PLP Project Paper

Team 2

Dmytro Daniuk, Lorenz Engelsberger

June 24, 2024

Preparations

Based on the assumption of the model of Czech mathematician Oldrich Vasicek from 2002 a loan defaults if the value of the borrower's assets at the loan maturity T falls below the contractual value B of its obligations payable. Vasicek lets A_i be the value of the i-th borrower's assets, described by the process

$$dA_i = \mu_i A_i dt + \sigma_i A_i dx_i$$

The asset value at T is than represented as

$$log A_i(T) = log A_i + \mu_i T - \frac{1}{2}\sigma^2{}_i T + \sigma_i \sqrt{T} X_i$$

where X_i is a standard normal variable. The variables X_i are jointly standard normal with equal pairwise correlations ρ^2, and are therefore represented as

$$X_i = Y\sqrt{\rho^2} + Z_i\sqrt{1-\rho^2}$$

where $Y, Z_1, Z_2, \ldots, Z_n$ are mutually independent standard normal variables (property of the equicorrelated normal distribution). The variable Y can be interpreted as a portfolio common factor, such as an economic index, over the interval (0,T). Then the term $Y\sqrt{\rho^2}$ is the company's exposure to the common factor and the term $Z_i\sqrt{1-\rho^2}$ represents the company specific risk. In Vasicek's original model of 2002, the correlation parameter was ρ, not ρ^2, but it was later proved that under certain circumstances the impact of economic factors can be simultaneously positive for some subjects and negative for others, and thus a negative value of ρ arises from which it is impossible to obtain the root. To avoid such cases, the indicator has to be squared.

Based on this formula, we can prove that $X_i \sim N(0,1)$. This can be done by proving that $E[X_i] = 0$, $Var(X_i) = 1$ and $corr(X_i, X_j) = \rho^2$, using fact that $Y, Z_1, Z_2, \ldots, Z_n$ are mutually independent standard normal variables:

$$E[X_i] = E[Y\sqrt{\rho^2} + Z_i\sqrt{1-\rho^2}] = \sqrt{\rho^2}E[Y] + \sqrt{1-\rho^2}E[Z_i] = \sqrt{\rho^2} \times 0 + \sqrt{1-\rho^2} \times 0 = 0$$

$$\begin{aligned} Var(X_i) &= E[X_i{}^2] - (E[X_i])^2 = E[X_i{}^2] - 0 = E[X_i{}^2] = E[(Y\sqrt{\rho^2} + Z_i\sqrt{1-\rho^2})^2] \\ &= E[Y^2\rho^2 + 2(Y\sqrt{\rho^2}Z_i\sqrt{1-\rho^2}) + Z_i{}^2(1-\rho^2)] = E[Y^2\rho^2 + 2(YZ_i\sqrt{\rho^2(1-\rho^2)}) + Z_i{}^2(1-\rho^2)] \\ &= \rho^2E[Y^2] + 2\sqrt{\rho^2(1-\rho^2)}E[YZ_i] + (1-\rho^2)E[Z_i{}^2] = \rho^2E[Y^2] + 2\sqrt{\rho^2(1-\rho^2)}E[Y]E[Z_i] + (1-\rho^2)E[Z_i{}^2] \\ &= \rho^2 \times 1 + 2\sqrt{\rho^2(1-\rho^2)} \times 0 \times 0 + (1-\rho^2) \times 1 = \rho^2 + 0 + 1 - \rho^2 = 1 \end{aligned}$$

$$\begin{aligned} corr(X_i, X_j) =& \frac{Cov(X_i, X_j)}{\sigma(X_i) \times \sigma(X_j)} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)} \times \sqrt{Var(X_j)}} = \frac{Cov(X_i, X_j)}{\sqrt{1} \times \sqrt{1}} = Cov(X_i, X_j) \\ &= E[X_iX_j] + E[X_i]E[X_j] = E[X_iX_j] + 0 \times 0 \\ &= E[X_iX_j] = E[(Y\sqrt{\rho^2} + Z_i\sqrt{1-\rho^2})(Y\sqrt{\rho^2} + Z_j\sqrt{1-\rho^2})] \\ &= E[Y^2\rho^2 + YZ_j\sqrt{\rho^2(1-\rho^2)} + YZ_i\sqrt{\rho^2(1-\rho^2)} + Z_iZ_j(1-\rho^2)] \\ &= \rho^2E[Y^2] + \sqrt{\rho^2(1-\rho^2)}E[Y]E[Z_j] + \sqrt{\rho^2(1-\rho^2)}E[Y]E[Z_i] + (1-\rho^2)E[Z_iZ_j] \\ =&\rho^2E[Y^2] + \sqrt{\rho^2(1-\rho^2)}E[Y]E[Z_j] + \sqrt{\rho^2(1-\rho^2)}E[Y]E[Z_i] + (1-\rho^2)E[Z_i]E[Z_j] \\ & \rho^2 \times 1 + 2 \times (\sqrt{\rho^2(1-\rho^2)} \times 0 \times 0) + (1-\rho^2) \times 0 \times 0) = \rho^2 + 0 + 0 = \rho^2 \end{aligned}$$

The probability of default of the i-th loan is then

$$p_i = P[A_i(T) < B_i = P[X_i < c_i] = N(c_i)$$

Where N is the cumulative normal distribution function and c_i is a critical treshold that in Vasicek model is calculated as

$$c_i = \frac{logB_i - logA_i - \mu_i T + \frac{1}{2}\sigma_i^2 T}{\sigma_i \sqrt{T}}$$

In our case the probabilities p_i are already known, so the critical treshold c_i can be represented as

$$c_i = N^{-1}(p_i)$$

Thus, if the risk factor X_i of a certain asset crosses a critical threshold c_i, the asset defaults.

Our bank's portfolio consists of two types of loans, namely personal loans to individuals and corporate loans to companies. Both of these types of loans have their own peculiarities, which had to be taken into account in our model. First of all, we conducted a study of open information to select the necessary parameters on the basis of which the simulation will be conducted.

Personal loans are typically smaller in size and granted to individual consumers for various purposes such as purchasing a car, financing education, or consolidating debt. The simulation of personal loan losses often considers factors like credit scores, income levels, employment status, and debt-to-income ratios to assess the creditworthiness of borrowers. Our team already had probabilities of default of individuals based on the analysis of historical data from Team 1a. These default probabilities became our input data for further simulation. In addition, other input factors were set. For all individuals, the maximum loan amount was set at EUR 11,500, which according to Investopedia was the average loan amount for an individual in 2023. This loan amount is small relative to the size of loans that legal entities can afford, but personal loans are much more risky.

Regarding the possibility of granting a loan to an individual client, it was decided that an individual could receive such a loan if the probability of his or her default was less than 25%.

As for the correlation between individual loans, a value of 0.22 was assumed, which when squared equals 0.05, showing that individuals are very little dependent on the overall economic situation and much more dependent on their own risks. According to research, rho values between 0.02 and 0.1 are best for personal loans, so our value of 0.05 should have fit the model well. However, to investigate the impact of this indicator on the final result, the following values were also tested: 0.01, 0.2, 0.3, 0.9.

As for LGD for personal loans, our team made the following decision. Personal loans can be secured or unsecured. If the loan is collateralized, then it is highly likely that in case of default, almost the entire loan amount will be reimbursed, and therefore LGD tends to 0. If there is no collateral, then the probability of repayment of the loan in the event of default by the borrower is low, and therefore LGD tends to 1. Of course, there is a possibility that this indicator will lie between these two values, although the probability of this is small.

To make this indicator continuous, it is necessary to build an appropriate distribution. After a little research, the beta distribution was chosen as the best option. The beta distribution can be given by the formula:

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

In this distribution, the ratio of the two coefficients α and β can be used to set the mean value of the distribution,

$$\mu = \frac{\alpha}{\alpha + \beta}$$

and the size of these coefficients can be used to determine how close the values of the distribution will be to the mean. We decided to combine the two beta distributions into one bimodal distribution. The first distribution had coefficients of $\alpha = 0.5$ and $\beta = 2$ with a mean close to 0, and the second distribution had coefficients of $\alpha = 2$ and $\beta = 0.5$ with a mean close to 1.

Similarly, we had calculated default probabilities for corporate loans based on historical data from Team 1b. Unlike personal loans, establishing a base case for corporate loans is more challenging because they are more heterogeneous and the amount of a possible loan can vary greatly depending on the company's credit rating. After studying the information on procedures and methods for determining credit ratings from the 2015 Fitch Ratings report, we divided the enterprises into 6 clusters and established the following corresponding inputs for them.

If a company has a default probability of less than 0.01%, it is rated AAA and can borrow up to EUR 600,000. If the probability of default is between 0.01% and 0.07%, these are AA+, AA and AA- rated companies that can count on a loan of EUR 450,000. If the probability of default is between 0.07% and 0.19%, these are enterprises in the A+, A, A- and BBB+ cluster, and they can receive a loan of EUR 300,000. If the probability of default is between 0.19% and 5.36%, these are companies rated BBB, BBB-, BB+, BB, BB-, B+, B, B-, and they can be granted a loan of up to EUR 100,000. Enterprises rated CCC+, CCC and CCC- with a default probability of 5.36% to 25.23% are already quite risky and cannot count on more than a small loan of EUR 25,000. All companies with a default probability of more than 25.23% are classified as CC, C, Rd and D and are considered too risky to be granted a loan.

Also, unlike individuals, often even in the event of a default, businesses can often recover a significant portion of the funds they receive. In calculating LGDs for corporate loans, it is common practice for many financial institutions to use the rule of 45, which means that the average LGD for a company is 45%. However, recent studies, in particular those by S&P Global, show that this figure is lower, averaging 40%, which is what we decided to use as the average for our simulation. But, as mentioned above, corporate clients are more heterogeneous, and therefore, setting a single average was not enough. Having studied historical data, we found that some companies are much better at repaying their debts, even in the event of a default, returning much more than 40%. The number of such companies is quite large, so this should be taken into account in the simulation. For the simulation, we again chose to combine two beta distributions into one bimodal distribution. The first distribution had coefficients of $\alpha = 6$ and $\beta = 10$ with a mean close to 0.375, and the second distribution had coefficients of $\alpha = 4.5$ and $\beta = 2$ with a mean close to 0.7. A schematic representation of the LGD distributions for personal and corporate loans in all our simulations can be found in Appendix A. In this way, the LGD will be continuous, but still realistic, which will ensure good model accuracy.

As for the correlation parameter, enterprises, unlike individuals, are much more dependent on general economic risk. The company's own risk is still the main one, but the role of systematic risk is increasing significantly, as problems in the economy have a strong impact on all companies at the same time. For our simulation, we chose a value of 0.45, which, when squared, yields a value of 0.2, which is within the recommended range of 0.15 to 0.3 for corporate entities. However, to investigate the impact of this indicator on the final result, the following values were also tested: 0.02, 0.4, 0.5, 0.99.

In order to avoid major difficulties in the process of building the model, we did not set a maximum loan amount for our bank. The model had to calculate it by itself, taking into account all the input data. The only thing we chose was an asset split. To determine the impact of this parameter on the simulation results, it was decided that the following ratios of personal loans to corporate loans would be tested: 3:7, 2:3, 1:1, 3:2, 7:3.

Python Implementation

The implementation of the portfolio simulation project was carried out using Python, leveraging a suite of libraries to facilitate statistical simulations, visualizations, and numerical computations. The primary libraries utilized were 'scipy.stats' (for statistical functions such as 'norm' and 'beta' distributions), 'matplotlib' (for plotting), 'numpy' (for numerical operations), 'pandas' (for data manipulation), and 'seaborn' (for advanced visualization).

The code structure and workflow for the project were designed mainly for clarity and transparency of the method. Initially, the probability of default data from teams 1a and 1b was imported. Following this, fixed parameters for the simulations were established and a random seed was set to ensure the repeatability of the results to maintain consistency throughout the analysis.

The core of the simulation involved the creation of five distinct portfolios, each with different distributions of assets: 30/70, 40/60, 50/50, 60/40, and 70/30. Samples were drawn from the imported data frames to assign the probability of default (PD) for all loans within these portfolios. Loan amounts were then allocated based on the PD values as previously described in the methodology.

To explore the impact of correlation on the portfolios, the simulation was run 25 times for each of the five different correlation values and five different portfolio distributions. After running the simulations, key figures and metrics were calculated from the results, providing insights into the performance and risk characteristics of each portfolio configuration.

By structuring the code in this systematic manner, the project not only achieved its objective of simulating various portfolio scenarios but also ensured that the results were reliable, reproducible and comprehensive.

Interpreting The Simulation Results

In our portfolio simulation project, we utilized the Vasicek modeling method to simulate losses across a range of portfolio distributions and rho correlation values. Our primary focus is on the results from Simulation 31 (with correlation values of rho_personal3 = 0.22 rho_business3 = 0.45, and a loan distribution of 30% private 70% business), which we assume to be a middle of the road scenario.

Simulation 31, with a portfolio value of €168,715,000.00 comprising €1,840,000.00 in personal loans and €166,875,000.00 in business loans, yielded the following key figures: an Expected Loss (EL) of €44,316.36, an Unexpected Loss (UL) of €113,549.23, a Value at Risk (VaR) at 99% confidence level of €379,406.64, and a VaR at 95% confidence level of €107,336.57. These figures are instrumental in understanding the risk profile of this particular portfolio distribution.

When examining the trends across the different simulations, we observed that portfolios with higher proportions of personal loans tended to have slightly higher Expected Losses. For instance, Portfolio 5, which had the highest percentage of personal loans, showed an EL of €61,759.33. This trend is consistent across the various simulations, indicating a direct correlation between the proportion of personal loans and the Expected Loss.

Unexpected Losses exhibited more variability, particularly in portfolios with different rho values. For example, Portfolio 5 showed a substantial increase in UL, reaching €643,884.23 in Simulation 51, compared to €113,549.23 in Simulation 31. This suggests that portfolios with higher rho values are more susceptible to larger deviations from the expected loss, emphasizing the importance of considering correlation in risk assessments.

Value at Risk (VaR) metrics also displayed significant variation across simulations. The VaR at 99% confidence level for Portfolio 31 was €379,406.64, whereas Portfolio 5's VaR reached €1,021,210.20 in Simulation 55. Similarly, the VaR at 95% confidence level for Portfolio 31 was €107,336.57, compared to €342,877.42 for Portfolio 5 in the same simulation. These results underscore the impact of portfolio composition and correlation on potential extreme losses.

In conclusion, our findings from the Vasicek model simulations reveal critical insights into the risk characteristics of different portfolio distributions. Simulation 31 serves as a benchmark for understanding these dynamics, highlighting the effects of varying rho values and portfolio compositions on Expected Loss, Unexpected Loss, and Value at Risk.

The figure in Appendix B shows the cumulative distribution of our bank's losses. Each subplot in this set represents the cumulative distribution function (CDF) of the simulated losses. The CDF graphs help in understanding where the bulk of losses lie, highlighting the tail risk and the likelihood of extreme losses. The CDF curves for different correlation settings show how the tail risk changes with varying rho values and how the bulk of losses can shift or change its shape based on the input parameters.

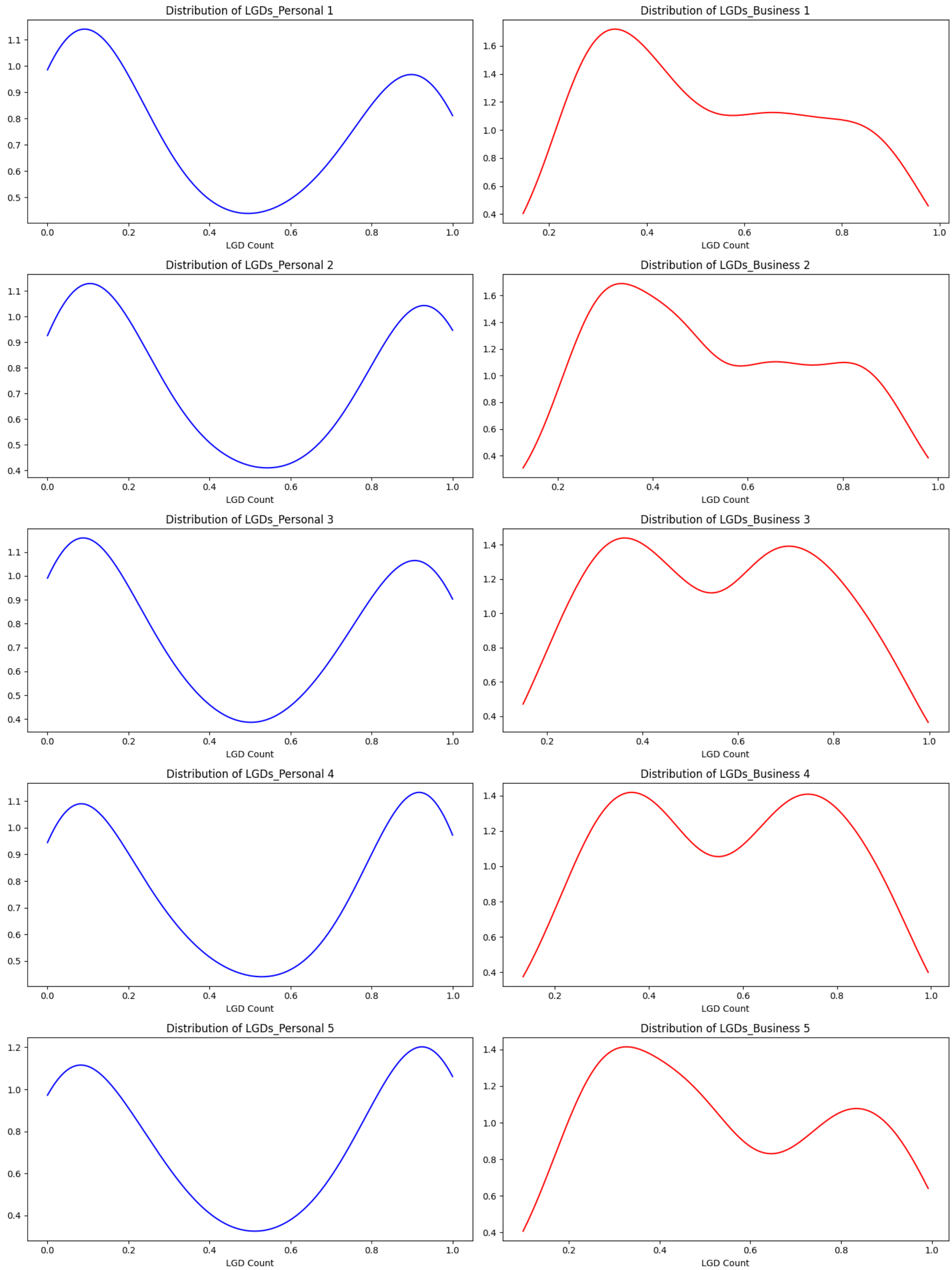
The graph in Appendix C plots the mean values of Expected Loss (EL), Unexpected Loss (UL), and VaR at 95% and 99% confidence levels against different combinations of rho_personal and rho_business. The trends in the plot show how the mean risk metrics vary with changing correlation values. EL appears to have the least variation, while VaR 99% shows significant sensitivity to changes in rho values. The CDF and mean risk metric plots highlight the impact of varying correlation values on the portfolio's risk profile. Higher correlations generally lead to increased tail risk, as evidenced by the higher VaR metrics. This underscores the importance of considering correlation effects when assessing risk.

In Appendix D each subplot displays the histogram of simulated losses along with key risk metrics such as Expected Loss (EL), Unexpected Loss (UL), and Value at Risk (VaR) at 95% and 99% confidence levels. These graphs are organized for different combinations of personal and business loan correlations (rho_personal and rho_business). The histograms indicate the frequency of loss amounts, showing a concentration of losses around the lower end of the spectrum. The vertical lines indicate the calculated EL, UL, VaR 95%, and VaR 99%, providing a visual representation of these risk measures relative to the distribution of losses. The histograms indicate that losses are predominantly concentrated towards the lower end, but the presence of long tails suggests the potential for significant extreme losses.

Appendices

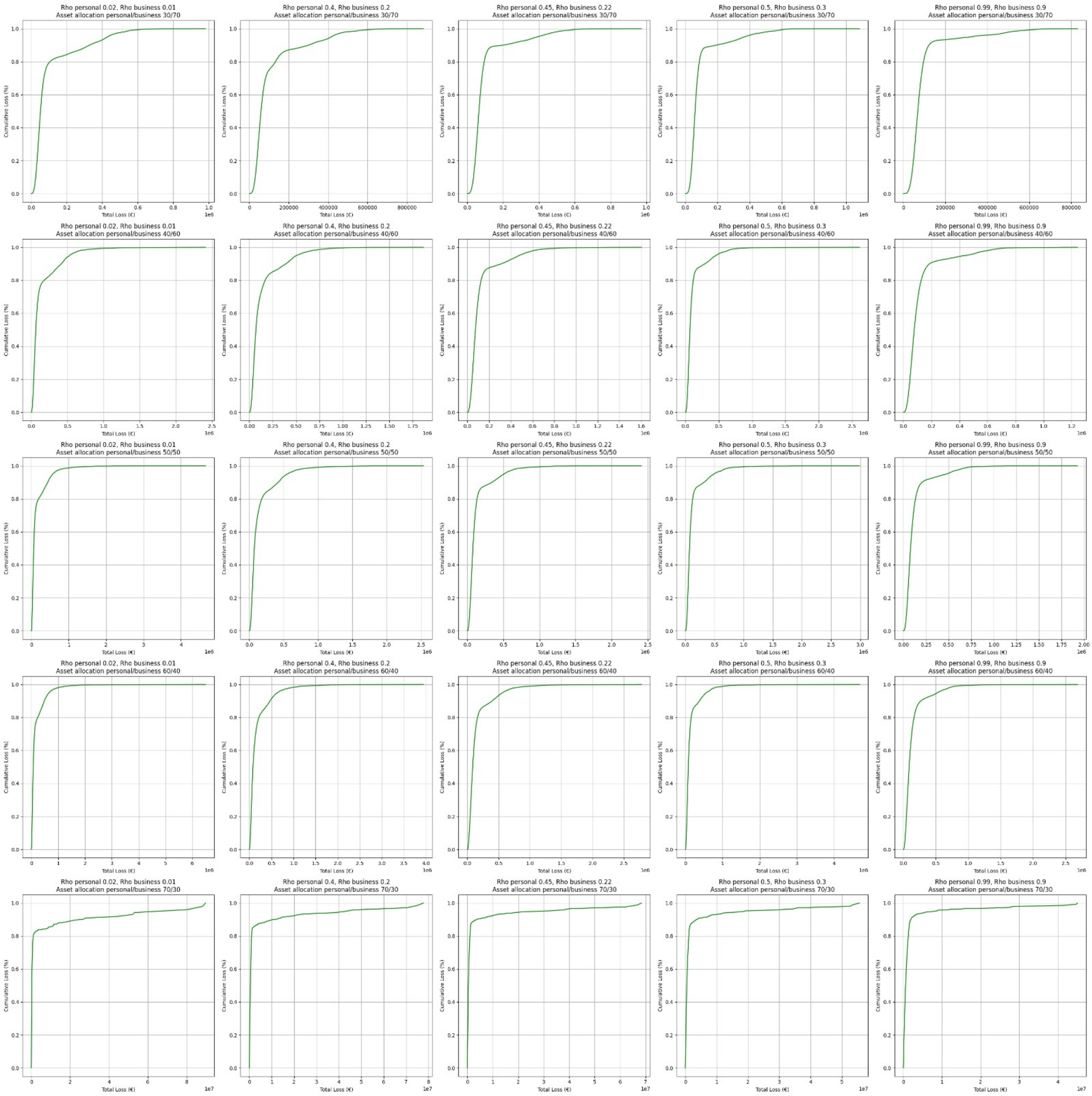
Appendix A

Distributions of LGD for personal and corporate loans in the simulation



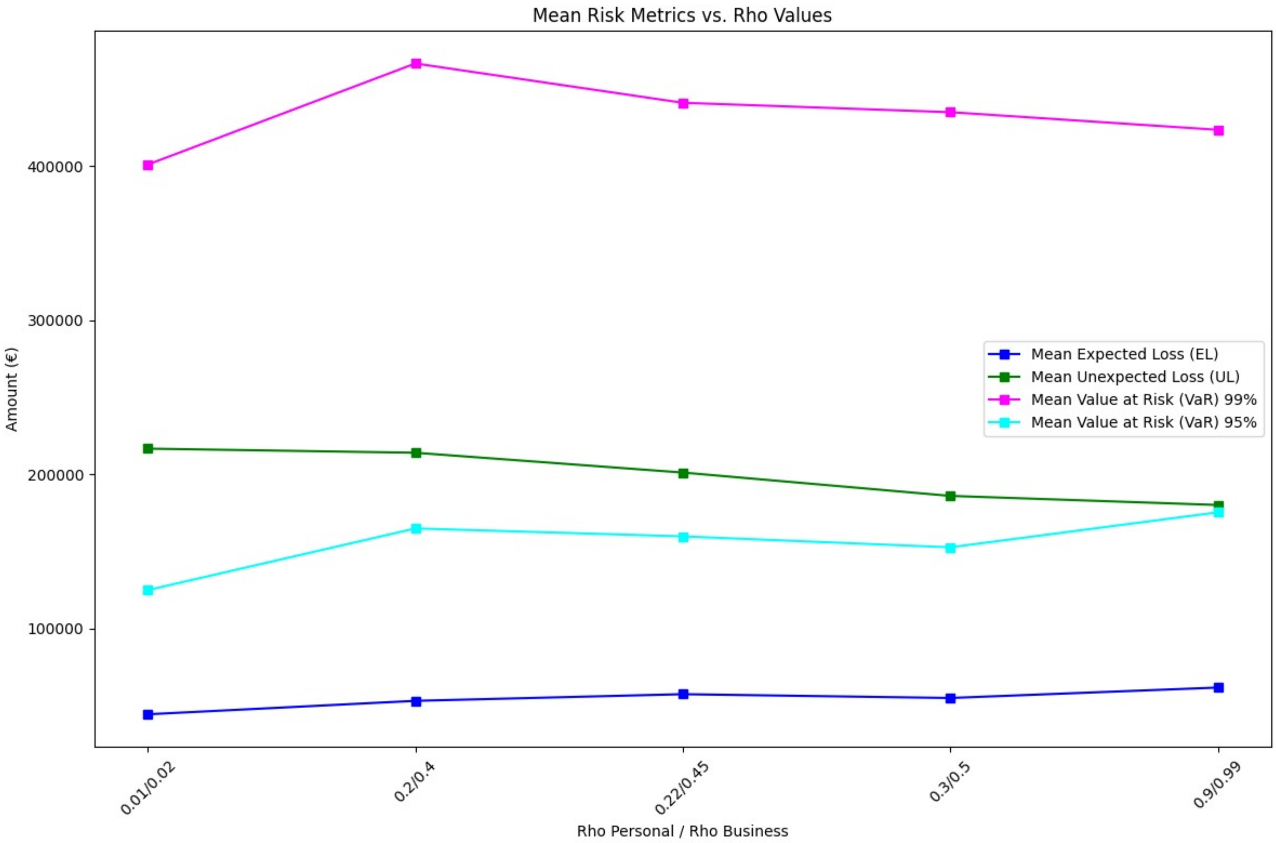
Appendix B

Cumulative loss distribution



Appendix C

Mean risk metrics vs rho values



Appendix D

Portfolio loss distribution