Master in Actuarial and Financial Sciences

MASTER THESIS

Analysis of PD-LGD correlation effects on the minimum capital requirement

Author: Víctor Pérez Torre

Tutor: Montserrat Guillen Estany

Academic year: 2017-2018

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Màster de Ciències Actuarials i Financeres

Faculty of Economics and Business

Universitat de Barcelona

Master thesis

Master in Actuarial and Financial Sciences

Analysis of PD-LGD correlation effects on the minimum capital requirement

Author: Víctor Pérez Torre

Tutor: Montserrat Guillen Estany

Abstract

The minimum capital requirement in the Basel II IRB approach implicitly assumes that the risk factors involved in PD and LGD are independent. This thesis analyses and quantifies the effects on the minimum capital requirement under the presence of correlation between the factors affecting PD and LGD. The same portfolio is simulated with different PD-LGD correlation and the minimum capital requirement in the IRB approach is computed with two different sets of risk factors: the real an correlated PD and LGD and the PD and LGD that will be estimated with the usual modeling approach. The main conclusions is that as the dependency between PD and LGD grows, the minimum capital is more underestimated.

Keyworkds: Probability of Default, Loss Given Default, correlation, IRB, Basel II.

Contents

Chapter 1

Introduction

After the disastrous effects that the financial crises started in 2008 caused, many people began to question how a crisis with such serious consequences could have happened despite the amount of international financial regulation that already existed at that time. The financial regulation tries to avoid and reduce the consequences of the crisis that many times are produced by the cyclical dynamics of the economy. This is done defining and elaborating rules and recommendations that financial institutions must apply. Unfortunately, economy and finance are not exact sciences and human behavior, specially society behavior, is not easy to predict. Under this circumstances, it is not easy to decide which are the correct rules and recommendations, those that will contribute to reduce the consequences of financial crises, even more given the amount of different points of view and interests that exists around everything related to finance. Anyway, once the necessity of the financial regulation is accepted, then it is important to analyze, understand and question as many aspects as possible related to the underlying hypothesis.

The Basel Committee on Banking Supervision has the purpose to establish recommendations about legislation and regulation for banks, one of the main actors on the financial system and one of the main protagonist of the crisis that started in 2008. The Basel II accord, published in 2004, establishes the regulatory capital (the risk sensitive capital requirement) a bank should hold.

Focusing on credit risk and simplifying, the Internal Rating-Based approach models the loss of a portfolio as the product of two risk factors: the probability of default

(PD) and the loss given default (LGD). The PD is the average percentage of the portfolio that will not be able to honor theirs obligations in one year, this percentage depends on many different factors and it evolves with time. Not all of those who default on theirs obligations will not pay again and there are ways to recover part of the defaults, so the LGD is the average percentage that will not be recovered from those who default on theirs obligations. Both risk factors have their own deep literature about their modeling techniques, but they are treated as two independent factors that do not affect each other.

The independence between PD and LGD is an implied assumption in the IRB approach, and the author, as many others did before, does not believe this is a correct assumption given the nature of the credit loss and the risk factors. The objective of this thesis is to analyze and quantify which are the effects on the IRB capital requirement given a portfolio where PD and LGD are not independent. In order to introduce the dependency between PD and LGD, a portfolio with different correlations between PD and LGD is simulated. For each different correlation, the capital requirement is computed with the estimated PD and LGD and it is compared with the computed with the real factors.

The rest of the thesis is structured in 4 chapters that follow this introduction. In chapter 2 there is a review of the financial regulation, Basilea's framework and everything related to the IRB approach is deeply treated, such as the economic foundations, the asymptotic single factor model and the risk parameters PD and LGD. In chapter 3 the methodology of the simulation is presented: how the portfolio is build, how it is modeled and how the results are analyzed and evaluated. The results of the simulation for different correlation are presented in chapter 4 and finally, in chapter 5 the conclusions of the thesis are summarized.

Chapter 2

Basilea's background

The first Basel Accord of 1988, also known as Basel I, laid the basis for international minimum capital standard and banks became subject to regulatory capital requirements, coordinated by the Basel Committee on Banking Supervision (BCBS). This committee has been founded by the Central Bank Governors of the Group of Ten at the end of 1974.

In June 2004 the BCBS released a Revised Framework on International Convergence of Capital Measurement and Capital Standards (Basel II). The rules officially came into force on January 1st, 2008, in the European Union.

Basel II is structured in a three pillar framework. Pillar one sets out details for adopting more risk sensitive minimal capital requirements, so called regulatory capital, for banking organizations (credit risk, market risk, operational risk). Pillar two lays out principles for the supervisory review process of capital adequacy and Pillar three seeks to establish market discipline by enhancing transparency in bank's financial reporting.

2.1 Credit Risk

Credit risk is the risk of a loss arising from a failure of a counterparty to honor its contractual obligations. The management of credit risk at financial institutions involves a range of tasks. To begin with, an enterprise needs to determine the capital it should hold to absorb losses due to credit risk, for both regulatory and economic capital purposes.

There are two possible methods, according to the Basel II regulation, to calculate the minimum capital for credit risk:

- Standard approach: it conserves the structure of the 1988 agreement. For each operation, there is an standard weight that depends on its degree of riskiness (4 different borrower categories: state, bank, mortgages, companies and retail) and then a capital of the 8% of this weighted assets is required. In order to achieve a greater sensitivity, the new regulation admits the use of external rating provided by rating agencies (like S&P's, Moody's or Fitch Ratings), to amplify the number of weights and to substantially reduce the weights for mortgage, retail, small and medium enterprises. The main advantage of this approach lies in its simplicity, so it can be applied for every kind of entities, but on the other hand, this approach is quite conservative.

- Internal ratings-based approach (IRB): for this approach a regulatory model has been developed. It is permitted that the entity internally estimates some risky factors or inputs of the model. In order to use this approach, the entity should have an internal rating system that allows to classify their clients in an enough number of homogeneous categories (buckets), the entity should accomplish some minimum requirements and the entity should have the explicit supervisory approval.

2.1.1 IRB approach

The risky factors that the entity can internally compute for each category are:

- Probability of default (PD): Likelihood that a loan will not be repaid in the period of one year.

- Exposure at default (EAD): Potential exposure measured in currency.

- Loss given default (LGD): Magnitude of likely loss on the exposure as a percentage of the exposure or, equivalently, the loss the entity will face in the case the obligor defaults.

- Maturity (M): Time to expiration date

The IRB approach includes two alternatives, the basic and the advanced. In the basic IRB approach, the entities can use the PD of each category to calculate the minimum capital with a model that weights the risks as a function of the value of PD. The LGD, EAD and M are fixed by the supervisor. In the advanced IRB approach, the entities should also use their internal estimations of LGD, EAD and M. Other risky factors, such as the asset's category correlation, are fixed by the supervisor in both cases.

Once the different risky factors are estimated for each category, the Risk Weighted Assets Formula is used to compute the capital requirements. The minimum capital for each category is the 8% of the RWA calculated. The total minimum capital is obtained as the sum of the minimum capital for each category.

2.1.1.1 Economic foundations

In the credit business, there are always some borrowers that default on their obligations. The losses that may arise in a particular year vary from year to year, depending on the variations of PD and LGD. The variation in realized losses over time leads to a distribution of losses. It is not possible to know in advance the losses the bank will suffer in a particular year, but the bank can forecast a reasonably level of average losses it can expect and thus can easily manage them trough the pricing and provisioning. This forecast of reasonably average losses is called Expected Losses (EL).

It can happen that the final losses are greater than the forecasted EL. Such peaks do not occur every year, but when they occur the bank's capital should cover the risk of such peak loses and provide a buffer to protect the bank's debt holders. Losses above EL are called Unexpected Losses. In Figure 2.1 we can observe the EL, the UL and the Loss Distribution.

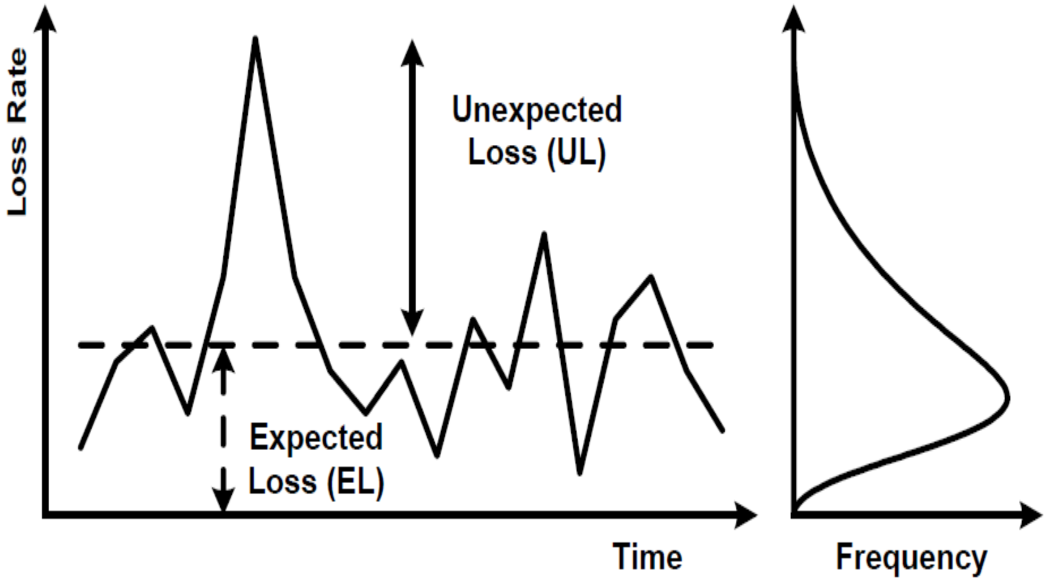

Figure 2.1: Loss Distribution. Source: Basel Committee on Banking Supervision (2005a).

On the one hand, banks have an incentive to minimize the capital they hold to free up resources that can be used in their business. On the other hand, the less capital a bank holds, the greater is the likelihood that they become insolvent when the peak losses can't be covered with the profit and the available capital.

The IRB approach focuses on the frequency of bank insolvencies arising from credit losses that the supervisors are willing to accept. In other words, capital to hold is set to ensure that unexpected losses will exceed this capital with a very low and fixed probability.

In Figure 2.2 we can observe the probability density function of the Loss Distribution. The likelihood that a bank will become insolvent, i.e.the realized losses exceed the EL and the UL, is equal to the black area under the curve. The confidence level is 100% minus this likelihood and represents the probability that the bank will not become insolvent. Finally, the corresponding threshold is called Value-At-Risk (VaR) and if the realized loss is greater than the VaR, then the bank will become insolvent.

Expected Loss can also be seen as the result of its components, that is:

$$EL = PD \cdot LGD \cdot EAD.$$

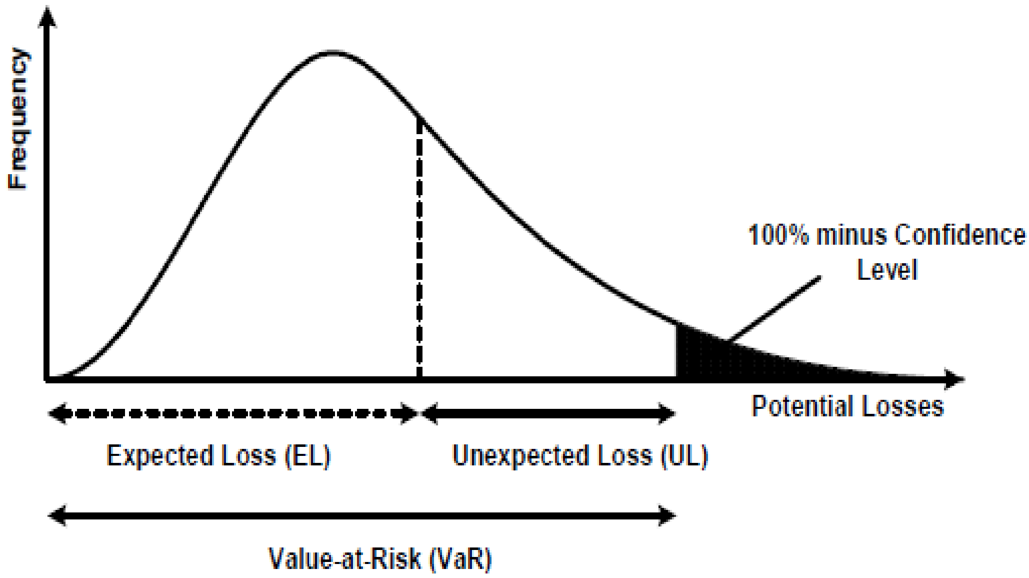

Figure 2.2: Density Loss Distribution. Source: Basel Committee on Banking Supervision (2005a).

2.1.1.2 The ASRF framework

Model specification is subject to an important restriction, the model should be portfolio invariant, i.e. the capital required for any given loan should be independent of the portfolio it is added to. This specification attends to well diversified portfolio and it has been deemed vital in order to make the IRB approach applicable to a wide range of credit institutions across the countries. In the specification process of the Basel II model, it turned out that portfolio invariance was a property with a strong influence on the structure of the portfolio model. Gordy (2003) showed that essentially only Asymptotic Single Risk Factor models are portfolio invariant, which are derived from ordinary credit portfolio models by the law of large numbers.

Unexpected Losses can be distinguished in systematic risk and idiosyncratic risk. The idiosyncratic risk arises from dependencies across individual obligors in the portfolio and from common shocks in the environment, while the systematic risk arises from obligor specific shocks. When a portfolio consists of a large number of small exposures and it is well diversified, idiosyncratic risks associated with individual exposures tend to cancel out one-another and only the systematic risk have an effect on portfolio losses.

In the ASRF framework, all systematic risk are modeled with only one systematic risk factor. Given an appropriately conservative value of the single systematic risk factor, category's bank-reported average PD is transformed into a conditional PD using a supervisory mapping function that allows to compute the conditional expected loss of that category.

Under the ASRF model, the total economic resources (capital for the UL and provisions for the EL) a bank must hold for an exposure is equal to that exposure's conditional expected loss. However, provisions for EL are outside the Basel II framework because banks are expected to cover this on an ongoing bases and the Risk Weight Formulas only computes the capital for the UL. In order to do that, in the formula to calculate capital requirements, expected loss under normal circumstances is diminished from the conditional expected loss.

Capital Requirement (K) for a category is calculated as a percentage of the EAD with the following formula:

$$K = \left[LGD \cdot N \left(\frac{G(PD)}{(1-R)^{0.5}} + \left(\frac{R}{1-R} \right)^{0.5} G(0.999) \right) - PD \cdot LGD \right] \frac{1 + (M - 2.5) \cdot b(PD)}{1 - 1.5 \cdot b(PD)}. \tag{2.1}$$

Where:

- **PD**: Category's reported average Probability of Default (or pooled PD). It will be developed in the following section.
- **LGD**: Downturn Loss Given Default of the category. It will be developed in another section.
- **R**: Assets correlation in the category. It reflects the dependence of individual exposures to the rest of the exposures in the category and links the total exposure to the systematic risk factor. It is determined by the asset class in the bucket and in short, the correlation could be described as the dependence of the asset value of a borrower on the general state of the economy. If there is a high assets correlation, like in a large corporate loan portfolio, interactions between borrowers are high and defaults are strongly linked to the status of the economy. On the other hand, if there is a low assets correlation, like in

retail exposures, defaults tend to be more idiosyncratic and less dependent on the economic cycle.

- $N(\frac{G(PD)}{(1-R)^{0.5}}+(\frac{R}{1-R})^{0.5}G(0.999))$: Conditional Probability of Default of the bucket. It will be developed in another section.

- **PD · LGD**: Expected Loss under normal circumstances. It diminishes the Conditional Expected Loss in order to remove EL from capital requirement.

- **M**: Maturity. Credit portfolio consist of instruments with different maturities. Long-term credits are riskier than short-term credits, as a consequence, capital requirement should increase with maturity. The last fraction of formula (2.1) is for the maturity adjustment, but this is out of the scope of this thesis and won't be explained.

2.1.1.3 Pooled PD

It is helpful to start drawing a distinction between the concept of a default probability linked to an individual obligor and the PD assigned to a risk bucket or category. The PD associated with a concrete obligor is a measure of the likelihood that this obligor will default during the following year, while the PD assigned to a risk bucket is a measure of the average level (the mean) of the PDs of obligors within that bucket, that is the so called pooled PD.

The Revised Framework specifies that the pooled PD should be a long-run estimation, i.e. that it does not depend on the moment of the economic cycle. It is allowed three different approaches to quantify pooled PDs:

- Historical approach: the PD is estimated using historical data on the frequency of observed defaults among obligors assigned to that bucket.With a minimum of 5 years of history, the PD calculated is called long-run default frequency.

- Statistical model approach: predictive statistical models are used to estimate a PD for each obligor of the bucket. Then the bucket's pooled PD is calculated as the average of the obligor-specific PDs.

- External mapping approach: First the mapping function establishes a link between each of the bank's internal risk buckets with an external rating grade, finally external data is used to calculate pooled default probabilities of the external rating grades.

2.1.1.4 Conditional PD

The conditional PD is obtained as the result of a mapping function depending on the Reported average PD and the predetermined systematic risk factor value (the predetermined supervisory confidence level, that is 99.9%). This function is derived from an adaptation of Merton (1974) single asset model to credit portfolios, that assumes that a default will arise when the assets (modeled with a Normal distribution) of the borrower is lower than the due amount. Vasicek (2002) showed that under certain conditions, Merton's model can be extended to a specific ASRF credit portfolio model. Since this thesis is restricted to an analysis of PD-LGD correlation, the conditional PD development is not presented here.

2.1.1.5 LGD

The economic-downturn LGD reflects adverse economic scenarios and is expected to exceed those LGD that arise during typical business conditions. It can be obtained with two different approaches: On the one hand, it is possible to apply a mapping function (as it is done to obtain the conditional PD) to bank-reported average LGDs. On the other hand, banks could provide downturn LGD based on their internal assessments of LGDs during adverse economic conditions. The Basel Committee decided that given the evolving nature of bank practices in the are of LGD qualification, it would be inappropriate to apply a single supervisory LGD mapping function to link average LGD to downturn LGD. Three different methods are used to measure LGD:

- Workout LGD: it is based on the discount of future cash flows resulting from the workout process from the date of default to the end of the recovery process.
- Market LGD: it is derived from the observation of market prices on defaulted bonds.
- Implied market LGD: it is derived from non-defaulted risky bond prices by means of an asset pricing model.

Chapter 3

Simulation study

The main objective of this thesis is to analyze the effects of the correlation between PD and LGD, for different levels of correlation, in a given portfolio. Even if it would be more reliable to compute the effects with real data, it would not be possible to decide a different level of correlation to analyze the effects and this is the main reason for simulating all the data used in this thesis. Another important reason is that it would not be easy to obtain data from a risk department.

There are three different phases in the simulation study. The first step is to build the dataset that the bank risk department would have with all the information of the portfolio. The second step is to estimate the models to compute the IRB risk parameters as the risk department would do and finally, the third step is to analyze the effects of not taking into account the correlation between PD and LGD in the second step, concretely in the capital requirements.

It is important to remark that the modeling of the PD that will be implemented is the ”Statistical model approach” (section 2.1.1.3), where the pooled PD is computed as the mean of the individuals PD. Although, the modeling of the LGD that is implemented is simpler than the one in the IRB approach (section 2.1.1.5) because of two reasons. First, the economic-downturn conditions are not included and second, the method that is used to measure LGD depends on a linear regression and this is not any of the three different methods that is possible to use to measure LGD in the IRB approach. The modeling with the linear regression is done in order to simplify the simulation process.

All the tables and figures in this chapter correspond to the same simulation study, with $n = 10000$ individuals and correlation between PD and LGD equal to $\rho = 0.5$.

3.1 Building the dataset

The aim is to build a dataset for a portfolio of n individuals with the following 4 values for each individual:

- D: Default response. It can be non-default (0) or indicates default (1).
- Z_1: Summarized intrinsic factor for the PD.
- LGD: Loss Given Default as a percentage.
- Z_2: Summarized intrinsic factor for LGD.

3.1.1 Introducing the correlation

In order to achieve that there exists a correlation between PD and LGD in the portfolio and that this correlation is equal to ρ, the correlation is induced in two uniform random variables u_1 and u_2. These two variables will lead to the PD and LGD, respectively.

The approach is to generate data from the Gaussian copula with an appropriate correlation coefficient such that the Spearman's ρ (ρ_S) corresponds to the desired correlation for the uniform random variables. ρ_S for a Gaussian copula is obtained as:

$$\rho_S = \frac{6}{\pi} arcsin \frac{\rho_P}{2},$$

where ρ_P is the Pearson's correlation coefficient, and $\rho_S = \rho_P$ for uniform distributions. The Cholesky decomposition of the covariance matrix is used to simulate the Gaussian copula.

At the end, the simulation process to obtain u_1 and u_2 with correlation ρ is the following:

Step 1. Compute the Pearson's coefficient (ρ_P), the covariance matrix (P) and its Cholesky decomposition (A):

$$\rho_P = 2 \cdot sin(\frac{\rho \cdot \pi}{6}),$$

$$P = \begin{bmatrix} 1 & \rho_P \\ \rho_P & 1 \end{bmatrix},$$

$$A = Chol(P).$$

Step 2. Generate n pairs $w_i = (w_{1i}, w_{2i})$ from an independent standard normal distribution.
Step 3. Add the correlation to each pair $w_i^* = w_i \cdot A$.
Step 4. Compute the uniform variables as the inverse of the normal distribution: $(u_{1i}, u_{2i}) = F^{-1}(w_{1i}^*, w_{2i}^*)$. This is a standard simulation procedure.

3.1.2 PD factor and default

Given a uniform random variable u_1 in the interval $(0, 1)$ and assuming that all possible information about the characteristics that might affect the PD of the individual are summarized in one intrinsic factor called Z_1 which is normally distributed, it is possible to compute the PD's factor (Z_1) as the inverse of the normal distribution $N(\mu_1, \sigma_1)$ where μ_1 is the mean of the factor and σ_1 is its standard deviation:

$$Z_1 = \mu_1 + F_{Z_1}^{-1}(u_1) \cdot \sigma_1$$

The resulting distribution of Z_1 is shown in Figure 3.1.

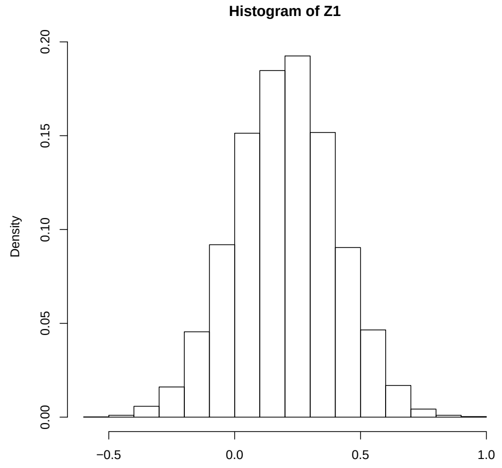

Figure 3.1: Distribution of Z_1 with $\mu_1 = 0.2$ and $\sigma_1 = 0.2$.

Once the factor Z_1 is calculated, the linear predictor of PD (S_1) is computed as $S_1 = \beta_0 + \beta_1 \cdot Z_1$ where β_0 is the independent term and β_1 is the coefficient of the factor. Straightaway, given the linear predictor S_1, the probability of default (P_1) is computed with the logistic distribution:

$$P_1 = \frac{1}{1 + e^{-S_1}}$$

The resulting distribution of P_1 is shown in Figure 3.2.

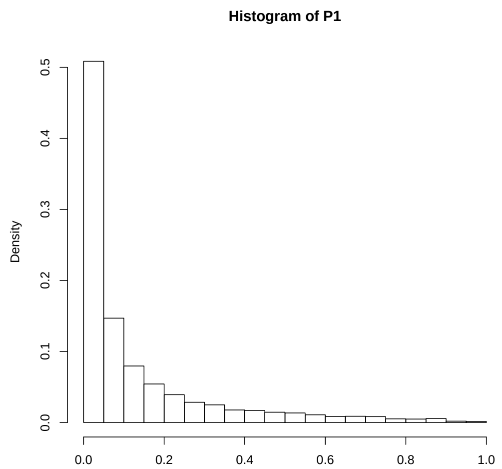

Figure 3.2: Distribution of P_1 with β_0 = -5 and β_1 = 10. Source: own elaboration.

Finally, the default (D) is calculated as the result of a Bernoulli random variable with probability P_1, so $P(D = 1) = P_1$ and $P(D = 0) = 1 - P_1$. By doing this procedure, the default rate of the portfolio is 12.76%.

In Figure 3.3 are shown the distribution of Z_1 for the default individuals (red), for the non-default individuals (blue) and for all the individuals (white). In the three cases the y-axis shows the percentage of the total population n.

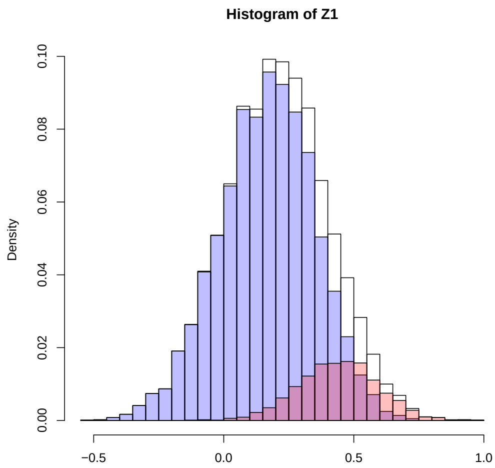

Figure 3.3: Distributions of Z_1 (white), Z_1 for $D = 1$ (red) and Z_1 for $D = 0$ (blue). Source: own elaboration.

A logistic regression is estimated with D as the dependent variable and Z_1 as the independent variable. If the simulation process is correct, then the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ should be really close to the ones that were defined. In table 3.1 it is shown the estimated coefficient for the logistic regression with β_0 = -5 and β_1 = 10.

Coefficient	Estimate	Std. Error	t-statistic	p-value
β_0	-4.920	0.104	-47.532	< 0.001
β_1	9.684	0.259	37.429	< 0.001

Table 3.1: Estimated coefficients of the logistic regression $D \sim Z_1$. Source: own elaboration.

3.1.3 LGD factor and LGD

Given a uniform random variable u_2 in the interval $(0, 1)$ and assuming that all possible information about the characteristics that might affect the LGD of the individual are summarized in one intrinsic factor called Z_2 which is normally distributed, it is possible to compute the LGD's factor (Z_2) as the inverse of the normal distribution $N(\mu_2, \sigma_2)$ where μ_2 is the mean of the factor and σ_2 is its standard deviation:

$$Z_2 = \mu_2 + F_{Z_2}^{-1}(u_2) \cdot \sigma_2$$

The resulting distribution of Z_2 is shown in Figure 3.4.

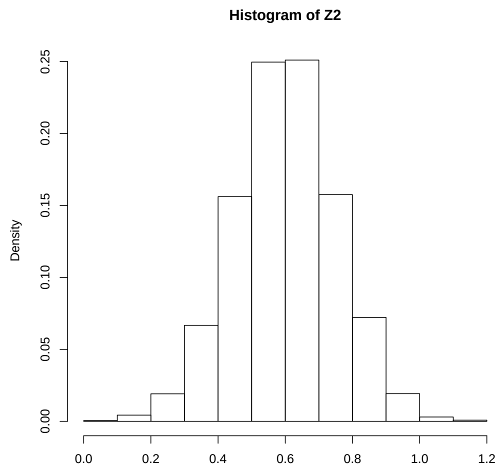

Figure 3.4: Distribution of Z_2 with $\mu_2 = 0.6$ and $\sigma_1 = 0.15$. Source: own elaboration.

Once the factor Z_2 is calculated, the linear predictor of LGD (Y) is computed as $Y = \alpha_0 + \alpha_1 \cdot Z_2$ where α_0 is the independent term and α_1 is the coefficient of the factor. Straightaway, given the linear predictor Y, the LGD is computed with a normal distribution of mean Y and standard deviation σ_{LGD}:

$$LGD = Y + N(0, \sigma_{LGD})$$

In order to include the economic foundations, the obtained values greater than one are censores to 1 and the values lower than zero are censored to 0. The resulting distribution of LGD (with $\sigma_{LGD} = 0.1$) is shown in Figure 3.5 and its scatter plot (LGD, Z_2) is shown in Figure 3.6.

The mean of the LGD for all the individuals is 60.11% and the mean for the defaulters is 69.34%.

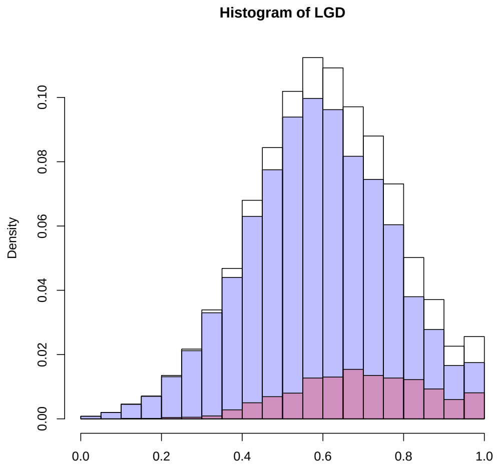

Figure 3.5: Distributions of LGD (white), LGD for $D = 1$ (red) and LGD for $D = 0$ (blue). Source: own elaboration.

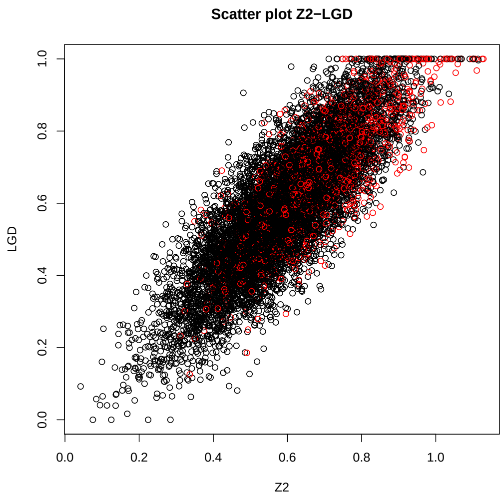

Figure 3.6: Scatter plot for Z_2 and LGD. $D = 1$ (red) and $D = 0$ (black). Source: own elaboration.

A linear regression is estimated with LGD as the dependent variable and Z_2 as the independent variable, but only for the defaulter individuals ($D = 1$). If the simulation process is correct, then the estimated coefficients $\hat{\alpha_0}$ and $\hat{\alpha_1}$ should be really close to the ones that were defined. In table 3.2 it is shown the estimated coefficient for the linear regression with $\alpha_0 = 0$ and $\alpha_1 = 1$.

Coefficient	**Estimate**	**Std. Error**	**t-statistic**	**p-value**
α_0	0.045	0.014	3.307	0.001
α_1	0.929	0.019	48.057	< 0.001

Table 3.2: Estimated coefficients of the linear regresion $LGD \sim Z_2$. Source: own elaboration.

3.1.4 Observed factors

The two intrinsic factors that produce the default (D) and the LGD are not totally observable. In the real world the observed factor would be biased. In order to get the observed factors, a bias in the intrinsic factors is introduced. This is done generating two uniform variables (u_3 and u_4) in the interval $(0.9, 1.1)$:

$$Z_1^o = Z_1 \cdot u_3$$
$$Z_2^o = Z_2 \cdot u_4$$

3.2 Estimation of the risk parameters

It is assumed that the dataset that is build in the previous section (n individuals with theirs D, Z_1^o, LGD and Z_2^o) will be similar to the input that a bank risk department would dispose. The first step is to assign a PD between 0 and 1 for each individual. This is done by fitting a logistic regression of the default D against the observed factor Z_1^o. The results on the estimation for the same dataset that the one that is used in section 3.1.2 are shown in Table 3.3. Due to the characteristics of the logistic regression, the mean of the estimated PD is equal to the percentage of defaulters in the portfolio, 12.76%.

Coefficient	Estimate	Std. Error	t-statistic	p-value
β_0	-4.851	0.101	-47.822	< 0.001
β_1	9.483	0.253	37.424	< 0.001

Table 3.3: Estimated coefficients of the logistic regresion $D \sim Z_1^o$. Source: own elaboration.

The second step is to assign a LGD between 0 and 1 for each individual. This is done fitting a linear regression of the LGD against the observer factor Z_2^o only for the defaulted individuals (only the ones with $D = 1$), the results of the estimation on the same dataset are shown in table 3.4.

Coefficient	Estimate	Std. Error	t-statistic	p-value
α_0	0.090	0.014	6.361	< 0.001
α_1	0.865	0.020	43.450	< 0.001

Table 3.4: Estimated coefficients of the linear regresion $LGD \sim Z_2^o$. Source: own elaboration.

The obtained model is applied to all the portfolio and finally the mean of the estimated LGD for all the individuals is 61.01% and the mean for the defaulters is 69.31%.

3.3 Evaluation of the simulation

Two different analysis will be done: to analyze the effects of different correlation in the estimated coefficients and to quantify the impact in the capital requirement when the risk parameters that are used do not take into account this correlation.

3.3.1 Estimated models

Four different models play a roll in the simulation process: the logistic regression estimated on the factor Z_1 with estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$, the logistic regression estimated on the observed factor Z_1^o with estimated coefficients $\hat{\beta}_0^o$ and $\hat{\beta}_1^o$, the linear regression on the factor Z_2 with estimated coefficients $\hat{\alpha}_0$ and $\hat{\alpha}_1$ and the last one, the linear regression estimated on the observed factor Z_2^o with estimated coefficients $\hat{\alpha}_0^o$ and $\hat{\alpha}_1^o$.

It is interesting to analyze the effects on the estimated coefficients that the different correlations may induce. To analyze this effect, the quadratic error of the estimation and the total standard deviation is computed separately for the "real models" (the ones obtained with the intrinsic factors) and the "observed models" (the ones obtained with the observed factors):

$$\begin{aligned}
e_{real} &= (\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta}_1 - \beta_1)^2 + (\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\alpha}_1 - \alpha_1)^2 \\
e_{obs} &= (\hat{\beta}_0^o - \beta_0)^2 + (\hat{\beta}_1^o - \beta_1)^2 + (\hat{\alpha}_0^o - \alpha_0)^2 + (\hat{\alpha}_1^o - \alpha_1)^2 \\
\sigma_{real} &= \sigma(\hat{\beta}_0) + \sigma(\hat{\beta}_1) + \sigma(\hat{\alpha}_0) + \sigma(\hat{\alpha}_1) \\
\sigma_{real} &= \sigma(\hat{\beta}_0^o) + \sigma(\hat{\beta}_1^o) + \sigma(\hat{\alpha}_0^o) + \sigma(\hat{\alpha}_1^o)
\end{aligned}$$

Following with the same example than before, the obtained quadratic error for the real models is 0.113 and for the observed models is 0.317, which make sense with the fact that the observed models were obtained from the observed factor, that were distorted from the intrinsic factors, the ones that were used to build the default and the LGD.

3.3.2 Risk parameters

To evaluate the effects of not taking into account the possible correlation between PD and LGD, the capital requirements according to Basilea will be computed with equation 2.1 applied to the portfolio mean of PD and LGD. On the one hand, the real capital requirement is obtained applying PD equal to the rate of default in the portfolio (PD_r) and LGD equal to the mean of the defaulters LGD (LGD_r). On the other hand, the estimated capital requirements is obtained with PD equal to the mean of the estimated PD (PD_e) and LGD with the mean of the estimated LGD (LGD_e) for all the portfolio.

Following the example of this chapter, the rate of default PD_r is 12.76% and is equal to the mean of the estimated PD, the PD_e, so all the difference between the real and the estimated capital requirement will arise from theirs differences in the LGD. The mean of the defaulters LGD (LGD_r) is 69.34% and the mean of the estimated LGD (LGD_e) is 61.01%. This will imply that the estimated capital requirement (K_e) will be a 12% lower than the real one (K_r).

Chapter 4

Results

The simulation procedure explained in the previous section is applied for 5 different levels of correlation: 0, 0.25, 0.5, 0.75 and 0.9. In order to summarize the simulation process, for each different ρ the following steps are followed:

- **Step 1:** Simulate two uniform random variables (u_1 and u_2) of size $n = 10000$ with correlation ρ.
- **Step 2:** Create the PD factor Z_1 with $\mu_1 = 0.2$ and $\sigma_1 = 0.2$.
- **Step 3:** Compute the probability of default P_1 from the logistic distribution of Z_1 with $\beta_0 = -5$ and $\beta_1 = 10$.
- **Step 4:** Compute the Default for each individual as the result of a Bernoulli random variable with probability P_1.
- **Step 5:** Estimate the logistic regression with the default against the real PD factor Z_1.
- **Step 6:** Create the LGD factor Z_2 with $\mu_2 = 0.6$ and $\sigma_2 = 0.15$.
- **Step 7:** Compute the LGD from a linear regression of Z_2 with $\alpha_0 = 0$, $\alpha_1 = 1$ and $\sigma_{LGD} = 0.1$.
- **Step 8:** Estimate the linear regression with the LGD against the real LGD factor Z_2.
- **Step 9:** Compute the observed factors Z_1^o and Z_2^o.

- **Step 10:** Estimate the logistic regression with the default against the observed PD factor Z_1^o.

- **Step 11:** Estimate the linear regression with the defaulters LGD against the observed LGD factor Z_2^o.

- **Step 12:** Compute the quadratic error and the standard deviation of the "real models" (e_{real} and σ_{real}) and the "observed models" (e_{obs} and σ_{obs}).

- **Step 13:** Compute the real risk parameters (PD_r and LGD_r) and the estimated risk parameters (PD_e and LGD_e).

- **Step 14:** Compute the percentage difference between the estimated capital requirement and the real one.

It should be noticed that the same random seed is fixed before starting step 1 for each different ρ. This implies that the simulation of u_1 and all PD terms will give the same values for each different ρ. This procedure ends with the same level of PD and the same estimated models (the logistic regression with the real factor Z_1 and the one with the observed factor Z_1^o) in each simulation for a different ρ. Therefore, all the differences in the results arise from the differences in the LGD terms.

4.1 Estimated models

The quadratic error and the standard deviation of the "real models" (e_{real} and σ_{real}) and the "observed models" (e_{obs} and σ_{obs}) are shown in table 4.1.

ρ	e_{real}	e_{obs}	σ_{real}	σ_{obs}
0.00	0.109	0.301	0.392	0.386
0.25	0.110	0.307	0.393	0.387
0.50	0.113	0.316	0.395	0.389
0.75	0.117	0.333	0.399	0.392
0.90	0.120	0.350	0.401	0.395

Table 4.1: Model's results of the simulation study for different PD-LGD correlation. Source: own elaboration.

Taking into account that the dependent variables (D and LGD) where computed from the real factors, it is normal that the observed factors produce a worse fitting than the real ones. This can be established from the fact that the quadratic error of the estimated coefficients of the real models (e_{real}) is always lower than the quadratic error of the estimated coefficients of the observed models (e_{obs}), meaning that the estimated coefficients of the real models are closer to the real parameters. On the contrary, the standard deviation of the estimated coefficients is higher for the real models (σ_{real}) than the observed ones (σ_{obs}), meaning that the variance of the estimated coefficients is higher in the real models.

The main conclusion related to the level of correlation is that the higher the level of correlation is, the greater the error and the standard deviation are, in both cases, the real and the observed models. As it was explained before, the increases in the error and the standard deviation arise from the differences in the estimation of the LGD models (the PD models obtain always the same estimation for each different ρ). The estimated coefficients for the LGD linear regressions (with the real factor and with the observed factor) are shown in table 4.2. Summarizing, for a higher level of correlation, the estimated LGD models are more biased (increase in the quadratic error, ie, increase in the distance from the estimated coefficients to the original values α_0 and α_1) and less precise (increase in σ).

ρ	$\hat{\alpha}_0^r$	$\hat{\alpha}_1^r$	$\hat{\alpha}_0^o$	$\hat{\alpha}_1^o$
0.00	0.025	0.953	0.053	0.907
0.25	0.034	0.943	0.069	0.890
0.50	0.045	0.929	0.090	0.865
0.75	0.059	0.914	0.120	0.831
0.90	0.068	0.903	0.146	0.802

Table 4.2: LGD linear regression estimated coefficients of the simulation study for different PD-LGD correlation, with $\alpha_0 = 0$ and $\alpha_1 = 1$. Source: own elaboration.

The main issue is to understand where is the origin of these differences. For every different ρ, the u_2 is always a uniform distribution between 0 and 1, so the distribution of Z_2 is equal for each ρ (they are not exactly equal due to the simulation, but at least they are really similar). So far, the distribution of Z_2 is equal for any ρ but then, where do the differences come from? It is important to remember than the linear regression for the LGD is done only with the default individuals. The key is to understand what is happening with the distribution of the defaulters in relation to Z_2.

Even the distribution of Z_2 is equal for each ρ, the position of the defaulters changes for every different ρ. On the one hand, when there is no correlation ($\rho = 0$) the defaulters are randomly distributed in Z_2 (there will be defaulters with low Z_2 and there will be defaulters with high Z_2, this implies that the red points in figure 3.6 will be distributed in all the diagonal) because neither u_2 or Z_2 have relation with the Default. When the LGD linear regression is estimated with the defaulters, there are enough defaulters for all the values of Z_2 to recover the diagonal of the original linear regression that was used to compute the real LGD.

On the other hand, the defaulters are not randomly distributed in Z_2 when there is some level of correlation. Defaulters tend to have higher level of Z_2 so they tend to concentrate in the highest levels of Z_2 (this can be seen in figure 3.6, where the red points are concentrated at the right of the diagonal). When the LGD linear regression is estimated with the defaulters, the points are closer to a cloud than a diagonal and the estimated intercept increases while the estimated slope decreases. Finally, the more correlation there is between PD and LGD, more concentrated will be defaulters in the highest levels of Z_2, increasing the difficulty to achieve the original linear regression coefficients.

4.2 Estimated risk parameters

The different risk parameters and the percentage variation of the estimated capital requirement from the real capital requirement are shown in table 4.3. As it was explained before and due to the simulation process, all the PD terms are equal for any different ρ. Furthermore, due to the characteristics of the logistic regression, the real PD and the estimated PD are equal in each simulation.

When there is no correlation, the real LGD and the estimated LGD are really similar. On the other hand, as the correlation grows, even both LGD (real and estimated) grow they do it at a different rate, arising a big gap between both LGD as correlation increases. So far, the higher the level of correlation, the higher the underestimation of the estimated LGD from the real one. In the case of a correlation of 0.9, the underestimation of the LGD will be 17.4%.

ρ	PD_r	PD_e	LGD_r	LGD_e	ΔK
0.00	0.128	0.128	0.603	0.599	-0.663%
0.25	0.128	0.128	0.649	0.604	-6.935%
0.50	0.128	0.128	0.693	0.610	-11.977%
0.75	0.128	0.128	0.736	0.619	-15.897%
0.90	0.128	0.128	0.759	0.627	-17.391%

Table 4.3: Risk parameters results of the simulation study for different PD-LGD correlation. Source: own elaboration.

The origin of the gap between both LGD is the fact that the estimated LGD is the expected value of the portfolio, while the real LGD is the observed value on the defaulters. The difference arises again from the distribution of the defaulters in Z_2. When there is no correlation, the defaulters are randomly distributed in Z_2 so the observed value is close to the expected value. On the contrary, when there is correlation the defaulters are concentrated in the highest levels of Z_2 so, even the expected value remains more or less equal to the one with no correlation (due to the fact that the distribution of Z_2 do not change), the observed value is higher because the defaulters are concentrated.

4.3 Identifying the correlation

So far the underestimation of LGD has been detected to be higher as higher is the correlation between PD and LGD. The remain issue is to discover where it is possible to detect the correlation in order to be aware of the possible underestimation. In table 4.4 are shown the observed correlation (ρ_o) between the observed factors (Z_1^o and Z_2^o) and the estimated correlation (ρ_e) between the prediction of PD and the prediction of LGD.

It can be seen that the observed correlation and the estimated correlation grow as the correlation between PD and LGD grow, even if they do not exactly reflect the original level of correlation. This implies that in the case of having the two factors, it is possible to discover if the estimated LGD is going to be underestimated.

ρ	ρ_o	ρ_e
0.00	0.016	0.014
0.25	0.270	0.225
0.50	0.517	0.430
0.75	0.751	0.624
0.90	0.885	0.735

Table 4.4: Correlations of the simulation study for different PD-LGD correlation. Source: own elaboration.

Chapter 5

Conclusions

The presence of correlation between the factors affecting PD and LGD underestimates the minimum capital requirement in the IRB approach. It should be taken into account that the numerical results of the simulation are obtained using a limited modeling of the risk parameter LGD. In fact, the LGD that should be introduced in the IRB formula is the economic-downturn LGD, that will always be greater than the punctual LGD that the portfolio faces any year, but it was not posible to introduce this characteristic in the simulation of this thesis. Despite this limitation and many others around the metodology that is followed to introduce the dependence via correlation in the intrinsic factors or in the modeling process that is followed, the main conclusion of the simulation is clear: as the dependency between PD and LGD grows, the minimum capital is more underestimated. This conclusion is in line with those in Meng et al. (2010), and it is also one of the main reasons to use the economic-downturn LGD, to guarante that LGD is enough conservative to cover the PD-LGD correlation effect an any other misspecification effect that can exists.

Due to the limitations of the metodology, it is not possible to accept the numerical results of the impact that concludes the simulation, because part of the capital underestimation may be covered with the economic-downturn LGD, and as it is not introduced in the simulation it is not posible to quantify the real underestimation. However, an other important issue must be taken into account: if some correlation between PD and LGD in a potfolio is detected, an alarm should be turned on to warn about the possible consequences of not be using the correct LGD and to be even more conservative.

The main contributions that this thesis incorporates in relation to the acquired knowledge during the master are: the ability to understand the financial regulation, the use of LaTeX editor to write this thesis and the ability to programme advanced routines in the software R due to the simulation process that is followed.

Bibliography

Banco de España (2006). *Implantación y validación de enfoques avanzados de Basilea II en España.*

Basel Committee on Banking Supervision (2001). *Consultative Document: the New Basel Capital Accord.* Bank for International Settlements.

Basel Committee on Banking Supervision (2005a). *An Explanatory Note on the Basel II IRB Risk Weight Functions.* Bank for International Settlements.

Basel Committee on Banking Supervision (2005b). *Working Paper No 14. Studies on the Validation of Internal Rating Systems. Revised version.* Bank for International Settlements.

Basel Committee on Banking Supervision (2006). *Basel II: International Convergence of Capital Measurement and Capital Standards. A Revised Framework-comprehensive version.* Bank for International Settlements.

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

Meng, Q., Levy, A., Kaplin, A., Wang, Y., and Hu, Z. (2010). Implications of PD-LGD correlation in a portfolio setting. *Moody's Analytics.*

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

Ortiz, L. (2017). Advanced Risk Quantification [Lecture notes]. Master of Actuarial and Financial Science. *Universitat de Barcelona.*

Vasicek, O. (2002). The distribution of loan portfolio value. *Risk*, 15(12):160–162.

Appendix A

Simulation process example (R script)

```
### Parameters
n <- 10000      # Portfoloo size

mu1 <-  0.2     # Mean PD factor
sigma1 <- 0.2   # S.d. PD factor
beta0 <- -5     # Constant coefficient logistic PD
beta1 <- 10     # Factor coefficient logistic PD

mu2 <-  0.6     # Mean LGD factor
sigma2 <- 0.15  # S.d. LGD factor
alpha0 <- 0     # Constant coefficient linear regression LGD
alpha1 <- 1     # Factor coefficient linear regression LGD
sigmaLGD <- 0.1 # S.d error term in the LGD linear regression

dibuj <- TRUE   # Auxiliar parameter to plot results

results <- as.data.frame(matrix(NA, nrow = 1, ncol = 1))
colnames(results) <- c("rho")

rho <- 0.5      # Correlation for the procedure
results$rho <- rho

### Begins the simulation process
set.seed(121)

  ### Step 1: Simulate two uniform random variable
  rho2 <- 2 * sin(rho * pi/6)     # Pearson correlation
  P <- toeplitz(c(1, rho2))       # Correlation matrix
  d <- nrow(P)                    # Dimension
  U <- pnorm(matrix(rnorm(n*d), ncol = d) %*% chol(P))
  u1 <- U[, 1]
  u2 <- U[, 2]

  ### Step 2: Create the PD factor
  Z1 <- mu1 + qnorm(u1)*sigma1

  if(dibuj){
    h = hist(Z1, plot = FALSE)
    h$density = h$counts/sum(h$counts)
    plot(h,freq=FALSE, xlab = "")
  }

  ### Step 3: Compute the probability of default P1
  P1 <- 1/(1+exp(-(beta0 + beta1*Z1)))

  if(dibuj){
    h = hist(P1, plot = FALSE)
    h$density = h$counts/sum(h$counts)
    plot(h,freq=FALSE, xlab = "")
  }

  ### Step 4: Compute the default as a Bernoulli.
  D <- rbinom(n = n, size = 1, prob = P1)

  if(dibuj){
    hist3 <- hist(Z1, breaks = 30, plot = F)
    hist1 <- hist(Z1[which(D == 0)], breaks = hist3$breaks, plot = F)
    hist2 <- hist(Z1[which(D == 1)], breaks = hist3$breaks, plot = F)
    hist3$density <- hist3$counts/sum(hist3$counts)
    hist2$density <- hist2$counts/sum(hist3$counts)
    hist1$density <- hist1$counts/sum(hist3$counts)
    plot(hist3, col = rgb(0, 0, 0, 0), xlab = "", main = "Histogram of Z1", freq = F)
    plot(hist1, col = rgb(0, 0, 1, 1/4), add = T, freq= F)
    plot(hist2, col = rgb(1, 0, 0, 1/4), add = T, freq = F)
  }

  ### Step 5: Estimate the logistic regression with the Default agains Z1
  datasetReal <- as.data.frame(cbind(D, Z1))
  modelZ1 <- glm(D ~.,family=binomial(link='logit'),data=datasetReal)
  summary(modelZ1)$coefficients

  results$paramZ10 <- modelZ1$coefficients[1]
  results$paramZ11 <- modelZ1$coefficients[2]
  results$sdZ10 <- summary(modelZ1)$coefficients[1, 2]
  results$sdZ11 <- summary(modelZ1)$coefficients[2, 2]

  ### Step 6: Create the LGD factor Z2
  Z2 <- mu2 + qnorm(u2)*sigma2

  if(dibuj){
    h = hist(Z2, plot = FALSE)
    h$density = h$counts/sum(h$counts)
    plot(h,freq=FALSE, xlab = "")
  }
```

```
  ### Step 7: Compute the LGD from a linear regression of Z2
  u5 <- rnorm(n, 0, sigmaLGD)
  Y <- alpha0 + alpha1 * Z2
  LGD <- Y + u5
  LGD[LGD>1] <- 1
  LGD[LGD<0] <- 0

  if(dibuj){
    hist3 <- hist(LGD, breaks = 30, plot = F)
    hist1 <- hist(LGD[which(D == 0)], breaks = hist3$breaks, plot = F)
    hist2 <- hist(LGD[which(D == 1)], breaks = hist3$breaks, plot = F)
    hist3$density <- hist3$counts/sum(hist3$counts)
    hist2$density <- hist2$counts/sum(hist3$counts)
    hist1$density <- hist1$counts/sum(hist3$counts)
    plot(hist3, col = rgb(0, 0, 0, 0), xlab = "", main = "Histogram of LGD", freq = F)
    plot(hist1, col = rgb(0, 0, 1, 1/4), add = T, freq= F)
    plot(hist2, col = rgb(1, 0, 0, 1/4), add = T, freq = F)

    colLGD <- rep("black", n)
    colLGD[D==1] <- "red"
    plot(Z2, LGD, col = colLGD, main = "Scatter plot Z2-LGD")
  }

  ### Step 8: Estime the linear regression with LGD agains Z2
  datasetReal <- as.data.frame(cbind(LGD[D==1], Z2[D==1]))
  colnames(datasetReal) <- c('LGD', 'Z2')
  modelZ2 <- lm(LGD ~.,data=datasetReal)
  summary(modelZ2)$coefficients

  results$paramZ20 <- modelZ2$coefficients[1]
  results$paramZ21 <- modelZ2$coefficients[2]
  results$sdZ20 <- summary(modelZ2)$coefficients[1, 2]
  results$sdZ21 <- summary(modelZ2)$coefficients[2, 2]

  ### Step 9: Compute the observed factors
  u3 <- runif(n, 0.9, 1.1)
  Z1obs <- Z1*u3

  u4 <- runif(n, 0.9, 1.1)
  Z2obs <- Z2*u4

  ### Step 10: Estimate the logistic regression with Defalt against Z1obs
  datasetObs <- as.data.frame(cbind(D, Z1obs))
  modelZ1obs <- glm(D ~.,family=binomial(link = 'logit'),data=datasetObs)
  summary(modelZ1obs)$coefficients

  p1estim <- 1/(1+exp(-(modelZ1obs$coefficients[1] + modelZ1obs$coefficients[2]*Z1obs)))

  results$paramZ1obs0 <- modelZ1obs$coefficients[1]
  results$paramZ1obs1 <- modelZ1obs$coefficients[2]
  results$sdZ1obs0 <- summary(modelZ1obs)$coefficients[1, 2]
  results$sdZ1obs1 <- summary(modelZ1obs)$coefficients[2, 2]

  ### Step 11: Estimate the linear regression with LGD against Z2obs
  datasetObs <- as.data.frame(cbind(LGD[D == 1], Z2obs[D == 1]))
  colnames(datasetObs) <- c("LGD", "Z2obs")
  modelZ2obs <- lm(LGD ~ ., data=datasetObs)
  summary(modelZ2obs)$coefficients

  estimLGDobs <- modelZ2obs$coefficients[1] + modelZ2obs$coefficients[2] * Z2obs
  estimLGDobs[estimLGDobs > 1] <- 1
  estimLGDobs[estimLGDobs < 0] <- 0

  results$paramZ2obs0 <- modelZ2obs$coefficients[1]
  results$paramZ2obs1 <- modelZ2obs$coefficients[2]
  results$sdZ2obs0 <- summary(modelZ2obs)$coefficients[1, 2]
  results$sdZ2obs1 <- summary(modelZ2obs)$coefficients[2, 2]

  ### Step 12: Compute the quadratic error and the S.d. of real and observed
  results$errorReal <- (beta0 - results$paramZ10)^2 + (beta1 - results$paramZ11)^2 + (alpha0 - results$paramZ20)^2 +
(alpha1 - results$paramZ21)^2
  results$errorObs <- (beta0 - results$paramZ1obs0)^2 + (beta1 - results$paramZ1obs1)^2 + (alpha0 -
results$paramZ2obs0)^2 + (alpha1 - results$paramZ2obs1)^2

  results$sdReal <- results$sdZ10 + results$sdZ11 + results$sdZ20 + results$sdZ21
  results$sdObs <- results$sdZ1obs0 + results$sdZ1obs1 + results$sdZ2obs0 + results$sdZ2obs1

  ### Step 13: Compute the real and the estimated risk parameters
  results$LGDreal <- mean(LGD[D==1])
  results$LGDest <- mean(estimLGDobs)

  results$PDreal <- mean(D)
  results$PDest <- mean(p1estim)
```

```
### Step 14: Compute the percentage difference between the estimated capital requirements
results$VarK <- (results$LGDest - results$LGDreal)/results$LGDreal*100

### Recovering Correlation
results$corrZ12obs[j] <- cor(Z1obs, Z2obs)
results$corrPredLin[j] <- cor(p1estim, estimLGDobs)
```