i] \,.$$

The computation of $\mathbb{Q}_\alpha[L]$ is more involved and depends on the joint distribution of the L_is. Precisely, we assume that all dependencies across credit events are driven by a set of S Gaussian risk factors $\boldsymbol{Y} = (Y_1, ..., Y_S)$, where $1 \leq S \leq I$. Gordy (2003) shows that, in this general framework, a closed form asymptotic approximation for $\mathbb{Q}_\alpha[L]$ can be found based on two core assumptions: (i) the portfolio is 'asymptotically fine-grained', meaning that $\sum_i \mathrm{EAD}_i = \infty$ and $\sum_i w_i^2 < \infty$ as $I \to \infty$ and (ii) the L_is are independent conditional upon the factors $\boldsymbol{Y}$. Under these two conditions, it follows from the strong law of large numbers that the idiosyncratic risks Z_i's can be fully diversified away, and that the conditional

[2] In the Basel Accords, credit exposures are defined as any types of loans, bonds, or debt instruments subject to credit risk in the banking book (BCBS, 2000).

distribution of L, knowing $\boldsymbol{Y}$, degenerates to a limiting loss distribution equal its conditional expectation, as $I \to \infty$:

$$L \xrightarrow{I \to \infty} L^{\infty}(\boldsymbol{Y}) := \mathbb{E}[L|\boldsymbol{Y}] = \sum_{i=1}^{I} w_i \, \mathbb{E}[L_i|\boldsymbol{Y}] \,, \tag{3}$$

this result is known as the *large pool* (LP) approximation.

From Proposition 2 in Gordy (2003), we also know that

$$\mathbb{Q}_\alpha[L] \xrightarrow{I \to \infty} \mathbb{Q}_\alpha[L^{\infty}(\boldsymbol{Y})] = \mathbb{Q}_\alpha\left[\sum_{i=1}^{I} w_i \, \mathbb{E}[L_i|\boldsymbol{Y}]\right] \tag{4}$$

where the above quantile can only be computed numerically.

Additionally, if $\boldsymbol{Y}$ is one-dimensional the replacement of L by $\mathbb{E}[L|Y]$ is known as the ASRF model. Under this single-factor structure, thanks to Proposition 4 in Gordy (2003), we know that if $\mathbb{E}[L|Y]$ is decreasing in Y, its quantile takes a simple and desirable form

$$\mathbb{Q}_\alpha[L^{\infty}(Y)] = L^{\infty}(\mathbb{Q}_{1-\alpha}[Y]) = L^{\infty}(-z_\alpha) \tag{5}$$

where we used the fact that $Y \sim \mathcal{N}(0,1)$ and $\mathbb{Q}_{1-\alpha}[Y] = -z_\alpha$, where $z_\alpha := \Phi^{-1}(\alpha)$, the α-quantile of the standard normal variable.

2.2. Gaussian one-factor model

The Basel ASRF, underlying the IRB approach, relies on a specific Gaussian one-factor model, first introduced by Vasicek (1991) and Vasicek (2002) for homogeneous portfolios and later extended by Gordy (2003) to fit heterogeneous portfolios. Both factor model specifications are based on the seminal work of Merton (1974), i.e., default occurs when a latent random variable X_i describing the financial well-being of obligor i falls below a certain threshold δ_i at a fixed horizon (one year in the Basel ASRF). The default events are modelled using latent creditworthiness variables $X_1, \ldots, X_I$ defined as:

$$B_i := \mathbb{1}_{\{X_i \le \delta_i\}} \quad \text{where} \quad X_i := \sqrt{\rho_i}\, Y + \sqrt{1-\rho_i}\varepsilon_i \tag{6}$$

where $\mathbb{1}_{\{A\}}$ denotes the indicator function (taking the value of 1 if A is true and 0 otherwise) and $Y, \varepsilon_1, \ldots, \varepsilon_I$ are independent standard normal random variables. The dependence between the default events results from the common systematic risk factor Y and is controlled by the factor loadings $\sqrt{\rho_i}$. In Merton inspired default models, the X_i's can be understood as the standardised asset return of each individual obligor i. This is the reason why in the IRB approach $\sqrt{\rho_i}$ is referred to as the parameter controlling the asset correlation (see section 4.5 of BCBS, 2005). The asset correlation between any pair of Gaussian latent random variables reads

$$\mathbb{C}\text{orr}(X_i, X_j) = \sqrt{\rho_i \rho_j} \quad \text{for all } i \ne j \,. \tag{7}$$

Letting $\Phi(\cdot)$ denote the cumulative distribution function (cdf) of the standard normal variable, setting the default threshold to $\delta_i := \Phi^{-1}(p_i)$ ensures that

$$\mathbb{P}(B_i = 1) = \mathbb{P}(X_i \leq \delta_i) = \Phi(\delta_i) = p_i \ .$$

Following Tarashev and Zhu (2008), we focus on a Basel ASRF version in which the LGD is a latent random variable independent from both the systematic and idiosyncratic risk factors driving the default event, so that the expectation and variance of the LGDs are defined as

$$\lambda_i := \ \mathbb{E}[\Lambda_i] \quad \text{and} \quad \sigma_i^2 := \mathbb{V}[\Lambda_i] \ .$$

With these notations at hand, the EL on the credit portfolio can be decomposed as follows:

$$\mathbb{E}[L] = \sum_{i=1}^{I} w_i \, \mathbb{E}[L_i] = \sum_{i=1}^{I} w_i \, \mathbb{E}[\Lambda_i]\mathbb{P}(B_i = 1) = \sum_{i=1}^{I} w_i \lambda_i p_i \quad . \tag{8}$$

The independence between the default indicator and the LGD yields

$$\mathbb{E}[L_i|Y] = \mathbb{E}[\Lambda_i \times B_i|Y] = \lambda_i \times \mathbb{E}[B_i|Y] \tag{9}$$

and the conditional expectation of B_i is equal to the conditional default probability of obligor i,

$$\mathbb{E}[B_i|Y] = \mathbb{P}(B_i = 1|Y) = \mathbb{P}(X_i \leq \delta_i|Y) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}Y}{\sqrt{1-\rho_i}}\right) =: p_i(Y) \ . \tag{10}$$

In this particular case, thanks to (3) the limit loss takes the form

$$L^{\infty}(Y) = \sum_{i=1}^{I} w_i \lambda_i p_i(Y) \tag{11}$$

and thanks to (5), its quantile collapses to

$$\mathbb{Q}_{\alpha}[L^{\infty}(Y)] = L^{\infty}(-z_{\alpha}) = \sum_{i=1}^{I} w_i \lambda_i p_i(-z_{\alpha}) \ . \tag{12}$$

It is precisely the linearity of the expectation operator in (12) that guarantees portfolio-invariance: the quantile of L is computed asymptotically as a simple sum of terms associated with the individual exposures, being the product of three elements. As a result, under the LP approximation, $L \to L^{\infty}(Y)$ and the UL on the pool can be decomposed as the sum of the unexpected loss on each constituent. Using the distribution of Y,

$$K_{\alpha} \xrightarrow{I \to \infty} \mathbb{Q}_{\alpha}[L^{\infty}(Y)] - \mathbb{E}[L^{\infty}(Y)] = \sum_{i=1}^{I} w_i \big(\lambda_i p_i(-z_{\alpha}) - \lambda_i p_i\big) \ . \tag{13}$$

It is important to note that, thanks to (3), (11) remains a valid approximation of the distribution of L also when B_i is driven by multiple systematic risk factors. However, going from a single to a multi-factor structure, we cannot rely on Proposition 4 as in (5) and loose the analytical tractability of the VaR. The latter can only be computed via numerical simulation techniques.

2.3. The IRB model

The minimum capital requirement formulas associated with the IRB approach (BCBS, 2023d, §31.4-31.16) are all built upon (13) with $\alpha = 99.9\%$ but differ from it in various aspects. First, it is known that the independence between the LGD and the default event does not hold.[3] This has been formally shown in the pioneer work of Altman et al. (2005), who highlighted the existence of a decreasing relationship between default and recovery rates in the context of defaulted American bonds. To account for this, under the IRB approach, the expected LGD λ_i in both the first and second term is replaced by a downturn value, λ_i^*, which can be interpreted as a worst-case LGD (see §4.1 and §4.3-4.4 of BCBS, 2005), leading to[4]

$$K^{\mathrm{IRB}} := \sum_{i=1}^{I} w_i \big(\lambda_i^* p_i(-z_{0.999}) - \lambda_i^* p_i\big) \tag{14}$$

where under the foundation IRB (F-IRB) approach the λ_i^* is assigned a fixed value depending on the regulatory asset class (BCBS, 2023c, §32.6-32.14) while under the advanced IRB (A-IRB) approach, banks are allowed to estimate this parameter internally as an expected LGD in an *adverse macroeconomic scenario* (see §32.15-32.19 of BCBS, 2023c and §36.83-36.88 of BCBS, 2023a).

Second, the IRB approach provides specific guidelines about how to calibrate the parameters (see BCBS, 2023c and Section 7,8 and 9 of BCBS, 2023a). For instance, the loading factor ρ_i, is restricted to be a decreasing function of p_i, whose explicit form is determined by the regulatory asset class.[5] In addition, depending on the specific IRB approach used (F-IRB *vs.* A-IRB) banks have the freedom to estimate either the PD or the PD, LGD and EAD. In either case, the estimation procedures are subject to regulatory constraints such as the 0.05% floor on the PD (BCBS, 2023c, §32.4) and 25% floor on the LGD (BCBS, 2023c, §32.16).

[3]A complete discussion on the PD-LGD dependency within the Basel ASRF framework can be found in Barbagli and Vrins (2023).

[4]We ignore the maturity adjustment factor and the effective maturity (M) within the scope of this analysis. Indeed, the latter is simply added 'on top' of the resulting analytical formula coming from (14).

[5]The regulatory asset classes are divided into two main categories in BCBS (2023d): corporate, sovereign and bank exposures (§31.4 onwards) and retail exposures (§31.13 onwards).

3. Multi-factor Gaussian default model

As explained in Section 1, equation (14) does not allow to take into account concentration risk. Therefore, we rely on the multi-factor model proposed by Düllmann and Masschelein (2007), which building on Pykhtin (2004)'s work, extends the Gaussian factor model detailed in Section 2.2, to a sector multi-factor structure. The goal is to have a richer correlation structure compared to (7), as a multi-factor model allows for a more flexible and realistic modelling of financial interlinkages (Puzanova and Düllmann, 2013). The dependence between obligors is no longer driven by a single systematic risk factor Y but instead by multiple systematic risk factors $\boldsymbol{Y} = (Y_1, ..., Y_S)$, each of which is driven by the same set of K independent standard normal common factors $\boldsymbol{Z} = (Z_1, ..., Z_K)^\top$:

$$Y_s = \boldsymbol{\alpha}_s^\top \boldsymbol{Z} \quad \text{with} \quad \boldsymbol{\alpha}_s^\top \boldsymbol{\alpha}_s = \|\boldsymbol{\alpha}_s\|^2 = 1 \ , \quad \boldsymbol{\alpha}_s \in \mathbb{R}^K \text{ and } s \in \{1, ..., S\} \ . \tag{15}$$

In this setup, each obligor i is uniquely assigned to a sector s and S is the number of sectors. In the sequel, we assume that the number of systematic risk factors S defines the multi-factor setup, e.g., a two-factor setup implies exactly two systematic risk factors. We denote $Y_{s(i)}$ as the systematic risk factor of obligor i in sector s. Additionally, we impose $K = S$ such that the coefficients $\alpha_{s,k}$ can be obtained via a Cholesky decomposition of the inter-sector correlation matrix. This choice delivers a parsimonious model, reducing considerably the number of risk factors, the computational burden while having only a limited impact on the model's accuracy. In contrast to Düllmann and Masschelein (2007), obligors remain free to feature their own individual PD, EAD and LGD while we still impose a sector dependence for the parameter $\rho_{s(i)}$ controlling the link between the default events.

In this setup, L is still defined as in (2) while the definition of the creditworthiness variables in (6) changes to

$$X_i := \sqrt{\rho_{s(i)}}\, Y_{s(i)} + \sqrt{1 - \rho_{s(i)}} \varepsilon_i \ , \tag{16}$$

where $Y_{s(i)}, \varepsilon_1, ..., \varepsilon_I$ are independent standard normal random variables and $s(i)$ stands for the sector which obligor i belongs to.

In this setup, the asset correlation between two distinct[6] creditworthiness variables is

$$\mathbb{C}\text{orr}(X_i, X_j) = r_{s(i),s(j)} \quad \text{for all} \quad i \neq j \quad \text{where} \quad r_{s,t} := \sqrt{\rho_s \rho_t} \boldsymbol{\alpha}_s^\top \boldsymbol{\alpha}_t \ , \tag{17}$$

and where $s, t \in \{1, ..., S\}$ represent two distinct sectors.

This dependency structure allows to distinguish between various notions of correlations: (i) the *intra-sector asset correlation* which reduces to ρ_s (correlation between X_i and $X_{j \neq i}$ of obligors i, j in the same sector s), (ii) the *inter-sector asset correlation* which is precisely

[6] Of course, it is one otherwise.

$r_{s,t}$ found in (17), and (iii) the *inter-sector correlation* which is computed as

$$\mathbb{C}\mathrm{orr}(Y_s, Y_t) = \gamma_{s,t} := \boldsymbol{\alpha}_s^\top \boldsymbol{\alpha}_t \ . \tag{18}$$

From the above, it is clear that the asset dependency structure is thus modelled by a multivariate normal distribution with zero mean and whose covariance matrix has a block structure. Dependencies between obligors arise both from their affiliation with a business sector and from the existing correlations between the sector systematic risk factors.

In this setup, the asymptotic loss thus reads:

$$L^\infty(\boldsymbol{Y}) = \sum_{i=1}^{I} w_i \lambda_i \tilde{p}_i(Y_{s(i)}) \quad \text{where} \quad \tilde{p}_i(y) := \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_{s(i)}}y}{\sqrt{1-\rho_{s(i)}}}\right) \ . \tag{19}$$

4. Approximate multi-factor VaR

Starting from the work of Gourieroux et al. (2000) and Martin and Wilde (2002), Pykhtin (2004) provides a methodology to find an analytical approximation of the multi-factor VaR. This is done by approximating the distribution of L by the distribution of a comparable one-factor model $\hat{L}$ plus an error term ϵU, where $U := L - \hat{L}$ and $\epsilon \in \mathbb{R}$ controls the magnitude of the error. Taking a second-order Taylor expansion around $L_\epsilon := \hat{L} + \epsilon U$, Pykhtin (2004) shows that the approximation of the quantile of L in the multi-factor model detailed in Section 3 decomposes into

$$\mathbb{Q}_\alpha[L] \approx \mathbb{Q}_\alpha[\hat{L}] + \Delta q_\alpha \quad \text{where} \quad \Delta q_\alpha = \Delta q_\alpha^\infty + \Delta q_\alpha^{GA} \tag{20}$$

where $\mathbb{Q}_\alpha[\hat{L}]$ is a linear component computed in a comparable ASRF (C-ASRF) setup, Δq_α^∞ is a systematic multi-factor adjustment (MFA) term representing sector concentration and Δq_α^{GA} is a multi-factor granularity adjustment term. When the portfolio size grows to infinity, Δq_α^{GA} vanishes and the MFA is simply equal to Δq_α^∞. In the sequel, to distinguish between the multi-factor and the corresponding *comparable one-factor model*, we use the notation $(\hat{\ })$.

In Section 4.1, we detail how to build this comparable one-factor model on which to compute $\mathbb{Q}_\alpha[\hat{L}]$. Once the $Y_1, ...Y_S$ are resumed into a comparable single-factor $\hat{Y}$, the X_i in (16) are no more independent conditionally on $\hat{Y}$. This gives rise to a conditional asset correlation which we briefly present in Section 4.2. In Section 4.3, we present the formulas to compute the MFA terms in (20), which require the conditional asset correlation presented in Section 4.2.

4.1. Comparable one-factor model

The key idea of Pykhtin (2004) is to approximate the portfolio loss L in the multi-factor model with the portfolio loss $\hat{L}$ in a *comparable one-factor model*. Following the same idea, we propose a slightly different route, allowing to circumvent the need to introduce a series of intermediary factor models.

First, let's recall that in the multi-factor model considered in Section 3, the creditworthiness variables $\boldsymbol{X} = (X_1, \ldots, X_I)$ have the following functional form

$$X_i = \sqrt{\rho_{s(i)}}\, Y_{s(i)} + \sqrt{1-\rho_{s(i)}}\varepsilon_i \quad \text{with} \quad Y_{s(i)} = \boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{Z} \tag{21}$$

where $Z_1, \ldots, Z_K, \varepsilon_1, \ldots, \varepsilon_I$ are independent standard normal variables. In this context, the *comparable one-factor model* shall be based on creditworthiness variables $\hat{\boldsymbol{X}} = (\hat{X}_1, \ldots, \hat{X}_I)$ with dependence structure

$$\hat{X}_i = \sqrt{\gamma_i}\, \hat{Y} + \sqrt{1-\gamma_i}\hat{\varepsilon}_i \quad \text{with} \quad \hat{Y} = \boldsymbol{\beta}^\top \boldsymbol{Z} \tag{22}$$

for some well chosen $\boldsymbol{\beta} \in \mathbb{R}^K$, $\|\boldsymbol{\beta}\|^2 = 1$ and where $Z_1, \ldots, Z_K, \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_I$ are i.i.d. standard normal variables.[7]

Second, given a combination vector $\boldsymbol{\beta}$, it is necessary to set the loadings $\sqrt{\gamma_i}$ in such a way that the single-factor model (22) provides a good approximation of the multi-factor one (21). Pykhtin (2004) proposes to use as condition that the expected loss of obligor i conditional upon $\hat{Y}$ are the same under both approaches. This amounts to force the default probabilities of obligor i conditional upon $\hat{Y}$ to be the same in both models.

We are thus looking for the coefficients $\sqrt{\gamma_i}$ such that

$$\mathbb{P}(B_i = 1|\hat{Y}) = \mathbb{P}(\hat{B}_i = 1|\hat{Y}) \quad \text{where} \quad B_i = \mathbb{1}_{\{X_i \leq \Phi^{-1}(p_i)\}} \quad \text{and} \quad \hat{B}_i = \mathbb{1}_{\{\hat{X}_i \leq \Phi^{-1}(p_i)\}}\ .$$

The second expression is known from Section 2.2, replacing ρ_i by γ_i and Y by $\hat{Y}$:

$$\mathbb{P}(\hat{B}_i = 1|\hat{Y}) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\gamma_i}\hat{Y}}{\sqrt{1-\gamma_i}}\right)\ .$$

Differently from Pykhtin (2004), to find these coefficients $\sqrt{\gamma_i}$, we do not specify an intermediary factor model linking the $Y_1, \ldots, Y_S$ with $\hat{Y}$, nor need to introduce interdependent idiosyncratic variables in the model driving X_i in (21). Alternatively, we rely on the work of Vrins (2018) to compute the conditional default probability in the multi-factor model conditional upon $\boldsymbol{\beta}^\top \boldsymbol{Z} = \hat{Y}$. Using the law of iterated expectations and exploiting that, in (21), $Y_{s(i)} = \boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{Z}$ is independent from $\varepsilon_i \sim \mathcal{N}(0,1)$, one finds:

[7]Note that there is no reason for the portfolio loss L in the multi-factor setup built from the X_i to agree with the loss $\hat{L}$ in the comparable single-factor model built from the $\hat{X}_i$. This is because the dependence structure in the latter is more restrictive than that of the former, unless $\boldsymbol{\alpha}_1 = \ldots = \boldsymbol{\alpha}_S$ in which case the multi-factor model collapses to the single-factor whenever $\boldsymbol{\beta} = \boldsymbol{\alpha}$ and $\gamma_i = \rho_{s(i)}$.

$$
\begin{aligned}
\mathbb{P}(B_i = 1|\hat{Y}) &= \mathbb{P}\big(\sqrt{\rho_{s(i)}}\, Y_{s(i)} + \sqrt{1-\rho_{s(i)}}\varepsilon_i \le \Phi^{-1}(p_i) |\, \boldsymbol{\beta}^\top \boldsymbol{Z} = \hat{Y}\big) \\
&= \mathbb{E}\left[\mathbb{P}\left(\varepsilon_i \le \frac{\Phi^{-1}(p_i) - \sqrt{\rho_{s(i)}}\boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{Z}}{\sqrt{1-\rho_{s(i)}}} \middle| \boldsymbol{Z} \right) \middle| \boldsymbol{\beta}^\top \boldsymbol{Z} = \hat{Y} \right] \\
&= \mathbb{E}\left[\Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_{s(i)}}\boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{Z}}{\sqrt{1-\rho_{s(i)}}} \right) \middle| \boldsymbol{\beta}^\top \boldsymbol{Z} = \hat{Y} \right] \\
&= \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\hat{\rho}_{s(i)}}\hat{Y}}{\sqrt{1-\hat{\rho}_{s(i)}}} \right) \\
&=: \hat{p}_i(\hat{Y})
\end{aligned} \tag{23}
$$

where the penultimate equality results from Lemma 7, found in Appendix B, with $\sqrt{\hat{\rho}_{s(i)}} := \sqrt{\rho_{s(i)}}\boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{\beta}$.

We conclude that the loadings $\sqrt{\gamma_i}$ one must take in the *comparable one-factor model* are simply

$$
\sqrt{\hat{\rho}_{s(i)}} = \mathbb{C}\mathrm{orr}(X_i, Y_{s(i)}) \times \mathbb{C}\mathrm{orr}(Y_{s(i)}, Y) = \sqrt{\rho_{s(i)}}\boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{\beta} \; . \tag{24}
$$

From (24) it is clear that, differently from the ASRF in Section 2.2, the loadings in the C-ASRF model, considered here, are only sector dependent.

Fourth, to insure that $\hat{Y}$ is indeed able to optimally resume the multi-factor structure, Pykhtin (2004) proposes to chose $\boldsymbol{\beta}$ in order to maximize the 'weighted' correlation, using weights d_i, between the sector systematic risk factors $Y_1, ..., Y_S$ and the single comparable $\hat{Y}$, precisely:

$$
\max_{\boldsymbol{\beta}\in\mathbb{R}^K} \left(\sum_{i=1}^{I} d_i \boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{\beta} \right) \qquad \text{subject to} \qquad \|\boldsymbol{\beta}\|^2 = 1 \; . \tag{25}
$$

The solution to this optimisation problem leads to[8]

$$
\beta_k := \frac{\sum_{i=1}^{I} d_i \alpha_{s(i),k}}{\sqrt{\sum_{k=1}^{K} (\sum_{i=1}^{I} d_i \alpha_{s(i),k})^2}} \; . \tag{26}
$$

It is important to note that the choice of the weight d_i is not unique. An intuitive development by Pykhtin (2004), leads to the following expression

$$
d_i := w_i \lambda_i \tilde{p}_i(-z_\alpha) \; , \tag{27}
$$

where $\tilde{p}_i(-z_\alpha)$ is given in (19) and the intuition behind this choice is that obligors with a high exposure in terms of VaR should have a large weight in the maximization problem whereas obligors with a small VaR should have a minor impact.

[8]We have chosen the value of the Lagrange multiplier in order to satisfy the constraint.

At last, using the LP approximation detailed in Section 2.1, the C-ASRF model reads

$$\hat{L}^{\infty}(Y) = \sum_{i=1}^{I} w_i \lambda_i \hat{p}_i(Y) \ , \tag{28}$$

where $\hat{p}_i(\cdot)$ is given in (23) and its quantile is computed as

$$\mathbb{Q}_\alpha[\hat{L}] \xrightarrow{I\to\infty} \mathbb{Q}_\alpha[\hat{L}^{\infty}(\hat{Y})] = \hat{L}^{\infty}(-z_\alpha) = \sum_{i=1}^{I} w_i \lambda_i \hat{p}_i(-z_\alpha) \ . \tag{29}$$

4.2. Conditional asset correlation

Let us add $\sqrt{\hat{\rho}_{s(i)}}\hat{Y} - \sqrt{\hat{\rho}_{s(i)}}\boldsymbol{\beta}^\top \boldsymbol{Z} = 0$ to the right-hand side of (21). Noting that $Y_{s(i)} = \boldsymbol{\alpha}_{s(i)}^\top \boldsymbol{Z}$ and rearranging the terms yields

$$X_i = \sqrt{\hat{\rho}_{s(i)}}\hat{Y} + \left(\sqrt{\rho_{s(i)}}\boldsymbol{\alpha}_{s(i)}^\top - \sqrt{\hat{\rho}_{s(i)}}\boldsymbol{\beta}^\top\right)\boldsymbol{Z} + \sqrt{1-\rho_{s(i)}}\varepsilon_i \ . \tag{30}$$

From the above it is clear that, differently from the one-factor model featured in Section 2.2, the default events are not independent conditionally on the single $\hat{Y}$. They are independent only conditionally on $\boldsymbol{Z}$. The latter phenomenon gives rise to a conditional asset correlation between two distinct borrowers

$$\mathbb{C}\text{orr}(X_i, X_{j\neq i}|\hat{Y}) = \hat{r}_{s(i),s(j)} \quad \text{where} \quad \hat{r}_{s,t} := \frac{\sqrt{\rho_s}\sqrt{\rho_t}\boldsymbol{\alpha}_s^\top\boldsymbol{\alpha}_t - \sqrt{\hat{\rho}_s}\sqrt{\hat{\rho}_t}}{\sqrt{1-\hat{\rho}_s}\sqrt{1-\hat{\rho}_t}} \ , \tag{31}$$

and where $s, t \in \{1, ..., S\}$ represent two distinct sectors.

Further details are provided in Appendix A.

4.3. Multi-factor adjustment terms

According to Pykhtin (2004), the MFA in (20) is computed as:

$$\Delta q_\alpha := -\frac{1}{2l'(y)}\left[v'(y) - v(y) \times \left(\frac{l''(y)}{l'(y)} + y\right)\right]\Bigg|_{y=-z_\alpha} \tag{32}$$

since we evaluate the above expression at $\mathbb{Q}_{1-\alpha}[\hat{Y}] = -z_\alpha$ and where $l'(y)$ and $l''(y)$ are the first and second derivative of (28), $v(y) := \mathbb{V}[L|\hat{Y} = y]$ is the conditional variance of L in the multi-factor model and $v'(y)$ its first derivative. Moreover, thanks to the law of total variance $v(y)$ decomposes into the sum of a systematic and an idiosyncratic part

$$v(y) := \mathbb{V}\left[\mathbb{E}[L|\boldsymbol{Z}]|\hat{Y} = y\right] + \mathbb{E}\left[\mathbb{V}[L|\boldsymbol{Z}]|\hat{Y} = y\right] = v_\infty(y) + v_{GA}(y) \tag{33}$$

where $v_\infty(y)$ describes the difference between the multi-factor and single-factor loss distribution in an infinitely granular portfolio, and $v_{GA}(y)$ is the granularity adjustment, which measures the influence of single name concentration. Replacing $v(y)$ in (32) by either $v_\infty(y)$ or $v_{GA}(y)$, allows to split the MFA into Δq_α^∞ and Δq_α^{GA} as shown in (20).

In the sequel, we provide the expressions for $v_\infty(y)$ and $v_{GA}(y)$ adapted to our model, where the latent creditworthiness variables driving L in (2) are expressed as in (30).

Lemma 1. *The systematic part of the variance of L conditional upon $\hat{Y}$ is*

$$v_\infty(y) := \sum_i \sum_j w_i w_j \lambda_i \lambda_j \left[\Phi_2\left(\Phi^{-1}(\hat{p}_i(y)), \Phi^{-1}(\hat{p}_j(y)); \hat{r}_{s(i),s(j)}\right) - \hat{p}_i(y)\hat{p}_j(y)\right] \tag{34}$$

where $\Phi_2(a,b;r)$ is the standard normal bivariate cdf with correlation r, and $\hat{r}_{s(i),s(j)}$ is the conditional asset correlation found in (31).

Lemma 2. *The first derivative of the conditional variance found in Lemma 1 becomes*

$$v'_\infty(y) := 2\sum_i \sum_j w_i w_j \lambda_i \lambda_j \hat{p}'_i(y) \times \left[\Phi\left(\frac{\Phi^{-1}(\hat{p}_j(y)) - \hat{r}_{s(i),s(j)}\Phi^{-1}(\hat{p}_i(y))}{\sqrt{1-(\hat{r}_{s(i),s(j)})^2}}\right) - \hat{p}_j(y)\right] \tag{35}$$

where $\hat{r}_{s(i),s(j)}$ is the conditional asset correlation found in (31).

Lemma 3. *The idiosyncratic part of the variance of L conditional upon $\hat{Y}$ is*

$$v_{GA}(y) := \sum_i w_i^2 \times \left(\lambda_i^2\left[\hat{p}_i(y) - \Phi_2\left(\Phi^{-1}(\hat{p}_i(y)), \Phi^{-1}(\hat{p}_i(y)); \hat{r}_{s(i),s(i)}\right)\right] + \sigma_i^2 \hat{p}_i(y)\right) \tag{36}$$

where $\Phi_2(a,b;r)$ is the standard normal bivariate cdf with correlation r, and $\hat{r}_{s(i),s(i)}$ is the conditional intra-sector asset correlation of sector $s(i)$ in (31).

Lemma 4. *The first derivative of the conditional variance found in Lemma 3 becomes*

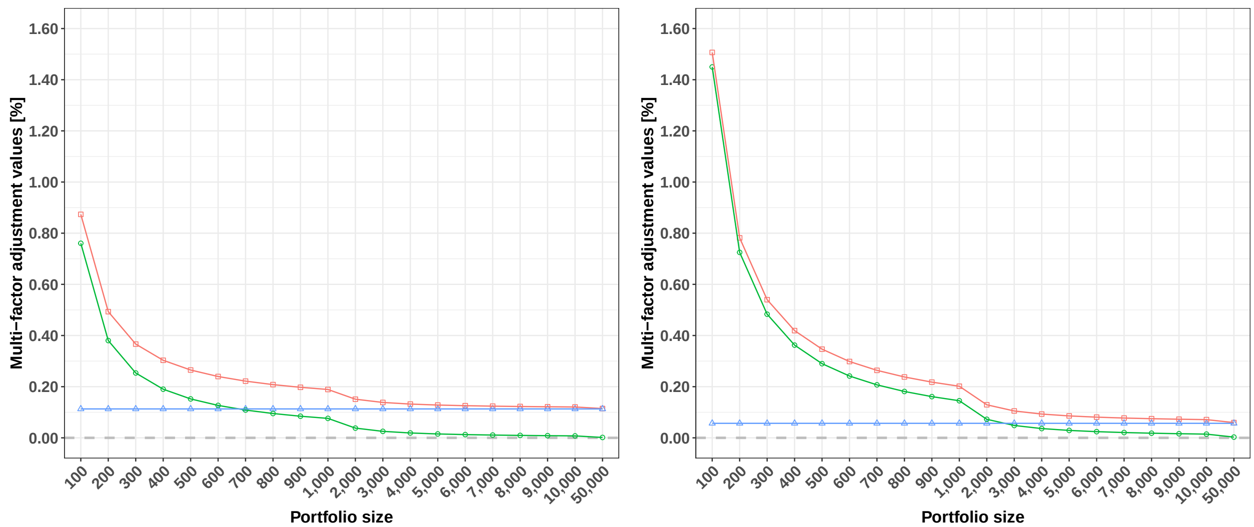
$$v'_{GA}(y) := \sum_i w_i^2 \hat{p}'_i(y) \times \left(\lambda_i^2\left[1 - 2\Phi\left[\frac{\Phi^{-1}(\hat{p}_i(y)) - \hat{r}_{s(i),s(i)}\Phi^{-1}(\hat{p}_i(y))}{\sqrt{1-(\hat{r}_{s(i),s(i)})^2}}\right]\right] + \sigma_i^2\right) \tag{37}$$

where $\hat{r}_{s(i),s(i)}$ is the conditional intra-sector asset correlation of sector $s(i)$ in (31).

Figure 1 depicts the different values of the MFA terms as a function of the portfolio size I, for two distinct portfolio compositions, the same two synthetic portfolios chosen in the work of Pykhtin (2004). As expected, when $I \to \infty$, Δq_α^{GA} vanishes and $\Delta q_\alpha \to \Delta q_\alpha^\infty$. At the same time, we observe that for smaller portfolio sizes, $\Delta q_\alpha^{GA} > \Delta q_\alpha^\infty$, as in these setups, it is the gap between quantiles in the asymptotic and the finite pool which mainly drives the multi-factor adjustment, rather than the presence of a multi-factor structure.

4.4. Performance of the approximations

It is important to note that the convergence speed of the VaR to the LP approximation detailed in Section 2.1 is a function of credit risk profile. The latter is defined by the value of



(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.

(b) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

Figure 1: MFA terms: Δq_α (□), Δq_α^∞ (△), Δq_α^{GA} (○), in a two-factor setup, with equal repartition of the obligors across the sectors, as a function of I and for a fixed value of the inter-sector correlation $\gamma_{1,2} = 0.5$. The LGD is distributed as a $Beta(2, 3)$, leading to $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups. We only show the MFA values for a set of informative I sizes; hence, the drops are not real, they are caused by the change of scale on the x-axis.

'risk components' controlling the credit risk of the pool in the Basel framework, i.e., PD, EAD and LGD (BCBS, 2023b).[9] It is true to say that, in general, for well diversified portfolios the convergence occurs generally for $I = 1,000$ (Tarashev and Zhu, 2008). Nevertheless, there exist some settings in which the convergence occurs much later. The most extreme case being the one of a credit pool composed by two (or more) homogeneous sub-pools, i.e., each sub-pool assigning the same PD, LGD and EAD to its constituents. Indeed, this setting is the hardest to model via a single systematic risk factor.

A similar phenomenon occurs with the multi-factor approximation proposed by Pykhtin (2004). First, the quality of the approximation decreases as one approaches the same homogeneous sub-pools portfolio as detailed above, where each sub-pool corresponds to a different sector (bucket). Second, the performance of the approximation slightly differs between a homogeneous and heterogeneous pool. For a homogeneous pool, the performance is the worst when the EAD is equally partitioned across the sectors (each sector having the same relative weight). For a heterogeneous pool, the accuracy decerases as the EAD of the largest exposure in the portfolio becomes a larger fraction of the entire portfolio EAD. In both these setups, for all choices of sector weights, the performance of the approximation improves as the

[9]The correlation parameter ρ linking the default indicators is a function of the PD only in the Basel framework, hence it does not directly appear in this list. Additionally, we ignore the effective maturity (M) within the scope of this analysis, as this component does not derive from the Gaussian factor model setup underlying the IRB approach.

inter-sector correlations and the number of the systematic factors increase. It is important to note that, in general, the multi-factor approximation performs better when faced with a heterogeneous pool compared to a homogeneous pool.

From the above paragraphs, it is evident that there are scenarios in which a discrepancy between the multi-factor approximation and the exact MC quantile occurs. Therefore, this highlights the necessity of developing and testing methodologies to decrease the computational burden of numerical estimation strategies. This is the reason why, in this paper, we propose a control variate estimator of the multi-factor VaR.

5. Variance reduction

When wanting to measure value-at-risk, we are considering ways to capture tail-events for which the convergence of multi-factor model rates are quite low: a large number of simulations is typically required, which is a major practical problem when both the number of exposures and the number of factors are large.

5.1. Control variate

The control variate method is a powerful technique that aims at improving the efficiency of estimators through variance reduction. More specifically, it allows to decrease the variance of an unbiased estimator without introducing any bias. We briefly recall below this method; more details can be found in Glasserman (2004).

Let θ be an unknown parameter of interest, for which we know an unbiased estimator $\hat{\theta}$ called the base estimator, i.e., $\mathbb{E}[\hat{\theta}] = \theta$. Suppose moreover that we derive another statistic $\hat{\eta}$, called the control estimator, with expectation $\mathbb{E}[\hat{\eta}] = \eta$. Then, we can define a control variate estimator as

$$\tilde{\theta} := \hat{\theta} - b(\hat{\eta} - \eta) \tag{38}$$

for some constant b.

The statistic $\tilde{\theta}$ is another unbiased estimator of θ whose mean and variance are

$$\mathbb{E}[\tilde{\theta}] = \mathbb{E}[\hat{\theta} - b(\hat{\eta} - \eta)] = \mathbb{E}[\hat{\theta}] - b\,\underbrace{\mathbb{E}[\hat{\eta} - \eta]}_{=0} = \mathbb{E}[\hat{\theta}] = \theta\ , \tag{39}$$

$$\mathbb{V}[\tilde{\theta}] = \mathbb{V}[\hat{\theta} - b(\hat{\eta} - \eta)] = \mathbb{V}[\hat{\theta}] + b^2\,\mathbb{V}[\hat{\eta}] - 2b\,\mathrm{Corr}(\hat{\theta}, \hat{\eta})\sigma_{\hat{\theta}}\sigma_{\hat{\eta}} \tag{40}$$

where $\sigma_{\hat{\theta}}$ and $\sigma_{\hat{\eta}}$ represent the standard deviation of the base and control estimator, respectively.

For some choices of $(b, \hat{\eta})$ the variance of the control variate estimator $\tilde{\theta}$ might be smaller than the variance of the original estimator $\hat{\theta}$. The value of b minimizing the variance of $\tilde{\theta}$ is

$$b^\star = \frac{\mathbb{C}\mathrm{ov}(\hat{\theta}, \hat{\eta})}{\mathbb{V}[\hat{\eta}]}\ . \tag{41}$$

For $b = b^\star$, the variance of $\tilde{\theta}$ shinks to

$$\mathbb{V}[\tilde{\theta}] = \left(1 - \mathbb{C}\mathrm{orr}(\hat{\theta}, \hat{\eta})^2\right) \mathbb{V}[\hat{\theta}] \tag{42}$$

and shows the benefit of choosing a control whose estimator $\hat{\eta}$ is highly correlated with the base one, $\hat{\theta}$.

5.2. Monte Carlo estimate with C-ASRF control

In this work, we are interested in estimating $\theta = \mathbb{Q}_\alpha[L]$ of the multi-factor model presented in Section 3, within a Monte Carlo framework. The application of the control variate method in this context is not straightforward, for several reasons. First, most of the quantile estimators are biased (even though often asymptotically unbiased, i.e., when the number of simulations tends to infinity)(Pitera and Schmidt, 2018). Second, it is not clear how to chose the control estimator, as it needs to be a random variable with: (i) known expectation, and (ii) correlated with the quantile estimator in the multi-factor model.

Quite naturally, we chose as base estimator for $\mathbb{Q}_\alpha[L]$ the empirical quantile $q_\alpha(L, m)$ of the relative portfolio losses in a multi-factor model, obtained from a finite number of m MC simulations of the relative portfolio losses, i.e.,

$$\hat{\theta} := q_\alpha(L, m) \tag{43}$$

where the latter is obtained by sampling the random variables $Z_1, ..., Z_K, \varepsilon_1, ..., \varepsilon_I, \Lambda_1, ..., \Lambda_I$ and choosing as $q_\alpha(L, m)$ the order statistic out of m simulated relative portfolio losses.[10]

The key idea here relies on choosing as control estimator, the empirical quantile $q_\alpha(\hat{L}, m)$ of $\mathbb{Q}_\alpha[\hat{L}]$ in the comparable one-factor model featured in Section 4.1, obtained from a finite number of m MC simulations of the relative portfolio losses, i.e.,

$$\hat{\eta} := q_\alpha(\hat{L}, m) \ . \tag{44}$$

where the latter is obtained by sampling the random variables $Z_1, ..., Z_K, \hat{\varepsilon}_1, ..., \hat{\varepsilon}_I, \Lambda_1, ..., \Lambda_I$ and choosing as $q_\alpha(\hat{L}, m)$ the order statistic out of m simulated relative portfolio losses.

Indeed, given that the single-factor $\hat{Y}$ driving $\hat{L}$ is built in order to optimally summarise the correlation structure found in the multi-factor framework, the latter is naturally highly correlated with $\hat{\theta}$, allowing to reach a variance reduction when using $\tilde{\theta}$ compared to $\hat{\theta}$ alone. It is interesting to note that both $\hat{\theta}$ and $\hat{\eta}$ are biased estimators of θ and η for a finite m, while they are asymptotically unbiased, when $m \to \infty$ (Glasserman, 2004).

A first possibility for a control variate estimator of the α-quantile of the relative portfolio

[10] The simulated losses are sorted in ascending order and the $[\alpha m]$-th smallest value is selected, corresponding to the loss level that is exceeded with probability approximately $1 - \alpha$.

loss computed in a multi-factor framework reads

$$q_\alpha(L,m) - b\left(q_\alpha(\hat{L},m) - \mathbb{E}\left[q_\alpha(\hat{L},m)\right]\right) ,$$

Unfortunately, $\mathbb{E}[q_\alpha(\hat{L},m)]$ is typically unknown for a finite m. We circumvent this issue by replacing $\mathbb{E}[q_\alpha(\hat{L},m)]$ by its asymptotic version $\mathbb{E}[\hat{L}|\,\mathbb{Q}_{1-\alpha}[\hat{Y}]]$, for which we have an analytical expression detailed in (29). The latter replacement guarantees that our control variate estimator is unbiased, only when both $I, m \to \infty$, i.e.,

$$\begin{aligned}\tilde{\theta} &= q_\alpha(L,m) - bq_\alpha(\hat{L},m) + b\hat{L}^\infty(-z_\alpha)\\ &= b\hat{L}^\infty(-z_\alpha) + \left(q_\alpha(L,m) - bq_\alpha(\hat{L},m)\right) .\end{aligned} \tag{45}$$

A naive choice consists in using $b = 1$. The latter leads to the variance reduction, stemming from the MC method, merely being used to estimate the gap between the quantiles of the multi-factor and the large pool C-ASRF models, instead of the quantile of the multi-factor model directly. [11]

Alternatively, we can estimate the optimal b numerically for every MC simulation. Precisely, we partition the set of m portfolio losses for L and $\hat{L}$ into N subsets, each of equal size $\widetilde{m} := m/N$. For each subset $i = 1, ..., N$, we compute the empirical α-quantiles $q^i_\alpha(L,\widetilde{m})$ and $q^i_\alpha(\hat{L},\widetilde{m})$ each based on $\widetilde{m}$ simulated losses. Collecting these quantiles into two vectors $\boldsymbol{q}_\alpha(L,\widetilde{m}) = \left(q^1_\alpha(L,\widetilde{m}), ..., q^N_\alpha(L,\widetilde{m})\right)$ and $\boldsymbol{q}_\alpha(\hat{L},\widetilde{m}) = \left(q^1_\alpha(\hat{L},\widetilde{m}), ..., q^N_\alpha(\hat{L},\widetilde{m})\right)$, we can estimate b as

$$\hat{b}^\star = \frac{\mathbb{C}\mathrm{ov}\left(\boldsymbol{q}_\alpha(L,\widetilde{m}), \boldsymbol{q}_\alpha(\hat{L},\widetilde{m})\right)}{\mathbb{V}\left[\boldsymbol{q}_\alpha(\hat{L},\widetilde{m})\right]} . \tag{46}$$

Obviously, one should draw pairs $(L_i, \hat{L}_i)$ from the right marginal distributions (that are known, see above), but maximizing the correlation between them. Pykhtin (2004)'s parametrisation offers an elegant solution to this problem. As it is evident in (21) and (22), the same set of $Z_1, ..., Z_K$ and Λ_i are used to draw samples from L_i and $\hat{L}_i$, respectively. The latter introduces a link between the two relative portfolio losses generated *run-by-run* in both underlying models. This methodology guarantees L_i and $\hat{L}_i$ to be highly correlated, the same would apply for their quantiles, hence maximising the variance reduction.

Notice that one can further increase the correlation between L_i and $\hat{L}_i$ by using the same idiosyncratic variables in both models (i.e., $\varepsilon_i = \hat{\varepsilon}_i$ in (21) and (22), respectively). However, in spite of the fact that this way of coupling the two models has the highest impact in terms of correlation between X_i and $\hat{X}_i$, it is, by far, the systematic factors that drive the correlation

[11] Note that a poor choice of b could result in a variance increase compared to the plain Monte Carlo. We verified empirically a variance reduction when $b = 1$, although it is unlikely to be the optimal value.

between the quantiles. This can be understood by the large pool effect: as $I \to \infty$, the impact of the idiosyncratic variables averages out. Hence, considering the same idiosyncratic variables in either models has little impact on the portfolio loss quantiles.

5.3. Bias reduction

Our control variate estimator suffers from a little drawback because we replaced the true expectation of $q_\alpha(\hat{L}, m)$ by its large-pool version. Therefore, it is likely that $\tilde{\theta}$ is biased (even asymptotically as $m \to \infty$), in contrast with the plain Monte Carlo estimator.[12]

To mitigate this effect, one needs to improve our estimation of the expected value of the quantile. To this end, one can replace the large-pool approaximation by an enhanced version, relying on the granularity adjustment proposed by Gordy and Lutkebohmert (2013). For a one-factor default model, it reads as

$$\mathrm{GA} := \mathbb{Q}_\alpha[L] - \mathbb{Q}_\alpha[L^\infty(Y)] \approx -\frac{1}{2\phi(-z_\alpha)} \frac{d}{dy} \left[\frac{\phi(y)v(y)}{l'(y)} \right] \Bigg|_{y=-z_\alpha} \tag{47}$$

where $\phi(\cdot)$ is the density of the standard normal and $v(y) = \mathbb{V}[L|Y = y]$. It is important to note that the latter formula is an approximation, obtained via Taylor expansion, of the exact gap between the quantile $\mathbb{Q}_\alpha[L]$ and its asymptotic version $\mathbb{Q}_\alpha[L^\infty(Y)]$. Adapting the above general formula to the Basel ASRF setup, one gets

Lemma 5. *The granularity adjustment adapted to the Basel ASRF framework detailed in Section 2.2 reads*

$$\mathrm{GA} = -\frac{1}{2l'(y)} \left[v'(y) - v(y) \left(\frac{l''(y)}{l'(y)} + y \right) \right] \Bigg|_{y=-z_\alpha} \tag{48}$$

and

$$l'(y) = \sum_{i=1}^{I} w_i \lambda_i p_i'(y) \quad , \quad l''(y) = \sum_{i=1}^{I} w_i \lambda_i p_i''(y)$$

with $p'(y)$ and $p''(y)$ being the first and second partial derivatives, with respect to y, of the conditional PD featured in (10), and are expressed as

$$p_i'(y) = -\frac{\sqrt{\rho_i}}{\sqrt{1-\rho_i}} \times \phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i} y}{\sqrt{1-\rho_i}} \right) ,$$
$$p_i''(y) = -\frac{\rho_i}{1-\rho_i} \times \left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i} y}{\sqrt{1-\rho_i}} \right) \times \phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i} y}{\sqrt{1-\rho_i}} \right) .$$

[12] Of course, $\tilde{\theta}$ is unbiased "jointly asymptotically", i.e., when both m and I tend top infinity

The conditional variance and its first derivative are given by

$$v(y) = \sum_{i=1}^{I} w_i^2 p_i(y) \left(\sigma_i^2 + \lambda_i^2(1 - p_i(y))\right) \ ,$$
$$v'(y) = \sum_{i=1}^{I} w_i^2 p_i'(y) \left(\sigma_i^2 + \lambda_i^2(1 - 2p_i(y))\right) \ .$$

The proof is given in Appendix C.

When Lemma 5 is used to find the GA for the C-ASRF quantile in (29), it is important to note that it is necessary to replace ρ_i by $\hat{\rho}_{s(i)}$, where the latter is computed as in (24). Thus, the GA of the C-ASRF quantile will be close to the Δq_α^{GA} of Pykhtin (2004), as they both aim at approximating the exact gap between quantiles in the asymptotic compared to the finite pool. Nevertheless, Δq_α^{GA} approximates the gap in a multi-factor setting, while GA approximates the gap in a single-factor setup (where, for the C-ASRF, the factor loadings are computed in a specific way). This difference is clearly visible when comparing the formulas for the conditional variances $v(y) = \mathbb{V}[L|\hat{Y} = y]$ in (33) and $v(y) = \mathbb{V}[\hat{L}|\hat{Y} = y]$ in Lemma 5.

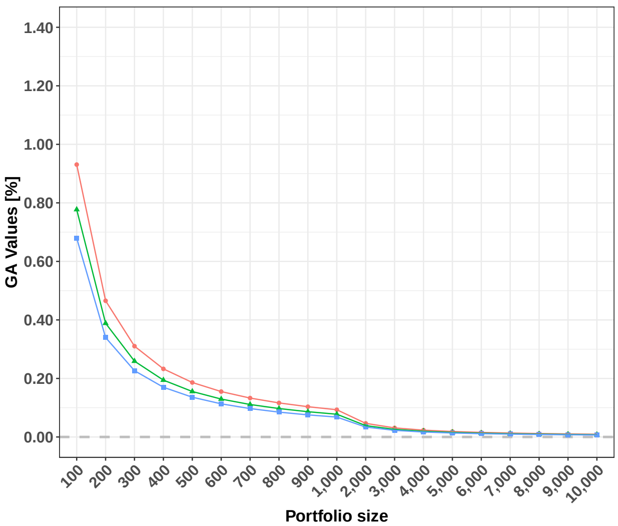
The control variate estimator adjusted for GA thus becomes

$$\begin{aligned} \tilde{\theta}_{GA} &= q_\alpha(L, m) - \hat{b}^\star q_\alpha(\hat{L}, m) + \hat{b}^\star \left(\hat{L}^\infty(-z_\alpha) + \text{GA}\right) \\ &= \hat{b}^\star \left(\hat{L}^\infty(-z_\alpha) + \text{GA}\right) + \left(q_\alpha(L, m) - \hat{b}^\star q_\alpha(\hat{L}, m)\right) \ . \end{aligned} \tag{49}$$

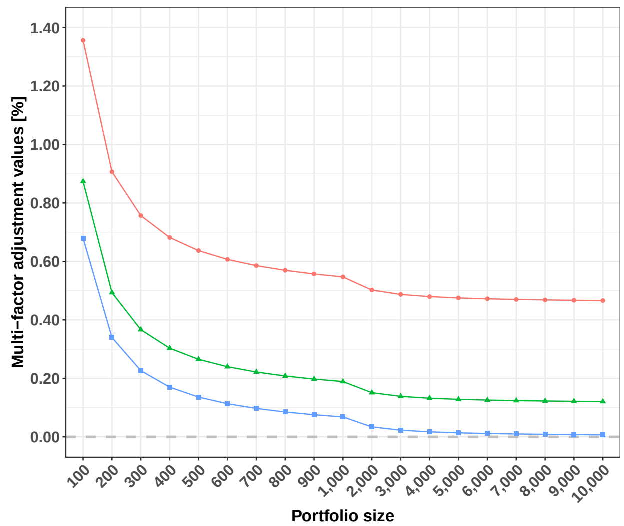
The panels of Figure 2, compare the GA and MFA values as a function of the portfolio size I and for different values of the inter-sector correlation. These comparisons are drawn for two specific portfolio compositions, mirroring those selected in the study of Pykhtin (2004). As expected, the GA shrinks to 0 as $I \to \infty$, while the MFA does not. Indeed, independently of the pool finite size, the MFA also captures the presence of the multi-factor structure via Δq_α^∞, which does not shrinks to 0 as the pool size increases. As a sanity check, we verify that these two granularity adjustments are indeed identical only when the inter-sector correlation is equal to one. The impact of the granularity adjustment on bias reduction is clearly visible in the numerical results discussed in Section 6.2.

6. Empirical analysis

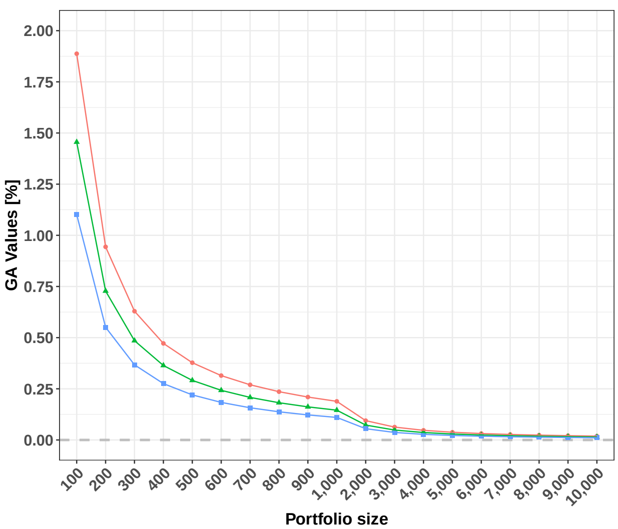
In this section, we conduct an empirical analysis of our mixed MC method, based on two synthetic portfolios and setting the confidence level to $\alpha = 99.9\%$. The main goal is to verify the performance of our control variate estimator $\tilde{\theta}_{GA}$ in (49) compared to the benchmark MC estimator $q_\alpha(L, m)$ in (43). For completeness, we also verify the performance of $\tilde{\theta}$ without the GA, as in (45) and the ability of the C-ASRF quantile $\hat{L}^\infty(-z_\alpha)$ in (29) to



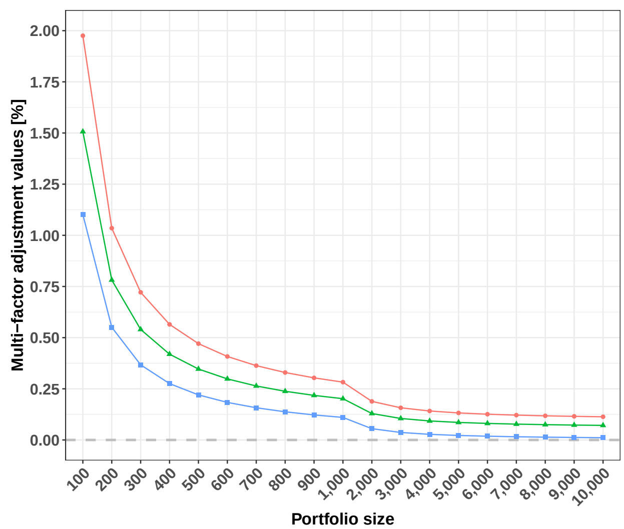
(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.



(b) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.



(c) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.



(d) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

Figure 2: Comparison between the GA (a,c) and MFA (b,d) values, for different inter-sector correlation values $\gamma_{1,2} = 0$ (●), 0.5 (■), 1 (▲), in a two-factor setup, with equal repartition of the obligors across the sectors, as a function of I. In the heterogeneous setup, the pool parameters are only defined at the sector level, hence $p_{s(i)} = p_s, \rho_{s(i)} = \rho_s$, for all obligor i in sector $s(i)$. The LGD is distributed as a $Beta(2, 3)$, leading to $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups. We only show the GA and MFA values for a set of informative I sizes; hence, the drops are not real, they are caused by the change of scale on the x-axis.

capture the MC benchmark, with and without the MFA in (32). For consistency, and ease of comparability with previous works, we select the same pool parameters as in Pykhtin (2004), leading to use a two systematic factors setup where the credit exposures are grouped into two sectors (buckets) $s \in \{1,2\}$. The pool parameters are only defined at the sector level, hence $p_i = p_{s(i)}, \rho_{s(i)}, \lambda_i = \lambda_{s(i)}, \sigma_i = \sigma_{s(i)}$ for all obligor i in sector $s(i)$. We are faced with a:

(i) *Homogeneous pool*: where $p_1 = p_2 = p = 0.5\%, \rho_1 = \rho_2 = \rho = 0.25$;

(ii) *Heterogeneous pool*: where $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$;

while the LGD is distributed according to a $Beta(2,3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$, in both the homogeneous and heterogeneous pools. All exposures in a given sector have the same EAD; hence the same bucket weight: $w_i = w_{s(i)}$ where $w_{s(i)} = \mathrm{EAD}_{s(i)} / \sum_i \mathrm{EAD}_{s(i)}$. As a starting point, we assume the case of equally weighted sectors, i.e., $\mathrm{EAD}_1 = \mathrm{EAD}_2 = 1000$, where 1000 is chosen as an arbitrary value for the exposure principal amount. As explained in Section 3, the $\alpha_{s,k}$ coefficients are obtained by Cholesky decomposition of the inter-sector correlation matrix, whose off-diagonal elements are controlled by the value $\gamma_{1,2}$ defined in (18).

To perform this comparison, we randomly generate M point estimates of the VaR for the benchmark $\boldsymbol{q}_\alpha(L,m) = \big(q_\alpha^1(L,m), ..., q_\alpha^M(L,m)\big)$ and for the comparable $\boldsymbol{q}_\alpha(\hat{L},m) = \big(q_\alpha^1(\hat{L},m), ..., q_\alpha^M(\hat{L},m)\big)$. From $\boldsymbol{q}_\alpha(L,m)$, we compute an average which serves as a benchmark for our accuracy rate in (50).

We then compare this average to another average computed from a series of 'alternative quantiles' which includes the randomly generated $\tilde{\boldsymbol{\theta}}_{GA} = (\tilde{\theta}_{GA}^1, ..., \tilde{\theta}_{GA}^M)$, $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}^1, ..., \tilde{\theta}^M)$, and the analytically computed $\hat{L}^\infty(-z_\alpha)$ and $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$. Precisely, for every M-th simulation, we follow (46) and compute $\hat{b}^\star$ numerically by setting $N = M$. This results in a vector $\hat{\boldsymbol{b}}^\star = (\hat{b}_1^\star, ..., \hat{b}_M^\star)$ which is then used to compute M point estimates of $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}^1, ..., \tilde{\theta}^M)$ and $\tilde{\boldsymbol{\theta}}_{GA} = (\tilde{\theta}_{GA}^1, ..., \tilde{\theta}_{GA}^M)$ computed as in (45) and (49), respectively.

Defining $\tilde{\boldsymbol{q}} = (\tilde{q}_1, ..., \tilde{q}_M)$ as a vector of M point estimates taken from this set of alternative quantiles, we use the following metrics for comparison:

1. Average accuracy [%]:

$$1 - \frac{\left|\frac{1}{M}\sum_{i=1}^M \tilde{q}_i - \frac{1}{M}\sum_{i=1}^M q_\alpha^i(L,m)\right|}{\frac{1}{M}\sum_{i=1}^M q_\alpha^i(L,m)} \tag{50}$$

where $\tilde{q}_i$ are values from a single type amongst $\tilde{\theta}_{GA}^i$, $\tilde{\theta}^i$, $\hat{L}^\infty(-z_\alpha)$ or $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$.[13]

[13] Obviously, for the analytical values $\hat{L}^\infty(-z_\alpha)$ or $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$, the mean reduces to a single value.

2. Variance reduction (VR) [%]:

$$100 \times \frac{\mathbb{V}[\tilde{\boldsymbol{\theta}}_{GA}] - \mathbb{V}[\boldsymbol{q}_\alpha(L,m)]}{\mathbb{V}[\boldsymbol{q}_\alpha(L,m)]} \,. \tag{51}$$

3. Coefficient of variation [%]:

$$100 \times \frac{\sqrt{\mathbb{V}[\tilde{\boldsymbol{q}}]}}{\mathbb{E}[\tilde{\boldsymbol{q}}]} \tag{52}$$

where $\tilde{\boldsymbol{q}}$ ca be either or $\tilde{\boldsymbol{\theta}}_{GA}$ or $\boldsymbol{q}_\alpha(L,m)$.

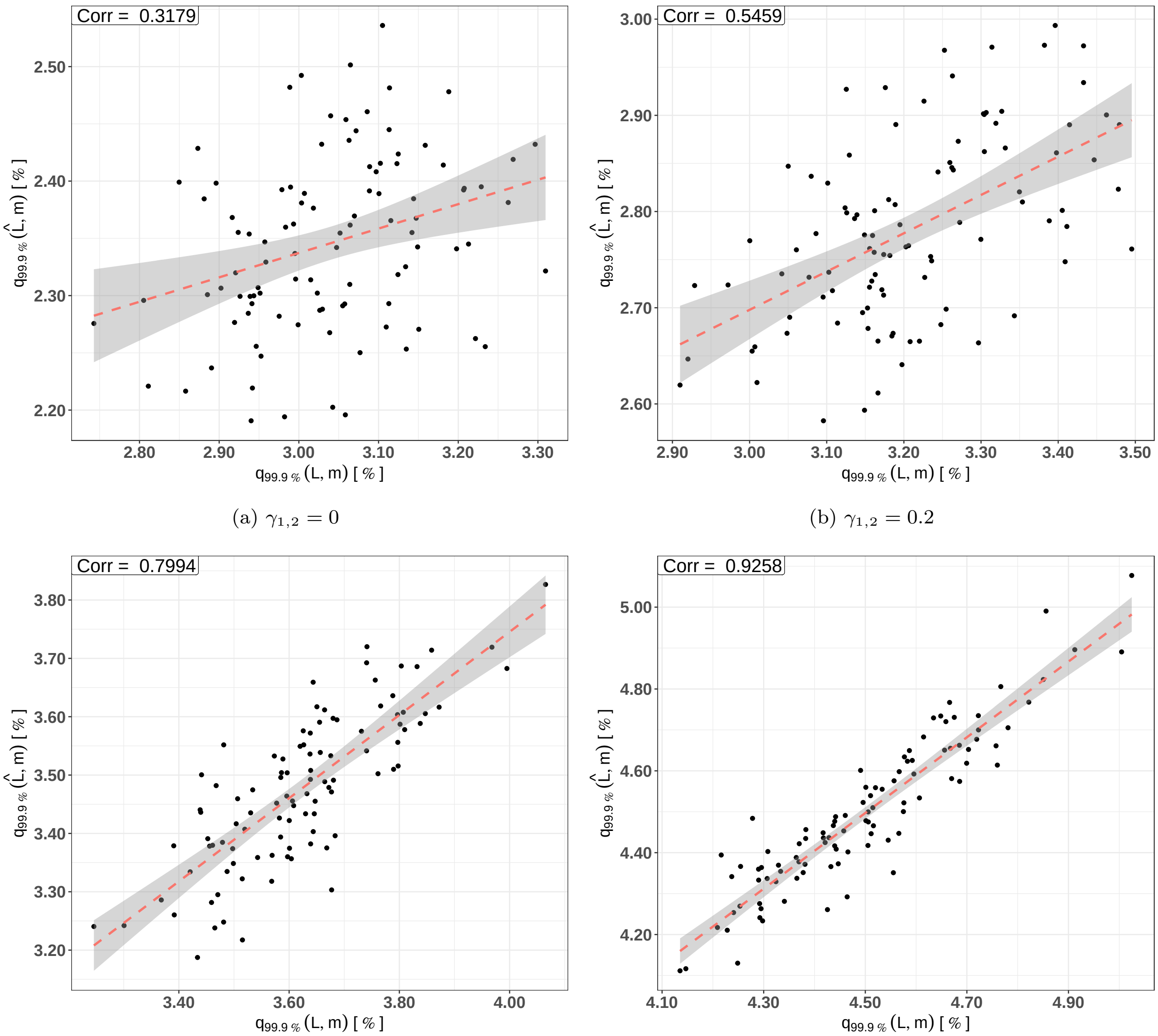
In the sequel, our numerical experiments are conducted by computing the sector factors $Y_1, ..., Y_S$ and the common factor $\hat{Y}$ by using the same set of sampled $Z_1, ...Z_K$ for each i-th simulation out of the m runs. We also tried sampling the $\boldsymbol{Z}$ vector conditionally upon $\boldsymbol{\beta}^\top \boldsymbol{Z} = \hat{Y}$. This involves initially drawing a value for $\hat{Y}$ in each i-th simulation run and then sampling the $Z_1, ...Z_K$ as detailed in Appendix B. We found nearly identical results for this two-factor exercise.[14] Additionally, as explained at the end of Section 5.2, we use the same idiosyncratic variables in both models, i.e., $\varepsilon_i = \hat{\varepsilon}_i$ in (21) and (22), respectively.

6.1. Quantile correlations

From Section 5, it is clear that the performance of our control variate estimator crucially depends on the correlation between our base ($q_\alpha(L,m)$) and control ($q_\alpha(\hat{L},m)$) estimators. The latter naturally depends on the pool parameters, most notably the inter-sector correlation which is noted as $\gamma_{1,2}$, in the two-factor setup at hand. Indeed, the higher the inter-sector correlation, the better the multi-factor structure in L is captured by the comparable single-factor structure in $\hat{L}$. This is the reason why we begin our analysis by looking at figures 3 and 4, which illustrate the scatter plots between these quantiles for both the homogeneous and heterogeneous portfolios, respectively. Each panel depicts the relationship for various levels of the inter-sector correlation $\gamma_{1,2}$.

The correlation coefficients between the quantiles, indicated at the top of each panel, increase notably from panel (a) where $\gamma_{1,2} = 0$, to panel (d) where $\gamma_{1,2} = 0.9$. These inter-quantile correlations span from approximately 31% to 100% in the homogeneous setting, and from around 42% to 100% in the heterogeneous setting. This highlights that for both portfolio compositions, the inter-quantile correlation increases with the inter-sector correlation $\gamma_{1,2}$. Notably, there is a slightly higher initial correlation between the quantiles in the heterogeneous (42.21%) *vs.* the homogeneous case (31.79%), when $\gamma_{1,2} = 0$.

[14] We expect results to differ when the number of considered sectors S increases.



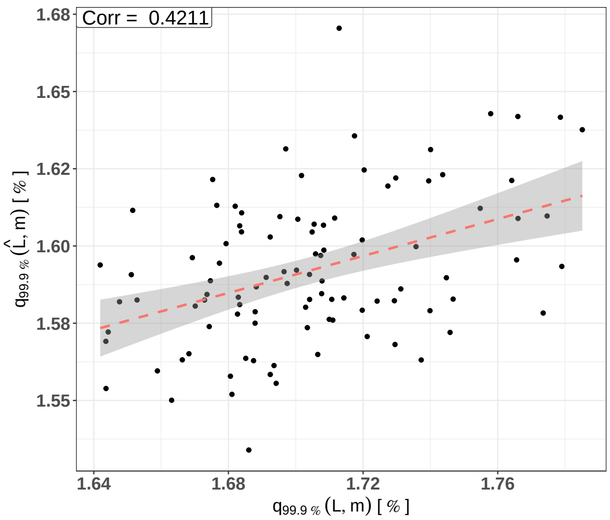
(a) $\gamma_{1,2} = 0$

(b) $\gamma_{1,2} = 0.2$

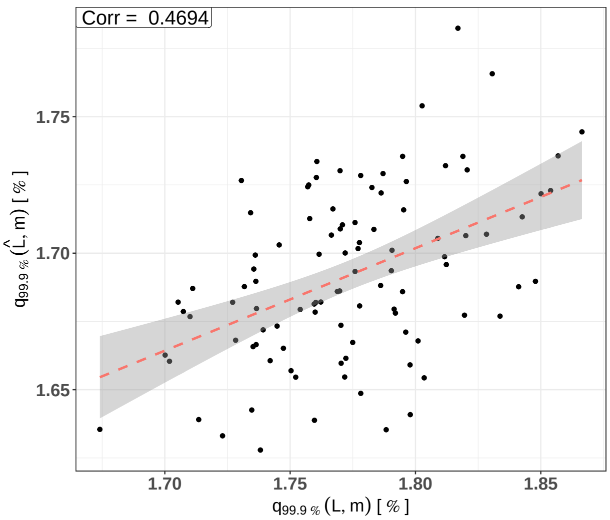
(c) $\gamma_{1,2} = 0.5$

(d) $\gamma_{1,2} = 0.9$

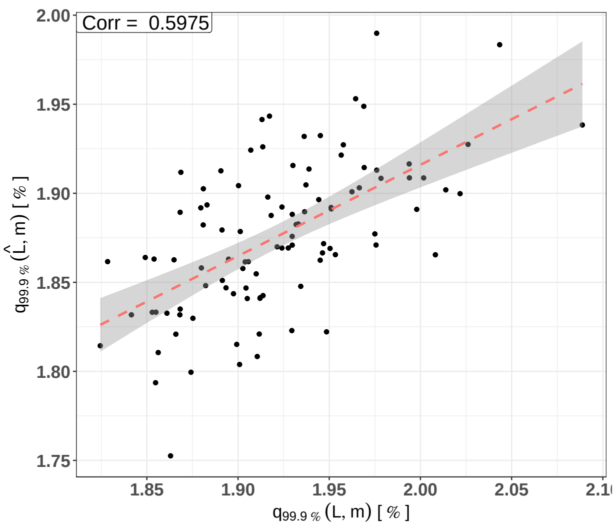
Figure 3: Scatter plots of the multi-factor quantile $q^i_\alpha(L, m)$ *vs.* the comparable one-factor quantile $q^i_\alpha(\hat{L}, m)$ in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. Each panel is drawn for a single value of the inter-sector correlation $\gamma_{1,2}$. The portfolio is a homogeneous pool with parameters $p = 0.5\%, \rho = 0.25, \lambda = 40\%, \sigma = 20\%$ and a finite size of $I = 1000$. The MC inputs are set to $m = 50,000$ and $M = 100$. In the upper corner of each scatter plot, we also include the correlation (noted as Corr) between the two quantiles. The trend lines (dashed lines) are computed using linear regression and the grey area represents the 95% confidence level interval for predictions from the linear model.



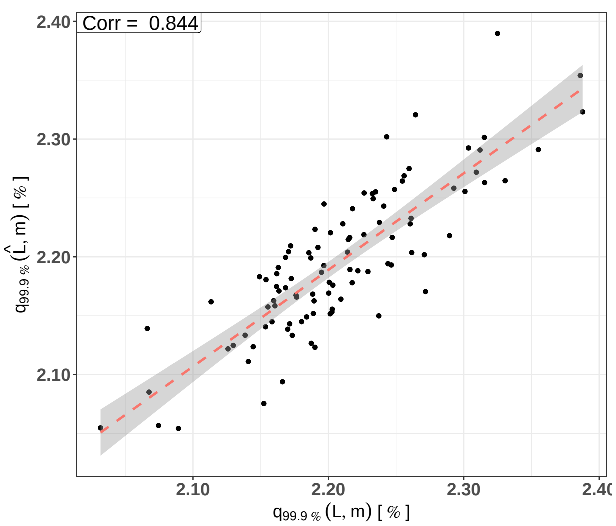
(a) $\gamma_{1,2} = 0$



(b) $\gamma_{1,2} = 0.2$



(c) $\gamma_{1,2} = 0.5$



(d) $\gamma_{1,2} = 0.9$

Figure 4: Scatter plots of the multi-factor quantile $q^i_\alpha(L, m)$ *vs.* the comparable one-factor quantile $q^i_\alpha(\hat{L}, m)$ in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. Each panel is drawn for a single value of the inter-sector correlation $\gamma_{1,2}$. The portfolio is a heterogeneous pool with parameters only defined at the sector-level: $(p_1, p_2) = (0.1\%, 2\%)$, $(\rho_1, \rho_2) = (0.25, 0.04)$ and $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$, with a finite size of $I = 1000$. The Monte Carlo inputs are set to $m = 50,000$ and $M = 100$. In the upper corner of each scatter plot, we also include the correlation (noted as Corr) between the two quantiles. The trend lines (dashed lines) are computed using linear regression and the grey area represents the 95% confidence level interval for predictions from the linear model.

6.2. Accuracy and variance reduction

In this section, we examine the performance of our set of alternative quantiles, outlined in Section 6, for a fixed value m of Monte Carlo simulations and across values of the inter-sector correlation $\gamma_{1,2}$.

Precisely, we compute the performance metrics featured in (50), (51) and (52) with $M = 100$ point estimates, each one obtained from $m = 50,000$ simulations, for varying values of $\gamma_{1,2}$. The results are shown in tables 1 and 2 for both the homogeneous and heterogeneous pools, respectively. Each row in these tables depicts the performance metrics computed from $M = 100$ point estimates, each one obtained from $m = 50,000$ simulations, at unique values of the inter-sector correlation $\gamma_{1,2}$. The columns beneath part (a) display the average estimated quantiles.[15] The columns beneath part (b) present the average accuracy rates as computed in (50). From these two sets of columns, we quickly notice the validity and usefulness of the bias reduction technique detailed in Section 5.3, underscoring the importance of using $\tilde{\theta}_{GA}$ rather than $\tilde{\theta}$, to capture $q_\alpha(L, m)$.[16]

For the homogeneous pool, the results in Table 1 show that, for $\gamma_{1,2}$ values ranging from 0 to 0.5, our control variate estimator $\tilde{\theta}_{GA}$ significantly outperforms the analytical approximation $\hat{L}^\infty(-z_\alpha)$, both with or without the MFA Δq_α. For these same lines, we observe a variance reduction spanning from -10.11% to -63% (when $\gamma_{1,2} = 0.5$). Hence, for the scenarios in which the gap between the approximation and the benchmark quantile is the largest, our control variate estimator effectively increases the precision of the point estimates while reducing its variance. This phenomenon is reinforced by looking at the columns below part (c), where we notice that the coefficient of variation for $\tilde{\theta}_{GA}$ is always smaller than that for $q_\alpha(L, m)$, this across all lines. The latter signals that $\tilde{\theta}_{GA}$ exhibits lower variability and higher level of precision and stability in its quantile estimates compared to $q_\alpha(L, m)$.

For the heterogeneous pool, the results in Table 2 show that $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$ captures very well the quantile $q_\alpha(L, m)$, this across all levels of $\gamma_{1,2}$. However, the precision of our control variate estimator $\tilde{\theta}_{GA}$ is superior across nearly all values of $\gamma_{1,2}$, the only exceptions are for $0.3, 0.7$ and 0.8. Notably, this increase in precision is again observed at lower inter-sector correlation values. For these same lines, we observe a variance reduction spanning from -17.73% to -35.71%, increasing up to -100% when $\gamma_{1,2} = 1$, which is accompanied by a consistently lower coefficient of variation for $\tilde{\theta}_{GA}$ compared to that for $q_\alpha(L, m)$.

At last, we observe that in both tables, when $\gamma_{1,2} = 1$ the coefficient of variation of our control variate estimator drops to 0. The latter should not come as a surprise, as when $\gamma_{1,2} = 1$, we have that $\sqrt{\hat{\rho}_{s(i)}} = \sqrt{\rho_{s(i)}}$ and since we imposed $\varepsilon_i = \hat{\varepsilon}_i$, we get $q_\alpha(L, m) = q_\alpha(\hat{L}, m)$ for every one of the M runs.[17] Hence, the covariance in (46) is also 1, leading to the $\tilde{\theta}_{GA}$ in (49) being only equal to: $\hat{L}^\infty(-z_\alpha) + \text{GA}$.

[15] Of course, for $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$, the average reduces to a single value.

[16] This is further supported by the panels (a) and (c) of Figure 2, which illustrates that, for the chosen pool parameters and $I = 1000$, the GA is not yet 0.

[17] This is not the case if we sample different idiosyncratic terms.

$\gamma_{1,2}$	(a) *Average quantiles* $q_\alpha(L,m)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(b) *Average accuracy* $\hat{L}^\infty(-z_\alpha)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(c) *Coefficient of variation* $q_\alpha(L,m)$	$\tilde{\theta}_{GA}$	VR
0	3.04	2.80	3.00	3.04	74.21	92.20	98.60	99.95	3.75	3.56	-10.11
0.10	3.11	2.93	3.06	3.11	79.51	94.15	98.54	99.97	3.53	3.27	-14.43
0.20	3.21	3.07	3.14	3.21	83.95	95.59	97.98	99.98	4.13	3.46	-29.81
0.30	3.33	3.22	3.27	3.33	87.83	96.93	98.16	99.94	4.02	3.25	-34.55
0.40	3.47	3.40	3.40	3.47	90.99	97.94	98.15	99.94	4.03	2.89	-48.60
0.50	3.62	3.58	3.55	3.62	93.71	98.93	97.97	99.90	4.05	2.44	-63.91
0.60	3.80	3.79	3.73	3.80	95.68	99.53	98.06	99.81	3.98	2.06	-73.28
0.70	4.01	4.00	3.93	4.00	97.06	99.88	98.17	99.90	4.19	2.03	-76.53
0.80	4.24	4.24	4.17	4.24	97.79	99.88	98.39	99.92	4.08	1.62	-84.29
0.90	4.50	4.48	4.42	4.48	98.02	99.66	98.23	99.66	4.16	1.58	-85.70
1	4.77	4.75	4.68	4.75	98.00	99.43	98.00	99.43	4.63	0	-100

Table 1: Average estimated quantile (a), average accuracy rate (b) and coefficient of variation (c) and variance reduction (VR), in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. The portfolio is a homogeneous pool with parameters $p = 0.5\%, \rho = 0.25, \lambda = 40\%, \sigma = 20\%$ and a finite size of $I = 1000$. The Monte Carlo inputs are set to $m = 50,000$ and $M = 100$. All values are expressed as percentages, except the inter-sector correlation $\gamma_{1,2}$.

$\gamma_{1,2}$	(a) *Average quantiles* $q_\alpha(L,m)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(b) *Average accuracy* $\hat{L}^\infty(-z_\alpha)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(c) *Coefficient of variation* $q_\alpha(L,m)$	$\tilde{\theta}_{GA}$	VR
0	1.70	1.69	1.59	1.71	82.50	99.08	93.53	99.90	2.00	1.81	-17.73
0.10	1.74	1.72	1.63	1.74	84.04	99.31	94.07	99.99	2.10	1.89	-18.44
0.20	1.77	1.77	1.67	1.77	85.62	99.66	94.25	99.93	2.19	1.94	-22.03
0.30	1.82	1.82	1.72	1.81	87.17	99.97	94.83	99.93	2.25	2.00	-20.91
0.40	1.87	1.87	1.78	1.86	88.55	99.78	95.36	99.87	2.43	2.19	-18.97
0.50	1.92	1.93	1.82	1.92	89.79	99.72	94.55	99.83	2.56	2.05	-35.71
0.60	1.98	1.99	1.88	1.98	90.97	99.71	94.59	99.81	2.50	1.87	-44.45
0.70	2.06	2.06	1.94	2.06	91.82	99.95	94.20	99.91	2.75	1.89	-52.91
0.80	2.13	2.13	2.02	2.13	92.88	99.87	94.74	99.80	2.77	1.76	-59.82
0.90	2.21	2.20	2.10	2.20	93.89	99.76	95.18	99.76	2.95	1.58	-71.24
1	2.29	2.28	2.17	2.28	94.94	99.75	94.94	99.75	3.15	0	-100

Table 2: Average estimated quantile (a), average accuracy rate (b) and coefficient of variation (c) and variance reduction (VR), in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. The portfolio is a heterogeneous pool with parameters only defined at the sector level: $(p_1, p_2) = (0.1\%, 2\%)$, $(\rho_1, \rho_2) = (0.25, 0.04)$ and $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$, with a finite size of $I = 1000$. The Monte Carlo inputs are set to $m = 50,000$ and $M = 100$. All values are expressed as percentages, except the inter-sector correlation $\gamma_{1,2}$.

m	(a) *Average quantiles* $q_\alpha(L,m)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(b) *Average accuracy* $\hat{L}^\infty(-z_\alpha)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(c) *Coefficient of variation* $q_\alpha(L,m)$	$\tilde{\theta}_{GA}$	VR
50,000	3.61	3.58	3.54	3.61	93.92	99.15	97.86	99.82	2.88	2.50	-24.80
100,000	3.62	3.58	3.56	3.63	93.72	98.94	98.34	99.88	2.83	1.99	-50.68
150,000	3.63	3.58	3.55	3.62	93.58	98.79	98.00	99.88	2.05	1.52	-45.23
200,000	3.62	3.58	3.56	3.62	93.81	99.04	98.28	99.98	1.95	1.25	-59.19
250,000	3.62	3.58	3.56	3.62	93.79	99.01	98.27	99.91	1.69	1.15	-53.91
300,000	3.62	3.58	3.56	3.62	93.67	98.89	98.16	99.85	1.63	1.18	-47.68
350,000	3.62	3.58	3.56	3.62	93.67	98.88	98.14	99.91	1.47	0.98	-56.04
500,000	3.61	3.58	3.56	3.62	93.96	99.19	98.45	99.87	1.21	0.77	-59.31

Table 3: Average estimated quantile (a), average accuracy rate (b) and coefficient of variation (c) and variance reduction (VR), in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. The portfolio is a homogeneous pool with parameters $p = 0.5\%, \rho = 0.25, \lambda = 40\%, \sigma = 20\%, \gamma_{1,2} = 0.5$ and a finite size of $I = 1000$. The results are computed from setting $M = 100$ and $\gamma_{1,2} = 0.5$. All values are expressed as percentages, except the number of simulations m.

6.3. Reduced number of simulations

In this section, we build upon the analysis from Section 6.2 by further evaluating the performance of the same set of alternative quantiles. This time, we fix the value of $\gamma_{1,2}$ at a prototypical value of 0.5, and vary the m number of MC simulations.

Results are presented in tables 3 and 4 for the homogeneous and heterogeneous pools, respectively. Each row in these tables displays the performance metrics computed from $M = 100$ point estimates, for a unique value of m. Consistent with the findings in Section 6.2, we observe a variance reduction spanning between -35.11% and -61% in the homogeneous pool, and spanning between -36.04% and $-26,74\%$ in the heterogeneous pool. These VR values align closely with the VR values found in tables 1 and 2 for the line featuring $\gamma_{1,2} = 0.5$.

Moreover, the accuracy of $\tilde{\theta}_{GA}$ is always superior to that of the analytical approximation $\hat{L}^\infty(-z_\alpha) + \Delta q_\alpha$, across all values of m, and the coefficient of variation for $\tilde{\theta}_{GA}$ is always strictly smaller than that for $q_\alpha(L, m)$.

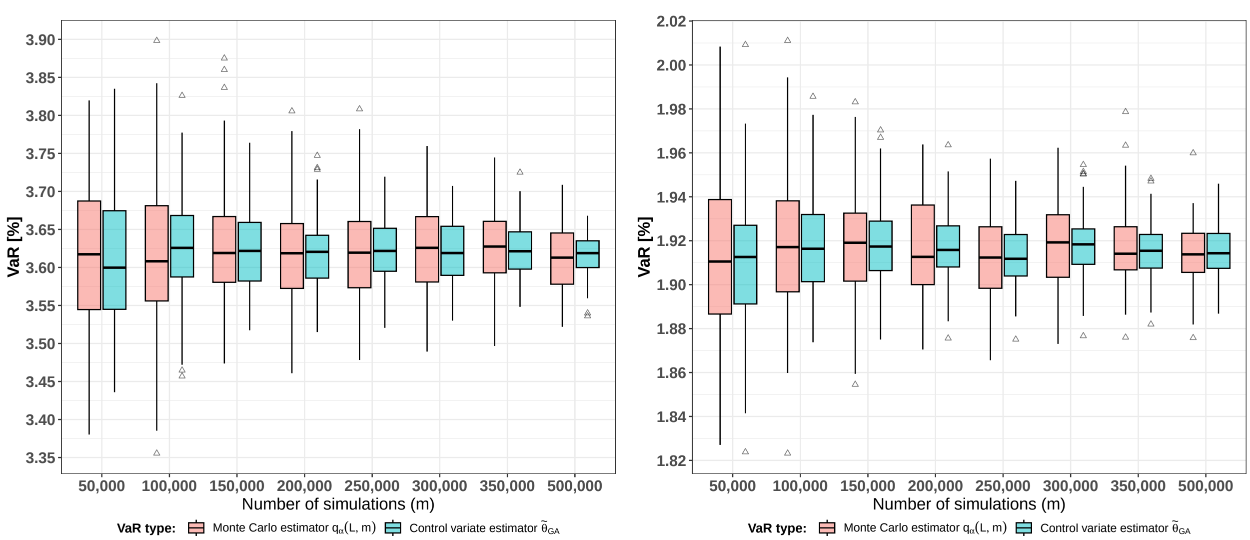
For a graphical illustration of these dynamics, refer to Figure 5, the boxplots of which visually depict the variance reduction, accuracy, and consistency of the performance metrics across varying m MC simulations.

6.4. Complexity vs. simulation time

Evaluating the performance of our control variate (CV) estimator $\tilde{\theta}_{GA}$ in (49) compared to the benchmark MC estimator $q_\alpha(L, m)$ in (43), solely in terms of variance for a fixed number m of MC simulations, does not provide an exhaustive assessment. Indeed, both estimators have different degrees of complexity as, most notably, $\tilde{\theta}_{GA}$ requires the computation of both $q_\alpha(L, m)$ and $q_\alpha(\hat{L}, m)$. The latter difference leads to different simulation times, between both estimators, to reach the same variance of the M VaR point estimates. The increased precision from variance reduction may come at the expense of longer simulation times. Therefore, it is essential to investigate the trade-off between the increased complexity of the CV estimator and the alternative approach of simply increasing the simulation time for the benchmark MC estimator.

m	(a) *Average quantiles* $q_\alpha(L,m)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(b) *Average accuracy* $\hat{L}^\infty(-z_\alpha)$	$\hat{L}^\infty(-z_\alpha)+\Delta q_\alpha$	$\tilde{\theta}$	$\tilde{\theta}_{GA}$	(c) *Coefficient of variation* $q_\alpha(L,m)$	$\tilde{\theta}_{GA}$	VR
$50,000$	1.91	1.93	1.78	1.91	90.21	99.25	93.00	99.88	1.93	1.68	-24.35
$100,000$	1.92	1.93	1.80	1.92	90.00	99.49	93.98	99.96	1.70	1.26	-45.46
$150,000$	1.92	1.93	1.80	1.92	89.97	99.52	93.63	99.99	1.38	1.01	-46.67
$200,000$	1.92	1.93	1.81	1.92	90.02	99.46	94.30	99.97	1.19	0.85	-49.09
$250,000$	1.91	1.93	1.80	1.91	90.23	99.23	94.03	99.97	1.02	0.72	-49.37
$300,000$	1.92	1.93	1.81	1.92	89.99	99.50	94.30	99.99	1.08	0.75	-51.86
$350,000$	1.92	1.93	1.80	1.92	90.02	99.46	94.14	99.95	0.92	0.64	-50.84
$500,000$	1.91	1.93	1.81	1.92	90.21	99.25	94.57	99.89	0.73	0.64	-23.41

Table 4: Average estimated quantile (a), average accuracy rate (b) and coefficient of variation (c) and variance reduction (VR), in a two-factor setup, with equal repartition of the obligors across the sectors and for $\alpha = 99.9\%$. The portfolio is a heterogeneous pool with parameters only defined at the sector level: $(p_1, p_2) = (0.1\%, 2\%)$, $(\rho_1, \rho_2) = (0.25, 0.04)$ and $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$, with a finite size of $I = 1000$. The results are computed from setting $M = 100$ and $\gamma_{1,2} = 0.5$. All values are expressed as percentages, except the number of simulations m.



(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.

(b) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

Figure 5: VaR boxplots, as a function of the number of simulations m, in a two-factor setup, with equal repartition of the obligors across the sectors. The panels are drawn for a fixed value of the inter-sector correlation of $\gamma_{1,2} = 0.5$. The LGD is a $Beta(2, 3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups. For every m on the x-axis, the boxplots are composed of $M = 100$ VaRs. In the panels, the triangles ($\triangle$) depict the outliers.

First, we examine the standard deviations of both estimators as a function of the number of simulations m. This relationship is shown in Figure 6 for both the considered homogeneous and heterogeneous pools. The dots and triangles represent the actually computed values of the standard deviation (SD), for both estimators, across a set of simulation sizes m (the same as in tables 3 and 4). Theoretically, we know that the SD of an estimator decreases proportionally to the inverse square root of the sample size (here simulation size m). Therefore, without running additional MC simulations, we can interpolate the full relationship thanks to a log-log linear model of the SD as a function of the simulation size m, i.e., $\ln(\text{SD}) = a + b\ln(m)$, where a, b are the model coefficients.[18] The fitted values fill the gap between $m = 250,000$ and $m = 500,000$, providing a relationship that is not subject to simulation noise, particularly for smaller sizes of m.[19] In both cases, the standard deviation of our CV estimator is consistently smaller than that of the benchmark MC estimator. This difference is more pronounced in the homogeneous case compared to the heterogeneous case.

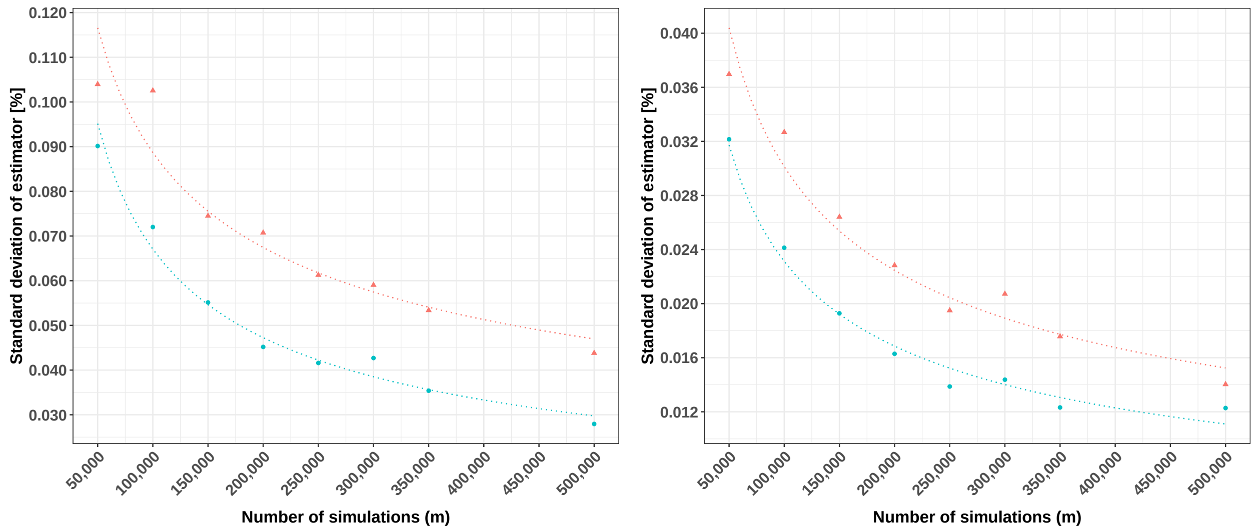
From these graphs, we determine the equivalent number of MC simulations m required by the benchmark MC estimator to achieve the same standard deviation as the CV estimator. This relationship is illustrated in Figure 7, where the dashed diagonal represents the scenario where both estimators require the same number of MC simulations m to reach the same level of standard deviation. Ideally, all points should lie above the diagonal, indicating that the benchmark MC estimator requires more simulations m to achieve the same standard deviation as the CV estimator. In the homogeneous pool, the benchmark MC estimator needs approximately double the number of simulations. When $m = 250,000$, we see that not even the double number of MC simulations for $q_\alpha(L, m)$ are sufficient to reach the same standard deviation as in $\tilde{\theta}_{GA}$. As our largest simulation is executed with $m = 500,000$, we can resort to the same interpolation method explained in the previous paragraph (or cubic spline alternatively) to estimate that for $m = 250,000$, the equivalent number of simulations required by the benchmark MC estimator is approximately $m = 700,000$. In the heterogeneous pool, the increase is more moderate, but all points remain well above the diagonal.

Finally, we evaluate the impact of this increased number of simulations on the required computation time. Figure 8 compares the simulation times needed to obtain 1 VaR as a function of the simulation size m for both estimators.[20] As expected, the CV estimator requires more time than the benchmark MC estimator due to its 'greater' complexity. However, a proper comparison should not be based on the simulation times for the same value of m. Instead, one should compare the simulation times required to achieve the same level of precision across both estimators, here chosen to be the same level of SD. To this end, we combine the information from Figure 7 and Figure 8. For the homogeneous pool, we see

[18] We verify empirically that the slope b of the fitted model is close to -0.5 which is consistent with the theoretical relationship.

[19] We also tried interpolation via cubic splines and obtained very similar results.

[20] The simulation times do not differ between the homogeneous and heterogeneous pools.



(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.

(b) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

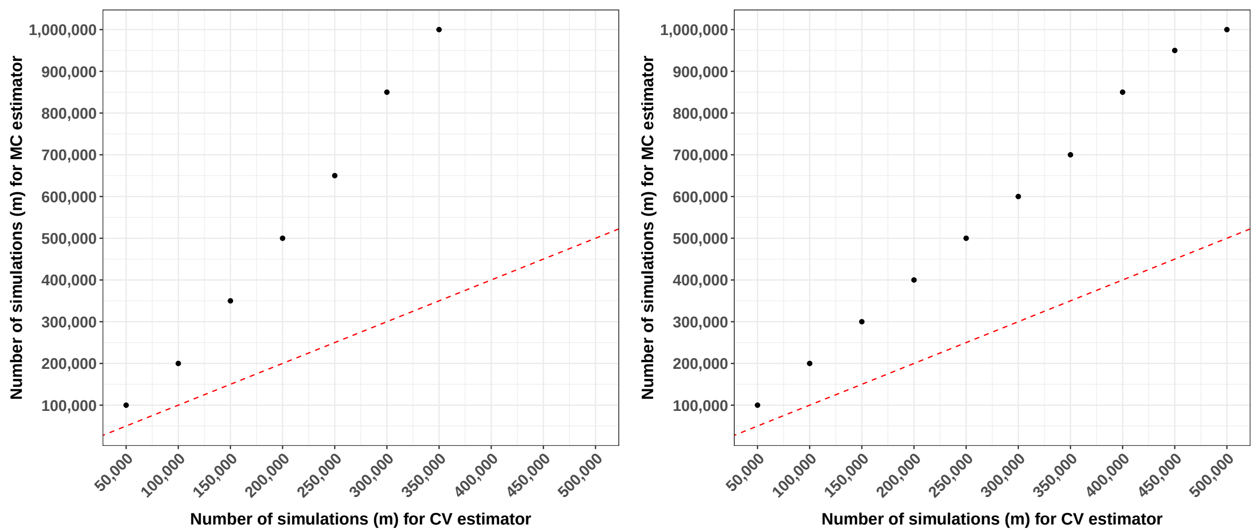
Figure 6: Standard deviation as a function of the simulation size m for the benchmark MC estimator $q_\alpha(L, m)$ (▲) and the CV estimator $\tilde{\theta}_{GA}$ (●). The dotted lines are obtained by interpolation using a log-log linear regression model of the standard deviation as a function of the simulation size m. The standard deviation is computed out of $M = 100$ VaRs. The panels are drawn for a fixed value of the inter-sector correlation of $\gamma_{1,2} = 0.5$. The LGD is a $Beta(2, 3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups.

that to reach the precision obtained by setting $m = 250,000$ (approx. 1,5 minutes) in $\tilde{\theta}_{GA}$, we need approximately $m = 650,000$ in $q_\alpha(L, m)$ (approx. 3 minutes), effectively more than doubling the simulation time. This demonstrates that while the control variate estimator is more complex, it offers significant computational efficiency when the same level of precision is required. A similar, albeit less intense, phenomenon is observed for the heterogeneous pool. To reach the precision obtained by setting $m = 250,000$ (approx. 1,5 minutes) in $\tilde{\theta}_{GA}$, we need approximately $m = 500,000$ in $q_\alpha(L, m)$ (approx. 2,25 minutes). Figure 9 extends this analysis across all our simulation sizes m for our CV estimator. This figure shows the simulation times required by both estimators, to reach the same level of SD, when the CV estimator is run with m simulations. This comparison underscores the trade-off between complexity and computation time, highlighting the efficiency of the control variate estimator in achieving precise results more quickly.

7. Conclusion

The credit risk arising from concentration of exposures to a group of borrowers, pertaining to same geographic region or industrial sector is arguably the single most important cause of major problems for a typical commercial bank with a medium to a large size banking book.

The Basel Committee recognises the significance of this type of credit concentration risk. However, to maintain the *simplicity* of the minimum capital requirements formulas under



(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.

(b) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

Figure 7: Number of simulations m required to achieve the same standard deviation between the benchmark MC estimator $q_\alpha(L, m)$ (on the y-axis) and CV estimator $\tilde{\theta}_{GA}$ (on the x-axis). The black points are obtained by interpolation from the dotted lines in Figure 6. The panels are drawn for a fixed value of the inter-sector correlation of $\gamma_{1,2} = 0.5$. The LGD is a $Beta(2, 3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups.

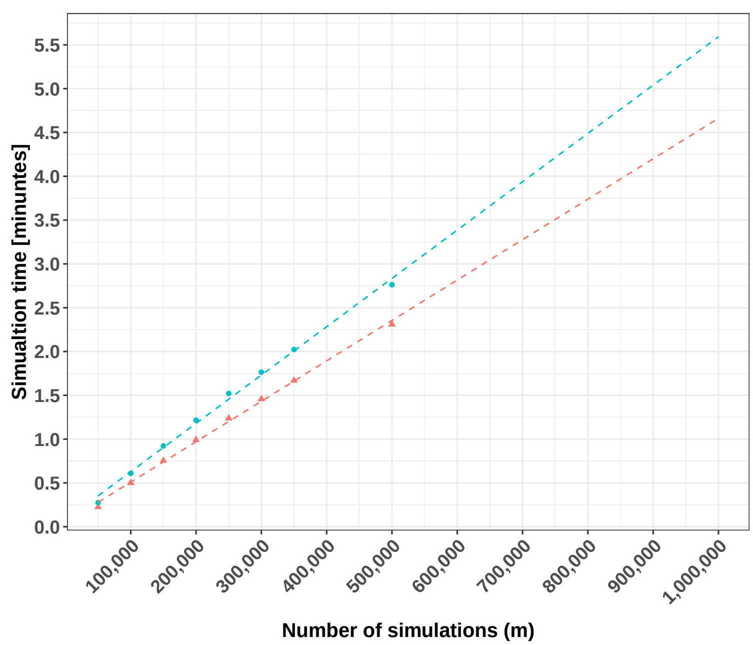
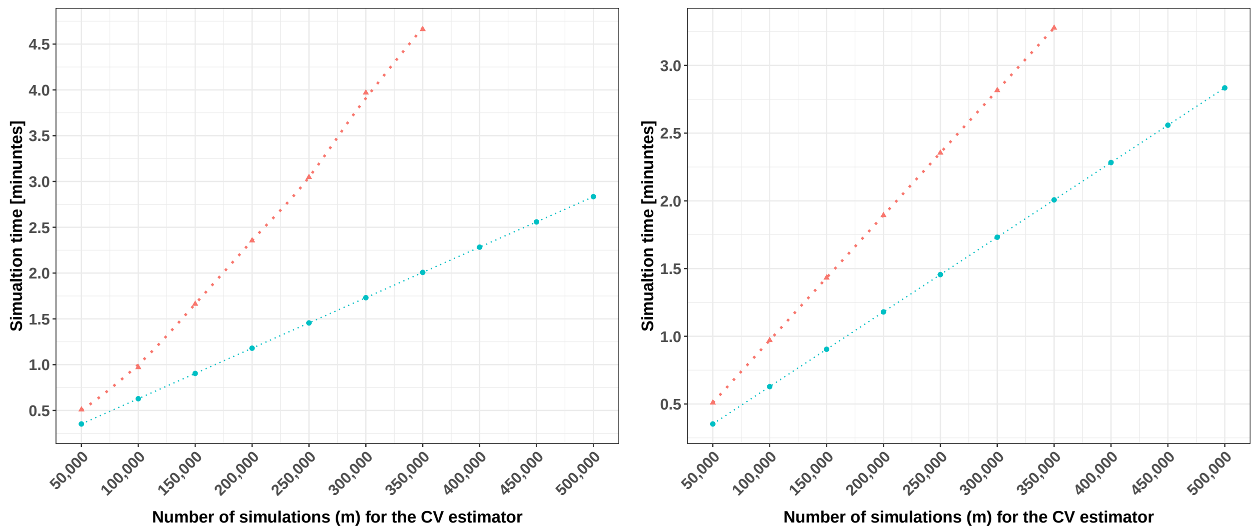


Figure 8: Required simulation time for the benchmark MC estimator $q_\alpha(L, m)$ (▲) and for the CV estimator $\tilde{\theta}_{GA}$ (●) to get $M = 1$ VaR, as a function of the simulation size m. The simulation times are the same for both the homogeneous and the heterogeneous pools considered here. The points and triangle correspond to the computed times, while the dotted lines are obtained by linear interpolation. The panels are drawn for a fixed value of the inter-sector correlation of $\gamma_{1,2} = 0.5$. The LGD is a $Beta(2, 3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups.



(a) Homogeneous pool: $p = 0.5\%, \rho = 0.25$.

(b) Heterogeneous pool: $(p_1, p_2) = (0.1\%, 2\%)$ and $(\rho_1, \rho_2) = (0.25, 0.04)$.

Figure 9: Required simulation time for the benchmark MC estimator $q_\alpha(L, m)$ (▲) and for the CV estimator $\tilde{\theta}_{GA}$ (●) to get $M = 1$ VaR, as a function of the simulation size m for the CV estimator. In panel (a), the dotted line is obtained by cubic spline, while in panel (b) it is obtained by linear interpolation. The panels are drawn for a fixed value of the inter-sector correlation of $\gamma_{1,2} = 0.5$. The LGD is a $Beta(2, 3)$, leading to set $\lambda_1 = \lambda_2 = \lambda = 40\%$ and $\sigma_1 = \sigma_2 = \sigma = 20\%$ in both the homogeneous and heterogeneous setups.

the Pillar 1 section of the Accords, the Committee has chosen not to include this specific subcategory of credit risk within the regulatory model. Precisely, under the IRB approach found in Pillar 1, credit risk is measured via an asymptotic single risk factor (ASRF) model which postulates that dependence between default events only arises due to the presence of a single systematic risk factor. The practical benefits of this setup are twofold: a closed-form expression is obtained for the MCR and the model exhibits the so-called portfolio-invariance property asymptotically, which means that the total portfolio VaR is the sum of individual marginal contributions that solely depends on the risk-profile of each borrower. The price to pay, however, is that the ASRF model fails to capture concentration risk.

The systematic component driving the portfolio losses may be sensitive to the economic conditions in which the exposures operate, requiring to insert multiple systematic risk factors. Hence, the joint default likelihood of two borrowers may not be uniquely determined by their common sensitivity to a single systematic risk factor representing the overall state of the economy. The quantitative measurement of credit concentration risk is relegated to the Pillar 2 section, where differently than for the measurement of credit risk under Pillar 1, no regulatory formula or credit risk model is prescribed. National competent authorities must conduct an individual assessment of capital adequacy to concentration risk bank-by-bank and model-by-model according to the guidelines outlined in Pillar 2.

Our research contributes to the sound measurement of credit concentration risk, by providing a stress-testing or back-testing tool suitable for assessing capital adequacy to credit concentration risk fitting the Pillar 2 guidelines. To model a richer correlation structure than the one available in the single-factor model we rely on a well acknowledged Gaussian multi-factor extension. In such a setup, the multi-factor VaR can either be computed numerically or be approximated via a comparable one-factor model plus an adjustment term. The numerical estimate is precise but requires considerable amount of time and computational power, while the performance of the analytical approximation varies depending on the credit risk profile of the pool at hand. For this reason, we introduce a control variate estimator blending the power of the analytical approximation of the multi-factor VaR with the precision of the Monte Carlo estimator.

An empirical analysis is conducted to evaluate the performance of our control variate estimator against a series of alternative quantiles. To that end, we select two synthetic portfolios aligned with the existing literature and featuring a two systematic risk factor structure. For both the considered homogeneous and heterogeneous pools, we achieve a consistent variance reduction and increased precision across values of the inter-sector correlation, for a fixed number of Monte Carlo simulations. We run a similar analysis for a fixed value of the inter-sector correlation and for an increasing number of Monte Carlo simulations and the results hold.

8. Acknowledgments

The work of Frédéric Vrins is supported by the Fonds de la Recherche Scientifique - (F.S.R.-FNRS) grant CDR J.0225.24. Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the F.R.S.-FNRS under convention 2.5020.11 and by the Walloon Region.

Appendix

A. Conditional asset correlation

To compute the conditional asset correlation featured in (31), we have to express X_i in (16) as driven by the single factor $\hat{Y}$. To that end, start by recalling that in (16), the latent random variable driving the default event is expressed as

$$X_i = \sqrt{\rho_{s(i)}} Y_{s(i)} + \sqrt{1-\rho_{s(i)}}\varepsilon_i$$

and from the definition of our comparable single risk factor recover that

$$\hat{Y} = \sum_{k=1}^{K} \beta_k Z_k \;\Leftrightarrow\; \hat{Y} - \sum_{k=1}^{K} \beta_k Z_k = 0 \;\Leftrightarrow\; \sqrt{\hat{\rho}_{s(i)}} \left(\hat{Y} - \sum_{k=1}^{K} \beta_k Z_k \right) = 0$$

leading to

$$X_i = \sqrt{\rho_{s(i)}} \left(Y_{s(i)} + \underbrace{\sqrt{\hat{\rho}_{s(i)}} \left(\hat{Y} - \sum_{k=1}^{K} \beta_k Z_k \right)}_{=0} \right) + \sqrt{1-\rho_{s(i)}}\varepsilon_i \, .$$

Following, we also express $Y_{s(i)}$ in terms of the same i.i.d. $Z_1, ..., Z_K$ that are driving $\hat{Y}$, leading to

$$\begin{aligned} X_i &= \sqrt{\rho_{s(i)}} \left(\sum_{k=1}^{K} \alpha_{s(i),k} Z_k + \underbrace{\sqrt{\hat{\rho}_{s(i)}} \left(\hat{Y} - \sum_{k=1}^{K} \beta_k Z_k \right)}_{=0} \right) + \sqrt{1-\rho_{s(i)}}\varepsilon_i \\ &= \sqrt{\hat{\rho}_{s(i)}}\hat{Y} + \sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\alpha_{s(i),k} Z_k - \sqrt{\hat{\rho}_{s(i)}} \sum_{k=1}^{K} \beta_k Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i \\ &= \sqrt{\hat{\rho}_{s(i)}}\hat{Y} + \sum_{k=1}^{K} \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k \right) Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i \, , \end{aligned} \tag{A.1}$$

where $f(Z_k) := \sqrt{\hat{\rho}_{s(i)}}\hat{Y}$ is independent of $g(Z_k) := \sum_{k=1}^{K} \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k \right) Z_k$.

To prove this relationship, it is sufficient to recall that $(f(Z_k), g(Z_k))$ are bivariate normal. Thus, to prove independence, it is sufficient to show that their covariance is equal to 0. Precisely,

$$
\begin{aligned}
\mathbb{C}\text{ov}\left[f(Z_k)|g(Z_k)\right] &= \mathbb{C}\text{ov}\left[\sqrt{\hat{\rho}_{s(i)}}\hat{Y}, \sum_{k=1}^{K}\left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k\right) Z_k\right] \\
&= \mathbb{C}\text{ov}\left[\sqrt{\hat{\rho}_{s(i)}}\sum_{k=1}^{K}\beta_k Z_k, \sum_{k=1}^{K}\left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k\right) Z_k\right] \\
&= \sum_{k=1}^{K}\sum_{t=1}^{K}\mathbb{C}\text{ov}\left[\sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k, \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),t} - \sqrt{\hat{\rho}_{s(i)}}\beta_t\right) Z_t\right] \\
&= \sum_{k=1}^{K}\mathbb{C}\text{ov}\left[\sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k, \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k\right) Z_k\right] \\
&= \sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k \times \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k\right)\underbrace{\mathbb{V}[Z_k]}_{=1} \\
&= \sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k \times \left(\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sqrt{\hat{\rho}_{s(i)}}\beta_k\right) \\
&= \sum_{k=1}^{K}\beta_k\sqrt{\hat{\rho}_{s(i)}}\sqrt{\rho_{s(i)}}\alpha_{s(i),k} - \sum_{k=1}^{K}\hat{\rho}_{s(i)}\beta_k^2 \\
&= \sqrt{\hat{\rho}_{s(i)}}\underbrace{\sum_{k=1}^{K}\beta_k\sqrt{\rho_{s(i)}}\alpha_{s(i),k}}_{=\sqrt{\hat{\rho}_{s(i)}}} - \sum_{k=1}^{K}\hat{\rho}_{s(i)}\beta_k^2 \\
&= \hat{\rho}_{s(i)} - \hat{\rho}_{s(i)}\underbrace{\sum_{k=1}^{K}\beta_k^2}_{=1} \\
&= 0\ .
\end{aligned}
$$

We can now compute the analytical expression of the conditional asset correlation as

$$
\mathbb{C}\text{orr}(X_i, X_j|\hat{Y}) = \frac{\mathbb{C}\text{ov}(X_i, X_j|\hat{Y})}{\sqrt{\mathbb{V}(X_i|\hat{Y})} \times \sqrt{\mathbb{V}(X_j|\hat{Y})}}\ . \tag{A.2}
$$

A.1. Covariance

The conditional covariance is computed as

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(X_i, X_j | \hat{Y}) &= \mathbb{C}\mathrm{ov}\left(\sqrt{\hat{\rho}_{s(i)}}\hat{Y} + \sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\alpha_{s(i),k} Z_k - \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i, \right. \\
&\qquad \left. \sqrt{\hat{\rho}_{s(j)}}\hat{Y} + \sum_{k=1}^{K} \sqrt{\rho_{s(j)}}\alpha_{s(j),k} Z_k - \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(j)}}\beta_k Z_k + \sqrt{1-\rho_{s(j)}}\varepsilon_j \,\middle|\, \hat{Y}\right) \\
&= \mathbb{C}\mathrm{ov}\left(\sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\alpha_{s(i),k} Z_k - \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i, \right. \\
&\qquad \left. \sum_{k=1}^{K} \sqrt{\rho_{s(j)}}\alpha_{s(j),k} Z_k - \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(j)}}\beta_k Z_k + \sqrt{1-\rho_{s(j)}}\varepsilon_j\right) \\
&= \left(\sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\sqrt{\rho_{s(j)}}\alpha_{s(i),k}\alpha_{s(j),k}\right) \mathbb{V}(Z_k) \\
&\quad - \left(\sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\alpha_{s(i),k} \times \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(j)}}\beta_k\right) \mathbb{V}(Z_k) \\
&\quad - \left(\sum_{k=1}^{K} \sqrt{\rho_{s(j)}}\alpha_{s(j),k} \times \sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(i)}}\beta_k\right) \mathbb{V}(Z_k) \\
&\quad + \left(\sum_{k=1}^{K} \sqrt{\hat{\rho}_{s(i)}\hat{\rho}_{s(j)}}\beta_k^2\right) \mathbb{V}(Z_k) \\
&= \left(\sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\sqrt{\rho_{s(j)}}\alpha_{s(i),k}\alpha_{s(j),k}\right) - \left(\sqrt{\rho_{s(i)}}\sqrt{\hat{\rho}_{s(j)}}\sum_{k=1}^{K} \alpha_{s(i),k}\beta_k\right) \\
&\quad - \left(\sqrt{\rho_{s(j)}}\sqrt{\hat{\rho}_{s(i)}}\sum_{k=1}^{K} \alpha_{s(j),k}\beta_k\right) + \left(\sqrt{\hat{\rho}_{s(i)}\hat{\rho}_{s(j)}}\underbrace{\sum_{k=1}^{K} \beta_k^2}_{=1}\right) \\
&= \left(\sum_{k=1}^{K} \sqrt{\rho_{s(i)}}\sqrt{\rho_{s(j)}}\alpha_{s(i),k}\alpha_{s(j),k}\right) + \sqrt{\hat{\rho}_{s(i)}}\sqrt{\hat{\rho}_{s(j)}}\;, \qquad \text{(A.3)}
\end{aligned}$$

where we use the fact that $\sum_{k=1}^{K} \alpha_{s,k}\beta_k = \sqrt{\hat{\rho}_s}/\sqrt{\rho_s}$.

A.2. Variance

The conditional variance is computed as

$$
\begin{aligned}
\mathbb{V}(X_i|\hat{Y}) &= \mathbb{V}\left(\sqrt{\hat{\rho}_{s(i)}}\hat{Y} + \sum_{k=1}^{K}\sqrt{\rho_{s(i)}}\alpha_{s(i),k}Z_k - \sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i \middle| \hat{Y}\right) \\
&= \mathbb{V}\left(\sum_{k=1}^{K}\sqrt{\rho_{s(i)}}\alpha_{s(i),k}Z_k - \sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k + \sqrt{1-\rho_{s(i)}}\varepsilon_i\right) \\
&= \mathbb{V}\left(\sum_{k=1}^{K}\sqrt{\rho_{s(i)}}\alpha_{s(i),k}Z_k\right) + \mathbb{V}\left(\sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k Z_k\right) \\
&\quad -2\left(\sum_{k=1}^{K}\sqrt{\rho_{s(i)}}\alpha_{s(i),k} \times \sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k\right)\underbrace{\mathbb{C}\mathrm{ov}\left(Z_k, Z_k\right)}_{=\mathbb{V}(Z_k)} \\
&= \left(\sum_{k=1}^{K}\sqrt{\rho_{s(i)}}\alpha_{s(i),k}\right)^2 \mathbb{V}\left(Z_k\right) + \left(\sum_{k=1}^{K}\sqrt{\hat{\rho}_{s(i)}}\beta_k\right)^2 \mathbb{V}\left(Z_k\right) \\
&\quad -2\left(\sqrt{\rho_{s(i)}}\sum_{k=1}^{K}\alpha_{s(i),k} \times \sqrt{\hat{\rho}_{s(i)}}\sum_{k=1}^{K}\beta_k\right)\mathbb{V}(Z_k) \\
&= \rho_{s(i)} + \hat{\rho}_{s(i)} + 1 - \rho_{s(i)} - 2\left(\sqrt{\rho_{s(i)}}\sqrt{\hat{\rho}_{s(i)}}\sum_{k=1}^{K}\alpha_{s,k}\beta_k\right) \\
&= \rho_{s(i)} + \hat{\rho}_{s(i)} + 1 - \rho_{s(i)} - 2\left(\sqrt{\rho_{s(i)}}\sqrt{\hat{\rho}_{s(i)}}\sqrt{\gamma_{s(i)}}\right) \\
&= \rho_{s(i)} + \hat{\rho}_{s(i)} + 1 - \rho_{s(i)} - 2\left(\sqrt{\rho_{s(i)}}\sqrt{\hat{\rho}_{s(i)}}\frac{\sqrt{\hat{\rho}_{s(i)}}}{\sqrt{\rho_{s(i)}}}\right) \\
&= \hat{\rho}_{s(i)} + 1 - 2\left(\sqrt{\hat{\rho}_{s(i)}}\sqrt{\hat{\rho}_{s(i)}}\right) \\
&= \hat{\rho}_{s(i)} + 1 - 2\hat{\rho}_{s(i)} \\
&= 1 - \hat{\rho}_{s(i)}\ ,
\end{aligned}
\tag{A.4}
$$

where we use the fact that $\sum_{k=1}^{K}\alpha_{s,k}^2 = 1$ and $\sum_{k=1}^{K}\beta_k^2 = 1$.

B. Distributions

Thanks to the work of Vrins (2018), we are taught how to derive the $(K-1)$ dimensional distribution corresponding to a K-dimensional i.i.d. standard Normal vector $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_K)^\top$ subject to the weighted sum constraint $\boldsymbol{w}^\top \boldsymbol{Z} = c$, where $\boldsymbol{w} = (w_1, ..., w_K)^\top$ and $w_i \neq 0$. This is exactly the setup, we are faced in Section 4.1, when we need to sample the $\boldsymbol{Z}$ vector, knowing that $\boldsymbol{\beta}^\top \boldsymbol{Z} = y$, where $\hat{Y} = y$. From the above, we introduce the following propositions

Lemma 6. *Let $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_K, \boldsymbol{I}_K)$, $\boldsymbol{\alpha} \in \mathbb{R}^K$ and $\boldsymbol{\beta} \in \mathbb{R}_0^K$ where $\mathbb{R}_0^K$ is the set of K-dimensional vectors with non-zero real entries. Then,*

$$\boldsymbol{\alpha}^\top \boldsymbol{Z} \big| \, \boldsymbol{\beta}^\top \boldsymbol{Z} = y \quad \sim \quad \mathcal{N}\left(y \frac{\alpha_K}{\beta_K} + \tilde{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{\mu}}, \sqrt{\tilde{\boldsymbol{\alpha}}^\top \widetilde{\Sigma} \tilde{\boldsymbol{\alpha}}} \right)$$

where the first term in $\mathcal{N}(\cdot, \cdot)$ represents the mean vector and the second term is square root of the variance covariance matrix.

The entries of $\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\mu}} \in \mathbb{R}^{K-1}$ and $\widetilde{\Sigma} \in \mathbb{R}^{K-1} \times \mathbb{R}^{K-1}$ are defined as

$$\tilde{\alpha}_i = \frac{\alpha_i}{\beta_i} - \frac{\alpha_K}{\beta_K}\,, \quad \tilde{\mu}_i = \frac{y\beta_i^2}{\|\boldsymbol{\beta}\|^2} \quad \text{and} \quad \widetilde{\Sigma}_{i,j} = \frac{\beta_i^2}{\|\boldsymbol{\beta}\|^2} \left(\mathbb{1}_{\{i=j\}} \left(\|\boldsymbol{\beta}\|^2 + \beta_j^2 - \beta_i^2 \right) - \beta_j^2 \right)\,.$$

Proof. From Vrins (2018), we know that the random vector $\boldsymbol{Z} = (Z_1, Z_2, ..., Z_K)^\top$ knowing $\boldsymbol{\beta}^\top \boldsymbol{Z} = y$, is distributed as $\tilde{\boldsymbol{Z}} = (\tilde{Z}_1, ..., \tilde{Z}_{K-1}, y - \sum_{k=1}^{K-1} \tilde{Z}_k)^\top$, with $\tilde{Z}_k = \beta_k Z_k$, where the density of $\tilde{\boldsymbol{Z}}$ is a multivariate normal

$$\tilde{\boldsymbol{Z}} \Big| \, \boldsymbol{\beta}^\top \boldsymbol{Z} = y \quad \sim \quad \mathcal{N}\left(\tilde{\boldsymbol{\mu}}, \sqrt{\widetilde{\Sigma}} \right)$$

where the entries of $\tilde{\boldsymbol{\mu}} \in \mathbb{R}^{K-1}$ and $\widetilde{\Sigma} \in \mathbb{R}^{K-1} \times \mathbb{R}^{K-1}$ are defined as

$$\tilde{\mu}_i = \frac{c\beta_i^2}{\|\boldsymbol{\beta}\|^2} \quad \text{and} \quad \widetilde{\Sigma}_{i,j} = \frac{\beta_i^2}{\|\boldsymbol{\beta}\|^2} \left(\mathbb{1}_{\{i=j\}} \left(\|\boldsymbol{\beta}\|^2 + \beta_j^2 - \beta_i^2 \right) - \beta_j^2 \right)\,.$$

From the above it is trivial to show that

$$\tilde{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{Z}} \Big| \, \boldsymbol{\beta}^\top \boldsymbol{Z} = y \quad \sim \quad \mathcal{N}\left(\tilde{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{\mu}}, \sqrt{\tilde{\boldsymbol{\alpha}}^\top \widetilde{\Sigma} \tilde{\boldsymbol{\alpha}}} \right)\,,$$

then it is sufficient to express

$$
\begin{aligned}
\boldsymbol{\alpha}^\top \boldsymbol{Z} &= \sum_{k=1}^{K-1} \alpha_k Z_k + \alpha_K Z_K \\
&= \sum_{k=1}^{K-1} \frac{\alpha_k}{\beta_k} \beta_k Z_k + \frac{\alpha_K}{\beta_K} \beta_K Z_K \\
&= \sum_{k=1}^{K-1} \frac{\alpha_k}{\beta_k} \tilde{Z}_k + \frac{\alpha_K}{\beta_K} \tilde{Z}_K \\
&= \sum_{k=1}^{K-1} \frac{\alpha_k}{\beta_k} \tilde{Z}_k + \frac{\alpha_K}{\beta_K} \left(y - \sum_{k=1}^{K-1} \tilde{Z}_k \right) \\
&= y \frac{\alpha_K}{\beta_K} + \sum_{k=1}^{K-1} \left(\frac{\alpha_k}{\beta_k} - \frac{\alpha_K}{\beta_K} \right) \tilde{Z}_k \\
&= y \frac{\alpha_K}{\beta_K} + \sum_{k=1}^{K-1} \tilde{\alpha}_k \tilde{Z}_k
\end{aligned}
$$

concluding the proof.

□

Lemma 7. *From Lemma 6, it holds that*

$$
\mathbb{E}\left[\Phi \left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_{s(i)}} \boldsymbol{\alpha}^\top \boldsymbol{Z}}{\sqrt{1 - \rho_{s(i)}}} \right) \middle| \hat{Y} \right] = \Phi \left(\frac{\Phi^{-1}(p_i) - \sqrt{\hat{\rho}_{s(i)}} \hat{Y}}{\sqrt{1 - \hat{\rho}_{s(i)}}} \right)
$$

where

$$
\hat{Y} = \boldsymbol{\beta}^\top \boldsymbol{Z} \quad \text{and} \quad \sqrt{\hat{\rho}_{s(i)}} = \sqrt{\rho_{s(i)}} \boldsymbol{\alpha}^\top \boldsymbol{\beta}
$$

and $p_i, \rho_{s(i)}, \hat{\rho}_{s(i)} \in [0, 1]$.

Proof. First, observe that the expectation in the Lemma is $\mathbb{E}\left[\Phi \left(a + b \boldsymbol{\alpha}^\top \boldsymbol{Z} \right) | \hat{Y} \right]$ with

$$
a = \frac{\Phi^{-1}(p_i)}{\sqrt{1 - \rho_{s(i)}}} \quad \text{and} \quad b = -\frac{\sqrt{\rho_{s(i)}}}{\sqrt{1 - \rho_{s(i)}}} .
$$

We know from Lemma 6 that $a + b\boldsymbol{\alpha}^\top \boldsymbol{Z} | \boldsymbol{\beta}^\top \boldsymbol{Z} \sim \mathcal{N}(a + b\mu, b\sigma)$ with

$$
\mu = \boldsymbol{\beta}^\top \boldsymbol{Z} \frac{\alpha_K}{\beta_K} + \tilde{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{\mu}} \quad \text{and} \quad \sigma^2 = \tilde{\boldsymbol{\alpha}}^\top \tilde{\Sigma} \tilde{\boldsymbol{\alpha}}
$$

and where $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\mu}}$ and $\tilde{\Sigma}$ are given in Lemma 6. This shows that using $Z \sim \mathcal{N}(0,1)$,

$$\mathbb{E}\left[\Phi\left(a + b\boldsymbol{\alpha}^\top \boldsymbol{Z}\right)|\hat{Y}\right] = \mathbb{E}\left[\Phi\left(a + b\mu + b\sigma Z\right)\right] = \Phi\left(\frac{a+b\mu}{\sqrt{1+(b\sigma)^2}}\right) .$$

Next, observe that

$$\mu = \hat{Y}\frac{\alpha_K}{\beta_K} + \tilde{\boldsymbol{\alpha}}^\top \tilde{\boldsymbol{\mu}} = \hat{Y}\boldsymbol{\alpha}^\top\boldsymbol{\beta} = \hat{Y}\frac{\sqrt{\hat{\rho}_{s(i)}}}{\sqrt{\rho_{s(i)}}} \quad \Rightarrow \quad a + b\mu = \frac{\Phi^{-1}(p_i) - \sqrt{\hat{\rho}_{s(i)}}\hat{Y}}{\sqrt{1-\rho_{s(i)}}} .$$

On the other hand,

$$1 + b^2\sigma^2 = 1 + \frac{\rho_{s(i)}}{1-\rho_{s(i)}}\sigma^2 = \frac{1-\rho_{s(i)}(1-\sigma^2)}{1-\rho_{s(i)}}$$

and it remains to prove that $1-\sigma^2 = \frac{\hat{\rho}_{s(i)}}{\rho_{s(i)}} = (\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2$. Observe that

$$\boldsymbol{\alpha}^\top\boldsymbol{\beta} = \sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2 + \frac{\alpha_K}{\beta_K} \quad \Rightarrow \quad (\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 = \left(\sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2\right)^2 + 2\frac{\alpha_K}{\beta_K}\sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2 + \frac{\alpha_K^2}{\beta_K^2} .$$

Developing σ^2,

$$\begin{aligned} 1-\sigma^2 &= 1 - \sum_{i=1}^{K-1}\tilde{\alpha}_i^2\beta_i^2 + \left(\sum_{j=1}^{K-1}\tilde{\alpha}_i\beta_i^2\right)^2 \\ &= 1 - \sum_{i=1}^{K-1}\tilde{\alpha}_i^2\beta_i^2 + \left(\boldsymbol{\alpha}^\top\boldsymbol{\beta}\right)^2 - \left(2\frac{\alpha_K}{\beta_K}\sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2 + \frac{\alpha_K^2}{\beta_K^2}\right) . \end{aligned}$$

The proof is concluded by noting that

$$1 - \sum_{i=1}^{K-1}\tilde{\alpha}_i^2\beta_i^2 = \frac{\alpha_K^2}{\beta_K^2} + 2\frac{\alpha_K}{\beta_K}\sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2$$

when replacing $\tilde{\alpha}_i$ by its expression, using the unit norm of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and exploiting that

$$\sum_{i=1}^{K-1}\alpha_i\beta_i = \boldsymbol{\alpha}^\top\boldsymbol{\beta} - \alpha_K\beta_K = \sum_{i=1}^{K-1}\tilde{\alpha}_i\beta_i^2 + \frac{\alpha_K}{\beta_K}(1-\beta_K^2) .$$

□

C. Basel ASRF granularity adjustment

Adapted to the ASRF setup underlying the Basel IRB approach, the granularity adjustment formula proposed in Gordy and Lutkebohmert (2013) becomes

$$\mathrm{GA} := \mathbb{Q}_\alpha[L] - \mathbb{E}[L|Y = \mathbb{Q}_{1-\alpha}[Y]] \approx -\frac{1}{2\phi(\mathbb{Q}_{1-\alpha}[Y])}\frac{d}{dy}\left[\frac{\phi(y)v(y)}{l'(y)}\right]\Bigg|_{y=\mathbb{Q}_{1-\alpha}[Y]} \tag{C.1}$$

where $l'(y)$ is the first partial derivative of $l(y)$ with respect to y:

$$l'(y) := \frac{d}{dy}l(y) = \frac{d}{dy}\left(\sum_{i=1}^{I} w_i\lambda_i p_i(y)\right) = \sum_{i=1}^{I} w_i\lambda_i p_i'(y) \tag{C.2}$$

and

$$l''(y) := \frac{d^2}{dy^2}l(y) = \frac{d}{dy}\left(\sum_{i=1}^{I} w_i\lambda_i p_i(y)\right) = \sum_{i=1}^{I} w_i\lambda_i p_i''(y)\ . \tag{C.3}$$

Since w_i and λ_i are constants, the first partial derivative of the conditional probability of default is computed as

$$\begin{aligned}
p_i'(y) &= \frac{d}{dy}p(y) \\
&= \frac{d}{dy}\Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \times \frac{d}{dy}\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \\
&= \phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \times -\frac{\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\ ,
\end{aligned} \tag{C.4}$$

and the second partial derivative, with respect to y, as

$$\begin{aligned}
p_i''(y) &= \frac{d^2}{d^2y}p_i(y) \\
&= \frac{d}{dy}p_i'(y) \\
&= -\sqrt{\frac{\rho_i}{1-\rho_i}}\frac{d}{dy}\left(\phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right)\right) \\
&= -\sqrt{\frac{\rho_i}{1-\rho_i}}\left(\phi'\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \times -\sqrt{\frac{\rho_i}{1-\rho_i}}\right) \\
&= \frac{\rho_i}{1-\rho_i} \times \phi'\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \\
&= -\frac{\rho_i}{1-\rho_i} \times \phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right) \times \left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho_i}y}{\sqrt{1-\rho_i}}\right)\ .
\end{aligned} \tag{C.5}$$

The term $v(y)$ denotes the variance of the portfolio loss L conditional on Y, i.e., $\mathbb{V}[L|Y = y]$. Recalling that $L = \sum_i w_i L_i$ where $L_i := \Lambda_i \times B_i$ we can write

$$\begin{aligned}
\mathbb{V}[L|Y] &= \mathbb{E}[L^2|Y] - (\mathbb{E}[L|Y])^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{I} L_i\right)^2 \middle| Y\right] - \left(\mathbb{E}\left[\sum_{i=1}^{I} L_i \middle| Y\right]\right)^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{I} w_i \Lambda_i B_i\right)^2 \middle| Y\right] - \left(\mathbb{E}\left[\sum_{i=1}^{I} w_i \Lambda_i B_i \middle| Y\right]\right)^2 ,
\end{aligned}$$

The second expectation in the above subtraction shrinks to

$$\begin{aligned}
\left(\mathbb{E}\left[\sum_{i=1}^{I} w_i \Lambda_i B_i \middle| Y\right]\right)^2 &= \left(\sum_{i=1}^{I} \mathbb{E}\left[w_i \Lambda_i B_i | Y\right]\right)^2 \\
&= \left(\sum_{i=1}^{I} w_i \, \mathbb{E}[\Lambda_i] \, \mathbb{E}\left[B_i | Y\right]\right)^2 \\
&= \left(\sum_{i=1}^{I} w_i \lambda_i p_i(Y)\right)^2 \\
&= l(Y)^2
\end{aligned}$$

where we noted $l(Y)^2 := (l(Y))^2$. The first expectation is equal to

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{I} L_i\right)^2 \middle| Y\right] &= \mathbb{E}\left[\sum_{i=1}^{I} L_i^2 + \sum_{i=1}^{I}\sum_{j\neq i} L_i L_j \middle| Y\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{I} L_i^2 \middle| Y\right] + \mathbb{E}\left[\sum_{i=1}^{I}\sum_{j\neq i} L_i L_j \middle| Y\right] \\
&= \sum_{i=1}^{I} \mathbb{E}[L_i^2|Y] + \sum_{i=1}^{I}\sum_{j\neq i} \mathbb{E}[L_i L_j|Y] \\
&= \sum_{i=1}^{I} w_i^2 \, \mathbb{E}[(\Lambda_i B_i)^2|Y] + \sum_{i=1}^{I}\sum_{j\neq i} \mathbb{E}[L_i L_j|Y] \\
&= \sum_{i=1}^{I} w_i^2 \, \mathbb{E}[\Lambda_i^2] \, \mathbb{E}[B_i|Y] + \sum_{i=1}^{I}\sum_{j\neq i} \mathbb{E}[L_i|Y] \, \mathbb{E}[L_j|Y] \\
&= \sum_{i=1}^{I} w_i^2 \, \mathbb{E}[\Lambda_i^2] p_i(Y) + \sum_{i=1}^{I}\sum_{j\neq i} w_i \lambda_i p_i(Y) \times w_j \lambda_j p_j(Y) ,
\end{aligned}$$

where can decompose further the above double sum as

$$\begin{aligned}
\sum_{i=1}^{I}\sum_{j\neq i} w_i\lambda_i p_i(Y)\times w_j\lambda_j p_j(Y) &= \sum_{i=1}^{I} w_i\lambda_i p_i(Y)\times\left(\sum_{j\neq i} w_j\lambda_j p_j(Y)\right)\\
&= \sum_{i=1}^{I} w_i\lambda_i p_i(Y)\times\left(l(Y)-w_i\lambda_i p_i(Y)\right)\\
&= l(Y)\sum_{i=1}^{I} w_i\lambda_i p_i(Y)-\sum_{i=1}^{I} w_i\lambda_i p_i(Y)\times w_i\lambda_i p_i(Y)\\
&= l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\ ,
\end{aligned}$$

leading to write

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{I} L_i\right)^2\middle| Y\right] &= \sum_{i=1}^{I} w_i^2\,\mathbb{E}[\Lambda_i^2]p_i(Y)+l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\\
&= \sum_{i=1}^{I} w_i^2(\mathbb{V}[\Lambda_i]+\mathbb{E}[\Lambda_i]^2)p_i(Y)+l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\\
&= \sum_{i=1}^{I} w_i^2(\sigma_i^2+\lambda_i^2)p_i(Y)+l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\\
&= \sum_{i=1}^{I}(w_i^2\sigma_i^2+w_i^2\lambda_i^2)p_i(Y)+l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\\
&= \sum_{i=1}^{I} w_i^2\sigma_i^2 p_i(Y)+\sum_{i=1}^{I} w_i^2\lambda_i^2 p_i(Y)+l(Y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(Y)\right)^2\ .
\end{aligned}$$

The conditional variance becomes

$$\begin{aligned}
v(y) := \mathbb{V}[L|Y=y] &= \left(\sum_{i=1}^{I} w_i^2\sigma_i^2 p_i(y)+\sum_{i=1}^{I} w_i^2\lambda_i^2 p_i(y)+l(y)^2-\sum_{i=1}^{I}\left(w_i\lambda_i p_i(y)\right)^2\right)-l(y)^2\\
&= \sum_{i=1}^{I} w_i^2\sigma_i^2 p_i(y)+\sum_{i=1}^{I} w_i^2\lambda_i^2 p_i(y)-\sum_{i=1}^{I} w_i^2\lambda_i^2(p_i(y))^2\\
&= \sum_{i=1}^{I} w_i^2 p_i(y)\left(\sigma_i^2+\lambda_i^2(1-p_i(y))\right)\ . \qquad \text{(C.6)}
\end{aligned}$$

The first partial derivative, with respect to y, of the conditional variance is

$$\begin{aligned}
v'(y) &= \frac{d}{dy} v(y) \\
&= \sum_{i=1}^{I} w_i^2 \sigma_i^2 p_i'(y) + \sum_{i=1}^{I} w_i^2 \lambda_i^2 p_i'(y) - \sum_{i=1}^{I} \left(2(w_i \lambda_i p_i(y)) \times (w_i \lambda_i p_i'(y))\right) \\
&= \sum_{i=1}^{I} w_i^2 \sigma_i^2 p_i'(y) + \sum_{i=1}^{I} w_i^2 \lambda_i^2 p_i'(y) - 2 \sum_{i=1}^{I} \left((w_i \lambda_i p_i(y)) \times (w_i \lambda_i p_i'(y))\right) \\
&= \sum_{i=1}^{I} w_i^2 \sigma_i^2 p_i'(y) + \sum_{i=1}^{I} w_i^2 \lambda_i^2 p_i'(y) - 2 \sum_{i=1}^{I} \big((w_i^2 \lambda_i^2 p_i(y) p_i'(y)\big) \\
&= \sum_{i=1}^{I} w_i^2 p_i'(y) \left(\sigma_i^2 + \lambda_i^2 (1 - 2 p_i(y))\right) \ . \qquad \text{(C.7)}
\end{aligned}$$

Finally, using $\phi'(y) = -y\phi(y)$, we find that

$$\begin{aligned}
\frac{d}{dy}\left[\frac{\phi(y) v(y)}{l'(y)}\right] &= \frac{(\phi(y) v(y))' \, l'(y) - \phi(y) v(y) l''(y)}{(l'(y))^2} \\
&= \frac{\phi(y)}{l'(y)} \left(v'(y) - v(y)\left(y + \frac{l''(y)}{l'(y)}\right)\right) \qquad \text{(C.8)}
\end{aligned}$$

concluding the proof.

References

Altman, E. I., Brady, B., Resti, A., and Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *The Journal of Business*, *78*(6), 2203–2228. DOI: https://doi.org/10.1086/497044.

Barbagli, M., and Vrins, F. (2023). Accounting for PD-LGD dependency: A tractable extension to the Basel ASRF framework. *Economic Modelling*, *125*, 106321. DOI: https://doi.org/10.1016/j.econmod.2023.106321.

BCBS (2000). *Principles for the management of credit risk*. Technical Report September. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/publ/bcbs75.htm.

BCBS (2005). *An explanatory note on the Basel II IRB risk weight functions*. Technical Report July. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/bcbs/irbriskweight.htm.

BCBS (2006a). *International convergence of capital measurement and capital standards*. Technical Report June. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/publ/bcbs128.htm.

BCBS (2006b). *Working paper No. 15: Studies on credit risk concentration*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/publ/bcbs_wp15.htm.

BCBS (2009). *Range of practices and issues in economic capital frameworks*. Technical Report March. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/publ/bcbs152.htm.

BCBS (2014). *Consultative document - Capital floors: The design of a framework based on standardised approaches*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/bcbs/publ/d306.htm.

BCBS (2017). *Basel III: Finalising post-crisis reforms*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/bcbs/publ/d424.htm.

BCBS (2018). *Stress testing principles*. Technical Report October. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/bcbs/publ/d450.htm.

BCBS (2019). *Minimum capital requirements for market risk*. Technical Report February. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/bcbs/publ/d457.htm.

BCBS (2021). *Basel Framework - Scope and definitions - Global systemically important banks*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/SCO/40.htm.

BCBS (2023a). *Basel Framework - Calculation of RWA for credit risk - IRB approach: Minimum requirements to use IRB approach*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/CRE/36.htm.

BCBS (2023b). *Basel Framework - Calculation of RWA for credit risk - IRB approach: Overview and asset class definitions*. Technical Report January. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/CRE/30.htm?inforce=20230101&published=20200327.

BCBS (2023c). *Basel Framework - Calculation of RWA for credit risk - IRB approach: Risk components*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/CRE/32.htm?inforce=20230101&published=20200327.

BCBS (2023d). *Basel Framework - Calculation of RWA for credit risk - IRB approach: Risk weight functions*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/CRE/31.htm.

BCBS (2023e). *Basel Framework - Calculation of RWA for market risk*. Technical Report January. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/standard/MAR.htm?tldate=20240424.

BCBS (2023f). *Basel Framework - Calculation of RWA for market risk - Internal models approach: Capital requirements calculation*. Technical Report January. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/MAR/33.htm?inforce=20230101&published=20200605#paragraph_MAR_33_20230101_33_18.

BCBS (2023g). *Basel Framework - Supervisory review process - Credit risk*. Technical Report December. Bank for International Settlements. Basel, Switzerland. URL: https://www.bis.org/basel_framework/chapter/SRP/32.htm?inforce=20230101&published=20200327.

Düllmann, K., and Masschelein, N. (2007). A tractable model to measure sector concentration risk in credit portfolios. *Journal of Financial Services Research*, *32*(1-2), 55–79. DOI: 10.1007/s10693-007-0014-3.

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. Springer. DOI: https://doi.org/10.1007/978-0-387-21617-1.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, *12*(3), 199–232. DOI: 10.1016/S1042-9573(03)00040-8.

Gordy, M. B., and Lutkebohmert, E. (2013). Granularity adjustment for regulatory capital assessment. *International Journal of Central Banking*, *9*(3), 33–71. URL: https://www.ijcb.org/journal/ijcb13q3a2.htm.

Gourieroux, C., Laurent, J. P., and Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of empirical finance*, *7*(3-4), 225–245. DOI: https://doi.org/10.1016/S0927-5398(00)00011-6.

Laurent, J.-P., Sestier, M., and Thomas, S. (2016). Trading book and credit risk: How fundamental is the Basel review? *Journal of Banking & Finance*, *73*, 211–223. DOI: https://doi.org/10.1016/j.jbankfin.2016.07.002.

Martin, R., and Wilde, T. (2002). Credit portfolio measuremenet unsystematic credit risk. *Risk Magazine*, *15*(11), 123–128.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, *29*(2), 449–470. DOI: https://doi.org/10.2307/2978814.

Pitera, M., and Schmidt, T. (2018). Unbiased estimation of risk. *Journal of Banking & Finance*, *91*, 133–145. DOI: https://doi.org/10.1016/j.jbankfin.2018.04.016.

Puzanova, N., and Düllmann, K. (2013). Systemic risk contributions: A credit portfolio approach. *Journal of Banking & Finance*, *37*(4), 1243–1257. DOI: https://doi.org/10.1016/j.jbankfin.2012.11.017.

Pykhtin, M. (2004). Multi-factor adjustment. *Risk Magazine*, *17*(3), 85–90. URL: https://www.risk.net/risk-management/credit-risk/1530333/multi-factor-adjustment.

Tarashev, N., and Zhu, H. (2008). Specification and calibration errors in measures of portfolio credit risk: The case of the ASRF model. *International Journal of Central Banking*, *4*(2), 129–173. URL: https://www.ijcb.org/journal/ijcb08q2a4.htm.

Vasicek, O. A. (1991). *Limiting loan loss probability distribution*. Technical Report August. Moodys KMV Company Research Paper. San Francisco, California, United States of America. URL: https://doi.org/10.1002/9781119186229.ch18.

Vasicek, O. A. (2002). Loan portfolio value. *Risk Magazine*, *15*(12), 160–162. URL: https://www.risk.net/risk-management/credit-risk/1500333/loan-portfolio-value#:~:text=It%20means%20that%20the%20equity,corresponds%20to%20the%20desired%20probability.

Vrins, F. (2018). Sampling the multivariate standard normal distribution under a weighted sum constraint. *Risks*, *6*(3), 64. DOI: https://doi.org/10.3390/risks6030064.