Measuring Credit Card Portfolio Risk: A Vasicek Model Approach with Monte Carlo Simulation for Value at Risk

Jordan Andrew, Ammy Lin, Farnoosh Memari, Tonantzin Real Rojas

December 10, 2025

Abstract

This project implements a Vasicek-model Monte Carlo simulation to quantify credit card portfolio risk, measuring Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) under varying default probabilities (3%-22.15%) and correlations (5%-15%). Unlike traditional approaches that assume independent defaults, our framework captures default dependency, critical for realistic stress testing. Results demonstrate that correlation structures dramatically impact tail risk, with VaR increasing up to 75% and CVaR increasing up to 93% when correlation rises from 5% to 15% at high default probabilities. The analysis reveals that while probability of default drives expected losses, correlation governs tail risk amplification, with maximum losses exceeding expected losses by over 900% in low Payment Default (PD)/high correlation scenarios. The study provides a practical tool for capital allocation and regulatory compliance, illustrating how systemic factors amplify losses in retail credit portfolios under stress conditions.

Contents

Technical Section

1 Motivation

Credit card portfolios represent one of the most significant and major exposure areas for retail banks, where accurately predicting potential defaults remains a fundamental challenge. Traditional risk models that assume independent defaults underestimate tail risk during economic crises, leaving financial institutions vulnerable to unexpected losses (Löffler and Posch, 2007). The 2008 financial crisis highlighted how interconnectedness can trigger cascading defaults, a phenomenon poorly captured by conventional methods (Amadeo, 2022). This project addresses this gap by developing a crisis simulation framework that answers the essential question: "In the worst-case scenarios, how much capital should a bank reserve to cover credit card losses?"

The practical implications extend directly to regulatory compliance and strategic decision-making. Financial institutions require robust risk measurement tools not only to meet Basel capital adequacy requirements but also to make informed lending decisions and maintain financial stability (Basel Committee on Banking Supervision, 2006). By implementing a Vasicek model with Monte Carlo simulation on credit card data, this project provides a more realistic approach to capturing correlated default behavior where economic crises trigger simultaneous customer defaults that simple probabilistic models might miss.

Ultimately, we examine how different correlation regimes and probability-of-default assumptions affect risk metrics, providing insights for both risk management education and practical banking applications. The resulting Value at Risk metrics offer concrete guidance for capital planning, while the methodology demonstrates how systemic risk factors and individual borrower characteristics interact to drive portfolio losses under stress conditions.

2 Introduction

The management of credit risk has become increasingly sophisticated as financial institutions seek to balance profitability with prudent risk-taking in an environment of heightened regulatory scrutiny and economic uncertainty. Credit portfolios, particularly those consisting of unsecured consumer loans such as credit cards, present unique challenges for risk measurement and capital allocation. Unlike traditional secured lending, where collateral provides a buffer against losses, unsecured retail credit exposes lenders to the full impact of borrower defaults. Understanding the probability, timing, and severity of these defaults—and more importantly, how they may cluster during periods of economic stress—is essential for maintaining financial stability and meeting regulatory requirements.

This study develops a comprehensive framework for measuring and analyzing credit risk in retail credit card portfolios using Monte Carlo simulation techniques. By combining theoretical models of default correlation with computational methods capable of capturing the complex interactions among thousands of individual borrowers, this research provides insights into the distribution of portfolio losses under various economic scenarios. The analysis draws on well-established credit risk frameworks, including the Vasicek model and Basel regulatory approaches, while emphasizing the practical challenges of implementing these models for large, granular portfolios. The remainder of this introduction reviews the fundamental concepts of credit risk measurement, the role of Monte Carlo simulation in financial modeling, and the theoretical foundations that underpin modern portfolio credit risk analysis.

2.1 Credit Risk

Credit risk refers to the possibility that a borrower fails to meet contractual financial obligations, resulting in an economic loss to the lender (Basel Committee on Banking Supervision, 2006). It is distinct from market risk, which arises from fluctuations in asset prices, and from liquidity risk, which reflects the inability to execute transactions or obtain funding at reasonable terms. In retail and commercial banking, credit risk is typically the most significant source of financial risk, particularly for institutions whose balance sheets are dominated by loans and unsecured consumer credit products.

2.1.1 Fundamental Components of Credit Risk Assessment

The quantitative assessment of credit risk relies on three fundamental components: the probability of default (PD), the loss given default (LGD), and the exposure at default (EAD). The PD measures the likelihood that a borrower defaults within a given time horizon. LGD represents the proportion of the exposure that is ultimately lost after accounting for recoveries, collateral liquidation, or other mitigating factors. EAD captures the total exposure that is at risk at the moment of default, including drawn balances and, in some cases, committed but undrawn credit (Basel Committee on Banking Supervision, 2006). Together, these elements determine the expected loss on a position or portfolio through the relationship:

$$EL = PD \times LGD \times EAD \tag{1}$$

This decomposition, which was formally codified in the Basel II framework in 2004, provides a consistent framework for pricing credit products, setting provisions, and evaluating the financial impact of potential borrower defaults (Basel Committee on Banking Supervision, 2006).

2.1.2 Credit Card Portfolio Characteristics

Credit card portfolios exhibit several characteristics that make credit risk measurement particularly important. These exposures are unsecured, meaning that lenders have no collateral to offset losses in the event of nonpayment. They are also revolving credit lines, giving borrowers the flexibility to draw additional funds over time, which affects the variability of EAD. Furthermore, credit card portfolios consist of a very large number of relatively small accounts, each with distinct behavioral and demographic profiles. Although this granularity offers potential diversification benefits, repayment performance can still be influenced by broad economic conditions such as income shocks, unemployment, and interest rate changes. As a result, understanding the drivers of PD, LGD, and EAD, and how they interact with borrower behavior, is essential for managing the inherent risk in credit card lending.

2.1.3 Forward-Looking Risk Assessment

In practice, credit risk analysis requires both forward-looking assessment of borrower solvency and quantitative estimation of expected losses under varying economic conditions. For unsecured consumer credit, such as credit cards, this task is especially sensitive to fluctuations in household finances and macroeconomic environments, underscoring the need for robust credit risk measurement frameworks (Joseph, 2013).

2.2 Monte Carlo Simulation

Monte Carlo simulation is a computational method used to approximate the behavior of complex systems by generating a large number of random scenarios and evaluating the outcomes they produce (Glasserman, 2003). Rather than relying on closed-form analytical solutions, Monte Carlo techniques use repeated random sampling to approximate probability distributions, expectations, and tail events. The method is particularly valuable when the underlying system involves nonlinearities, stochastic components, or interactions among multiple random variables that make analytical characterization intractable. By simulating thousands or millions of possible outcomes, Monte Carlo simulation provides a flexible way to study the distribution of results and to estimate quantities such as means, variances, and extreme quantiles.

2.2.1 Applications in Financial Modeling

In financial modeling, Monte Carlo simulation is widely used to evaluate risks and uncertainties that arise from market fluctuations, interest rate dynamics, or borrower behavior. Because many financial processes are inherently stochastic and path-dependent, analytical formulas are often unavailable or difficult to derive. Monte Carlo simulation circumvents this limitation by generating synthetic realizations of the underlying risk factors and computing the corresponding financial outcomes. This approach is especially useful for estimating the likelihood and severity of rare events, which play a central role in risk management, derivative pricing, and stress testing.

2.2.2 Framework for Credit Portfolio Risk

Within the context of credit risk modeling, Monte Carlo simulation provides a natural framework for approximating the distribution of portfolio losses (Glasserman and Li, 2005). Credit portfolios consist of many borrowers whose defaults are uncertain, and, in practice, lenders are concerned not only with expected losses but also with the behavior of the loss distribution under adverse economic conditions. The Monte Carlo approach for credit risk typically involves several key steps:

1. **Factor simulation**: Generate random realizations of systematic risk factors representing economic conditions (e.g., industry sector performance, macroeconomic indicators)
2. **Individual borrower outcomes**: Simulate borrower-specific default outcomes based on the systematic factors and idiosyncratic risk
3. **Loss calculation**: Compute losses for each borrower given defaults, incorporating loss given default (LGD) and exposure at default (EAD) estimates
4. **Portfolio aggregation**: Sum individual losses to obtain total portfolio loss for each simulation scenario
5. **Distribution construction**: Aggregate results across all simulations to build an empirical loss distribution

Monte Carlo methods allow the analyst to simulate borrower-specific default outcomes, recoveries, and exposures under different economic scenarios, producing an empirical loss distribution that reflects both idiosyncratic and systematic variability. The simulation generates random sets of future values of economic factors and specific risks, from which asset values and credit status derive directly.

2.2.3 Risk Measures and Capital Requirements

This empirical distribution can then be used to evaluate quantities such as VaR and CVaR, which summarize the potential tail losses that may occur in extreme but plausible circumstances. VaR represents the maximum loss at a specified confidence level (e.g., 99.9% for regulatory capital under Basel frameworks), while CVaR (also known as Expected Shortfall) captures the average loss in scenarios beyond the VaR threshold, providing a more comprehensive measure of tail risk.

Monte Carlo simulation provides several conceptual advantages over scenario-based analysis: it assigns probabilities to scenarios showing not only what could happen but how likely each outcome is; it ensures completeness by including all possible loss events; and it provides consistent risk allocation that is not sensitive to arbitrary scenario selection. For credit portfolios, the method is particularly effective in capturing default clustering and the correlation structure between obligors, features that are critical for accurate risk assessment but challenging to model analytically.

In the broader framework of this study, Monte Carlo simulation serves as the computational engine that translates assumptions about credit behavior into quantitative measures of portfolio risk. The method enables financial institutions to estimate Credit Value at Risk (Credit VaR) and economic capital requirements—the capital buffer needed to absorb unexpected losses with high confidence—which are essential components of regulatory compliance under Basel III and internal capital adequacy assessment processes.

2.3 Default Correlation

Although the simulations in this study treat the asset correlation ρ as an exogenous parameter, it is useful to understand how default correlation is theoretically linked to joint default behavior. The default correlation between two borrowers i and j is defined as (Li, 2000):

$$\rho_{ij}^{\text{default}} = \frac{P(D_i \wedge D_j) - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}, \tag{2}$$

where p_i and p_j are their unconditional default probabilities, and $P(D_i \wedge D_j)$ is the joint default probability.

In a Gaussian latent-factor framework—which forms the basis of the Vasicek model used in Basel II capital calculations—the joint default probability can be expressed using the bivariate normal cumulative distribution function (Vasicek, 2002):

$$P(D_i \wedge D_j) = \Phi_2(\Phi^{-1}(p_i), \Phi^{-1}(p_j); \rho_{ij}^{\text{asset}}), \tag{3}$$

where Φ_2 is the standard bivariate normal CDF with correlation ρ_{ij}^{asset}, and Φ^{-1} is the inverse of the standard normal CDF. This expression links observable joint default frequencies to the underlying asset correlation. Inverting the formula allows practitioners to estimate asset correlation from empirical data.

2.3.1 The Role of Systematic Risk

This relationship clarifies how systematic risk—captured by ρ in the Vasicek model—drives default clustering (Vasicek, 2002). Higher asset correlation increases the probability that many borrowers default simultaneously, which in turn amplifies tail losses in portfolio-level risk measures such as VaR and CVaR (Gordy, 2003). The Gaussian copula framework provides an elegant way to model this dependence structure, as it separates the marginal default behavior of individual obligors from their joint dependence through correlation (Li, 2000).

Empirical studies have demonstrated that default correlation tends to increase during economic downturns, exhibiting state dependence. This phenomenon is critical for stress testing and capital adequacy assessment, as it implies that portfolio losses in adverse scenarios may be substantially higher than simple models with constant correlation would predict.

2.4 Vasicek Model

The Vasicek model is a foundational framework in portfolio credit risk analysis that links individual default behavior to a common underlying economic factor (Vasicek, 1987, 2002). Developed as a one-factor latent variable model, it provides a tractable way to capture default correlation and to describe how systematic economic shocks influence the credit quality of all borrowers in a portfolio. In this framework, each borrower is assumed to possess an unobservable creditworthiness index that deteriorates when economic conditions worsen. A default occurs when this latent variable falls below a specified threshold, which is determined by the borrower's probability of default.

2.4.1 Model Structure and Key Innovation

The key innovation of the Vasicek model is its decomposition of default risk into two components: a systematic factor shared by all borrowers and an idiosyncratic component unique to each individual (Vasicek, 1987). Mathematically, the asset value or creditworthiness of borrower i is expressed as:

$$X_i = \sqrt{\rho} \cdot Z + \sqrt{1-\rho} \cdot \epsilon_i, \tag{4}$$

where $Z \sim N(0,1)$ represents the systematic factor (representing the state of the economy), $\epsilon_i \sim N(0,1)$ is the idiosyncratic shock, and $\rho \in [0,1]$ is the asset correlation parameter (Gordy, 2003). Default occurs when $X_i < \Phi^{-1}(PD_i)$, where Φ^{-1} is the inverse standard normal CDF. The correlation parameter ρ captures the extent to which individual asset values are driven by common systematic factors versus firm-specific factors.

By adjusting the relative weight of the systematic factor, the model naturally incorporates different degrees of default correlation. When the systematic factor plays a larger role (higher ρ), defaults become more synchronized; when idiosyncratic risk dominates (lower ρ), defaults tend to occur independently. This structure makes the Vasicek model particularly well suited for analyzing tail risk, as it emphasizes the joint impact of economic downturns on portfolio losses (Gordy, 2003).

2.4.2 Application to Large Granular Portfolios

For large retail credit portfolios, such as credit card portfolios, the Vasicek model provides several advantages. Because these portfolios contain many small and granular exposures, the law of large numbers mitigates idiosyncratic risk, making systematic risk the primary driver of portfolio losses (Vasicek, 2002). The model captures this reality by showing how even moderate levels of systematic dependence can generate heavy-tailed loss distributions with substantial losses in severe economic states. Under the large homogeneous portfolio (LHP) assumption, where the portfolio contains many identical small exposures, the conditional loss distribution given the systematic factor Z has a closed-form expression (Vasicek, 2002):

$$P(L \leq l | Z = z) = \Phi\left(\frac{\Phi^{-1}(l) - \sqrt{\rho} \cdot z}{\sqrt{1-\rho}}\right), \tag{5}$$

where L represents the portfolio loss rate. This analytical tractability allows for a clear mapping from borrower-level parameters, most notably the probability of default, to the distribution of aggregate losses, providing a foundation for risk measures such as VaR and CVaR (Gordy, 2003).

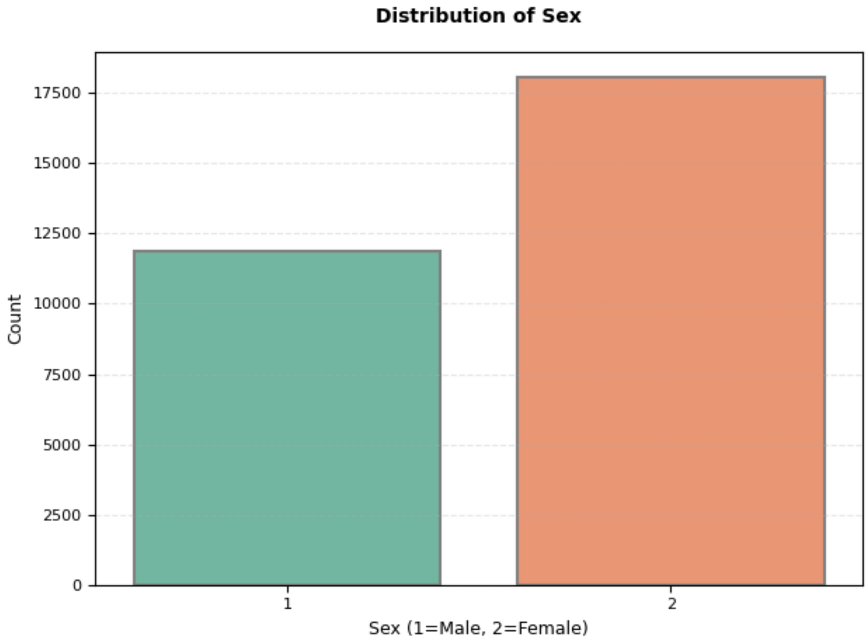

Figure 1: Sex Distribution

2.4.3 Regulatory Applications and Extensions

In addition to its theoretical appeal, the Vasicek model underlies many regulatory and industry practices in credit risk management. Variants of the model form the basis of the Basel internal ratings-based (IRB) approach, which uses the one-factor structure to determine capital requirements as a function of PD and assumed asset correlation (Basel Committee on Banking Supervision, 2006; Gordy, 2003). The Basel II capital formula for unexpected loss is directly derived from the Vasicek model, with regulatory asset correlations prescribed as a function of borrower type and PD. In this study, the Vasicek model provides the framework through which default probabilities and correlations are translated into simulated loss distributions for the credit card portfolio under examination.

3 Data

The data we used comes from Kaggle and contains information about default payments, credit data, history of payment, and bill statements of credit card clients in Taiwan from April 2005 to September 2005. The dataset contains 25 features. Given the nature of our project, we only used the credit limit for each user, we performed an exploratory data analysis to better understand the data.

3.1 Exploratory Data Analysis

The dataset exhibits distinct demographic characteristics across sex, education, and marital status. The sex distribution is mainly composed of female (Figure 1). In terms of educational background, the portfolio is predominantly composed of clients with graduate school and university education, indicating a highly educated user base, while those with Unknown or in the Others category form a smaller segment (Figure 2). Regarding marital status, single individuals constitute the largest group, followed by married individuals, with the Others category representing a small fraction of the total (Figure 3). This overall profile suggests the credit card portfolio is primarily composed of single, well-educated individuals.

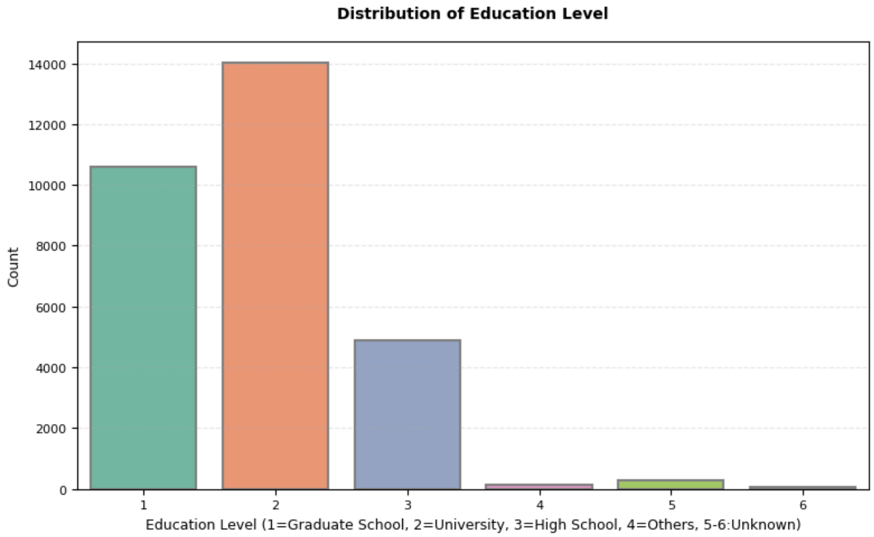

Figure 2: Education Level Distribution

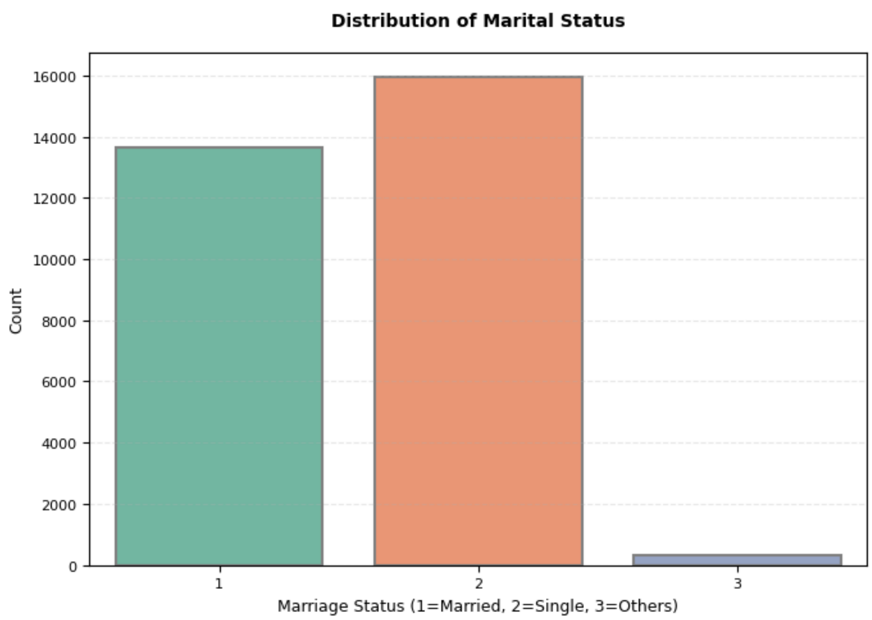

Figure 3: Marriage Distribution

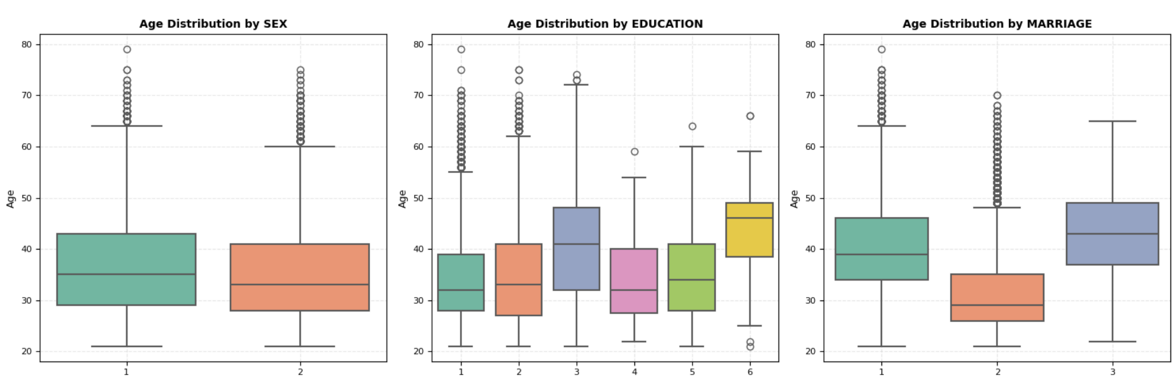

Figure 4: Age Distribution by Sex, Education and Marriage

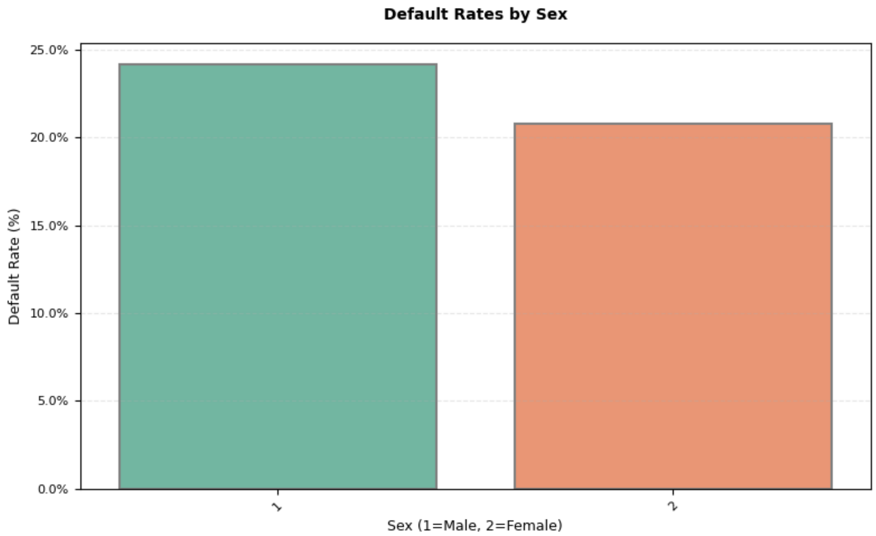

Figure 5: Default Distribution by Sex

Based on the Figure 4, several key patterns emerge regarding the age distribution across different demographic variables. The analysis by sex reveals that both male and female groups share nearly identical median ages of approximately 35 years, with very few individuals exceeding 65–70 years of age. When examining age by education level, each educational category displays a similar central range, generally spanning from about 25 to 45 years, though outliers are present in the 60–70 year range, suggesting a small presence of older individuals across all education groups. A more pronounced difference is observed in the age distribution by marital status: single individuals exhibit a noticeably lower median age compared to those who are married or belong to other marital categories. This pattern may reflect life-stage trends, where younger individuals are more likely to be unmarried. Overall, these visualizations highlight both consistency in age medians across sex and education, as well as meaningful variation by marital status.

According to the analysis of default rates across different demographic segments, several patterns emerge. The distribution of defaults by sex reveals that males exhibit almost a 25% default rate, 4 percentage points higher than females (Figure 5). Furthermore, we chose 22.15% as the maximum payment default for our analysis after computing the default average in our dataset. When examining education levels (Figure 6), default rates display a negative correlation with educational attainment; individuals with graduate school education demonstrate lower probability of default, followed by university graduates, while those with high school education show progressively higher default rates (more than 25%). Even though the Other and Unknown categories constitute the lowest default rate, we note that we do not possess sufficient information to draw conclusions. Similarly, in Figure 7 marital status shows distinct patterns in default behavior, with single individuals maintaining the lowest default rates (around 21%), married persons showing moderately higher rates, and the Other category exhibits the highest incidence of defaults.

The relationship between credit limits and default behavior reveals distinct patterns in portfolio risk composition. Customers who default consistently demonstrate concentration at the lower end of the credit spectrum, with their density distribution skewed toward more modest credit limits compared to non-defaulting counterparts (Figure 8). The overall credit limit distribution shows a right-skewed pattern, where the majority of accounts cluster below NT$ 200,000, while higher limits become progressively less frequent (Figure 9).

While the modeling approach for this project was designed around credit limit as the sole risk exposure variable, exploratory data analysis was conducted to deepen our understanding of the portfolio composition. This analysis examined how credit limits interact with client demographics and default behavior.

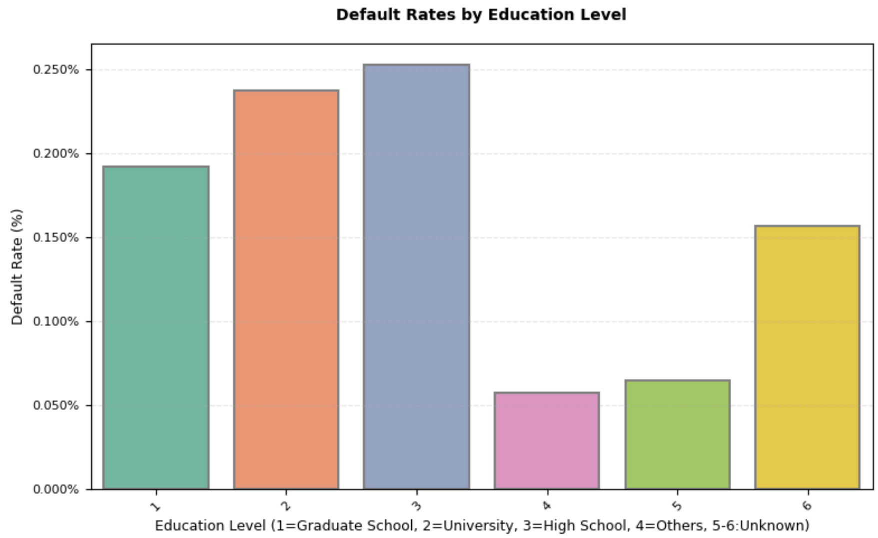

Figure 6: Default Distribution by Education

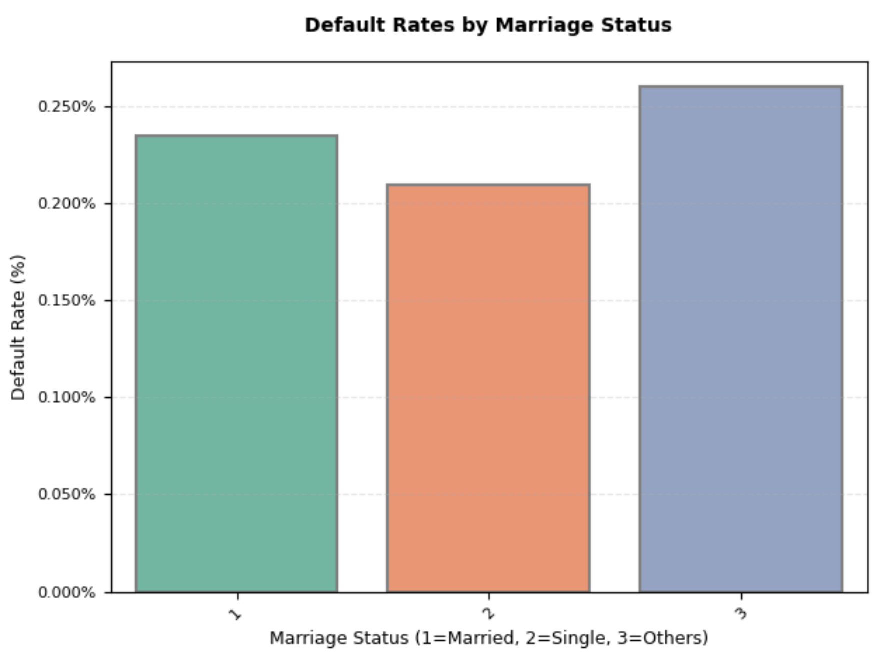

Figure 7: Default Distribution by Marriage

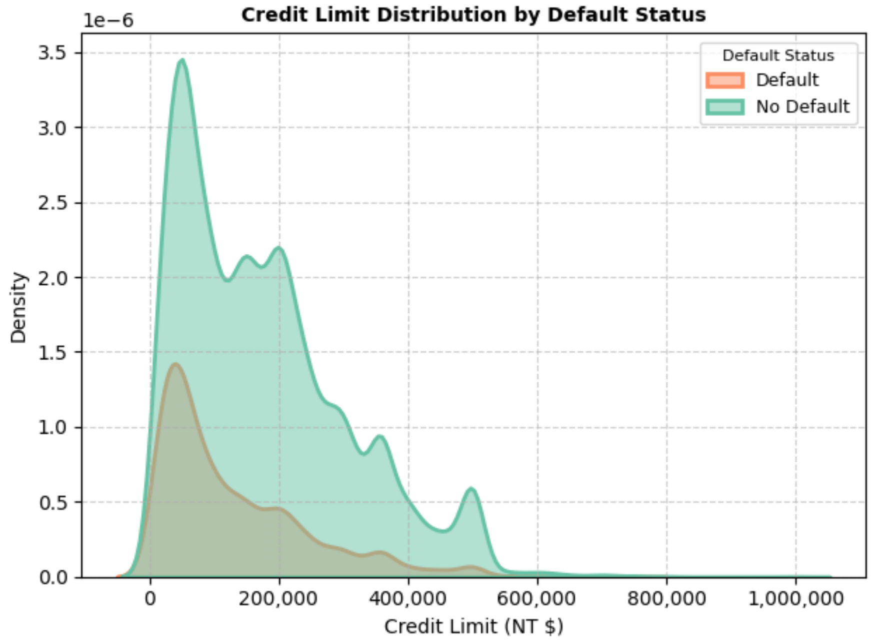

Figure 8: Credit Limit Distribution by Default Status

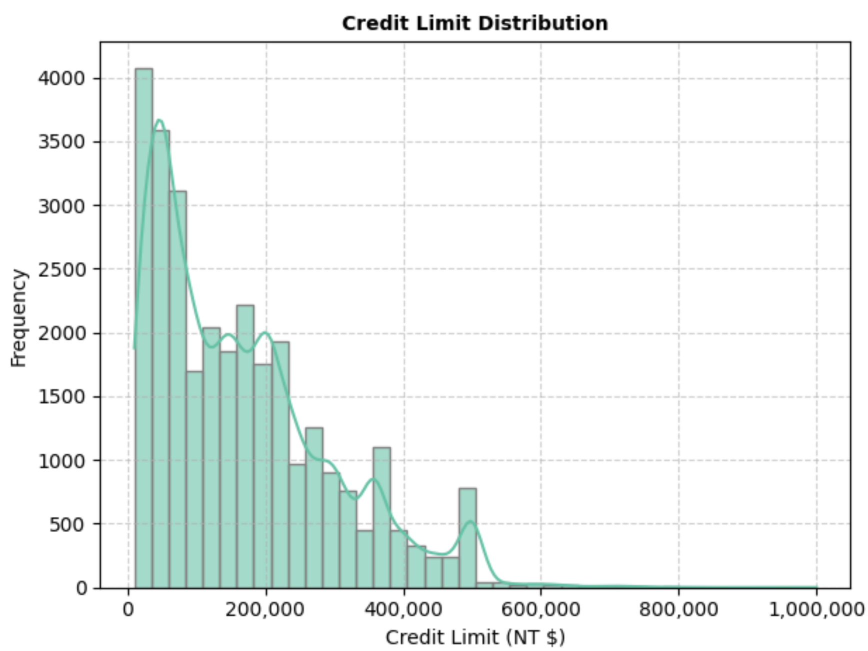

Figure 9: Credit Limit Distribution

4 Methodology

4.1 Core Framework

We conducted a Monte Carlo simulation study ($M = 10{,}000$ iterations) to analyze credit portfolio risk for $N = 30{,}000$ clients under the Vasicek single-factor model. This framework evaluates 12 distinct scenarios formed by the combination of four probability of default levels (PD $\in \{3\%, 9\%, 15\%, 22.15\%\}$) and three asset correlation values ($\rho \in \{0.05, 0.10, 0.15\}$). The Monte Carlo approach allows us to generate empirical loss distributions and estimate tail risk metrics that would be analytically intractable under realistic portfolio conditions.

We chose to work with the aforementioned default rates and correlation values following our research on financial risk assessment models (Basel Committee on Banking Supervision, 2006; Kang and Ma, 2007). The literature referred primarily to data from the United States, which showed default values from 3%–15% based on historical Federal Reserve statistics (Federal Reserve Bank of St. Louis, 2024). However, given that our data is from Taiwan, we added the average default value (22.15%) calculated from our dataset.

4.2 Key Assumptions

The model is built on several fundamental assumptions that simplify the complex reality of credit risk. We assume a single systematic risk factor $Z \sim \mathcal{N}(0,1)$ drives common movements across all borrowers, representing macroeconomic conditions or market-wide shocks. The asset correlation ρ is homogeneous across all borrower pairs, meaning the degree of co-movement in credit quality is constant throughout the portfolio. Individual risk factors $\varepsilon_i \sim \mathcal{N}(0,1)$ are mutually independent, capturing borrower-specific circumstances unrelated to systematic factors. Within each scenario, the probability of default is identical across all borrowers, reflecting a homogeneous portfolio assumption. Finally, we assume 100% loss given default, meaning the entire credit line exposure is lost upon default, which represents a conservative worst-case scenario.

4.3 Default Model

Default events are modeled through a latent variable framework that links observable defaults to unobservable creditworthiness. Each borrower i in simulation m has a latent creditworthiness measure:

$$X_{i,m} = \sqrt{\rho} \cdot Z_m + \sqrt{1-\rho} \cdot \varepsilon_{i,m} \tag{6}$$

This Vasicek formula decomposes credit risk into two components: a systematic component ($\sqrt{\rho} \cdot Z_m$) that affects all borrowers simultaneously through the common factor Z_m, and an idiosyncratic component ($\sqrt{1-\rho} \cdot \varepsilon_{i,m}$) unique to each borrower. The weights ensure that the correlation between any two borrowers' latent variables equals ρ.

Default occurs when creditworthiness falls below a critical threshold $\tau = \Phi^{-1}(\text{PD})$, where Φ^{-1} is the inverse standard normal CDF. This threshold is calibrated so that, unconditionally, each borrower has exactly a PD probability of defaulting. The default indicator is formally defined as:

$$D_{i,m} = 1\{X_{i,m} < \tau\} \tag{7}$$

where $1\{\cdot\}$ equals 1 if the condition holds and 0 otherwise. This threshold approach naturally incorporates correlation: when Z_m is extremely negative (bad economic state), many borrowers' $X_{i,m}$ values fall below τ simultaneously, creating clustered defaults.

4.4 Portfolio Loss Calculation

The total portfolio loss in each simulation run aggregates the credit exposures of all defaulting clients:

$$L_m = \sum_{i=1}^{N} D_{i,m} \cdot C_i \tag{8}$$

where C_i represents the credit limit for borrower i, obtained from the actual credit card dataset. This formulation weights defaults by exposure size, so larger credit lines contribute more to portfolio loss. Repeating this calculation across $M = 10{,}000$ simulations generates an empirical loss distribution $\{L_1, L_2, \ldots, L_M\}$ for each scenario, from which we can estimate both central tendencies and tail risks.

4.5 Risk Metrics

We quantify portfolio risk using four complementary metrics that capture different aspects of the loss distribution.

Expected Loss ($\text{EL} = \frac{1}{M}\sum_{m=1}^{M} L_m$) represents the average loss across all simulations, providing a baseline measure of typical portfolio performance. This metric is useful for pricing and provisioning but does not capture downside risk or extreme scenarios.

Maximum Loss (ML) represents the worst loss observed across all simulated portfolio outcomes, providing an upper bound on the loss distribution under the modeled assumptions. While VaR and CVaR summarize the behavior of the tail at specific confidence levels, ML captures the single most extreme realization generated by the Monte Carlo simulation. This metric is particularly informative in settings where default events may cluster under highly adverse economic conditions, creating outcomes that exceed even the 99% tail thresholds (Glasserman, 2003).

In practice, ML is obtained by computing the total portfolio loss for each simulation iteration and selecting the maximum among these values. Because it is driven entirely by the most extreme simulated scenario, ML is sensitive to both the number of simulations and the dependence structure generated by the asset correlation parameter. Higher correlation increases the likelihood of simultaneous defaults across the portfolio, which can substantially elevate the maximum observed loss even when expected losses remain relatively stable (Gordy, 2003).

Although ML is not typically used as a regulatory risk measure, it serves as an important diagnostic tool for stress testing and scenario analysis (Basel Committee on Banking Supervision, 2006). It highlights the theoretical upper tail risk embedded in a portfolio and helps illustrate how severe losses can become when probabilities of default and default correlations rise together. In the context of this analysis, ML complements Expected Loss, VaR, and CVaR by providing a complete picture of potential outcomes under extreme credit stress.

Value at Risk (VaR_α) identifies the loss threshold that is exceeded in only $(1-\alpha)\%$ of scenarios, computed at confidence levels $\alpha \in \{90\%, 95\%, 99\%\}$. For example, $\text{VaR}_{95\%}$ represents the minimum loss in the worst 5% of cases. VaR is widely used in regulatory capital calculations (Basel framework) and risk management because it provides a single, interpretable number for extreme loss potential. However, VaR has a limitation: it says nothing about how bad losses can be beyond the threshold.

Conditional Value at Risk ($\text{CVaR}_\alpha = E[L \mid L \geq \text{VaR}_\alpha]$) addresses this gap by measuring the expected loss conditional on exceeding VaR. Also known as Expected Shortfall, CVaR averages all losses in the tail beyond the VaR threshold, providing a more complete picture of extreme risk. CVaR is particularly important for understanding the severity of worst-case scenarios and is increasingly preferred in risk management because it accounts for tail shape, not just a single percentile.

5 Results

The following analysis focuses primarily on how portfolio losses behave as the probability of default (PD) increases, while correlation and confidence level serve as contextual parameters that shape the risk distribution.

The results allow us to understand how increasing PD influences both average losses and tail losses, and how this interacts with correlation and confidence level.

5.1 Effects of Increasing the PD at Fixed Correlation Levels

5.1.1 Correlation = 5%

At 5% correlation, increasing the probability of default (PD) from 3% to 22.15% produces a clear, monotonic increase across all loss measures, as shown in Tables 1 and 2.

Correlation	PD	Mean Default Rate	VaR 99% (NT$)	CVaR 99% (NT$)
5.00%	3.00%	2.99%	414,593,000	484,376,138
5.00%	9.00%	8.99%	1,025,199,617	1,141,292,693
5.00%	15.00%	15.05%	1,493,044,400	1,618,633,026
5.00%	22.15%	22.18%	2,014,251,803	2,185,773,155

Table 1: VaR and CVaR at Correlation = 5%

Both the VaR and CVaR demonstrate a considerable increase as the PD rises from 3% to 22.15%. For instance, noting from Table 1, the VaR increases nearly five times from nearly NT$415 million at 3% PD to about NT$2.01 billion at 22.15% PD. This sharp rise not only signifies a higher loss threshold but also suggests that the likelihood of exceeding this threshold during adverse market conditions becomes more pronounced.

Similarly, the CVaR increases from approximately NT$484 million to NT$2.19 billion across the same PD spectrum. The growth in CVaR indicates that average losses during extreme cases escalate, stressing the importance of understanding tail risk in portfolio management.

Correlation	PD	Mean Default Rate	Expected Loss (NT$)	Max Loss (NT$)
5.00%	3.00%	2.99%	150,431,748	860,577,680
5.00%	9.00%	8.99%	451,538,822	1,596,430,000
5.00%	15.00%	15.05%	756,066,316	2,062,626,000
5.00%	22.15%	22.18%	1,114,376,271	2,785,277,680

Table 2: Expected Loss and Maximum Loss at Correlation = 5%

As seen in Table 2, the expected loss exhibits a marked increase, escalating from about NT$150 million at a 3% PD to NT$1.11 billion at 22.15% PD, which represents over a seven-fold increase. This growth underscores how higher PD directly correlates with larger average losses anticipated over the portfolio's lifespan.

Likewise, the maximum loss figures indicate a similar trend regarding worst-case scenarios during adverse market conditions: it rises significantly from nearly NT$861 million at 3% PD to NT$2.79 billion at 22.15% PD. This near threefold increase highlights that as PD rises, the potential for extreme losses also grows considerably.

Given that expected losses are increasing at a faster pace than maximum losses at 5% correlation, we note the importance of focusing on average loss metrics in risk assessment and planning. It suggests that institutions can anticipate more frequent losses, which can strain cash flows, disrupt financial stability, and challenge budgetary forecasts. Thus, financial decision-makers must prioritize proactive risk management strategies that not only mitigate catastrophic losses but also address the cumulative impact of rising expected losses.

5.1.2 Correlation = 15%

Similar to at 5% correlation, Tables 3 and 4 show a clear positive relationship between increasing the PD (from 3% to 22.15%) and increasing risk metrics.

Correlation	PD	Mean Default Rate	VaR 99% (NT$)	CVaR 99% (NT$)
15.00%	3.00%	3.05%	746,879,600	948,810,032
15.00%	9.00%	8.90%	1,591,009,980	1,826,071,295
15.00%	15.00%	15.10%	2,241,624,900	2,547,059,071
15.00%	22.15%	22.12%	2,802,354,880	3,092,803,048

Table 3: VaR and CVaR at Correlation = 15%

In Table 3, we see that with a 15% correlation, VaR shows a substantial increase as the PD rises from approximately NT$747 million at 3% PD to NT$2.8 billion at 22.15% PD, representing nearly a fourfold increase. This upward trajectory is consistent with the trend observed at 5% correlation, where VaR increased approximately 4.88 times. The similarity in these patterns emphasizes that rising PD significantly heightens the potential loss threshold, signaling a growing risk environment across both correlation levels.

CVaR reinforces this trend, moving from about NT$949 million at 3% PD to nearly NT$3.1 billion at 22.15% PD at a 15% correlation, marking an increase of approximately 3.46 times. This sharp increase underscores the urgency of understanding tail risks, as the average loss during extreme scenarios escalates. At 5% correlation, CVaR rose about 4.65 times. Both scenarios indicate a pressing need for institutions to prepare for severe losses that could occur beyond typical expectations.

Correlation	PD	Mean Default Rate	Expected Loss (NT$)	Max Loss (NT$)
15.00%	3.00%	3.05%	153,359,063	1,725,656,000
15.00%	9.00%	8.90%	447,358,094	2,714,173,680
15.00%	15.00%	15.10%	758,844,876	3,498,532,000
15.00%	22.15%	22.12%	1,111,250,457	4,306,219,680

Table 4: Expected Loss and Maximum Loss at Correlation = 15%

Expected losses (Table 4 also exhibited a notable rise, climbing from about NT$153 million at 3% PD to NT$1.1 billion at 22.15% PD, representing over a seven-fold increase. This trend is mirrored at 5% correlation, where expected losses had a similar increase of approximately 7.39 times.

Maximum losses follow an upward trend as well, increasing from approximately NT$1.7 million at 3% PD to NT$4.3 million at 22.15% PD, a rise of about 2.53 times. In contrast, at 5% correlation, maximum losses progressed from about NT$861 million to almost NT$2.8 million, showcasing a more pronounced increase of approximately 3.26 times. This difference suggests that while extreme losses are a significant concern, the more frequent average losses warrant increasing attention for effective risk management.

Overall, these results highlight that higher correlation is associated with greater tail risk, as evidenced by the higher VaR and CVaR values at the 22.15% PD level. Greater correlation increases the likelihood of simultaneous defaults during periods of economic stress, thereby amplifying losses in severe scenarios. However, while correlation shifts the overall level of tail losses upward, the PD remains the primary driver of how rapidly these loss measures increase.

5.2 Values Across 90%, 95%, and 99% Confidence Intervals

This section examines how portfolio losses evolve across the 90%, 95%, and 99% confidence levels, and how these shifts depend on both the probability of default (PD) and the asset correlation. The results allow us to evaluate the steepness of the loss distribution tail, the degree of tail amplification caused by correlation, and the interaction between PD and confidence level. Table 5 displays the VaR and CVaR across our chosen correlation and PD values, and Table 6 shows these values for expected and maximum loss.

Correlation	PD	Mean Default Rate	VaR 90% (NT$)	VaR 95% (NT$)	VaR 99% (NT$)	CVaR 90% (NT$)	CVaR 95% (NT$)	CVaR 99% (NT$)
5%	3.00%	2.99%	255,722,000	301,864,500	414,593,000	323,867,383	371,622,071	484,376,138
5%	9.00%	8.99%	702,181,000	803,989,000	1,025,199,617	841,845,153	937,582,915	1,141,292,693
5%	15.00%	15.05%	1,107,440,000	1,230,335,296	1,493,044,400	1,277,444,693	1,391,620,873	1,618,633,026
5%	22.15%	22.18%	1,563,598,912	1,721,142,000	2,014,251,803	1,769,284,870	1,903,249,878	2,185,773,155
10%	3.00%	2.99%	298,145,600	378,324,500	569,897,383	414,294,096	493,974,866	674,738,702
10%	9.00%	9.00%	822,628,400	965,285,184	1,319,493,600	1,036,223,611	1,184,025,114	1,515,561,290
10%	15.00%	15.00%	1,273,699,680	1,476,231,384	1,862,155,077	1,538,892,300	1,714,110,589	2,068,319,429
10%	22.15%	22.30%	1,785,074,000	2,013,568,580	2,482,550,777	2,094,498,516	2,303,181,097	2,743,529,764
15%	3.00%	3.05%	341,106,768	452,173,000	746,879,600	515,251,131	637,160,585	948,810,032
15%	9.00%	8.90%	887,015,600	1,102,598,200	1,591,009,980	1,187,024,176	1,388,858,538	1,826,071,295
15%	15.00%	15.10%	1,399,558,768	1,669,167,884	2,241,624,900	1,770,508,914	2,018,879,892	2,547,059,071
15%	22.15%	22.12%	1,927,874,400	2,229,704,096	2,802,354,880	2,327,991,045	2,587,940,431	3,092,803,048

Table 5: VaR and CVaR Across Correlation and PD Levels

Correlation	PD	Mean Default Rate	Expected Loss (NT$)	Max Loss (NT$)
5%	3.00%	2.99%	150,431,748	860,577,680
5%	9.00%	8.99%	451,538,822	1,596,430,000
5%	15.00%	15.05%	756,066,316	2,062,626,000
5%	22.15%	22.18%	1,114,376,271	2,785,277,680
10%	3.00%	2.99%	150,036,296	1,266,616,000
10%	9.00%	9.00%	452,118,471	2,326,330,000
10%	15.00%	15.00%	753,974,697	2,696,746,000
10%	22.15%	22.30%	1,120,543,646	3,587,896,000
15%	3.00%	3.05%	153,359,063	1,725,656,000
15%	9.00%	8.90%	447,358,094	2,714,173,680
15%	15.00%	15.10%	758,844,876	3,498,532,000
15%	22.15%	22.12%	1,111,250,457	4,306,219,680

Table 6: Expected Loss and Max Loss Across Correlation and PD Levels

5.2.1 Percentage Change in VaR Between Confidence Levels

Correlation	PD	%Δ VaR 90$\to$95	%Δ VaR 95$\to$99
5%	3.00%	18.05%	37.33%
5%	9.00%	14.51%	27.52%
5%	15.00%	11.10%	21.36%
5%	22.15%	10.08%	17.03%
10%	3.00%	26.90%	50.64%
10%	9.00%	17.35%	36.69%
10%	15.00%	15.90%	26.14%
10%	22.15%	12.80%	23.29%
15%	3.00%	32.57%	65.18%
15%	9.00%	24.31%	44.30%
15%	15.00%	19.26%	34.30%
15%	22.15%	15.66%	25.68%

Table 7: Percentage Change in VaR Between Confidence Levels

5.2.2 Percentage Change in CVaR Between Confidence Levels

Correlation	PD	%Δ CVaR 90→95	%Δ CVaR 95→99
5%	3.00%	14.74%	30.28%
5%	9.00%	11.38%	21.76%
5%	15.00%	8.95%	16.31%
5%	22.15%	7.58%	14.83%
10%	3.00%	19.28%	36.65%
10%	9.00%	14.28%	28.09%
10%	15.00%	11.39%	20.64%
10%	22.15%	9.97%	19.14%
15%	3.00%	23.71%	48.99%
15%	9.00%	17.02%	31.49%
15%	15.00%	14.04%	26.18%
15%	22.15%	11.17%	19.52%

Table 8: Percentage Change in CVaR Between Confidence Levels

5.2.3 Percentage Change for Expected Loss to Max Loss

Correlation	PD	Expected Loss → Max Loss (%)
5%	3.00%	473.3%
5%	9.00%	253.6%
5%	15.00%	172.8%
5%	22.15%	150.0%
10%	3.00%	744.0%
10%	9.00%	414.6%
10%	15.00%	257.6%
10%	22.15%	220.1%
15%	3.00%	1025.0%
15%	9.00%	506.7%
15%	15.00%	361.1%
15%	22.15%	287.5%

Table 9: Percentage Change: Expected Loss to Maximum Loss

5.2.4 Analysis

The results in Tables 7, 8, and 9 reveal several clear patterns regarding how portfolio losses evolve across confidence levels for different combinations of PD and asset correlation. Across all parameter settings, both VaR and CVaR increase monotonically as the confidence level rises from 90% to 95% and then to 99%. This reflects the deeper exploration of the tail of the loss distribution at higher confidence levels. For instance, at 5% correlation and a 3% PD, VaR rises by 18.05% when moving from the 90% to the 95% confidence level and by an additional 37.33% from 95% to 99% (Table 7). CVaR exhibits a similar pattern, with increases of 14.74% and 30.28% over the same intervals (Table 8). These consistent upward shifts confirm that higher confidence levels systematically reveal more severe potential losses.

Probability of Default (PD) continues to be the dominant driver of the absolute magnitude of portfolio losses. For a fixed correlation level, higher PD values generate substantially larger VaR and CVaR. This is reflected not only in the raw loss values but also in the diminishing percentage changes as PD rises. For example, at 5% correlation, the percentage increase in VaR from 95% to 99% declines from 37.33% at 3% PD to 17.03% at 22.15% PD (Table 7). The same monotonic decline appears in

CVaR, where the corresponding increase falls from 30.28% to 14.83% (Table 8). Higher PD leads to a heavier baseline loss distribution, leaving proportionally less room for extreme tail amplification.

Correlation, by contrast, governs the shape and severity of the tail. At low PD levels, increases in correlation significantly steepen the tail, resulting in much larger proportional jumps in both VaR and CVaR when moving to higher confidence levels. For example, when the PD is 3%, the shift in VaR from 90% to 95% rises from 18.05% at 5% correlation to 26.90% at 10% correlation, and further to 32.57% at 15% correlation (Table 7). Similarly, the 95%→99% CVaR grows from 30.28% at 5% correlation to 36.65% at 10% and 48.99% at 15% (Table 8). These results illustrate how correlation amplifies tail risk by increasing the probability of clustered defaults, especially when defaults are otherwise rare.

Table 9 further highlights the disproportionate effect of correlation on extreme losses. At low PD levels, maximum loss can exceed expected loss by very large multiples. For example, at 15% correlation and 3% PD, the maximum loss is more than ten times the expected loss (1025%), underscoring the dominance of rare but highly correlated default events. As PD increases, this ratio declines because expected loss itself rises; however, for any given PD, higher correlation consistently produces a larger gap between expected and worst-case outcomes.

Overall, the tables collectively demonstrate a consistent interplay between PD and correlation. PD determines the baseline scale of losses, while correlation governs their extremity by shaping the tail of the distribution. The higher the confidence level, the more pronounced this nonlinear tail behavior becomes, especially in portfolios with low PD and high correlation. These findings highlight the importance of jointly modeling PD and correlation when assessing portfolio credit risk across multiple confidence levels.

5.3 VaR and CVaR 99% by Parameter Combination

5.3.1 VaR 99%

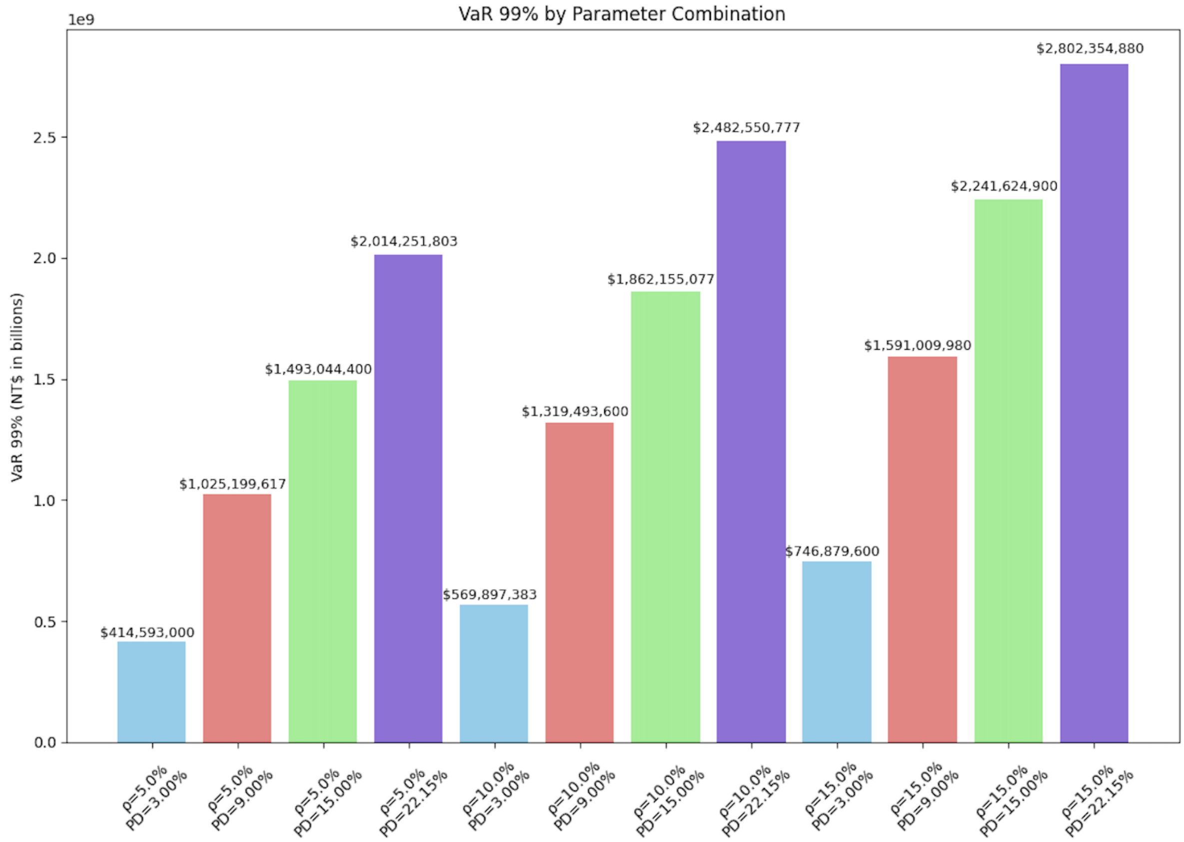

Figure 10: VaR 99% by Parameter Combination

5.3.2 CVaR 99%

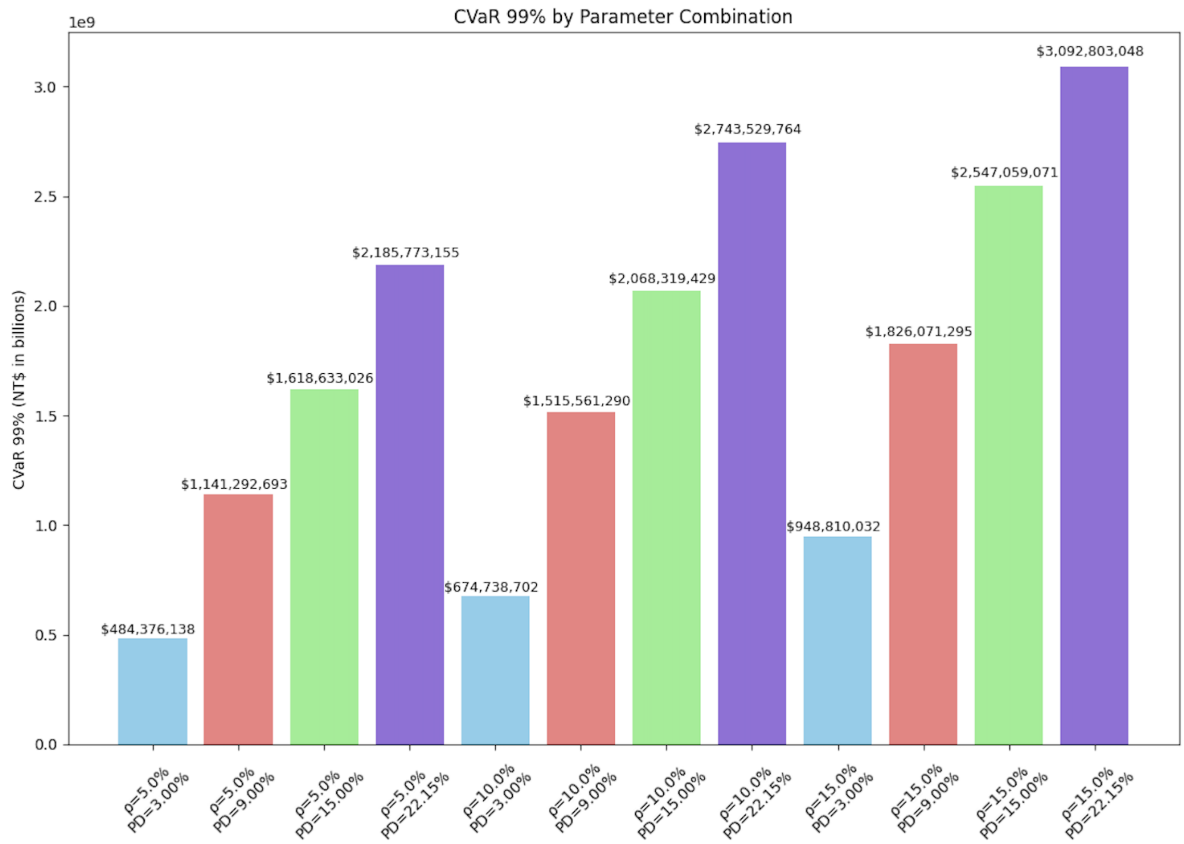

Figure 11: CVaR 99% by Parameter Combination

In Figures 10 and 11, we can see how the portfolio's risk changes under different combinations of default probability and default correlation. Figure 10 shows the VaR and Figure 11 shows the CVaR, both at the 99% confidence level. At this 99% confidence level, VaR tells us the loss threshold that will only be exceeded in the worst 1% of scenarios, and CVaR measures the average loss in those worst 1% of scenarios. This means that CVaR will always be higher than VaR, which is exemplified in these graphs, because for every combination of PD and correlation, the CVaR is higher than the corresponding VaR.

Through both of these figures, we can see a very clear pattern: higher probability of default leads to larger losses, because more borrowers default on average, and higher correlation dramatically amplifies extreme losses, as defaults tend to cluster together during bad economic states. When both PD and correlation are high, such as when correlation = 15% and PD = 22.15%, we get the largest tail losses, with VaR exceeding NT$2.8 billion and CVaR exceeding NT$3.0 billion (compared with VaRs and CVaRs of only a little over NT$414 million and NT$484 million, respectively, when correlation and PD are just 5% and 3%).

Together, Figures 10 and 11 make one thing incredibly clear: both the probability of default and the correlation between defaults are powerful drivers of risk, and when they rise together, tail losses increase explosively.

5.4 Expected Loss by Combination

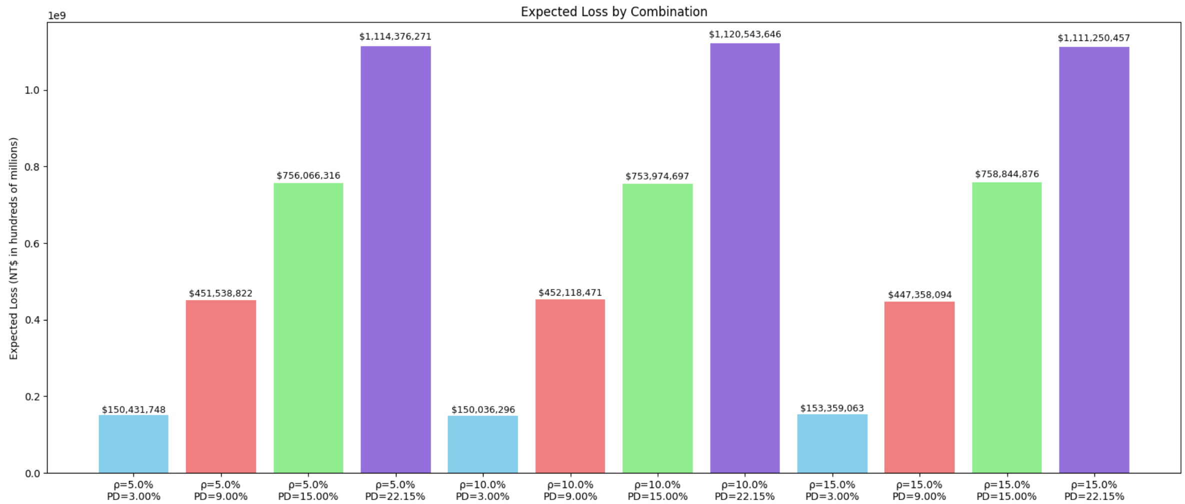

Figure 12: Expected Loss by Combination

Figure 12 displays the expected loss for each parameter combination of probability of default and correlation, providing a baseline measure of the portfolio's average credit loss before considering any tail-risk effects. Unlike the VaR and CVar results presented in Figures 10 and 11, which capture extreme outcomes in the 1% tail, expected loss reflects the typical loss the bank would anticipate over time.

As the figure shows, expected loss increases steadily with higher PD values but remains almost unchanged across different correlation levels, since correlation influences the clustering of defaults but not their long-run average. For example, when PD is 3%, expected loss stays close to NT$150 million regardless of whether correlation is 5, 10, or 15%. When PD rises to 22.15%, expected loss increases to roughly NT$1.1-1.2 billion across all correlation settings. This illustrates how expected loss is driven almost entirely by the underlying default probability, which is exactly what theory predicts, as the expected loss formula is just *probability of default*$\times$ *loss given default* $\times$*exposure*. Thus, the probability of default drives those average, everyday losses, while correlation drives the catastrophic ones.

5.5 Loss Distributions by Combination

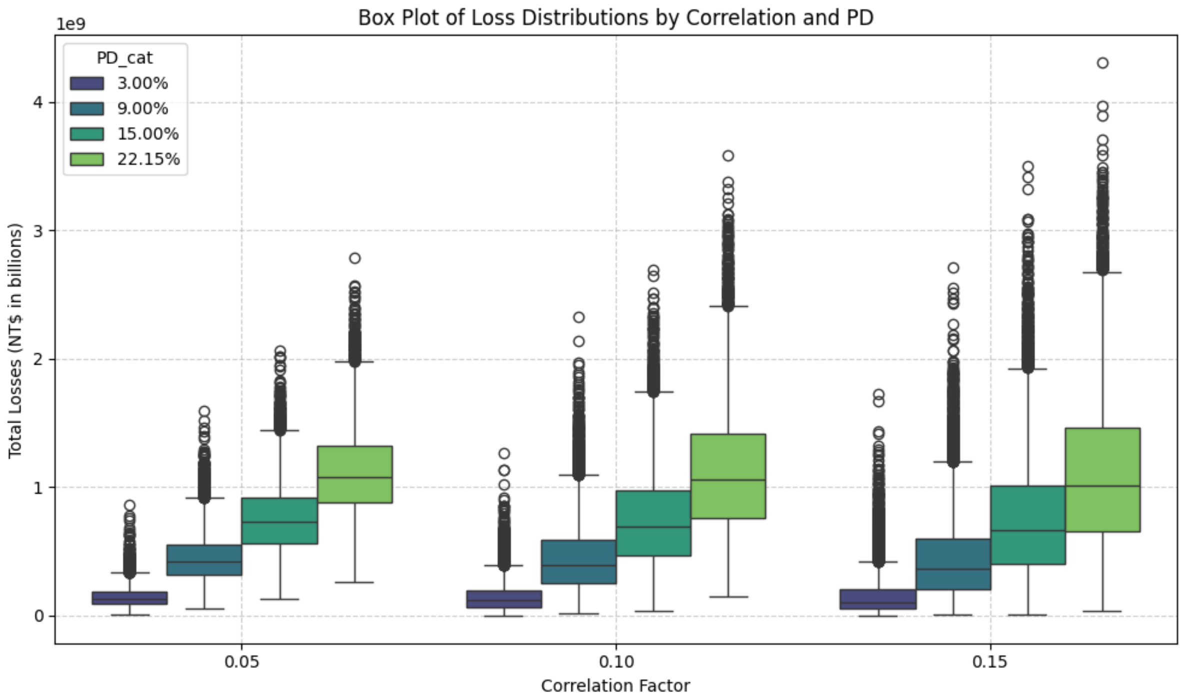

Figure 13: Loss Distributions by Combination

Figure 13 presents box plots of the full loss distributions at each correlation level (ρ = 5%, 10%, 15%) across all four PD categories, allowing us to see not only the average losses but also the spread, variability, and presence of extreme tail events. The 99% VaR and CVaR values shown earlier in Figures 10 and 11 summarize single points in the tail, but the box plots reveal the entire distribution shape behind those tail-risk measures. The box plots show that losses increase monotonically as PD rises; at ρ = 5%, median losses move from roughly NT$120 million at PD = 3% to over NT$1.0 billion at PD = 22.15%, illustrating how expected losses scale with default likelihood. At the same time, increasing correlation consistently stretches the upper tail: for PD = 15%, the upper whiskers grow from around NT$1.4 billion at ρ = 5% to more than NT$1.9 billion at ρ = 15%, reflecting a higher chance of large, clustered loss events. The thick concentration of outliers at the top of each distribution, especially at high PD and high correlation, shows that the portfolio can experience rare but extremely large losses, consistent with the elevated CVaR values observed in Figure 11.

Overall, Figure 13 makes it clear that correlation does not just affect the mean loss; it actually affects the entire distribution's spread, widening the range of possible outcomes and increasing the likelihood of catastrophic joint defaults. This visualization reinforces why high-confidence measures like the 99% VaR and CVaR are necessary: the loss distributions become heavily skewed at higher PD and correlation, making extreme outcomes far more impactful to risk management and capital planning.

5.6 Limitations and Future Research

While this study provides a comprehensive analysis of how probability of default drives portfolio losses, the exploration of correlation's role, though insightful, was limited to a range of 5% to 15%. To more fully isolate and understand the impact of the correlation factor itself, future work could implement a dedicated sensitivity analysis. This would involve holding the PD constant at a specific level (e.g., the portfolio's average of 22.15%) and systematically varying the correlation across a wider spectrum, such as from 1% to 30%. This approach would illustrate how the shape of the loss distribution—specifically the fatness of the tail and the gap between Expected Loss and VaR—changes purely as a function of default correlation, providing even deeper insight into the driver of catastrophic, clustered-loss scenarios.

6 Contributions

This project makes several key contributions to credit risk management:

Methodological Implementation: we developed a complete operational framework that translates the theoretical Vasicek model into a practical Monte Carlo simulation tool, demonstrating how financial institutions can implement correlated default modeling for retail portfolios.

Risk Parameter Analysis: our comprehensive sensitivity analysis provides clear evidence of how probability of default and correlation interact differently - showing that PD primarily drives expected losses while correlation governs tail risk amplification and extreme loss scenarios.

Practical Risk Metrics: the study delivers concrete, quantifiable risk measures (VaR, CVaR, Expected Loss) across multiple confidence levels and parameter combinations, providing financial institutions with actionable data for capital allocation decisions.

Educational Framework: this work serves as an applied learning demonstration that bridges theoretical finance concepts with practical implementation, showing how default correlation fundamentally changes risk assessment compared to traditional independent-default models.

Visualization Tools: we created intuitive visualizations that effectively communicate complex risk concepts, including loss distributions, tail risk amplification, and the interaction between PD and correlation parameters.

Non-Technical Section

1 Background and History of Capital One

Capital One began in 1988 when Richard Fairbank and Nigel Morris came up with a new idea for banking: using data and technology to tailor credit card offers to individual customers. Instead of offering one generic card to everyone, they believed they could analyze people's financial habits and create products that fit different needs. The company first operated as part of Signet Bank, and its data-driven approach worked so well that Capital One was spun off as its own company in 1994. From there, it quickly became known as one of the most innovative players in the credit card industry, using technology in ways that most banks at the time had not even considered.

As Capital One grew, it started expanding beyond credit cards. In the 2000s, the company bought several regional banks, which allowed it to move into full-service consumer and commercial banking. These acquisitions helped Capital One build a strong national presence and offer checking accounts, savings products, and auto loans. Along the way, the company continued to focus heavily on technology and customer experience, eventually becoming known for things like its mobile banking tools. Today, Capital One is one of the largest banks in the U.S., recognized for combining traditional banking with a tech-forward, data-driven mindset (Capital One Financial Corporation, nd).

2 Role: Principal Data Analyst - Credit Risk and Innovation

2.1 Responsibilities

The Principal Data Analyst leads analytical work that supports Capital One's credit risk strategy and drives the development of innovative data solutions.

Beyond general analytics, a major responsibility involves transforming raw data into decision-ready insights. Analysts work with large, complex datasets using Python, SQL, R, and Spark, build automated data pipelines, and create cloud-based workflows using AWS services. They also design and implement analytical methods that support credit policy evaluation, portfolio monitoring, segmentation, and risk measurement. This includes developing metrics, producing dashboards, and clearly communicating findings to business stakeholders and risk leaders.

A broader but equally important part of the role focuses on maintaining high standards for data governance, reliability, and analytical rigor. Analysts document data definitions, monitor data lineage, and ensure consistency across systems used for credit modeling and reporting. They collaborate closely with credit strategy teams, engineers, and product managers to define analytical requirements and validate datasets used for regulatory and strategic purposes. In addition, analysts contribute to organizational learning by mentoring junior team members, sharing best practices, and actively exploring new tools, open-source technologies, and analytical approaches that can improve efficiency, scalability, or insight generation across the credit risk organization.

2.2 Key Skills

The Principal Data Analyst position at Capital One demands a blend of robust technical, analytical, and business skills, particularly in credit risk and innovation. Candidates should be proficient in programming languages such as Python, R, SQL, and Spark, and have experience with large, complex datasets in both structured (e.g., data warehouses/SQL) and unstructured formats. These abilities are crucial for efficiently manipulating and extracting insights from extensive credit and transactional data, facilitating quicker and more accurate decision-making. Additionally, experience with AWS services and leveraging open-source tools to develop automated, self-service data solutions that promote scalability and reproducibility is highly valued.

Analysts must also excel in business intelligence and analytics. They should be capable of translating business requirements into actionable insights by designing dashboards, metrics, and analytical tools using visualization platforms such as Tableau or Power BI. These skills are vital for effectively communicating complex data findings to stakeholders and shaping credit strategies, and a solid foundation in statistical modeling and data analysis enables analysts to identify patterns, quantify risk, and support data-driven decision-making.

Likewise, data management and governance are essential aspects of the role. Candidates should implement data quality principles, including metadata, lineage, and business definitions, while monitoring and enforcing data access and security standards. This ensures that analyses are based on reliable, consistent data, which is particularly significant in the regulated financial sector. Experience in delivering data governance and quality management practices within financial services demonstrates familiarity with industry best practices and compliance requirements.

Familiarity with process improvement methodologies like Agile, Lean, or Six Sigma is beneficial, aiding analysts in managing projects, prioritizing tasks, and enhancing workflows. A deep understanding of credit risk, including concepts like probability of default and loss-given-default, enables accurate data interpretation and insights that directly influence business outcomes. Finally, strong communication and collaboration skills, coupled with a curiosity-driven mind, are key for working effectively with both technical and business teams, quickly learning new technologies, and contributing to innovative, data-driven solutions that enhance Capital One's credit risk management and strategic decision-making.

2.3 Suggested Preparation

To be competitive for this role at Capital One, several steps can be taken to build the required skills, experience, and professional profile:

1. Strengthen Technical Skills
    - **Python/R/SQL/Spark:** Complete advanced online courses, such as those on Coursera, DataCamp, or Udemy, focused on data analysis, scripting, and large-scale data processing; apply these skills in a mix of class projects, personal datasets, and simulations.
    - **AWS Services:** Take AWS training, such as the AWS Certified Data Analytics - Specialty, and practice deploying small data pipelines or storing large datasets in S3 and querying with Athena; explore open datasets to practice cloud-based analytics beyond our project.
    - **Business Intelligence Tools:** Learn Tableau or Power BI through tutorials and implement dashboards using class projects or publicly available datasets to strengthen visualization skills.
2. Deepen Knowledge of Credit Risk and Financial Analytics
    - **Credit Risk Models:** Take specialized courses in credit risk through CFA Institute, Moody's Analytics, or online platforms to understand PD, LGD, EAD, and Basel regulatory frameworks.
    - **Simulation Techniques:** Apply Monte Carlo or other stochastic methods to various datasets (not just our project) to understand risk modeling in different contexts.

- **Correlation Analysis:** Conduct sensitivity analyses in small portfolio datasets or public credit datasets to see how correlations impact tail risk and expected losses.

3. Gain Experience with Data Management and Governance

   - **Data Quality and Lineage:** Participate in internships or projects requiring cleaning, documenting, and integrating datasets; practice defining metadata and business rules for reproducibility.
   - **Security and Access Governance:** Learn fundamentals of data privacy and security in SQL and cloud environments; experiment with access controls on sample or anonymized datasets.

4. Develop Business Intelligence and Communication Skills

   - **Data Storytelling:** Present class projects or portfolio analyses in clear narratives emphasizing actionable insights for business decisions.
   - **Dashboarding and Metrics:** Build dashboards visualizing key metrics for multiple datasets, including but not limited to the Monte Carlo project, to practice translating data into insights.

5. Professional Development and Networking

   - **Internships/Co-ops:** Seek roles in financial services or fintech focusing on data analytics or credit risk to gain applied experience.
   - **Conferences and Seminars:** Attend webinars or events on data analytics, credit risk, and financial services to learn trends and network with professionals.
   - **Certifications:** Pursue CFA, FRM, or data analytics certifications to demonstrate domain knowledge and commitment.

2.4 Motivation for the Role

We are drawn to Capital One's focus on applying advanced analytics and data-driven insights to real-world credit risk challenges. This role offers the opportunity to leverage technical skills such as statistical modeling, Python/R programming, SQL, and cloud-based data management to build meaningful solutions. Our experience with Monte Carlo simulations for credit card portfolio risk, where we quantified VaR and CVaR under varying default probabilities and correlations, illustrates how analytical frameworks can directly inform business decisions. We are motivated by the chance to extend these techniques to large-scale, real-world datasets, develop actionable dashboards, and contribute to risk management strategies that have a tangible impact.

References

Amadeo, K. (2022). 2008 financial crisis causes, costs, and could it happen again? The Balance.

Basel Committee on Banking Supervision (2006). International convergence of capital measurement and capital standards: A revised framework (comprehensive version). Technical report, Bank for International Settlements, Basel, Switzerland. ISBN web: 92-9197-720-9.

Capital One Financial Corporation (n.d.). Our company.

Federal Reserve Bank of St. Louis (2024). Delinquency rate on credit card loans, all commercial banks. FRED, Federal Reserve Economic Data. Retrieved November 29, 2025.

Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering.* Springer.

Glasserman, P. and Li, J. (2005). Importance sampling for portfolio credit risk. *Management Science*, 51(11):1643–1656.

Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12(3):199–232.

Joseph, C. (2013). *Advanced Credit Risk Analysis and Management*. Wiley.

Kang, T. S. and Ma, G. (2007). Recent episodes of credit card distress in asia. *BIS Quarterly Review*.

Li, D. X. (2000). On default correlation: A copula function approach. *Journal of Fixed Income*, 9(4):43–54.

Löffler, G. and Posch, P. N. (2007). *Credit Risk Modeling Using Excel and VBA*. John Wiley & Sons, Chichester, UK.

Vasicek, O. (1987). Probability of loss on loan portfolio. Technical report, KMV Corporation.

Vasicek, O. (2002). Loan portfolio value. *Risk*, 15(12):160–162.