ACTSC 445/845: Final Project
Modelling Home Credit Default Risk

Guo, Tom (Student ID: 20927888)
Ye, Kyle (Student ID: 20999121)

Fall 2024

1. Introduction

Credit risk assessment is a critical component of risk management in the financial sector. It focuses on potential monetary losses due to borrowers failing to fulfill their debt obligations. This aspect of risk management is vital for maintaining the stability and solvency of financial institutions and ensuring the efficient allocation of resources in the broader economy. Given the complexity and stochastic nature of financial systems, advanced quantitative methods are essential for accurately modelling and managing credit risk.

The Monte Carlo simulation, known for its flexibility and robustness, has emerged as a powerful tool in financial risk assessment. By simulating a large number of potential scenarios, this method provides a probabilistic approach to understanding the dynamics of risk exposure. Our project aims to use the Monte Carlo method to simulate and analyze credit risk given the credit amount of five hundred individuals. Through this approach, we seek to capture realistic risk interactions among borrowers by incorporating dependency structures such as copulas or correlation matrices. This will enable us to provide insights into the potential distribution of portfolio losses by quantifying the two key risk metrics emphasized in this course: Value at Risk (VaR) and Expected Shortfall (ES). Furthermore, our project will examine the influence of critical parameters, namely, default probability and loss-given default. By exploring these factors, we aim to better understand their roles in shaping a financial portfolio's risk profile.

The relevance of this project to quantitative risk management lies in its application of course concepts to a practical problem in finance. We will use the R programming language to employ advanced techniques such as Monte Carlo simulation and dependency modelling, our study seeks to bridge the gap between theoretical knowledge and real-world applications, contributing valuable insights to academia and industry.

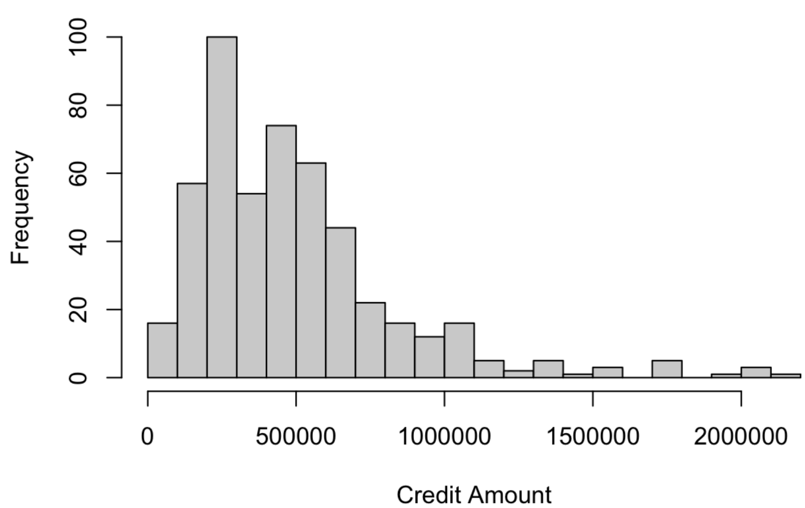
---

2. Data and Methodology

2.1 Data

The dataset we have chosen, titled *Home Credit Default Risk*, is sourced from Kaggle and comprises five hundred observations of individuals and their corresponding financial and demographic information. Each observation includes a unique identifier along with various attributes, such as gender, car and property ownership, details regarding the property, days of employment, annual income, and more. For this analysis, the *AMT_CREDIT* column, which represents the amount of credit assigned to each individual, is the primary variable of interest. The following is a summary of *AMT_CREDIT* and its histogram.

Figure 1: Histogram of AMT_CREDIT



```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 45000  254700  431505  495864  630000 2156400
```

We see that the histogram of *AMT_CREDIT* shows a right-skewed distribution of credit amounts among individuals in the dataset. The majority of credit amounts are concentrated in the lower range, with the highest frequency observed at just around 250,000. This indicates that most individuals have relatively small credit amounts. On the other hand, the distribution exhibits a long tail extending toward higher credit amounts, suggesting that large credit amounts are rare, but not nonexistent. It is important to note that the presence of bins in the upper range, particularly between 1,500,000 and 2,000,000, may indicate the existence of potential outliers or extreme values. Overall, this distribution highlights significant variability in credit amounts, with most values clustered at lower levels and a few large amounts creating a notable skew. For the sake of visualization, histograms of gender and education level were also made, refer to appendix, figures 2 and 3 respectively. We see that our data has a ratio of 2:1 in favour of women and the majority of our sample has secondary education.

2. 2 Methodology

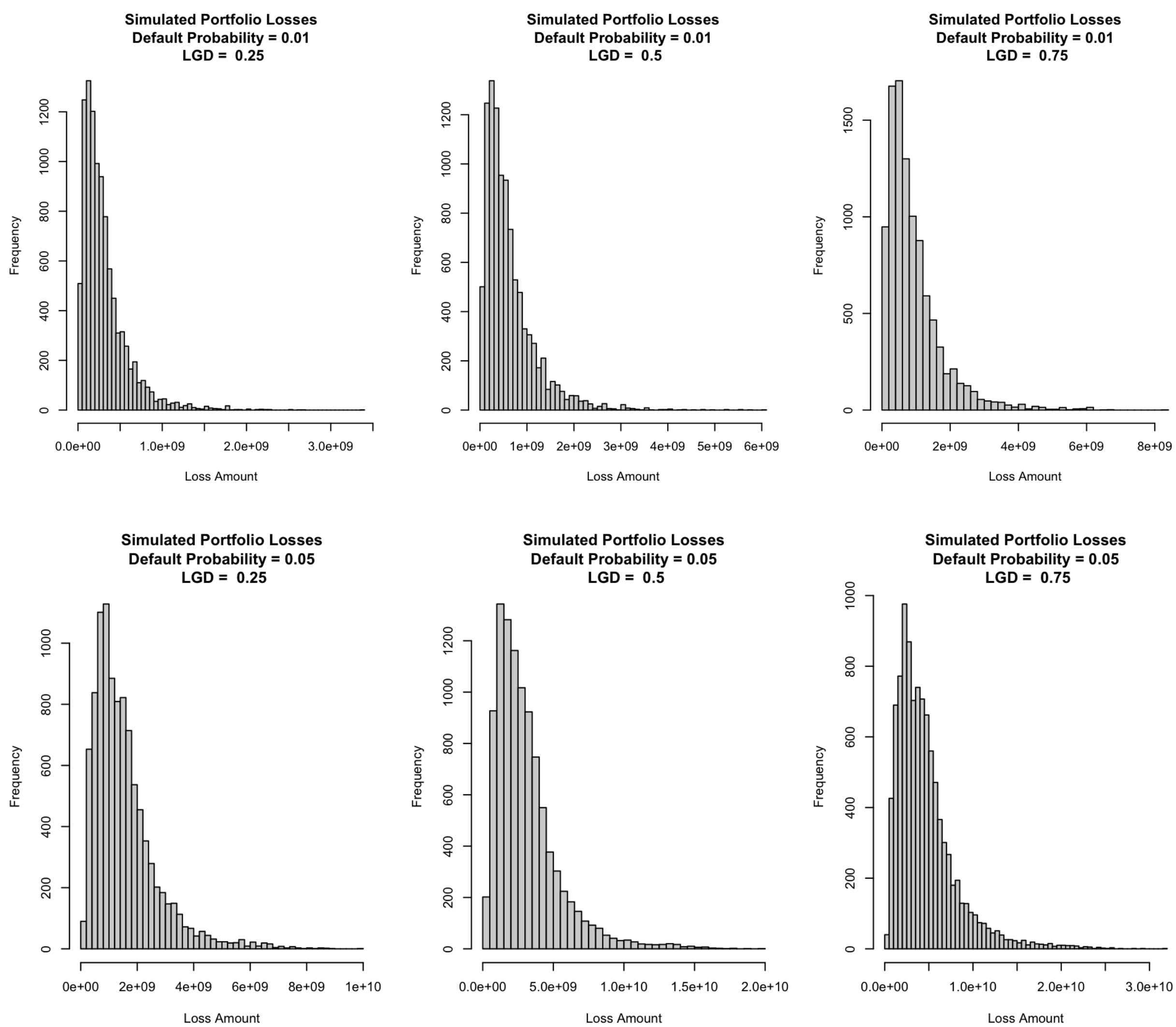
We will conduct 10,000 Monte Carlo simulations to assess the credit risk. We will evaluate different default probabilities: 0.01 (representing prime borrowers), 0.05 (a standard value), and 0.1 (representing borrowers with higher credit risk). These values reflect the likelihood that an individual borrower will default on their loan. Additionally, we will examine different loss-given default (LGD) values: 0.25 (representing low-risk assets), 0.5 (moderate-risk assets), and 0.75 (high-risk assets). LGD represents the proportion of the loan that is lost if a borrower defaults. Using the *rbinom* function, we will generate numbers from a binomial distribution to determine the number of defaults in each simulation run. Total losses will then be calculated using matrix algebra with the nominal loss method. Leveraging the copula package, we will then model the dependencies among variables. Specifically, the *normalCopula* function will be used to simulate the dependency structure, while the *fitCopula* function will fit the copula model to the dataset. Following the simulation of the credit risk distribution, we aim to compute VaR and ES to evaluate portfolio-level credit risk effectively.

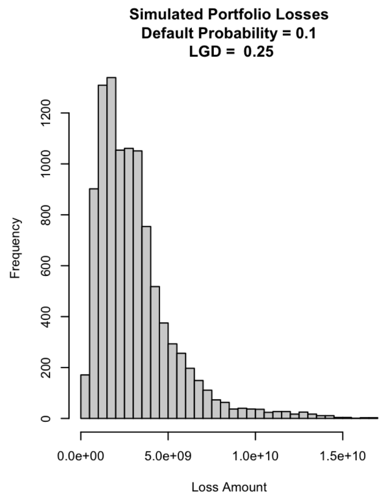
---

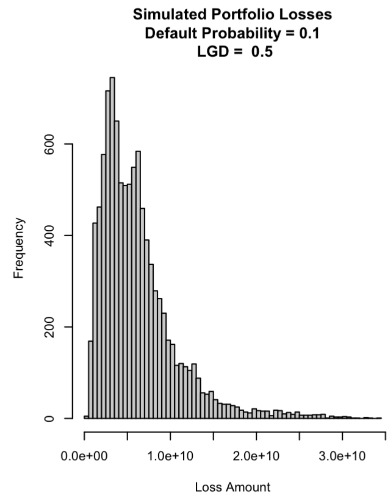
3. Analysis

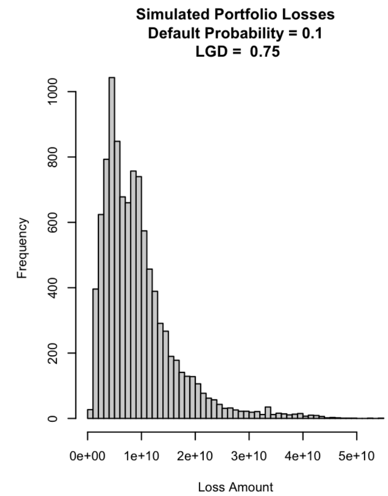
3.1 Loss Simulations

Below are the histograms of losses based. For each plot individually, refer to the appendix, figures 4 through 12.









As the default probability increases, the distribution of losses becomes more heavily skewed to the right, indicating a greater frequency of larger losses. This observation aligns with the expectation that higher default probabilities lead to more defaults, and consequently, higher portfolio losses. As for the LGD, an increase in its value results in more severe losses, which is consistent with its definition as each loss is simply worth more. Statistically, higher LGD values amplify the magnitude of losses for each default event, leading to a more pronounced right tail in the loss distribution. These trends are in line with the theoretical understanding of credit risk: an increase in the likelihood of default or the severity of loss per default directly correlates with a higher potential for greater portfolio losses.

The right-skewed distribution observed in the histogram is expected in the context of home credit default risk. In most loan portfolios, especially those focusing on an individual scale like ours, we would expect defaults to be relatively rare. This is the case as the distribution of losses is concentrated around low values, with a few large losses on the right-hand side of the distribution. These large losses represent the relatively infrequent but significant defaults that can have a substantial financial impact, even if they occur infrequently. This heavy right tail is a characteristic feature of credit risk, where the majority of loans perform as expected, but a few defaults lead to disproportionally large losses.

This distribution pattern is important for risk management, as it not only exhibits the heavy tail whose importance is constantly emphasized in class, but it also highlights the potential for rare but severe losses, which are essential to consider when calculating risk metrics.

3.2 Dependency Modelling

As not all columns of our data were necessary, we decided to only focus on credit amount, gender, if a car was owned, if realty is owned, annuity amount, type of income, education type, and days employed, as these were expected to be significant. These variables are then converted into factors, as they are categorical by nature. Using the *cor()* function to compute the correlation matrix for these numeric variables, results in the following matrix.

	AMT_CREDIT	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	AMT_ANNUITY	NAME_INCOME_TYPE	NAME_EDUCATION_TYPE	DAYS_EMPLOYED
AMT_CREDIT	" 1.00"	" 0.01"	" 0.08"	" 0.08"	" 0.78"	"-0.07"	"-0.13"	"-0.08"
CODE_GENDER	" 0.01"	" 1.00"	" 0.41"	"-0.03"	" 0.09"	" 0.07"	" 0.01"	"-0.13"
FLAG_OWN_CAR	" 0.08"	" 0.41"	" 1.00"	" 0.02"	" 0.12"	" 0.06"	"-0.08"	"-0.21"
FLAG_OWN_REALTY	" 0.08"	"-0.03"	" 0.02"	" 1.00"	" 0.09"	"-0.10"	"-0.04"	" 0.05"
AMT_ANNUITY	" 0.78"	" 0.09"	" 0.12"	" 0.09"	" 1.00"	"-0.05"	"-0.10"	"-0.15"
NAME_INCOME_TYPE	"-0.07"	" 0.07"	" 0.06"	"-0.10"	"-0.05"	" 1.00"	"-0.03"	"-0.33"
NAME_EDUCATION_TYPE	"-0.13"	" 0.01"	"-0.08"	"-0.04"	"-0.10"	"-0.03"	" 1.00"	" 0.17"
DAYS_EMPLOYED	"-0.08"	"-0.13"	"-0.21"	" 0.05"	"-0.15"	"-0.33"	" 0.17"	" 1.00"

Focusing on the credit amount, the correlation analysis of *AMT_CREDIT* with other variables reveals several key insights. The strongest correlation observed is between *AMT_CREDIT* and *AMT_ANNUITY*, with a value of roughly 0.78, indicating a relatively strong positive relationship. This suggests that individuals with higher credit amounts tend to have higher annuity amounts, which is logically sound. On the other hand, the correlation between *AMT_CREDIT* and *CODE_GENDER* is minimal, at just around 0.01, indicating virtually no relationship between credit amount and gender. Similarly, the correlation with *FLAG_OWN_CAR* and *FLAG_OWN_REALTY* is also weak, both at around 0.08, showing that car and realty ownership have very little impact on the credit amount. The correlation between *AMT_CREDIT* and *NAME_INCOME_TYPE, NAME_EDUCATION_TYPE,* and *DAYS_EMPLOYED* is very slightly negative, suggesting a very weak inverse relationship between credit amount and income type, education type, and days employed. In summary, the most significant relationship involving *AMT_CREDIT* is with *AMT_ANNUITY*, while the other correlations, such as those with gender, car ownership, and education type, are weak, suggesting that credit amount is not strongly influenced by these factors. Apart from just the credit amount, we can see that *DAYS_EMPLOYED* has a small, possibly significant correlation with *NAME_INCOME_TYPE*, *AMT_ANNUITY*, and *NAME_EDUCATION_TYPE* which aligns with socio-economic norms.

We then employed the normal copula to model the relationships separately from their marginal distributions. To ensure that the data is uniformly distributed, a necessity for copula modelling, variables were transformed into values between 0 and 1 that represent the rank of each data point within the variable. Finally, the model's fit is evaluated to assess how well the copula captures the dependency between the variables. Below is the output of the fit.

```
Call: fitCopula(cop_model, data = u[, 1:8])
Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
Normal copula, dim. d = 8
      Estimate Std. Error
rho.1   0.1924      0.021
The maximized loglikelihood is 15.63
Optimization converged
Number of loglikelihood evaluations:
function gradient
       9        9
```

The results from fitting the normal copula model with eight dimensions indicate a positive correlation between the variables with an estimate of 0.1924. This suggests a moderate level of dependency between the variables in the dataset. The small standard error of 0.021

indicates a precise estimate of the correlation parameter. The maximized log-likelihood value of 15.63 shows that the model has provided a good fit to the data and as the process successfully converged, the model was properly optimized. This result implies that the normal copula is a reasonable model to capture the dependence structure among the selected variables. For further analysis, we saw from the correlation matrix the annuity amount had the greatest correlation with the credit amount. The following is the output of the model with just the credit and annuity amount.

```
Call: fitCopula(cop_model, data = u[, c(1, 5)])
Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
Normal copula, dim. d = 2
      Estimate Std. Error
rho.1   0.8471      0.013
The maximized loglikelihood is 311.5
Optimization converged
Number of loglikelihood evaluations:
function gradient
      19       19
```

The estimated correlation parameter is 0.8471, with a standard error of 0.013, indicating a strong positive relationship between these two variables.

With all being said and done, given the moderate correlation, other copula models, such as Student's t or Gumbel should also be considered. The following are the results of the Student t and Gumbel copula fit for all variables of interest, succeeded by annuity amount and credit risk only.

```
Call: fitCopula(cop_model, data = u[, 1:8])
Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
t-copula, dim. d = 8
      Estimate Std. Error
rho.1   0.0396      0.014
df      4.2823         NA
The maximized loglikelihood is 106.3
Optimization converged
Number of loglikelihood evaluations:
function gradient
      34       34
```

```
Call: fitCopula(cop_model, data = u[, c(1, 5)])
Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
t-copula, dim. d = 2
       Estimate Std. Error
rho.1    0.8471      0.013
df     933.3562         NA
The maximized loglikelihood is 311.5
Optimization converged
Number of loglikelihood evaluations:
function gradient
      36       36
```

```
Call: fitCopula(cop_model, data = u[, 1:8])
Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
Gumbel copula, dim. d = 8
      Estimate Std. Error
alpha    1.121      0.014
The maximized loglikelihood is 28.7
Optimization converged
Number of loglikelihood evaluations:
function gradient
       6        6
```

```
Call: fitCopula(cop_model, data = u[, c(1, 5)])
Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
Gumbel copula, dim. d = 2
      Estimate Std. Error
alpha    2.449      0.108
The maximized loglikelihood is 265
Optimization converged
Number of loglikelihood evaluations:
function gradient
       7        7
```

For Student t copula fit to the model with all variables from the subset, we see that the rho estimate is much larger. This suggests that the distribution of all the variables in the subset likely follows a distribution with a heavier tail than Gaussian. Fortunately, the results of just the credit and annual amount for the student t copula align what that of the normal copula. As for the Gumbel copula, for the full model, the log-likelihood is low, suggesting that it does not fit the data as well as the other copulas. For the model fit to only the annual and credit amounts, the results are similar to that of both the student-t and normal copulas. This indicates that our data might have significant upper-tail dependence, for which the Gumbel copula would be appropriate.

3.3 Risks Metrics

The analysis of the portfolio's risk metrics reveals that the VaR at a 95% confidence level is approximately $22.14 billion, indicating that under normal market conditions, the portfolio is expected to incur losses of at most this amount with 95% confidence. On the other hand, the ES at the 95% confidence level is about $29.93 billion. This highlights the portfolio's vulnerability to severe losses when market conditions are adverse, providing a more conservative risk estimate compared to VaR alone.

---

4. Conclusion

The right-skewed loss distribution mirrors typical credit portfolios where the majority of loans perform well, but a few defaults can lead to very severe consequences. Our results suggest that the portfolio is very vulnerable to extreme losses, which are critical to account for, despite them being rare. The weak dependencies among the variables indicate that the portfolio's risks are relatively diversified and the occurrence of one risk is not strongly influenced by other risk factors such as education level.

In practice for this portfolio specifically and others similar, these results suggest that risk managers should be prepared for losses of all kinds, even rare ones. The weak dependencies imply that diversification strategies could reduce the overall risk of large-scale defaults, but the portfolio is still exposed to substantial tail risk.

5. References

Crenshaw, M. (2022, November 4). Home credit default risk. Kaggle. Retrieved November 17, 2024 from https://www.kaggle.com/datasets/megancrenshaw/home-credit-default-risk

Kenton, W. (2024, June 27). Monte Carlo Simulation: What it is, how it works, history, 4 key steps. Investopedia. Retrieved November 28, 2024 from https://www.investopedia.com/terms/m/montecarlosimulation.asp

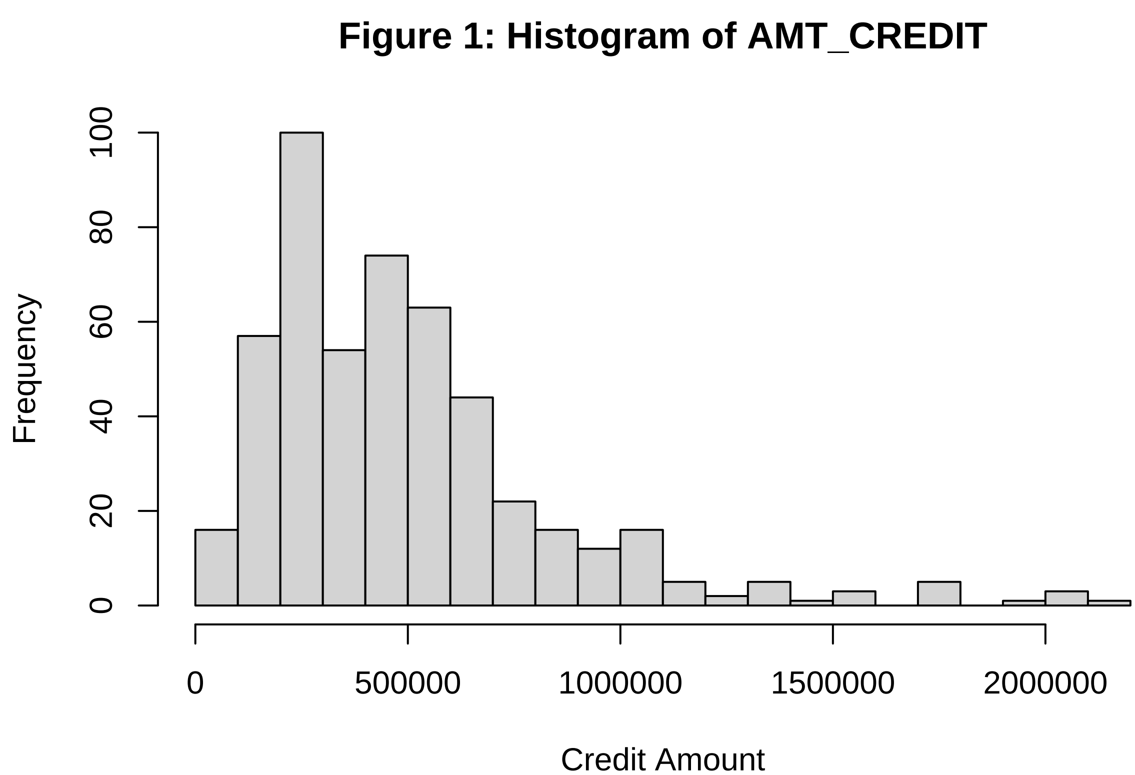
Monte Carlo Simulation of Credit Portfolios. (2022, November 21). Open Risk Manual. Retrieved December 4, 2024 from https://www.openriskmanual.org/wiki/index.php?title=Monte_Carlo_Simulation_of_Credit_Portfolios&oldid=35343

6. Appendix

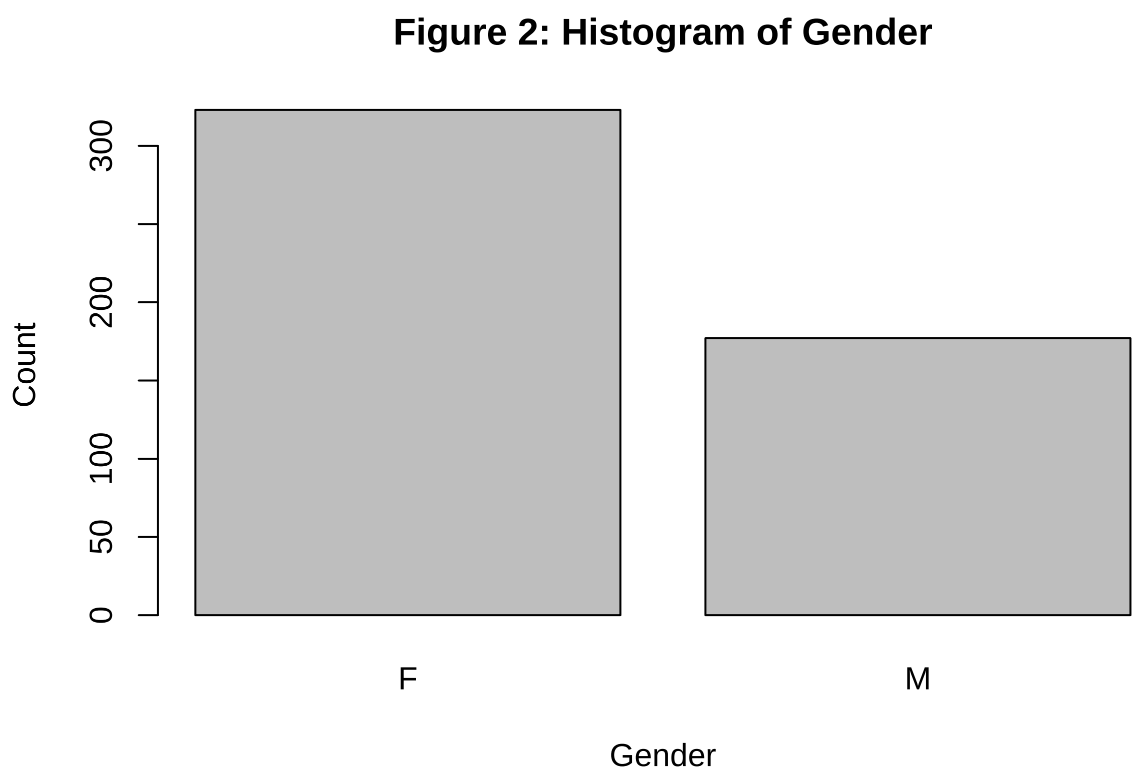
```r
## Load Required Libraries
library(copula)
library(dplyr)
```

```r
## Load Dataset
data <- read.csv("HomeCreditDefaultRisk.csv")
```

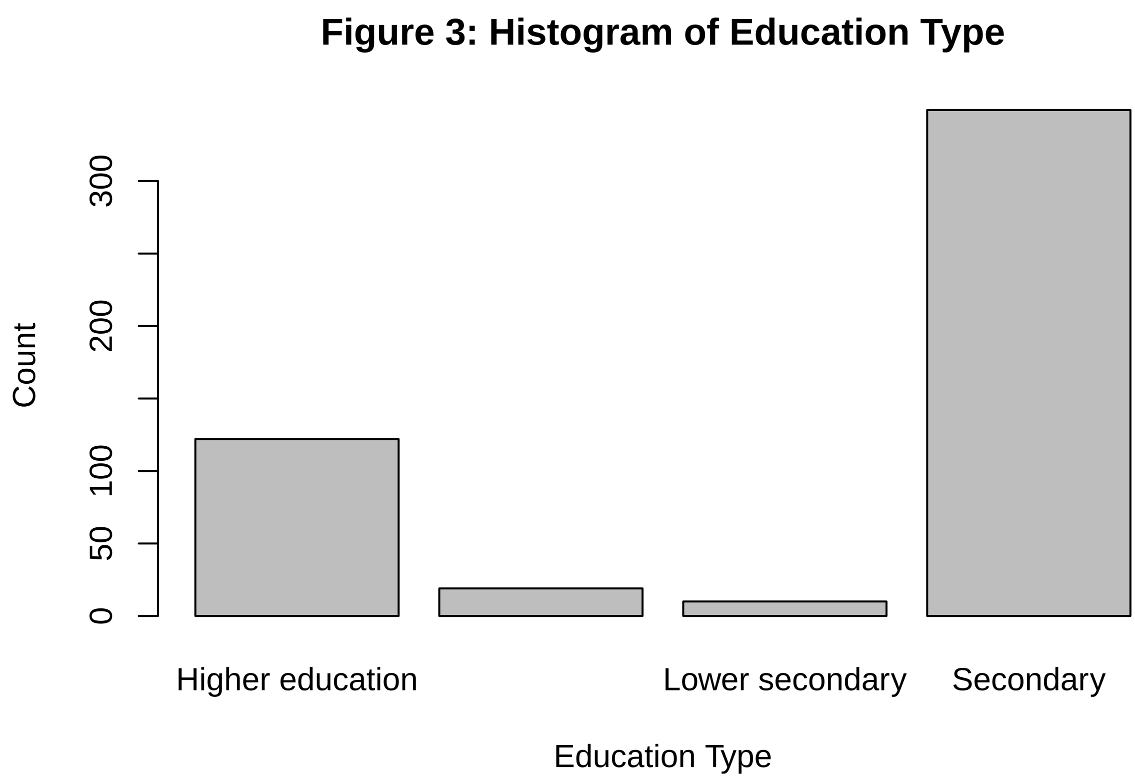
```r
## Plotting Histograms of Select Data
hist(
  data$AMT_CREDIT,
  breaks = 20,
  main = "Figure 1: Histogram of AMT_CREDIT",
  xlab = "Credit Amount",
  ylab = "Frequency"
)
```




```r
barplot(
  table(data$CODE_GENDER),
  main = "Figure 2: Histogram of Gender",
  xlab = "Gender",
  ylab = "Count"
)
```




```r
data$NAME_EDUCATION_TYPE[data$NAME_EDUCATION_TYPE == "Secondary / secondary special"] <- "Secondary"

barplot(
  table(data$NAME_EDUCATION_TYPE),
  main = "Figure 3: Histogram of Education Type",
  xlab = "Education Type",
  ylab = "Count",
)
```

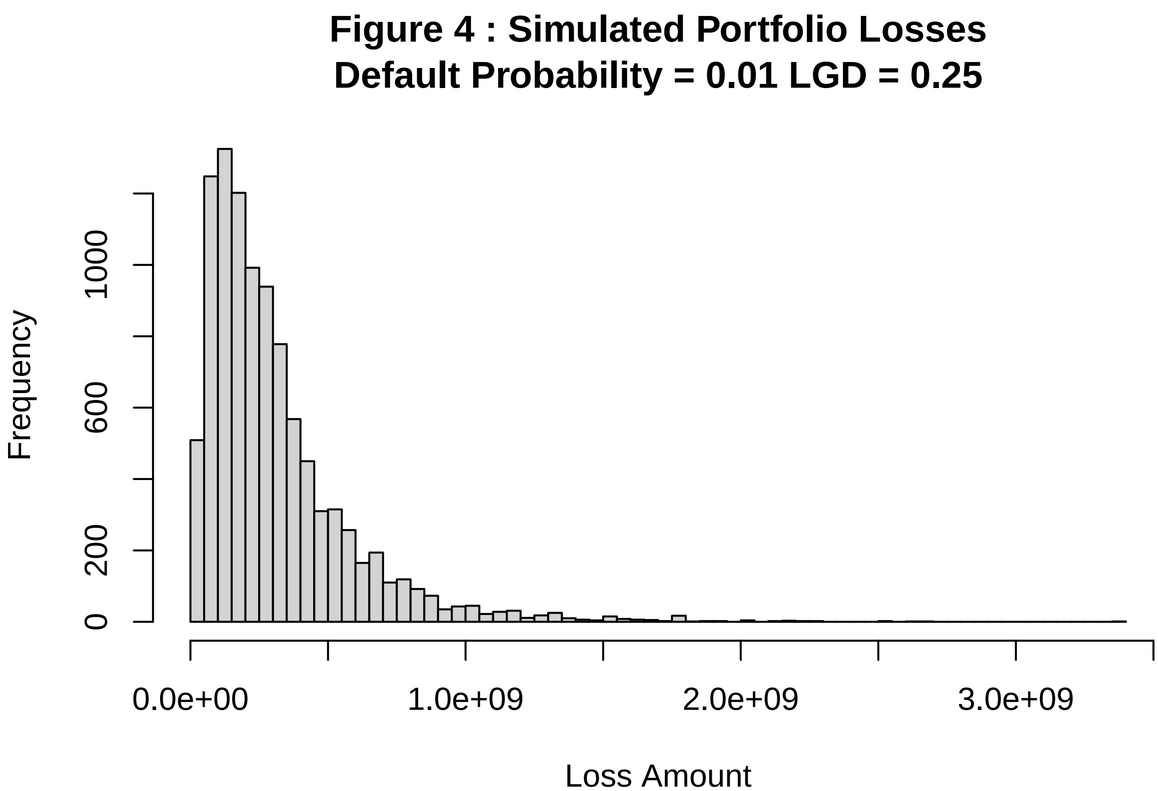



```r
## 5 Number Summary of Credit Amount
credit_amounts <- data$AMT_CREDIT
summary(credit_amounts)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   45000  254700  431505  495864  630000 2156400
```

```r
## Monte Carlo Simulation
set.seed(445)

n_simulations <- 10000
default_probabilities <- c(0.01, 0.05, 0.1)
loss_given_defaults <- c(0.25, 0.5, 0.75)
results <- list()

figure_counter <- 4

for (default_probability in default_probabilities) {
  for (lgd in loss_given_defaults) {
    defaults <- rbinom(n_simulations, size = length(credit_amounts), prob = default_probability)

    losses <- rowSums(matrix(defaults, nrow = n_simulations,
                             ncol = length(credit_amounts)) * credit_amounts * lgd)

    results <- append(results, list(losses))
```

```
    hist(losses, breaks = 50,
         main = paste("Figure", figure_counter,
                      ": Simulated Portfolio Losses\nDefault Probability =", default_probability,
                      "LGD =", lgd),
         xlab = "Loss Amount")

    figure_counter <- figure_counter + 1
  }
}
```



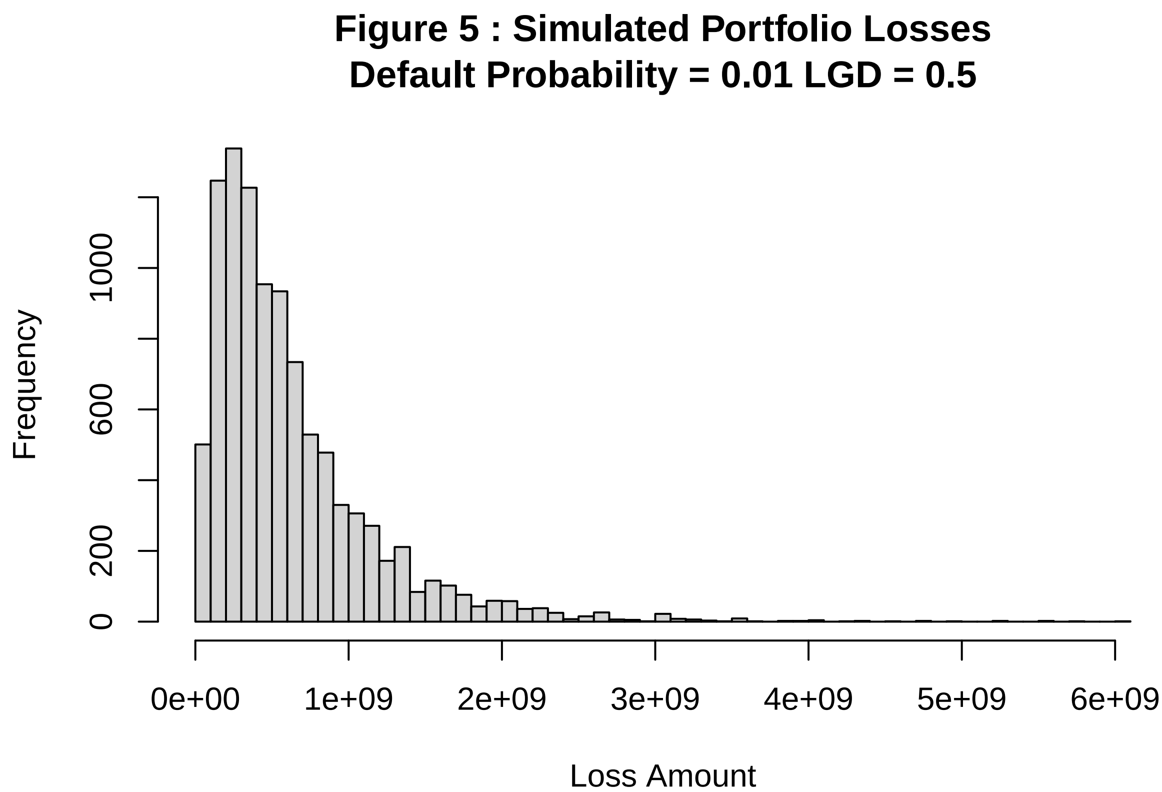


Figure 5 : Simulated Portfolio Losses
Default Probability = 0.01 LGD = 0.5
Frequency
0
200
600
1000
0e+00
1e+09
2e+09
3e+09
4e+09
5e+09
6e+09
Loss Amount

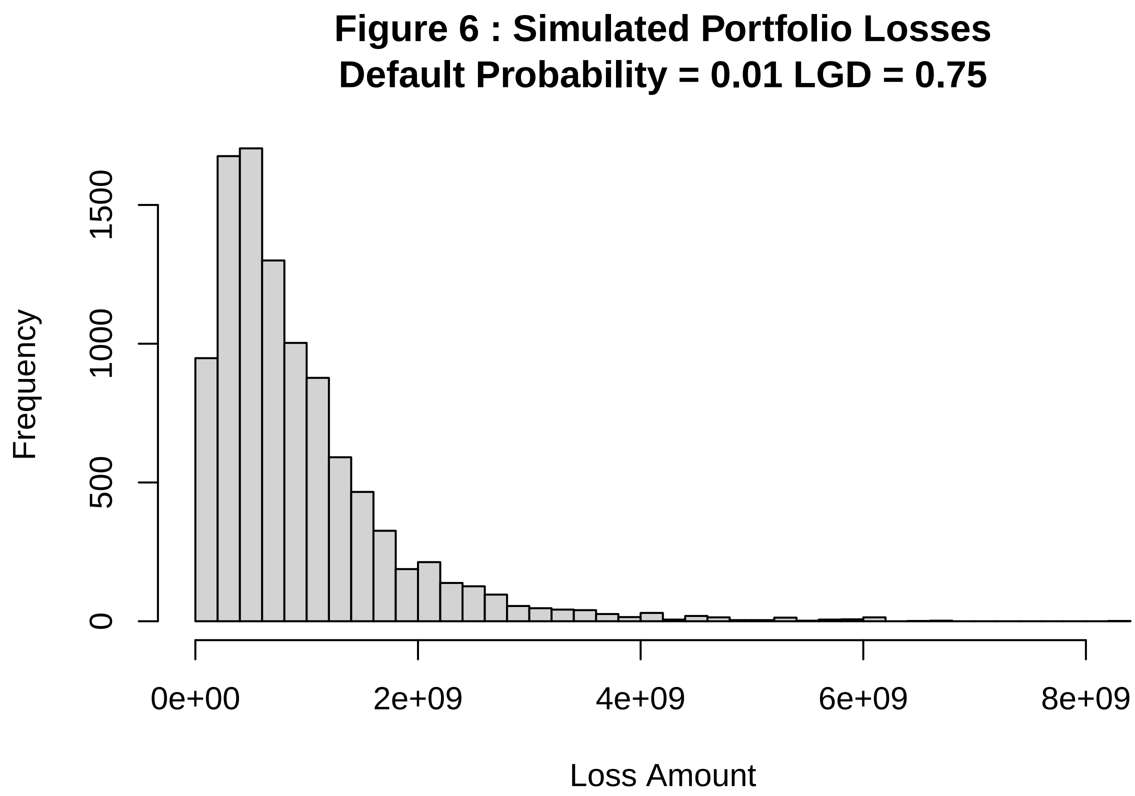
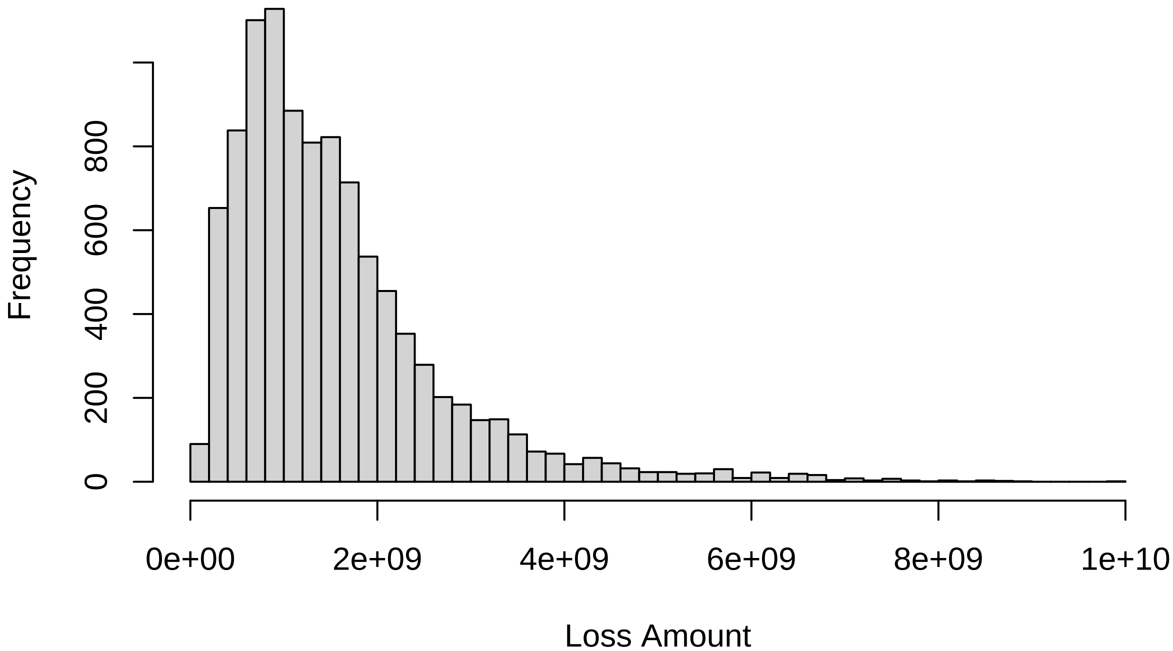


Figure 6 : Simulated Portfolio Losses
Default Probability = 0.01 LGD = 0.75
Frequency
0
500
1000
1500
0e+00
2e+09
4e+09
6e+09
8e+09
Loss Amount

Figure 7 : Simulated Portfolio Losses
Default Probability = 0.05 LGD = 0.25



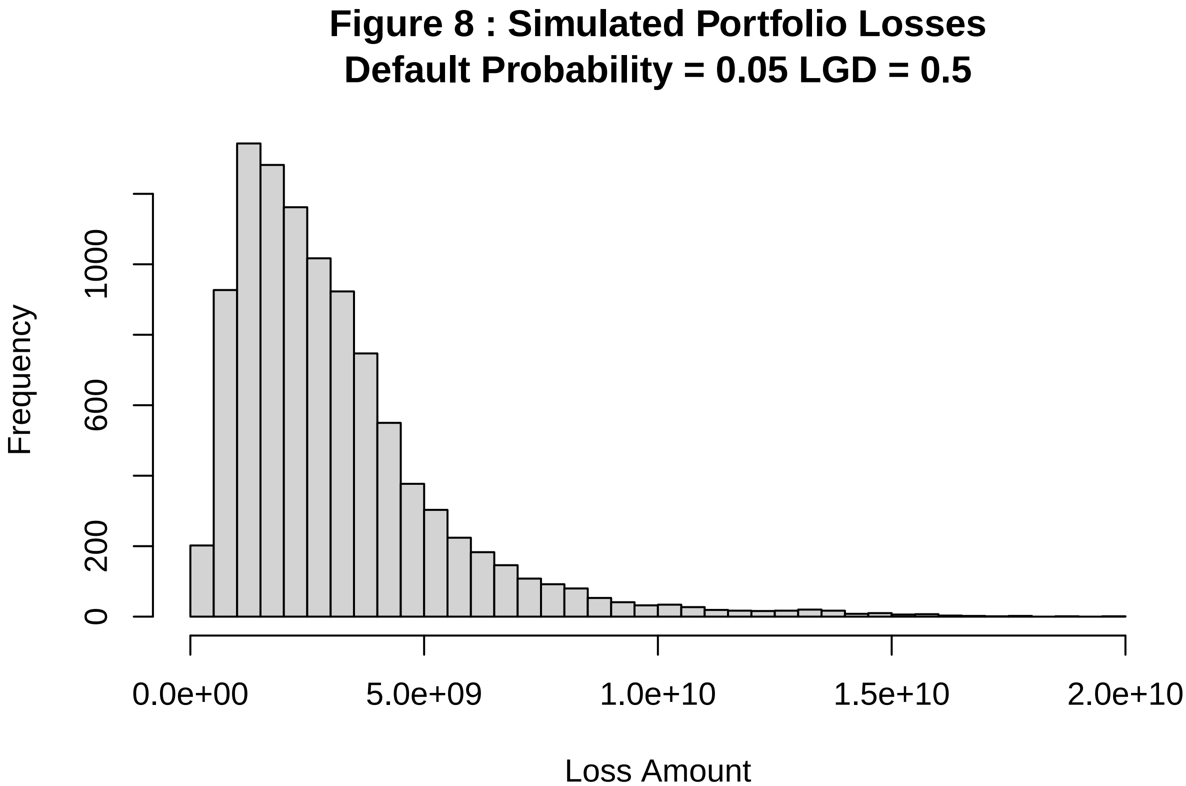


Figure 8 : Simulated Portfolio Losses
Default Probability = 0.05 LGD = 0.5
Frequency
0
200
600
1000
0.0e+00
5.0e+09
1.0e+10
1.5e+10
2.0e+10
Loss Amount

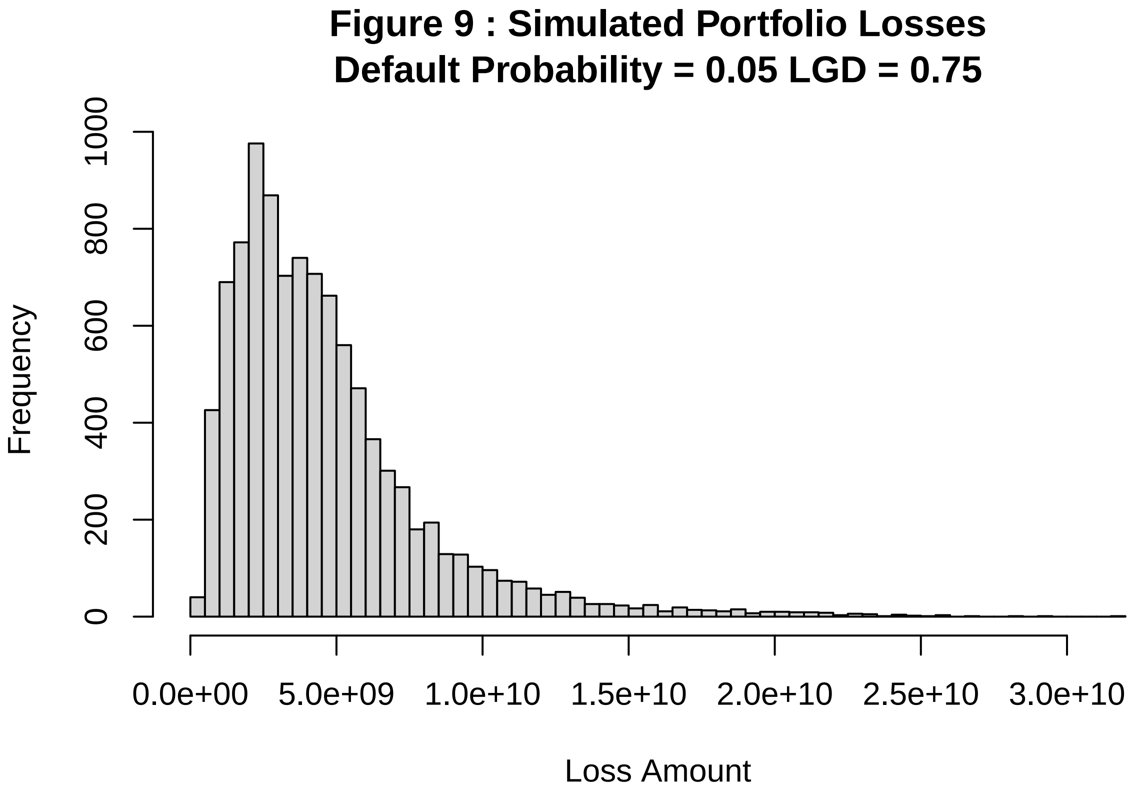


Figure 9 : Simulated Portfolio Losses
Default Probability = 0.05 LGD = 0.75
Frequency
0
200
400
600
800
1000
0.0e+00
5.0e+09
1.0e+10
1.5e+10
2.0e+10
2.5e+10
3.0e+10
Loss Amount

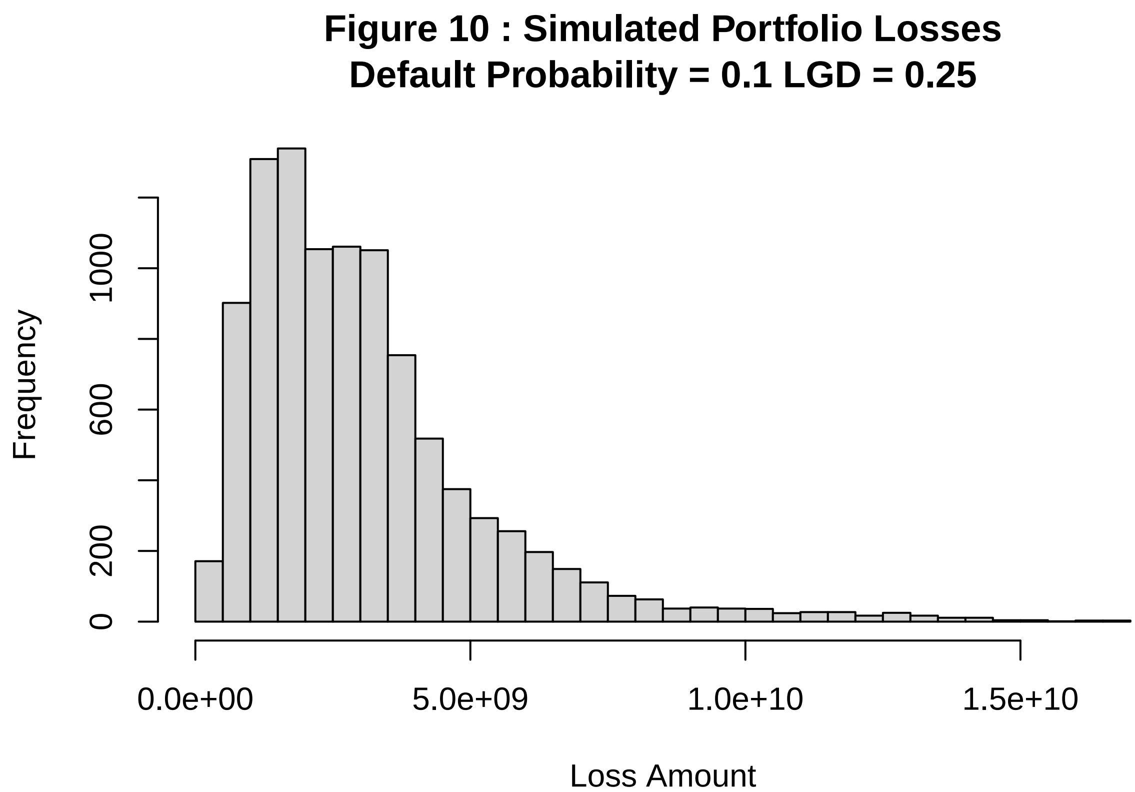


Figure 10 : Simulated Portfolio Losses
Default Probability = 0.1 LGD = 0.25
Frequency
0
200
600
1000
0.0e+00
5.0e+09
1.0e+10
1.5e+10
Loss Amount

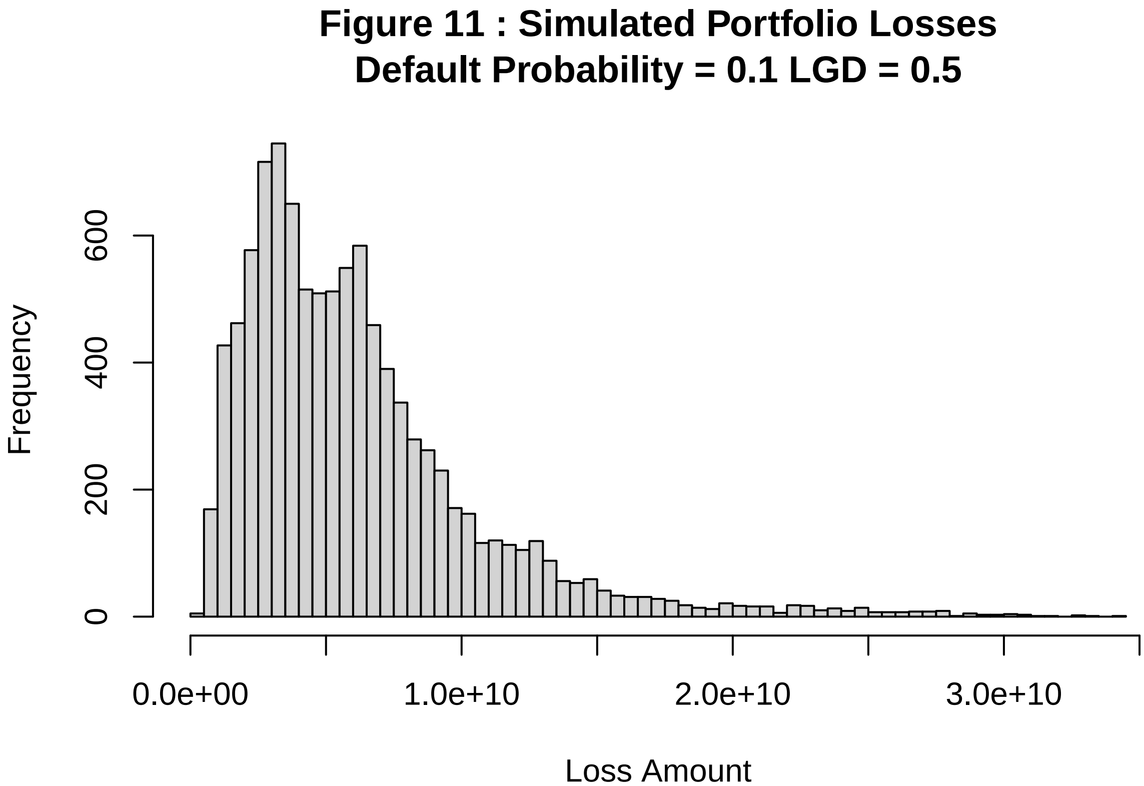
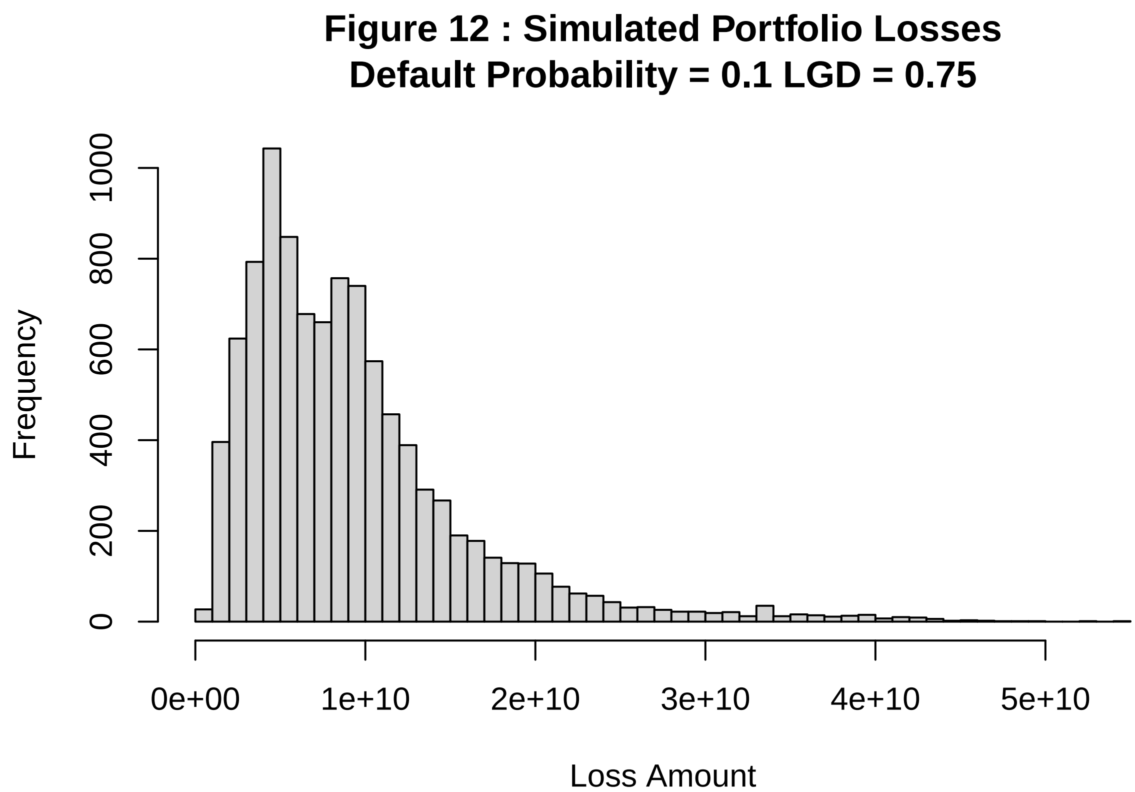


Figure 11 : Simulated Portfolio Losses
Default Probability = 0.1 LGD = 0.5
Frequency
0
200
400
600
0.0e+00
1.0e+10
2.0e+10
3.0e+10
Loss Amount




```r
## Initializing a Subset of Data for Variables of Interest
data_subset <- data %>%
  select(AMT_CREDIT, CODE_GENDER, FLAG_OWN_CAR, FLAG_OWN_REALTY, AMT_ANNUITY,
         NAME_INCOME_TYPE, NAME_EDUCATION_TYPE, DAYS_EMPLOYED)

## Making Data Usable for cor() Function
data_subset$CODE_GENDER <- as.factor(data_subset$CODE_GENDER)
data_subset$FLAG_OWN_CAR <- as.factor(data_subset$FLAG_OWN_CAR)
data_subset$FLAG_OWN_REALTY <- as.factor(data_subset$FLAG_OWN_REALTY)
data_subset$NAME_INCOME_TYPE <- as.factor(data_subset$NAME_INCOME_TYPE)
data_subset$NAME_EDUCATION_TYPE <- as.factor(data_subset$NAME_EDUCATION_TYPE)
data_subset_numeric <- data_subset %>%
  mutate(across(where(is.factor), as.integer))

## Calculating Correlation Matrix
cor_matrix <- cor(data_subset_numeric)
print(cor_matrix)
```

```
##                       AMT_CREDIT  CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY
## AMT_CREDIT            1.00000000  0.012893859   0.07718558      0.07743989
## CODE_GENDER           0.01289386  1.000000000   0.40803533     -0.03408573
## FLAG_OWN_CAR          0.07718558  0.408035334   1.00000000      0.01814809
## FLAG_OWN_REALTY       0.07743989 -0.034085728   0.01814809      1.00000000
## AMT_ANNUITY           0.78003537  0.085132715   0.11631480      0.08982221
## NAME_INCOME_TYPE     -0.06813547  0.065786254   0.06076792     -0.09981550
## NAME_EDUCATION_TYPE -0.13081245  0.005017126  -0.07653824     -0.03716783
```

```
## DAYS_EMPLOYED       -0.07856059 -0.126699349  -0.20919549      0.04528111
##                     AMT_ANNUITY NAME_INCOME_TYPE NAME_EDUCATION_TYPE
## AMT_CREDIT           0.78003537      -0.06813547        -0.130812449
## CODE_GENDER          0.08513271       0.06578625         0.005017126
## FLAG_OWN_CAR         0.11631480       0.06076792        -0.076538238
## FLAG_OWN_REALTY      0.08982221      -0.09981550        -0.037167829
## AMT_ANNUITY          1.00000000      -0.05415457        -0.095844559
## NAME_INCOME_TYPE    -0.05415457       1.00000000        -0.033350987
## NAME_EDUCATION_TYPE -0.09584456      -0.03335099         1.000000000
## DAYS_EMPLOYED       -0.14552812      -0.32503072         0.172149304
##                     DAYS_EMPLOYED
## AMT_CREDIT            -0.07856059
## CODE_GENDER           -0.12669935
## FLAG_OWN_CAR          -0.20919549
## FLAG_OWN_REALTY        0.04528111
## AMT_ANNUITY           -0.14552812
## NAME_INCOME_TYPE      -0.32503072
## NAME_EDUCATION_TYPE    0.17214930
## DAYS_EMPLOYED          1.00000000
```

```r
## Dependency Modeling
## Normal Copula Model for All Variables in Data_Subset
cop_model <- normalCopula(dim = 8)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, 1:8])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, 1:8])
## Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
## Normal copula, dim. d = 8
##       Estimate Std. Error
## rho.1   0.1924      0.021
## The maximized loglikelihood is 15.63
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##        9        9
```

```r
## Normal Copula Model for Just Annuity and Credit
cop_model <- normalCopula(dim = 2)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, c(1, 5)])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, c(1, 5)])
## Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
## Normal copula, dim. d = 2
##       Estimate Std. Error
## rho.1   0.8471      0.013
## The maximized loglikelihood is 311.5
## Optimization converged
```

```
## Number of loglikelihood evaluations:
## function gradient
##       19       19
```

```r
## Student's t Copula Model for All Variables in Data_Subset
cop_model <- tCopula(dim = 8)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, 1:8])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, 1:8])
## Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
## t-copula, dim. d = 8
##       Estimate Std. Error
## rho.1   0.0396      0.014
## df      4.2823         NA
## The maximized loglikelihood is 106.3
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##       34       34
```

```r
## Student's t Copula Model for Just Annuity and Credit
cop_model <- tCopula(dim = 2)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, c(1, 5)])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, c(1, 5)])
## Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
## t-copula, dim. d = 2
##       Estimate Std. Error
## rho.1   0.8471      0.013
## df    933.3562         NA
## The maximized loglikelihood is 311.5
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##       36       36
```

```r
## Gumbel Copula Model for All Variables in Data_Subset
cop_model <- gumbelCopula(dim = 8)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, 1:8])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, 1:8])
## Fit based on "maximum pseudo-likelihood" and 500 8-dimensional observations.
## Gumbel copula, dim. d = 8
```

```
##       Estimate Std. Error
## alpha    1.121      0.014
## The maximized loglikelihood is 28.7
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##        6        6
```

```r
## Gumbel Copula Model for Just Annuity and Credit
cop_model <- gumbelCopula(dim = 2)
u <- pobs(as.matrix(data_subset_numeric))

fit <- fitCopula(cop_model, u[, c(1, 5)])
summary(fit)
```

```
## Call: fitCopula(cop_model, data = u[, c(1, 5)])
## Fit based on "maximum pseudo-likelihood" and 500 2-dimensional observations.
## Gumbel copula, dim. d = 2
##       Estimate Std. Error
## alpha    2.449      0.108
## The maximized loglikelihood is 265
## Optimization converged
## Number of loglikelihood evaluations:
## function gradient
##        7        7
```

```r
## Calculate VaR
var_95 <- quantile(losses, 0.95)

## Calculate ES
es_95 <- mean(losses[losses > var_95])

cat("Value at Risk (95%):", var_95, "\n")
```

```
## Value at Risk (95%): 22135687847
```

```r
cat("Expected Shortfall (95%):", es_95, "\n")
```

```
## Expected Shortfall (95%): 29929824425
```