Zsuzsanna Tajti:

The methodological opportunities of quantifying the retail mortgage loan's LGD in Hungary

Institute of Finance and Accounting
Department of Investments and Corporate Finance

Tutor:
Katalin Vér Gáspár dr.

Corvinus University of Budapest
Management and Business Administration Doctoral School
Department of Investments and Corporate Finance

The methodological opportunities of quantifying the retail mortgage loan's LGD in Hungary

- Ph.D. dissertation -

Written by: Zsuzsanna Tajti

Budapest, 2011

Table of contents

List of Figures

List of Tables

List of Appendices

“Except digging a hole there are only a few things that you can start immediately from the top.”

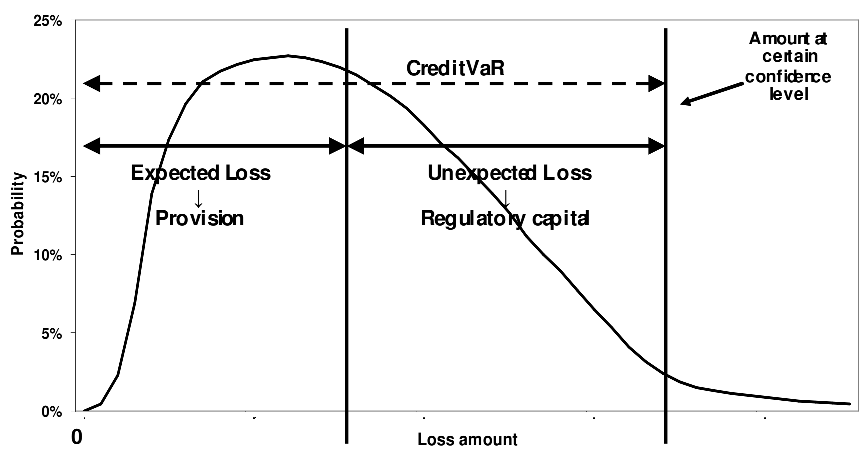
/ Harry Lorayne /

1. Introduction

The CRD (Capital Requirements Directive) founded on the Basel recommendations put the whole risk assessment of the banks on new bases. It is no exaggeration to state that it generated considerable changes on all risk-relevant areas of the activity of the credit institutions, concerning both the credit risk, the operational risk and the market risk.
However, this thesis focus only on a fairly narrow scope not disputing that the particularly complex system of the Basel rules does not enable the dissociation of the areas so categorically. I descend to the particulars only for the proper credit risk which can be defined as the so-called default risk, and I follow this interpretation in the framework of the whole thesis.

The CreditVaR concept serves as a basis for modelling the credit risk according to the Basel recommendations, on the basis of which different prescriptions refer to the assessment of the expected and the unexpected risks: while provision has to be formed for the first one, capital has to be allocated for covering the latter one. The task of the regulatory capital is to protect against the unexpected loss at a given confidence level. It can also be quantified as the difference between a given percentile of the loss distribution and the expected loss (
Figure 1). The term “unexpected loss at a given confidence level” derives from that.

Figure 1: The loss distribution of the credit risk



(Self-made figure)

With establishing the CRD it became possible for the credit institutions, if they use the Internal Rating Based (IRB) method regarding the credit risk, then they are allowed to apply their own calculations concerning certain credit risk parameters for quantifying the capital requirement, provided that they meet the assumptions and regulatory prescriptions of the Basel II.

The quantification and the measure of the credit risk are founded on the under-mentioned risk parameters in case of using the Internal Rating Based (IRB) models:

- Probability of Default (PD): the probability that the client becomes non-performing over a one year period.
- Loss Given Default (LGD): the ratio of the loss due to the default of the client to the exposure amount at default.
- Exposure at Default (EAD): the exposure at the default event.
- Maturity (M): the remaining time until the expiration of the deal.

In addition to serving for the objectives of managing the portfolio, the risk parameters also play an important role in calculating the expected and the unexpected loss as well as the Risk Weighted Assets (RWA) eventually.

1.1. Research antecedents and reasoning the theme

In the present dissertation I make known certain aspects of calculating the Loss Given Default (LGD), which is one of the most significant components referring to the calculation of the expected loss.

The rating system serving as a basis for the internal rating based method has to provide measurement of the credit risk, classifying and assigning the exposures to pools as well as quantification of the credit risk parameters belonging to them. The classification to grades and pools has to be based on assignment criteria, but the institutions have a relatively large liberty in defining them, because both the CRD and the *Government Decree No. 196/2007 on the Management and Capital Requirement of Credit Risk (Hkr.)* contain only very general prescriptions concerning them. It is expected that the credit institutions lean on the significant risk drivers during the calculations, but there are neither in the CRD nor in the Hungarian regulations any exact prescriptions relating to their scope, so their establishment is the certain institution’s task.

An overall requirement for assignment to pool is that the concentration should not be disproportionately high. The categories have to be defined and the number of categories has to be appointed in a way, which provides the assignment of homogenous exposures to the same pools, but the numbers of exposures in the certain pools should be sufficient to allow reliable quantification of the risks, enabling the exact and consistent quantification of loss characteristics at grade or pool level. So the regulation prescribes the credit institutions to choose the "golden mean".

Considering that the data series available for the majority of the Hungarian credit institutions are not old and accurate enough for carrying out appropriately consistent estimations, as well as the quantity of the default data is not adequate in many cases, therefore calculating the own LGD values comes up against numerous difficulties.

In the actual Hungarian practice the institutions are not able to take advantage of the theoretical opportunity given by the Basel rules in many cases yet, because the necessary conditions of the secondary market of loans and bonds do not exist. For that very reason the credit institutions have to focus on the historic collecting the internal data and on the basis of them on preparing the most possible accurate predictive models for the sake of exactly quantifying the credit risk parameters on the basis of them.

At the same time the opportunity has a great importance from the viewpoint of the present dissertation that, though the estimations have to be founded basically on the internal data, but external or even common data can be used as well, if it is provable that there is not any significant difference between the internal and external data regarding the assigning processes into grades or pools, as well as the composition of the data (risk profile), or if the differences can be adjusted properly for the sake of completing representativity.

While the literature of the credit risk has been paying notable attention to estimating Probability of Default (PD) for a long time, the quantification of Loss Given Default rate (LGD) has got much less emphasis. Only in the latest few years came modelling the LGD and the recovery rate into the limelight.

Concerning the corporate sector there is already a comprehensive literature about both the theoretical and the empirical LGD modelling, and the more so about modelling the recovery rates, while there are barely a few examples in case of the retail loans in spite of the fact that the retail loan outstandings in whole considerably exceed the amount of the corporate deals.

Considering that the scarcity of data means the largest barrier of the model-building in Hungary, the available database with larger quantity for the loan deals relating to the retail segment implicates more considerable potential in some respects, in comparison with the corporate sector. At the same time there is a rather narrow scope of the information, which is available for the credit institutions, and which can be used as influencing factor of the recovery rate and the LGD in the course of preparing the predictive models hereby.

The aim of my research was to study the characteristic features of the LGD parameter of the retail mortgage loans and to prepare a model for calculating the LGD, which enables the more exact and accurate quantification of this risk parameter than the actual, under domestic conditions. Considering that only the application of the workout LGD methodology has actually the reason for existence in Hungary recently (Chapter 3), I also grounded my empirical researches on it.

1.2. The examined Hypotheses

In the framework of this dissertation I studied the specialities of the LGD parameter of the retail mortgage loans, and I took steps to prepare a model with which more exact and more accurate LGD calculation will be possible.

1st Hypothesis: The LGD values of the loans with home purpose are lower than the LGD values of the mortgage equity withdrawals.

The object of my 1st Hypothesis was the connection between the purpose of the loan and the LGD. According to my anticipative expectations in the case of the deals, where the purpose of the loan is the construction or purchase of the real estate which serves as collateral, larger recoveries can be expected in comparison with the mortgage equity withdrawals. In addition to the preceding empirical results (for example *Grippa et al. [2005]*) the belief lies behind this that the clients presume less to take the risk of losing their home in the case, if they had decided to take up the loan exactly for the sake of its obtainment.

2nd Hypothesis: The purely collateral-based loans without income verification are characterized by higher LGDs than the loans based on income verification.

In the framework of my 2nd Hypothesis I investigated whether the LGD values of the purely collateral-based loans without income verification and the mortgage loans based on income verification differ from each other significantly. According to my presumption only lower recoveries can be expected from the deals which belong to the former group, following the occasional default event, because the income of the clients who have resort to this kind of loan is supposedly lower and less steady in comparison with the ones who are prepared to give free run of their income certificate to the bank at the application.

3rd Hypothesis: The type of the applied discount rate influences the calculated LGD value considerably.

In case of the basic model I used the contractual lending rate of each deal as discount rate, and in the course of investigating my 3rd Hypothesis I analyzed the effects of using the four following alternative discount rates: discount rate of 0%, the contractual Annual Percentage Rate of the given deals, the central bank base rate of the currency of the deal effective at the default, and the central bank base rate of the currency of the deal effective on 30th June 2011.

4th Hypothesis: The lowering of the materiality threshold used in the basic model does not affect the result of the LGD calculation considerably in case of the retail mortgage loans.

Considering that the credit institutions are also allowed to use criteria for materiality threshold which are different from the prescriptions *(Hkr. 68. § (5)-(7) Paragraph)*, if they are able to justify its necessity, reasonability, my 4th Hypothesis was directed towards investigating whether a considerable role can be put down to the use of the lower materiality thresholds from the viewpoint of the result of the LGD calculation. According to my anticipative expectations the low-amount arrears are quite rare in case of the mortgage loans, so it has a relatively small probability that the clients delay with an amount which is smaller than the materiality threshold described by the Hkr.

5th Hypothesis: The LGD values of the categories according to the closing type of the deals differ strongly from each other, and the elements of the two groups which have closed recovery process ("NoFurtherRec", "WorkoutEnd") can be properly separated with using logistic regression.

In the course of my 5th Hypothesis I investigated whether my anticipative expectation can be justified that the LGD values of the categories defined according to the closing type of the deals differ strongly from each other, and the logistic regression methodology can be successfully applied for carrying out the classification.

6th Hypothesis: With the linear regression models on the basis of the Hungarian Interbank LGD Database, the deals of the "NotClosed" category can also be involved in the calculation, and a more exact and more accurate deal level LGD estimation becomes possible.

My 6th Hypothesis was directed towards the examination of the Hungarian Interbank LGD Database: my goal was to make a survey of the factors which are able to predict statistically confidently the length of the period which is needed for the recoveries from selling the collateral or the debt, and to predict the recovery rate itself. I took steps with using the data of the Hungarian Interbank LGD Database to develop a complex model, with which the deals can also be involved in the calculations whose collection process has not been closed yet. However, it is its very important condition to prepare regressions, with which a precise prediction can be made referring to the expected length of the recovery period of the deals and the recovery rate deriving from the selling, on the basis of the data which are available at the default.

> ***7th Hypothesis: Different factors influence the LGD values of the deals with different closing types ("WorkoutEnd", "NoFurtherRec"), thus it is inappropriate to handle these categories together in the course of modelling the deal level LGD.***

In the framework of the 7th Hypothesis I investigated whether the influencing factors of the LGD values of the deals with different closing types differ considerably from each other.

The *"WorkoutEnd"* category contains the deals which are not in default status any more, because the client has paid back the delayed amount, the exposure has been written off or the property which served as underlying collateral has been sold. Contrary to that the *"NoFurtherRec"* category consists of the deals which are still in default status, since their becoming non-performing longer than 36 months duration has passed, and in case of which at least 90% of the exposure at the date of default has recovered.

1.3. The applied methods

In the framework of my 1st and 2nd Hypothesis I investigated whether the LGD values of the categories, worked up from the deals in the database examined by me, significantly differ from each other on the basis of the loan purpose and the type of the application. In the first step I compared the distributions on the basis of the descriptive statistics (mean values, indices of dispersion, kurtosis and skewness) and graphically illustrating with bar-chart, then in the next step I carried out Homogeneity Analysis regarding the equivalence of the LGD distributions. For the purpose of the Homogeneity Analysis I created 16 LGD bands (classes), but I did not define their broadness equally, instead I considered narrow intervals on the segments near 0% and 100%, and broader intervals on the middle section as separate LGD bands, moreover I worked up distinct classes for the LGD values of 0% and 100% with respect to the large quantity of the extreme values.

Despite the fact that the distributions notably differed from the normal distribution, regarding the considerably large quantity of elements I carried out asymptotic ***z***-tests to examine the equality of the average LGD values. Here and during the execution of the statistical tests of the further hypotheses (asymptotic ***z***-tests, ***t***- and ***F***-tests, Homogeneity Analyses) alike I applied a significance level of 5% and ***p***-value approach.

My 3rd Hypothesis was directed towards evaluating to what extent certain alternative discount rates divert the LGD values from the ones of the basic model, namely I compared the LGD values, which were calculated with the alternative discount rates, to the LGD values of the basic model in all cases. Following the investigation of the descriptive statistics and the graphical illustrations of the distributions I carried out Homogeneity Analysis pair-wise referring to the equivalence of the distributions using the 16 LGD bands worked up previously, and with regard to the considerably large quantity of elements I examined the equality of the LGD values calculated with the different discount rates with paired two-sample ***t***-tests.

In the framework of my 4th Hypothesis I investigated the effect of using four alternative thresholds, in addition to the materiality threshold in the basic model, on the results of the LGD calculation. I separated the "technical defaults", namely those which are not considered as non-performing according to the definition of the basic model, but they did according to the materiality threshold of 0 HUF. In the course of the examinations I compared the LGD values of this subportfolio with the LGD values in the basic model, using the same methodologies as during testing the 1st and the 2nd Hypothesis.

The subject of my 5th Hypothesis was the search for the features of the categories defined on the basis of the closing type of the deals, since according to my anticipative expectations the characteristics of the cases which compose the categories of the different closing types are insomuch diverse, that they are properly classifiable with using statistical methods.

I built the logistic regression with SAS Enterprise Miner™ 5.2 applying stepwise model selecting procedure. Testing numerous model types and transformation procedures I compared the performances of the regressions on the basis of fit statistics, and considering them I decided upon the model which applies logit link without any transformation. Following that I analyzed the results of the Maximum Likelihood estimation referring to the variables of the model also from the viewpoint of interpretability.

In the course of examining my 6^{th} Hypothesis my goal was to make a survey of the factors which are able to predict statistically confidently the length of the period which is needed for the recoveries from selling the collateral or the debt, and to predict the recovery rate itself. For the purpose of justifying my hypothesis I built separately linear regressions referring to the expected length of the recovery period and to the recovery rate deriving from the selling on the basis of the Hungarian Interbank LGD Database.

Following that in the framework of my 7^{th} Hypothesis I investigated whether the influencing factors of the LGD values of the deals with different closing types differ considerably from each other. In this case I also created linear regressions separately for the categories according to the closing types of the deals, and on the basis of them I searched the factors which proved to be significant.

I built the models, which served as a basis for the examination of my 6^{th} and 7^{th} Hypotheses, with stepwise procedure using SAS Enterprise Miner™ 5.2, then in case of the models whose adjusted coefficients of determination were rather low I made modifications on expert base for the sake of improving the explanatory power. During the model selection I considered the adjusted coefficients of determination and the results of the global Wald test, and I verified the relevance of each variable with using ***t***-test.

1.4. The most important results of the thesis

In the following I summarize the most important results of my research.

1st Hypothesis: The LGD values of the loans with home purpose are lower than the LGD values of the mortgage equity withdrawals.

According to the examinations, which were carried out, my 1^{st} Hypothesis did not prove to be true, the LGD values of the loans with home purpose seemed lower than the LGD values of the mortgage equity withdrawals at none of the popular significance levels, the results of the tests show just the opposite of that. The analyses also clarified that the LGD distributions of the two groups defined within the loans with home purpose (home building and home purchase) differ much less from each other than the LGD distributions of the mortgage loans with home purpose and the mortgage equity withdrawals, thus the separate treating has relevance only in the case of the two latter groups in the course of the categorization, the application of more detailed parcelling does not have any notable added value.

2^nd Hypothesis: The purely collateral-based loans without income verification are characterized by higher LGDs than the loans based on income verification.

The tests, which were carried out, uniformly seem to verify my 2^nd Hypothesis, since they show that the LGD values of the purely collateral-based loans without income verification and of the deals based on income verification differ from each other significantly, the graphical illustration of the distributions and the descriptive statistics clearly show that the LGD values of the latter category are lower in the examined portfolio.

Considering that the LGD values of the deals based on income verification proved to be significantly lower than the LGD values of the purely collateral-based loans without income verification, if the deals pertaining to the latter category dominate among the loans with home purpose, then this can partly explain why the statement which is composed in the 1^st Hypothesis did not prove to be watertight. However, since the average LGD values of the loans with home purpose are higher in case of both deal categories which are defined on the basis of the type of the application, in comparison with the ones of the mortgage equity withdrawals, it does not give any explanation why the statement composed in the 1^st Hypothesis did not pass the test. Moreover the fact that in the examined portfolio the purely collateral-based loans without income verification represent larger proportion within the group of the mortgage equity withdrawals, than within the category of the loans with home purpose, would reason intuitively exactly the fact that the mortgage equity withdrawals should be featured by higher LGD values.

3^rd Hypothesis: The type of the applied discount rate influences the calculated LGD value considerably.

My examinations showed that, though in the high LGD range large differences did not appear between the proportions of the LGD values, which are calculated with the given alternative discount rates, considerable deviations can be experienced by 0% and in the LGD bands which are near that. The use of the 0% discount rate and the contractual Annual Percentage Rate diverted the LGD values the most considerably from the ones of the basic model. Although the differences seemed to be smaller in the case of the two other discount rates, even in case of them the presumption of both the equivalence of the distributions and the equality of the averages had to be rejected at all the popular significance levels. From all these results it may be concluded that the used discount rate has an important LGD influencing role, so they support the statement composed in my 3rd Hypothesis.

4th Hypothesis: The lowering of the materiality threshold used in the basic model does not affect the result of the LGD calculation considerably in case of the retail mortgage loans.

For the sake of testing my hypothesis I compared the LGD values of the "technical defaults" with the LGD values in the basic model, and the results showed that the statement composed in my 4th Hypothesis, according to which using the lower materiality thresholds does not cause considerable affect on the result of the LGD calculation, can be accepted only at quite low significance levels.

5th Hypothesis: The LGD values of the categories according to the closing type of the deals differ strongly from each other, and the elements of the two groups which have closed recovery process ("NoFurtherRec", "WorkoutEnd") can be properly separated with using logistic regression.

As the result of the modelling I managed to configure two fairly strong models.

In the model which applies the logit link the reasons of the default (whether the deal is considered as non-performing because of death; whether the delay is the reason of the default status), the settlement type of the real estate which serves as collateral, some macroeconomic factors (the yearly average growth of the GDP and of the real wages from the origination of the deal to the default; the yearly real wage index at the default event), the ratio of the loan amount and the market value of the collateral at the origination as well as the paying history (the length of the period from the origination of the deal to the default event) proved to be key factors regarding the categorization of the default events.

In the model which applies the probit link the variables in connection with the rate of growth of the real wage do not appear, but as a quasi compensation the indices which measure the changing of the consumer prices proved to be significant. Similarly, the model which applies the probit link does not contain the ratio of the loan amount and the market value of the collateral at the origination, but the product type which is in tight connection with this variable, was qualified as significant. The industry of the client's employer, the region of the property and the amount of the first instalment occurred as further variables.

It is an important lesson that the reasons of the default, the settlement type of the real estate which serves as collateral, the yearly average rate of the GDP-growth in the period from the disbursement of the loan to the default, and the length of the period to the default proved to be significant in case of this model as well, and regarding these variables the direction of the connections are the same as the ones in the model with the logit link. Roughly speaking it can be stated that the two models show considerable cognateness concerning both the scale of the influencing factors and the direction of the connections.

On the basis of all these results my 5th Hypothesis can be considered as justified.

> ***6th Hypothesis: With the linear regression models on the basis of the Hungarian Interbank LGD Database, the deals of the "NotClosed" category can also be involved in the calculation, and a more exact and more accurate deal level LGD estimation becomes possible.***

In the first step I built a linear regression referring to the expected length of the recovery period. A part of the variables of the model prepared on the basis of the Maximum Likelihood estimation is in connection with the deal itself (the purpose of the loan; the proportion of the exposure at the default and the disbursement amount) or with the underlying collateral (the county; the ratio of the exposure at the default and the value of the collateral at the same time), whereas the other part of them consists of the macroeconomic changes in the period from the origination of the deal (the yearly average growth of the consumer prices and the unemployment rate from the origination of the deal to the default event) and the characteristics of the macroeconomic situation at the default (the consumer price index; the minimal wage).

In the next step I constructed a linear regression also for the recovery rate (the proportion of the recovery deriving from the selling discounted to the date of the default and the exposure at the default). It is conspicuous that numerous ones among the explanatory variables appear also in the model created for the length of the recovery period, namely there is a large overlapping between the factors of the two models: as a matter of fact very similar factors influence the length of the recovery period and the recovery rate deriving from the selling.

In this case as well, a part of the variables of the model prepared on the basis of the Maximum Likelihood estimation is in connection with the deal itself (the purpose of the loan) or with the underlying collateral (the county; the type of the settlement; the ratio of the exposure at the default and the value of the collateral at the same time; the quotient of the prior charges on the collateral and the realization value at the origination of the deal), whereas the other part of them consists of the macroeconomic changes in the period from the origination of the deal (the yearly average growth of the consumer prices and the unemployment rate from the origination of the deal to the default event), but the role of the characteristics of the macroeconomic situation at the default did not prove to be important. My results agreed in numerous respects with the results published in the studies of *Qi and Yang [2007; 2009]*.

It can be generally said that all the variables of the regressions can be interpreted logically easily, after all the explanatory power of the models is insofar low that it does not justify the statement composed in the 6th Hypothesis, since using the Hungarian Interbank LGD Database I did not manage to built a linear regression model which can is applicable for the purpose of prediction.

> ***7th Hypothesis: Different factors influence the LGD values of the deals with different closing types ("WorkoutEnd", "NoFurtherRec"), thus it is inappropriate to handle these categories together in the course of modelling the deal level LGD.***

The *"WorkoutEnd"* deal class is considerably heterogeneous, and it is not surprising that the regression, built with stepwise procedure using SAS Enterprise Miner™ 5.2, had a rather small explanatory power.

In the linear regression the factors which describe the macroeconomic situation at the default (the average default rate[1]; the average net income; the consumer price index; the yearly growth index of the real wages; the unemployment rate) as well as some deal and collateral characteristics (the county; the exposure at the date of the default; the proportion of the exposure at the default and the disbursed amount; the quotient of the prior charges on the collateral and the realization value at the origination of the deal) gained the most dominant role, and it is conspicuous that none of the client characteristics proved to be significant influencing factor.

I consider it necessary to emphasize the negative sign of the estimated parameter of the default rate (*avg_PD*), since we can usually read in the literature about the positive correlation between the LGD and the default rate (for example *Grunert and Weber [2005; 2009]*, *Brady et al. [2007]*, *Bellotti and Crook [2008]*), or in some cases about independency respectively (*Carey – Gordy [2003]*). However, in case of the other factors the results were not surprising.

[1] The default rate is actually the concrete realization of the PD (Probability of Default), however, I use these two terms as synonyms of each other, if it does not prevent the understanding.

According to my anticipative expectations, the *"NoFurtherRec"* category is much more homogeneous in comparison with the *"WorkoutEnd"* category, and the influencing factors of the deal level LGD can be better defined. This model proved to be much stronger indeed.

In this linear regression the client characteristics (the age of the client at the default; the landline phone) also played an important role in addition to the deal and collateral characteristics (the length of the period from the origination of the deal to the default event; whether the deal became non-performing because of delay; the region; the type of the settlement) as well as the macroeconomic factors (the growth of the real wages, of the consumer prices and of the GDP in the period from the origination of the deal to the default; the average default rate at the default date), contrary to the regression prepared for the *"WorkoutEnd"* deal category.

On the basis of the linear regression models developed for the *"WorkoutEnd"* and the *"NoFurtherRec"* deal categories, it can be said summing up that the results support the statement composed in my 7th Hypothesis according to which different factors influence the LGD values of the deals with different closing types, thus it is inappropriate to handle them together in the course of modelling the deal level LGD.

1.5. The structure of the dissertation

Following the Introduction I make known the regulation field relevant in terms of the dissertation in Chapter 2. In the course of that I give a survey of the legal frameworks concerning the credit institutions' risk management, the Basel recommendations, the system of the CRD and the most important changes of the regulation. On the one hand I notably focus on the elucidation of the terms relevant from the viewpoint of the thesis, and on the other hand I touch upon some special Hungarian aspects as well.

In Chapter 3 I deal with the theoretical models for LGD calculation, the opportunities of their application in Hungary as well as the systems for collecting data about recoveries from real estates. Considering that actually the application of the workout LGD methodology presents the only real opportunity in Hungary, so I give only a broad outline of cross tables, implicit historical LGD, market and implied market LGD methodology, emphasizing more the characterization of the workout LGD model. Concerning the systems for collecting data about recoveries from real estates I shortly make mention of the international scene, then I present the Hungarian Interbank LGD Database coordinated by the Hungarian Mortgage Association (HMA).

The theme of Chapter 4 is the econometric methodological bases of LGD calculation. For the sake of establishing my research, it is necessary to clarify some econometric concepts and to review certain proceedings, thus I outline the problems of data samples which characterize the databases serving as a basis for LGD calculation, and the possible manners of solving them, the most frequently used parameter estimating methodologies, the aspects of model selection and testing, then I briefly describe the model types which are relevant from the viewpoint of my thesis with distinguished attention to the logistic regression.

Following the treatment of the methodological questions I already focus on the empirical area in Chapter 5, in the framework of that I make known the examinations and consequences published in the literature. I outline empirical works dealing with LGD models which prepare forecasts on the basis of historical recovery and LGD data by using analytical procedures, particularly applying regression methodologies, modelling distributions. In the course of the presentation instead of emphasizing the values calculated and published by the researchers, I focus on methodological aspects such as for example specifying the factors, which influence the recovery rate, studying the distributions and the transformation procedures used in the course of modelling.

My empirical researches mean the theme of Chapter 6. In the framework of that I outline the bank database which serves as a basis for my analyses,[2] its content and the structure of the data used as well as the data deriving from the Hungarian Central Statistical Office's (HCSO) STADAT Database, then I make known the definitions and assumptions which I used and some methodological decisions, finally I present the concrete analyses and their results. During the expounding I do not focus on the calculated LGD values, but on introducing the influencing factors and the models, and valuating their performance.

As a closing Chapter 7 contains the Appendices, the figures and tables which illustrate my empirical research results.

[2] The other important data source is the Hungarian Interbank LGD Database which is presented by Chapter 3.5.

2. Regulatory background

In the past few years the more and more increasing market competition, the pressure of the owners' expectations and numerous other factors together led to the result that the financial institutes started increasingly risky activities, and took sight with their products and services at customers with higher risk level as well than earlier. These events include newer potential sources of losses, so it became essential to reform the regulation which can play in addition the role of protection for the banks from the consequences of their own "irresponsible behaviour". It is no exaggeration to state that the global crisis determines directions of the regulation.

In the past period, there occurred considerable changes in the bank regulation in five areas (*Terták [2010]*):

- increasing the severity of prudential rules,
- development of new European supervisory system and Single Rule Book,
- making up of procedures for handling the crisis and advancing highly responsible lending,
- reforming the corporate governance and the remuneration system,
- bank taxation and insurance of deposits.

In the present chapter I give a survey of the legal frameworks concerning the credit institutions' risk management, referring to the aspects in connection with the transformation of the supervisory system as well. However, I mention the general guiding principles only roughly, considering that the theme of the dissertation is actually closely credit risk specific.

2.1. The Basel recommendations and the Capital Requirements Directive (CRD)

The Basel Committee on Banking Supervision (BCBS) which was set up in 1974 as an association of 10 countries' central banks, worked out the Basel Agreement in 1988 with the aim of providing the long-run solvency, the prudential operation and the stability of the bank system. The recommendations from 1988, became famous as the name of Basel I, which has the title of *"International Convergence of Capital*

Measurement and Capital Standards" ordered a 8% capital adequacy requirement[3] (Cook-rate) uniformly referring to all institutes, but concerning only the credit risk, so the undertaking of market risk and operational risk was not regulated at all. Considering that it confronted all institutes with same conditions independently of the de facto undertaken risks as well as the individual characteristics, it was not directly conductive to working out the more developed risk management methods. On the other hand it is hardly disputable that it meant an important milestone from numerous aspects in spite of its deficiencies, since it promoted the harmonizing process of the uniform national regulations to a considerable degree, while the solvency rate became the generally used measurement of solvency all over the world within a few years.

In the following years the BCBS revised the recommendations from 1988 on the basis of the questions and problems arisen, and worked out new proposals with the aim of answering them. Even though in 1993 the *93/6/EEC directive* on the capital adequacy on investments firms and credit institutions *(EEC [1993])* put a uniform procedure for quantification the capital requirement in the future as well, but it covered a significantly broader risk range. As the next step in 1996 the Basel I was supplemented with the prescriptions concerning the quantification of capital requirement needed for covering the market risks, which have already made possible and necessary the adoption of individual risk management methods, the most fitting to the activity of the given institution.

Considering the further evolution a significant role can be put down to the fact as well that in December 2002 the Council of the European Union extended the Lámfalussy process, reforming the Union's market regulation, onto the whole European financial sector, in whose framework a four-level system was formed according to the proposal of the Committee of Wise Men led by Sándor Lámfalussy *(Soós [2011])*:

- Level I: joint decision-making process of the European Parliament and the ECOFIN (Council of Economics and Finance Ministers of the European Union.
- Level II: executive clauses given by the member states and the European Commission.
- Level III: strict supervisory cooperation.
- Level IV: provisions of the European Commission for the sake of carrying out the rules appropriately by the member states.

[3] The solvency rate is the quotient of the banks' or financial institutions' regulatory capital and their adjusted total assets, that is the proportion of their financial resources which has to be kept in reserve for covering risks and protecting the depositors.

The active contribution of the third-level boards, established in 2001 (CEBS – Committee of European Banking Supervisors; CESR – Committee of European Securities Regulators; CEIOPS – Committee of European Insurance and Occupational Pensions Supervisors) resulted in significant steps in the field of decision making regarding the financial sector and the supervisory cooperation. In the matters concerning the banking sector the CEBS gained the largest role among them, considerably contributing the reduction of the differences among the regulations, the supervisory methodologies and procedures, to approximating the practices pursued by the national authorities and to cooperation of the institutions by working out standards with the national supervisors jointly.

From this time on large momentum was given to the development of the Basel II recommendations, the proposals of BCBS renewing the capital requirement calculation of the internationally active banks, which already set the complex risk management covering a broader range of risks as an aim. The primary intention of establishing the new risk management systems is *"to approximate the economically needed capital and the regulatory capital requirement amount to each other, as well as to protect the financial sector's stability by means of among others introducing the comprehensive appraisal of credit risks" (Baranyi – Széles [2010], pp. 168.)*.

2.1.1. The adaptation of the Basel II proposals by the European Union: the CRD

In July 2004 the European Commission initiated the modification of the *2000/12/EC directive* relating to the taking up and pursuit of the business of credit institutions *(EPC [2000])*, and the *93/6/EEC directive* on the capital adequacy on investments firms and credit institutions *(EEC [1993])* by presenting an official motion for an amendment. Considering the conclusions of the study written by PricewaterhouseCoopers about the potential effects of the planned new capital requirements *(PWC [2004])*, the meeting of ECOFIN on 7th December 2004 ratified the proposal for the capital adequacy directive (Capital Requirement Directives – CRD)[4] which contains the new directions for capital requirement. Its primary target was to strengthen the global financial system and to create equal competitive conditions.

[4] The CRD is a regulation which joins the elements of CID (Credit Institutions Directive) relating to the banking book and of CAD (Capital Adequacy Directive) relating to trading book, and refers to both the credit institutions and the investment firms.

The documents, which were modified on the basis of the proposals of the European Parliament approved on 28th September 2005, were ratified by the ECOFIN in joint decision-making process on 11th October 2005 and following the juristic-linguistic discussion again on 7th June 2006.

Following numerous modifications after the ratification of the European Parliament and CEBS on 14th June 2006 the revised *2006/48/EC* and *2006/49/EC directives (EPC [2006a; 2006b])* based on the new Basel II recommendations about the capital adequacy of credit institutions, investment firms and other financial institutes were finally published in the Official Journal of the European Union. In spite of the long preparation numerous questions remained open, in which the right of decision was due to the member states or their supervisory authorities, making possible for them to form flexibly on their own authority the Union's prescriptions according to their regulatory culture, legal characteristics and market conditions. These are the so-called national discretions, which enable some preferential proceedings, promoting for example on the field of credit risk the introduction of internal based method for the certain credit institutions.

Since the full-scale use of the directives approved in 2006 became compulsory on 1st January 2008 in the financial sector, moreover the institutions which are parts of its scope could turn to the operation under the new rules already from 1st January 2007, certain countries started to implement the directives into their own national legal environment soon. However, the discussions, especially because of the financial crisis in the past period, continued intensively, as the result of which initiatives occurred for modifications relating to numerous fields.

The CRD (Capital Requirements Directives) is a regulation referring to the whole sector of credit institutions, which follows the principles of Basel II, the detailed proposals laid down by it. Its essential elements are the forward looking capital regulation, the appraisal of future risks, the advancing, increasing and spreading of the risk sensitivity of the credit institutions.

The main purpose of the prudential system is to provide the risk-based supervision, for the sake of which it is based on three pillars *(HFSA [2008b])*:

- Pillar I which means the quantification of the minimal required capital needed for supervisory adequacy, contains the uniform quantitative requirements such as for example the evaluation prescriptions, the definition of regulatory capital, the methodology of capital requirement calculation or just the rules of investments and concentration. Detailed prescriptions refer to all related fields, and derived from the circumstance that it raises the same requirements for all credit institutions (for example it requires the usage of uniform risk functions), it also makes comparison possible.

- In Pillar II, in which the target is the calculation of internal capital requirement in compliance with certain institutions' risk profile, the emphasis is laid on the individual methods chosen by the credit institutions. In its framework risks have to be considered as well, which are not at all or not properly handled by the Pillar I. Concerning the modelling techniques the institutions get quite large liberty, hereby it encourages the development and implementation of the new methods, methodologies, and advances the introduction of more developed and more efficient risk management systems. It also involves two elements which are in connection with each other in many respects:
  - It strongly emphasises the responsibility of the corporate governance and the building up of necessary functions (for example risk management, internal audit). It attaches distinguished importance to the development of the credit institutions' own risk identification, evaluation and management system (ICAAP – Internal Capital Adequacy Assessment Process), and the setting up of the regarding internal management and control rules, the support for the supervisory revision procedures. It prescribes that the strategies and the evaluation processes have to be subjected to regular revisions.
  - It draws up a standardized process for the sake of providing the uniform supervisory practice (SREP – Supervisory Review and Evaluation Process). It puts down for example the supervisory competences, the applicable quantitative and qualitative tools, puts into words the principle of proportionality[5] as well as the preventive and the corrective supervisory steps regarding to the break of the capital requirement level. The implementation of all these requires continuous connection and discussion between the Supervisor and the institutions.
- Pillar III, however, which really has gained attention only recently and is based on the *Government Decree No. 234/2007*, plays the largest part in the wording

[5] The principle of proportionality declares that the frequency and intensity of the supervisions have to consider the size, the importance of the organization, as well as the subtlety and complexity of the business activity of the credit institutions *(HFSA [2008c])*.

of publishing and reporting requirements, hereby advancing the appropriate information giving for the public, the increasing of transparency and the success of the market's control power.

The regulatory capital has to cover the lending, market and operational risks quantified in Pillar I as well as the capital requirement assessed in Pillar II based on the credit institutions' own risk disclosure, and the capital buffer has to be reserved as the result of the stress-tests and the SREP. In the course of capital adequacy calculation the possibility of extreme risks has to be taken into consideration as well.

Considering that from the viewpoint of recent dissertation primarily the prescriptions of Pillar I possesses relevance, I concentrate in the following subsection on the demonstration of these elements.

2.1.2. Applicable methods for calculating the capital requirement under Pillar I

Under Pillar I the capital requirement defined by the CRD is derived from the sum of the credit risk's, the market risk's and the operation risk's capital requirement, but the certain components can be quantified with using different methodologies. The CRD enables for the institutions to change certain elements of the model prescribed by the regulators to self-developed submodels, on condition that those have to be subjected to strict authorization processes and systematic revisions in all cases.

- Regarding the credit risk a standard (SA – Standardised Approach) and an internal rating based model (IRB – Internal Ratings Based) can be applied, and within the latter – excluding the retail sector – the regulation differentiates an "elementary IRB method" (F-IRB – Foundation IRB) and a "developed IRB method" approach (A-IRB – Advanced IRB).[6] For the retail exposures the credit institutions can choose between the SA and the A-IRB.
- The quantification of the market risk's (position risk's) capital requirement can occur according to a standard model or an internal model. The latter is typically VaR-based modular calculation.
- The broadest scale of alternatives is in the field of operational risk, namely the credit institutions can choose from three standard methods (BIA – Basic Indicator Approach, TSA – Standardised Approach, ASA – Alternative Standard

[6] In case of the Foundation IRB method only the PD of the credit risk parameters is based on own estimation, however the credit institution applying the Advanced IRB method appraises all risk parameters itself.

Approach) and an internal model approach (AMA – Advanced Measurement Approach).

Generally speaking the internal rating based models are able to consider the measure of risk undertaking better, hereby these enable to quantify realer capital requirement, it is apparent however, that the implementation and application of them claim remarkable resources for the institutions.

In case of internal models regarding the selection of methods, assumptions and the calculation of correlation the CRD provides liberty at different degree on the certain risk fields: on the one hand it offers a quite comprehensive modelling freedom for the operational and market risks, but on the other hand in the case of credit risk it allows only the risk parameters' self-made estimation, it does not provide possibility for using own proper default or credit risk models for the credit institutions.

It prescribes simple addition of the capital requirements calculated for the certain risk categories, it does not permit to consider the effects of diversification among the risk categories (correlation = 1), furthermore referring to that it does not allow for the credit institutions to apply their own individual methods and estimates based on them. Whereas the own calculations are generally permitted relating to the correlation within the same risk category, but the credit risk of banking book makes also an exception to this rule, in case of which this option is not even allowed.

2.1.3. Hungarian regulation: the implementation of CRD, the laws concerned

The 8% capital adequacy measure and the prescription of limitation of the undertakable risks appeared for the first time in Hungary in the *Act LXIX of 1991* on the monetary institutions. The *4/1993 (PK 17.) Disposal* of the Hungarian Banking Supervision meant the next step, in which it was put down that all monetary institutions are obliged to create and apply bylaws for lending, rating of obligors, investment, evaluation of collaterals, rating and provisioning.

It is hardly disputable that in Hungarian relation these regulations laid down the fundamentals for the more intensive improvement of the banks' risk management systems. Later on numerous modifications and developments became necessary, but it did not come to radical changes until the Basel II recommendations as well as the working out and implementation of the EU directives which are formed on the basis of the former ones.

From 2005 the Hungarian Financial Supervisory Authority (HFSA) has published numerous conceptual documents, on the basis of which widespread consultations happened with involving the institutions concerned. Finally the implementation of the regulations associated directly to the CRD into the national laws and orders was carried out in the second part of 2007. The modifications of the *Act CXII of 1996 on Credit Institutions and Financial Enterprises (Hpt.)* and the *Act CXX of 2001 on the Capital Market (Tpt.)*, which involved the implementation of the prescriptions concerning the Basel II regulation as well, came into force on 1st January 2008. Among others the *Act CXXIV of 1999 on the Hungarian Financial Supervisory Authority*, the *Act IV of 1959 on the Civil Code of the Republic of Hungary* and the *XXXVIII Act of 1992 on Public Finances* were altered. At the same time in the course of implementing the CRD into the national laws and orders it was an important aspect that only the most needed rules appear on the level of acts, since thus the carrying over of possible necessary modifications could be more flexible and faster afterward.

The under-mentioned regulations contain the fundamental prescriptions regarding the credit institutions' risk management:[7]

- *Government Decree No. 196/2007 on the Management and Capital Requirement of Credit Risk (Hkr.)*,
- *Government Decree No. 200/2007 on the Management and Capital Requirement of Operational Risk (Mkr.)*,
- *Government Decree No. 244/2000 on the Rules for Specifying the Capital Requirement Necessary as Collateral for Trading Book Positions and Risks and Currency Exchange Rate Risks and on the Detailed Rules for Maintaining a Trading Book (Kkr.)*, and
- *Government Decree No. 381/2007 on the Management of Credit Institution Counterparty Risk.*

In addition to harmonizing the listed regulations numerous further regulatory elements on the level of acts and decrees, as well as modifications became necessary by the national implementation of the CRD and its subsequent amendments. Supplementary

[7] I demonstrate only the aspects concerning the credit institutions and financial enterprises, but I note that the CRD include rules relating to the investment firms' activity as well.

decree pertains for example to the fulfilment of the credit institutions' publicizing requirement *(Government Decree No. 234/2007)*, to the capital requirement of securitization *(Government Decree No. 380/2007)* as well as several relating fields. The *Decree No. 13/2001 of the Minister of Finance on the reporting obligations of financial enterprises* needed to be altered as well.

For the sake of providing the availability of the information needed for the capital requirement calculation using the internal rating based method, the achievability of the minimum requirements and the legal enforceability of the instruments which can be accepted to mitigate the credit risk, further alterations became necessary as well, so modifications have been carried out on the *Act LXIII of 1992 on the Protection of Personal Data and Release of Data of Public Interest*, on the *Act XLIX of 1991 on the Bankruptcy Proceedings and Liquidation Proceedings*, on the *Decree No. 14/2001 of the Minister of Finance on the classification and evaluation criteria for outstanding receivables, investments, off-balance sheet items and collaterals*, as well as on the *Decree No. 45/2008 of the Minister of Finance on the Scope of Data to be Reported by Credit Institutions to the Hungarian Financial Supervisory Authority and the Manner of Reporting*.

The modification process of the regulation happened in the framework of comprehensive discussion, and primarily founded on the rules and proceedings involved in the EU Directives. It took the proposals of the Basel Committee as a basis concerning only those questions, in case of which the prescriptions of the Directive did not prove to be detailed enough. For the sake of the proper transparency the HFSA made public the most important information relating to its pursued practice, with distinguished attention to the validation procedures as well as the principles and methods applied during the supervisory process (for example: *HFSA [2005; 2008a; 2008b; 2008c; 2009; 2010]*).

The system of discretions is a considerably unusual element in the Hungarian legal system, therefore the majority of these questions also was solved in the framework of laws, moreover the number of the supervisory discretions is firmly low in internationally comparison as well.

2.2. The most important changes of the regulation

It was clearly proved in the past years that the regulation was not able to keep step with the dynamic growth of the evolution of the financial markets, with the deepening of the financial integrity, thus notable attention has to be paid to the harmonization of the regulatory environment. The crisis broken out in the autumn of 2008 worsened the situation more and added new problems to it, emphasizing these questions even more. The present subsection deals with the most important changes of the last period.

2.2.1. Alterations of the Basel recommendations and the CRD

In the past years it was unambiguously proved that the regulation was not able to keep up with the financial markets' dynamic evolution, with the deepening of the financial integrity, so the harmonization of the regulatory environment has to get distinguished attention. The crisis started in the autumn of 2008 continued to worsen the settled situation and added subsequent problems, emphasizing this scope of questions even more.

The accumulated observations shed light upon the necessity of a more robust and uniform, more harmonized prudential capital requirement system for the sake of consolidating the global financial system, the preparation for the "coming crises" as well as the providing of equal competitive conditions.

It can be mentioned as an important step that in October 2008 the European Commission entrusted the high-level independent group led by Jacques de Larosière with the task to make suggestions for strengthening the supervision of the European financial institutions and markets as well as the financial stability. As the result of the group's active work, the comprehensive proposal package relating to the new financial supervisory structure and the cooperation of the authorities, the *De Larosière Report* reached completion on February 2009 *(Soós [2011])*.

Since the potential differences deriving from the discretional decisions make the supervision on a consolidated basis of the global financial groups extraordinarily more difficult, occasionally they even threaten the harmony of application on the internal market, moreover in extreme cases could even lead to regulatory arbitrage, so comprehensive consultations have happened several times since the CRD was put into force, which are destined to balance between the proper flexibility and the uniform application of rules as well. In the past period under the coordination of CEBS there occurred successful steps in order to enhance the supervisory convergence: the number of national discretions, options, exceptions and derogation as well as the latitude of the member states significantly decreased *(Kardosné [2010])*.

Since the approval of the capital adequacy directives in 2006, based on the Basel principals, numerous problematic areas have been identified and interpretative questions have arisen during the practical application, furthermore not least in order to treat the imperfections which became clear as the consequence of the financial crisis, remarkable altering initiatives occurred, some of which have been already approved, moreover have been adapted into the national laws and orders as well *(HFSA [2010])*:

- The CRD I *(EC [2009a; 2009b])* was formed based on practical observations. The Capital Requirements Transposition Group (CRDTG), the professional workgroup founded by the European Commission, prepared it, then the European Commission approved it under comitology procedure, without the contribution of the Council and of the Parliament. During its preparation the European Commission initiated public consultations as well, finally based on the proposals which involved the gap-fillings, refinements and modifications relating to the technical questions of the capital requirement calculation the CRD took effect on 1st January 2011, and the national authorities had to implement the new disposals into their regulations and make them public by 31st October 2010.
- From the same time has to be applied the CRD II, which was approved by the European Parliament and the Council in May 2009 based on the proposals of the Commission in October 2008 *(EPC [2009])*, then it was published in the Official Journal of the European Union in October 2009. This alteration concerns the operational frameworks of the supervisors' cooperation (the activity and the division of the labour of the so-called supervisory colleges), standardizes the reporting requirements prescribed for the supervised institutions as well as the regulation concerning the regulatory capital (the criteria for acceptability of the core capital elements, the categorization of hybrid capital elements) and the high-risk undertaking, furthermore founded on the observations from the financial crisis enters restrictions for the institutions' liquidity requirements as well as for the risk management and capital rules concerning the securitization.
- In the first half of 2009 the European Commission published more recent consultative documents, then based on the discussions the CRD III *(EPC [2010a])* has been approved in July 2009 by the Commission, and came out in the Official Journal of the European Union on 14th December 2010. This time the modifications referred mainly to strengthening the capital requirement of the

trading book's items and the complex securitization (respectively re-securitization) positions (for example setting-in stress-conditions), to refinement of the SREP, and on the basis of the proposals of the *De Larosière Report (De Larosière [2009])* the alteration of the remuneration system in line with the powerful and effective risk management, and they are in force also from 1st January 2011.

As a consequence of the fact that the new Treaty of the European Union, the Treaty of Lisbon which came into force on 1st January 2009 rearranged the legislative competences in the Union *(Szájer [2010])*, the further modifications after the CRD I were not approved under comitology procedure any more, but these are pertaining to joint decision-making process of the Council and of the Parliament which expanded and expands the time-consumption of the approval procedure.

In the past period also the European financial supervisory system went through strong changes. Until the autumn of 2009 the European Commission carried out the draft regulations which were establishing the foundation of the new European financial supervisory structure proposed by the *De Larosière Report*, and these were finally approved by the European Council, the Commission and the Parliament in December 2010, so a new two-pillar supervisory structure started its operation in January 2011 which contains a separate macro- and micro-prudential subsystem. The purpose of the transformation is to advance the further integration of the European financial markets, as well as to improve the institutions' operational environment and competitiveness *(Soós [2011])*.

The consideration in the background of the newer and newer alterations of the CRD is that according to the opinion of the heads of states and prime ministers of the G20 Group, which involves the world's most considerable developed and developing countries, the sector's vulnerability has to be decreased for the sake of preventing the coming financial crises, and with the aim of emphasising this target they also put particular pledges into words on the summit meetings in London and Pittsburgh in 2009.

In the past period the rethinking of the Basel recommendations appeared on the agenda as well, paying great attention to the possibilities of remedy of the market, supervisory and regulatory failures which played role in evolving and deepening the crises. After comprehensive discussion on the basis of the consultative matters *(BCBS [2009b;*

2009c; 2009d; 2009e; 2009f]), the Basel III, the reform of the Basel recommendations, was finally approved in November 2010. The proposals refer on the one hand to enhancing the resistance of the individual institutions against the coming financial and economic stress situations, and on the other hand to treating the effects of the infections, the systemic risks and the procyclicality. Considering that there is considerably long time left until its coming into force on 1st January 2013, the evaluation of its potential affects is difficult for the time being. However, having considered the results of the finished quantitative impact studies (QISs – Quantitative Impact Studies)[8] as well, the Basel Commission approved a gradual implementation plan relating certain fields which lasts until 2019 so that the provisions keep back the financial sector's activity as little as possible *(Szombati [2010])*.

The European Commission is going to carry the relevant reforms included in the Basel III into effect in the framework of CRD IV, keeping in view the object of establishing and retaining the competition under equal terms regarding the regulatory environment, but numerous supplements, modifications and simplifications can be noticed, with respect to the characteristics of the European market. The European Union extends the effect of the capital rules included in the Directive to all credit institutions and investment firms that give reasons even in itself to certain corrections, as the Basel recommendations contain actually the principles relating to the internationally active large institutions.

For the sake of establishing the proposals directing towards the modification of the capital adequacy directive, the Commission initiated professional discussions during 2009 and 2010 *(EC [2009c; 2010a; 2010b; 2010c])*, having outlined the directions of the possible alterations for the institutions involved. The European Commission published the particular legislative drafts also harmonized with the Basel recommendations, named CRD IV, as well as the linking impact studies on 20th July 2011. The motion contains a rule package involving a Directive *(EC [2011c])* and a Regulation *(EC [2011d])* which take over the preceding *Directives 2006/48/EC* and *2006/49/EC*, moreover both the drafts are joined with impact study document each *(EC [2011e; 2011f])* as well, which are destined for supporting the reduction of the probability of evolving the systemic bank crisis.

[8] Over and above that the Macroeconomic Assessment Group *(MAG [2010])* which was established jointly by the Financial Stability Board (FSB) and the BCBS examined in details the temporary potential effects, and the Long-term Economic Impact *(LEI [2010])* workgroup of the BCBS the long-term consequences.

The proposal for the Regulation published by the Commission *(EC [2011c])* is dealing with the reregulation of the regulatory capital from quantitative and qualitative view as well, with prescribing uniform liquidity requirements, with introduction of a leverage limit, with strengthening the treatment of partner risk and accomplishment of the Single Rule Book, in addition to the general prudential requirements relating to the credit institutions and the investment firms. In contrast with this, the Directive *(EC [2011d])* treats those questions, concerning which the consideration of the characteristics of the member states' individual regulatory environment is essential. Here took place among others the aggravation of the requirements relating to the corporate governance systems and proceeding as well as the prescription of reserving the capital buffer.[9]

2.2.2. The regulation relating to the mortgage lending

Considering the mortgage lending the *77/780/EEC Directive on the taking up and pursuit of the business of credit institutions* (Coordination Directive I) *(EEC [1977])* can be mentioned as the first important regulatory item, which came out in 1977. This was followed by the *89/646/EEC Directive* (Coordination Directive II) *(EEC [1989])* in 1989, which enabled, for example, for a credit institution operating in one of the member states to give a citizen in another member state a loan secured by property.
Later, because of the dynamic growth of the cross-border lending, the modernization of the regulation has become necessary, which was carried out basically by the *2000/12/EC Directive (EPC [2000])*, then by the CRD. In the latter one the principle of reciprocal recognition *(EPC [2006a] Article 23)* was already drawn up, according to which the member states ensure that the activities listed in Annex I of the *2006/48/EC Directive* may be carried out within their territories by the establishment of a branch or by way of the provision of services.
For the sake of improving the market efficiency and the competitiveness, namely the achievement of the Lisbon Strategy, the European Commission make efforts to integrate the market of the retail financial services, that is why it handed out the *Green Paper (EC [2005])* in 2005 with the intent to appraise the potential significance of the interventions into the mortgage markets of the European Union. It was dealing with the questions of customer protection and law, as well as the aspects of the mortgage collaterals and the financing in this document, but has not made known any concrete

[9] A more detailed review of the elements of the CRD can be read for example in the study of *Tajti [2011]*.

steps and measures yet. Following that, as a result of the wide-spread discussion and detailed impact studies which occurred on the basis of the *Green Paper*, the *White Paper (EC [2007])* was published on the Integration of European Union Mortgage Credit Markets on 18th December 2007, which summarized the results of the review relating to the mortgage markets of European Union, and dealt with the problematical areas of direct relevance to responsible lending and borrowing (for example pre-contractual information, assessment of creditworthiness, early repayment and credit intermediation). In addition to that, it also reported a balanced package of measures in order to enhance the efficiency, integration and competitiveness of the market, and dealt with the potential barrier factors of the integrity. The under-mentioned objectives appeared in the *White Paper* as the most important elements:

- advancing the cross-border mortgage lending and financing,
- enlarging the product range,
- promoting consumer confidence, and
- encouraging the mobility of the customers.

For the sake of achieving them, the *White Paper* already drew up certain steps as well, which were specified in the *2008/48/EC Directive (EPC [2008*]) *on the credit agreements for consumers* for the first time *(Bodzási [2010])*. This covers in principle only customer credit loans from EUR 200 to EUR 75 000, but numerous member states apply it to mortgage credits as well.

So the coming into the limelight of the principle of responsible lending concerns also the mortgage lending considerably. It is worth mentioning as an important circumstance the *European Voluntary Code of Conduct on Pre-contractual Information for Home Loans* of March 2001 as well, which dealt with the pre-contractual information relating to the mortgage credit loans and with the European Standardised Information Sheet. The Code was endorsed by the Commission in the *Recommendation 2001/193/EC of 1st March 2001 on pre-contractual information to be given to consumers by lenders offering home loans (EC [2001])*, but its implementation was inconsistent, thus it did not fulfil the expectations.

The European Commission published *the proposal for a directive on credit agreements relating to residential property (EC [2011a])* on 31st March 2011, prescribing considerable aggravations in reference to the mortgage lending. On the one hand it draws on the provisions of the *2008/48/EC Directive (EPC [2008*]) *on the credit* *agreements for consumers*, but on the other hand considers the specific features of the mortgage credits.

The proposal discusses the notion of responsibility in dual approach, in its interpretation the responsible lending is the *"care taken by creditors and intermediaries to lend amounts that consumers can afford'*, and the responsible borrowing means that *'consumers provide relevant, complete and accurate information on their financial conditions, and make informed and sustainable borrowing decisions" (EC [2011b] pp. 5.).*

Basically the mortgage credits to consumers, together with the prudential and supervisory requirements for creditors and credit intermediaries are its subjects, and its scope contains the following deals *(EC [2011a] Article 2 (1))*:[10]

a) *Credit agreements which are secured either by a mortgage or by another comparable security commonly used in a Member State on residential immovable property or secured by a right related to residential immovable property.*
b) *Credit agreements whose purpose is to acquire or retain property rights in land or in an existing or projected residential building.*
c) *Credit agreements whose purpose is the renovation of the residential immovable property, a person owns or aims to acquire, which are not covered by Directive 2008/48/EC of the European Parliament and of the Council of 23rd April 2008.*

The proposal stipulates strict conditions (for example appropriate professional knowledge) for both creditors and credit intermediaries *(EC [2011a] Article 5-6)*, prescribes the general information obligation and providing the personalized information on the basis of the European Standardized Information Sheet *(EC [2011a] Article 9)*, and give orders relating to information to be delivered to the customer in case of changing the borrowing rate *(EC [2011a] Article 13)*. On the one hand it introduces general principles for marketing and advertising communications *(EC [2011a] Article 7-8)*, and on the other hand it prescribes strictly the use of annual percentage rate of charge (APRC) *(EC [2011a] Article 12)*. With reference to early repayment of the loans, it also puts down that the right of the customers has to be ensured to repay the credit before the expiry of the credit agreement *(EC [2011a] Article 18)*.

It prescribes for the creditors that the customer's ability to repay the credit has to be checked based on sufficient information and taking into account the personal

[10] However the member states are allowed to apply it also with a broader scope, for example they can extend it to the commercial property transactions.

circumstances of the customers, and the granting of credit has to be refused in case of negative result *(EC [2011a] Article 14)*. With reference to that, it also gives orders that the creditors can access information from relevant databases on a non-discriminatory basis *(EC [2011a] Article 16)*.

As antecedents for the proposal, the European Commission held lengthy and detailed consultations with the stakeholders, moreover the European Parliament and the European Economic and Social Committee have adopted numerous reports. The relating expertises and studies had been considered, and impact assessments had been also prepared before carrying out the concrete proposals. Lifting the barriers to the cross-border mortgage lending (for example by carrying out the European passport) is an important objective of the Commission, however, it is likely that several financial institutions will respond to the aggravations of the control conditions with reducing their activity. The background of aggravating the regulation is that in the view of the Commission the irresponsible lending and borrowing considerably contributed to evolving the conditions which led to the current financial crisis, and its repetition can be prevented only by encouraging the financial stability *(Kardosné [2010])*.

It is remarkable that considering the changes of the regulation more and more constructive elements can be experienced in the last period: a considerable part of the modifications refers actually not any more to the follow-up treatment of the crisis, but rather to the development, the advancing and necessarily the preparation for the "coming crises".

2.3. Prescriptions of the regulation field in the European Union relevant in terms of the dissertation

Concerning the Hungarian credit institutions the *Government Decree No. 196/2007 (Hkr.)* specifies basically the terms of calculating the credit risk's capital requirement according to the standard method *(Hpt. 76/A. §)* and the internal rating based method *(Hpt. 76/B-D. §)* as well as the application of the internal rating system. However, considering that the calculation of LGD is not a relevant question in case of the standard method, hereinafter I concentrate basically on the rules relating to the internal rating based method.

The credit institutions, which apply IRB method, are obliged to assign each exposure to one of the under-mentioned exposure classes according to the logic of the IRB approach *(EC [2011c] Article 142 (2); Hkr. 24-29. §)*:

- exposures to central governments or central banks,
- exposures to credit institutions and investment firms,
- exposures to corporates,
- retail exposures,
- equity exposures,
- securitisation positions, and
- other exposures non credit-obligation assets.

Different risk functions (capital functions) apply to each of the exposure classes in accordance with certain exposure class's risk features, so that it becomes possible to quantify the risk weighted exposure value and the relating capital requirement in a more sophisticated manner.

In the framework of the present dissertation I demonstrate the regulation connected to lending to the retail sector, within it the aspects of those deals' credit risk which are secured by residential property collateral.[11] I mention neither the prescriptions relating to the exposures of other sectors, nor the other activities of credit institutions, for example the themes of financial leasing services or the purchased claims.

In case of internal rating based method *"An exposure belongs to the retail exposure class, if*

> *a) it exists to natural person, micro, small or medium enterprise,*
>
> *b) the exposure can be classed into a group of a significant number of exposures which can be described with similar features and are treated consistently and in a similar manner in risk management, thus enabling the mitigation of risk relating to lending,*
>
> *c) in case of micro, small or medium enterprise the total amount of the debts of the obligor client or group of clients connected to each other to the institution and its undertakings and subsidiaries – including all past due exposures, but excluding claims secured by residential property collateral – do not exceed, to the knowledge of the credit institution, EUR 1 million or any equivalent amount, and the credit institution takes reasonable steps to get the information needed, and*
>
> *d) the exposure is not managed just as individually as exposure in the corporate exposure class" (Hkr. 27. § (1) Paragraph).*

[11] The detailed terms of the collaterals' acceptability, the prescribed minimum requirements and the calculation methodology of the effect of the credit risk mitigation are included in Chapter XIV-XVIII of Hkr.

This definition implicitly includes the acknowledgement of the risk mitigating effect of the diversified portfolio, while serves as a basis for establishing subclasses within the retail exposure category. Moreover the concerning articles of CRD and the *31. § (1)-(6) Paragraphs of Hkr.* contain even particular prescription regarding that all of the retail exposures have to be assigned to one of the under-mentioned subclasses:

i. retail exposures secured by residential property collateral,
ii. revolving retail exposures, and
iii. other exposures to retail sector.

The accuracy of the separation has great importance particularly in that case, if the credit institution uses different rating systems for certain subportfolios *(Hkr. 54. § (2) Paragraph).*

The Hungarian regulation differently from the Basel recommendations assigns to the (i) category not only the deals of the retail exposure class (including the small and medium enterprises as well) secured by residential property collateral, but those which are secured by commercial property collateral as well, provided that they meet all other requirements which are necessary to be classified into the retail exposure category. However, in the framework of the present dissertation I do not extend the analysis to these deals, concerning that the Hungarian Interbank LGD Database being presented in Chapter 3.5 does not include the necessary data.

In case of the credit products examined by me, in contrast with for example the credit cards, there can not be any amounts not drawn up from the limit which is available for the client, thus the on-balance exposure can be quantified as the gross value before forming impairment losses and risk provision.[12] Hereinafter I use the term “exposure” in this interpretation all the time last.

2.3.1. The definition of non-performing (default)

In connection with the retail exposures the *68. § (1) Paragraph of Hkr.* allows the identification of non-performing[13] on deal level, but it draws up the prescriptions on client level: it assigns a client as non-performing, if according to the information available for the credit institution, he/she is unlikely to fulfil his/her obligations to one of the members of the bank group, or the delay of its material loan obligations exists for more than 90 days or 3 months continuously. The existence of any listed conditions,

[12] Otherwise the correction with the so-called conversion factor (CCF) would be necessary.
[13] I use the term default in the paper as the synonym for the term non-performing.

independently from the other conditions, causes the getting into non-performing status, but this is not a "final" qualification, since if the existence of the conditions of non-performing comes to an end, the client gets into normal status again *(Hkr. 69. § (3) Paragraph).*

Concerning the term "unlikeliness to pay" the latitude of the institutions is relatively wide, as only the most important signs are listed in itemized manner in the Hkr.:

"There is a notice of the unlikeliness to fulfil the obligation, if

a) the credit institution has an interest claim on the client for at least ninety days or three months continuously,

b) the credit institution forms impairment or risk provision because of significant perceived deterioration in credit quality after the starting of exposure,

c) the credit institution sells the credit obligation with a material economic loss relating to the loan,

d) the credit institution agrees to distressed restructuring of the credit obligation because of financial difficulties, which results in forgiveness of principal, interest or fees, or in reduced financial obligation because of deferred payments,

e) there is a bankruptcy or liquidation process against the client under way, or

f) the credit institution initiated a liquidation proceeding against the client" (Hkr. 59. § (1) Paragraph).

The credit institutions have the opportunity to formulate further considerations typical of the client, deal or even the market as well, provided that they are able to support them with proper arguments, analysis.

I consider it as necessary to emphasise that, though according to the CRD in case of the internal rating based model the accrued interest means non-performing the deal, and the interest has to be accrued already after a 30-day delay in accordance with the 17. § of the *Government Decree No. 250/2000 on the Characteristics of the Annual Reporting and Bookkeeping Obligations of Credit Institutions and Financial Enterprises*, neither this in itself, nor the forming of impairment and provision does not cause the deals to get into non-performing status. Since the legal and the accounting concept of the non-performing can not be substituted for each other directly, so increased attention has to be paid on this during the calculation of the risk parameters.

On the other hand it also influences considerably the results of the parameter estimation, whether the credit institution takes the opportunity of cross-default, namely transmits the getting into non-performing status of the loan to the other deals of the client as well.

The establishment of the concept of materiality promotes the elimination of numerous technical default events, since therefore the delay of "insignificant amounts" does not result automatically in getting into non-performing status.
So concerning the retail exposures the delayed amount has to be basically considered significant, if it exceeds the lowest monthly minimum wage effective at the time of becoming delayed (absolute threshold) or 2% of the total obligations of the client or one monthly repayment instalment (relative threshold).
On the other hand it is true in this respect as well that the credit institutions are also allowed to use criteria for the materiality thresholds or just for the duration of delay which are different from the prescriptions *(Hkr. 68. § (5)-(7) Paragraph)*, if they are able to justify its necessity, reasonability. In such cases they are obliged to provide the opportunity of comparison with proper corrections, since for example the definition of the materiality threshold can make just a considerable effect on calculated values of the risk parameters.

2.3.2. Risk weights in the internal rating based method

The credit risk parameters play a cardinal role in quantifying the capital requirement of the institutions, which apply the internal rating based method, as the calculation of the risk weights are actually based on them, moreover the value of risk weighted assets (RWA) is the product of the risk weight and the exposure for the retail portfolio *(EC [2011c] Article 149 (1); Hkr. 31. § (1) Paragraph)*:

$$\mathrm{RWA} = \mathrm{RW} * \mathrm{EAD} \tag{2.1}$$

where: RW: Risk Weight,
EAD: Exposure At Default.

In contrast with the remarkably simplified prescriptions of the standard method, the internal rating based method does not define particular risk weights, but specify only the formula needed for the calculation for the credit institutions, into which they have to replace their own credit risk parameters.
The weights of the retail exposures are quantified according to the following formula *(EC [2011c] Article 149 (1); Hkr. 31. § (1) Paragraph)*:

$$\begin{aligned} &\textit{Risk weight} = \\ &= \left\{ LGD * N\left[\frac{1}{(1-R)^{0,5}} * G(PD) + \left(\frac{R}{1-R} \right)^{0,5} * G(0{,}999) \right] - LGD * PD \right\} * 12{,}5 * 1{,}06 \end{aligned} \tag{2.2}$$

where: N(x): the cumulative distribution function for a standard normal variable,

G(x): the cumulative distribution function for a standard normal variable.

In case of the retail exposures secured by residential property collateral a coefficient of correlation R of 0.15 and in case of the revolving asset class an R of 0.04 has to be used, while in turn for the retail exposures also this is a function of the credit risk parameters *(EC [2011c] Article 149 (1); Hkr. 31. § (1) Paragraph)*:

$$R = 0{,}03 * \frac{1 - e^{-35*PD}}{1 - e^{-35}} + 0{,}16 * \left(1 - \frac{1 - e^{-35*PD}}{1 - e^{-35}}\right) \tag{2.3}$$

Then after the computation of these, the expected loss (EL) can be calculated quite simply as the product of the probability of default (PD), the loss given default rate (LGD) and the exposure value (EAD):

$$EL = PD * LGD * EAD \tag{2.4}$$

The regulation prescribes a special formula, which grounds for several uncertainties, for the estimation of expected loss and risk weight of defaulted exposures (PD=1) *(EC [2011c] Article 149 (1); Hkr. 31. § (1) Paragraph)*:

$$\textit{Risk weight} = \max\left[0; 12{,}5 * \left(LGD - EL_{BE}\right)\right] \tag{2.5}$$

where: EL_{BE} is the credit institution's best estimate of the expected loss deriving from the exposure already in default.

The uncertainty is caused by the manner of the quantification of EL_{BE}, since the Hkr. puts down in connection with that only that *"For establishing the Loss Given Default for the exposures already in default the credit institution take as a basis the sum of the best estimate of expected loss given economic circumstances and exposure status and the additional unexpected losses which are possible to occur during the workout period" (Hkr. 74. § (8) Paragraph, EC [2011c] Article 177 (1) h).*

However, in spite of all difficulties, we can say that the quantification of the risk weights of the internal rating based method is founded on the credit risk parameters estimated by the credit institution, and because of this the different risk levels of the certain exposure classes prevail in it, so this methodology encourages the institutions to significantly stronger risk-consciousness in comparison with the standard method.

2.3.3. Minimum requirements of the internal rating based method

The CRD involves detailed rules among others for the conditions of using the internal models, the statistical and data quality norms, the evaluation, risk management and documentation procedures, the structure of the rating system, the criteria which have to be used for rating, the assignment to grades of the exposures and the integrity of assignment system into the risk management processes.

In case of using the internal rating based method the Supervisor can allow the credit institutions the self-made quantifying of risk parameters, if they comply with a reasonably extensive requirement system. Present subsection presents its most important element.

(a) General requirements to the rating system

The rating system serving as a basis for the internal rating based method has to provide measurement of the credit risk, classifying and assigning the exposures to pools as well as quantification of the credit risk parameters belonging to them *(EC [2011c] Article 166; Hkr. 54. § (1) Paragraph).*

The rating systems are playing role in quantification of the portfolio's risk parameters, which can be carried out in two ways basically:

- One of the alternatives is that at first the exposures are classified to grades or pools, then the risk parameters are assigned to them one by one in the framework of calibration.
- The other alternative is the direct estimation, in the course of which the risk parameters are appraised in a sole step from the exposure's characteristics.

Following the making up and establishment of the system the credit institutions systematically have to review both the proceeding and the terms of assigning to deal grades or pools from the point of view, whether the system applied by them is still suitable for describing their portfolio. Similarly, at least annually it is obliged to refresh the assignment of clients and deals, as well as the loss features of the risk pools *(EC [2011c] Article 169 (2); Hkr. 54. § (3) Paragraph; Hkr. 62. § (1), (3) Paragraph).*

The classification to grades and pools has to be based on assignment criteria, but the institutions have a relatively large liberty in defining them, because both the CRD and the Hkr. contain only very general prescriptions concerning them *(EC [2011c] Article 167 (1)-(2); Hkr. 56. § (1) and 58. § (1)-(4) Paragraph).*

An overall requirement for assignment to pool is that the concentration should not be disproportionately high. The categories have to be defined and the number of categories has to be appointed in a way, which provides the assignment of homogenous exposures to the same pools, but the numbers of exposures in the certain pools should be sufficient to allow reliable quantification of the risks, enabling the exact and consistent quantification of loss characteristics at grade or pool level *(EC [2011c] Article 166; Hkr. 57. § (1) Paragraph).* So the regulation prescribes the credit institutions to choose the "golden mean".

There is an expectation for the criteria, procedures and methods of assigning to pools to enable the obvious and consistent categorization of exposures within the rating system, and be in line with the credit institution's internal by-laws at the same time as well.

Although the Hkr. does not forbid the proper decision-makers to override the assignment of certain exposures, but it prescribes for the credit institution *"to analyse the performance of the exposures whose assignments have been overridden" (EC [2011c] Article 168 (3); Hkr. 61. § (2) Paragraph).* It says nothing about its manner, labels the procedure which is to be applied into the credit institution's competency.

(b) Statistical models

To mitigate the effects of subjectivity, usage of some reliable model or other mechanic method during the client and deal rating proceeding is a preferred procedure, but there are equally strict prescriptions concerning development, application and regular review of them.

It is obligatory to document the models' *"methodology, namely the theory, assumptions and mathematical empirical data of the assignment of estimates to grades, individual clients, exposures or pools (sets), the detailed outline of the data sources applied for the model, as well as those circumstances under which the model does not work effectively" (Hkr. 64. § (3) Paragraph; EC [2011c] Article 171 (4)).*

It is a basic requirement that the model's development has to be done on the basis of representative data, and later on both the continuous follow-up of the input data's accuracy, completeness and appropriateness, and the model's regular review, checking, correction (if necessary) and improvement has to be provided. It is an elementary requirement as well that the models have to possess good predictive power.

Using rating models obtained from a third-party vendor is also allowed, but also in this case the credit institution is responsible for satisfying the requirements.

Concerning the statistical models comprehensive supervision has to be made at least annually, which has to include the monitoring of the predictive power, the freeness of distortion and the stability, the review of specifications, the comparison of predicted and real realized results (Back Testing). For the objectivity and exploration of the model's deficiencies the requirement of a review by professional evaluation is a further prescription *(EC [2011c] Article 170; Hkr. 63. §).*

With a view to assess the capital adequacy the Hkr. prescribes the use of credit risk stress-tests in addition to the regular control of the rating system's appropriateness, and requires that the credit institution evaluates the re-ratings which are necessary because of the stress-test conditions. Concerning these, however, it puts down only general prescriptions *(EC [2011c] Article 173; Hkr. 67. §)*, so the power of the regulation actually manifests itself in the manner that it attaches the opportunity of using the proceedings to the Supervisor's approval.

(c) Risk parameters in the internal rating based method

The credit institutions have to use both the long-term historical experience and the empirical data, information for estimating the risk parameters[14] *(EC [2011c] Article 175 (1); Hkr. 70. § (1)-(3) Paragraph)*, but they can not rely on these solely during the calculation, moreover they have to respect the expectations as well. There is a further prescription that the review of the calculations and the so estimated risk parameters are required at least annually or more often if new considerable information becomes available.

The credit institutions are allowed to use different risk parameters for the quantification of capital requirement and for internal purposes only in that case, if they can justify with documents and calculations the reasonability of them *(EC [2011c] Article 175 (1); Hkr. 70. § (8) Paragraph).*

It is expected that the credit institutions lean on the significant risk drivers during the calculations, but there are neither in the CRD nor in the Hkr. any exact prescriptions relating to their scope, so their establishment is the certain institution's task.

[14] I expound only the PD relating and LGD relating prescriptions, because the CCF (Credit Conversion Factor) is not relevant in case of the loan products secured on residential property collateral.

It is an elementary requirement that the PDs of the certain deal categories or pools have to be estimated from lung-run averages of the one-year default rates, or have to be derived from real realised losses and the estimated LGDs *(EC [2011c] Article 176 (2); Hkr. 73. § (1) Paragraph).*

A similar rule refers to the calculation of Loss Given Default as well: the LGDs of the some deal categories or pools have to be estimated from their real long-run averages, as an average of loss rates concerning all observed non-performings weighted by the number of defaults, and for retail exposures there is also an opportunity for the credit institution to quantify the LGD based on the real realized loss data and the proper estimates of PD *(EC [2011c] Article 177 (2); Hkr. 74. § (1) and 76. § (1) Paragraph).*

In the course of LGD calculation, special aspects have to be considered as well, so *(EC [2011c] Article 177 (1); Hkr. 74. § (2)-(4) Paragraph)*:

- if the LGDs calculated by involving impacts of economic downturn are higher than the long-run average, than these more conservative (downturn) values have to be used to mitigate the cyclical impacts of the economic downturn to the capital requirement,
- while considering the collaterals:
  - all dependencies have to be considered which are between the risk of the client and the collateral or that of the collateral provider,
  - the estimates can not solely be based on the collateral's estimated market value, the impacts of the liquidity and other, non-itemized factors have to be involved as well, and
  - the currency mismatches between the obligations and the underlying collaterals also require special treatment.

It is prescribed as well that the exposure-weighted average LGD of the retail exposures secured by residential property collaterals, which do not belong under the effect of guarantees from central governments, can not be lower than 10% *(EC [2011c] Article 160 (4)).*

The exposures used for parameter estimation, the lending standards effective during the period which was the basis of the calculation, and other relevant characteristics have to be comparable with the current exposures and rules. If this assumption is not met, or the economic and market conditions that underlie the data are not relevant to current and foreseeable circumstances, and concerning that the credit institution is not able to use correction supported properly, than it can not apply its own estimates *(EC [2011c] Article 175 (1) d); Hkr. 70. § (4)-(5) Paragraph).*

(d) Data used for calculating risk parameters

The *Article 181 of EC [2011c]* prescribes that in case of using statistical models or other automated system with assigning purpose, the credit institutions have to be in possession proceedings, which assess the accuracy (reliability), completeness (the measure of data deficiency) and the appropriateness (freeness of distortion) of the data being used. The institutions also have to present that the data which are the basis of the model building are representative concerning the recent portfolio.

The estimation of risk parameters has to be based on long-run historical experience, so the estimates of loss features have to be founded on 5-year data at least, but if better forecast can be made from the more recent data, than the earlier ones should not be considered, or only with lower weight. The requirement concerning the length of the data series, which serve as a basis for the estimations, has to be fulfilled until the models' operational use is started. On the other hand, according to the temporary disposals prevailing in the framework of national discretions only a two-year-long data series has to be used in case of the internal rating based method, than this period has to be lengthened by one year each year until reaching the period of five years *(EC [2011c] Article 177 (2); Hkr. 76. § (4) Paragraph)*, furthermore the Supervisor can mitigate the expectation of three-year period concerning the appropriateness of the rating system and the use of the results derived from them for risk management purposes, if the certain institute meets all the other minimum requirements entirely.

The estimations have to be founded basically on the internal data, but external or even common data can be used as well, if it is provable that there is not any significant difference between the internal and external data regarding the assigning processes into grades or pools, as well as the composition of the data (risk profile), or if the differences can be adjusted properly for the sake of completing representativity.

From the point of view of my dissertation it is worth emphasising the *71. § (1) Paragraph of the Hkr.*, according to which *"A credit institution can use common data of some institutions in case of getting the permission from the Supervisor, if*

a) the rating systems and criteria of other institutions participating in using the common data are similar to those which are used by it, *b) the set of common data representatively reflects the portfolio, relating to which the common data are applied, and*

c) the common data are consistently used by the credit institution during a longer period" (Hkr. 71. § (1) Paragraph).

The fulfilment of this requirement also has to be kept in view by those credit institutions, which intend to use the data from the interbank LGD Database for estimating their own Loss Given Default rate.

2.4. Hungarian aspects

The CRD and the national laws, implementing it, require supervisory authorization for using the advanced methods for calculating the capital requirements, but the way to granting the permission is still partly unbeaten for the Hungarian credit institutions. As a matter of fact, it means large challenge that according to the credit institutions' job is to demonstrate to the Supervisor that they are able to answer the requirements. At the same time the difficulties are considerably mitigated by the national discretions relating to the minimum requirements of using the internal rating based method, for example the opportunity of implementing the advanced methods regarding the certain risk areas (credit, operational, market risk) at different date, the roll out within the certain risk areas *(Hpt. 76/B § (8)-(10) Paragraph)* and the permanent partial use *(Hpt. 76/D §)*. Further significant relief is that the Supervisor has the opportunity to ease up the prescriptions relating to the databases needed for the implementation,[15] to handle the data collected before the getting into force more flexible, and to decrease the required minimum 3-year-long period of the Experience test in the transitional period.

However, *"One important benefit of the application of advanced capital calculation approaches is the strengthening of the institution's risk awareness and the further improvement of their risk management systems. Therefore, the HFSA [considers it as high priority and] expects institutions, within the limits of financial reasonability, to develop their systems and internal operations in a way that enables the fastest possible fulfilment of IRB application requirements" (HFSA [2008c], pp. 27.).* The Supervisor unambiguously aims to encourage the development and implementation of the more advanced methods, basically supports the opportunity of using preferential procedures, but at the same time it stipulates strict conditions for the institutions.

[15] In case of IRB method it is sufficient that the institution has only 2-year-long time series.

Those institutions which would like to use the internal rating based method have to fulfil the minimum requirements of the CRD already before handing in the application, because it can be assured only in this way that the Supervisor may assess in all detail indeed the preparedness of the applicants in the six month period available for the validation, and it ascertains that these methods form part of a properly developed and integrated risk management process for the certain institutions. Since the so-called Use test criteria draws up the requirement that the rating systems and processes used for calculating the capital adequacy, as well as the risk parameter estimates have to be planted into the risk management, credit approval, internal capital allocation and corporate governance processes *(EC [2011c] Article 139; Hpt. 76/B § (2) b) and e); Hkr. 91. § (1)-(2) and 93. § (4) Paragraph).*

The Supervisor makes efforts to visit the institutions still in the pre-application phase, to review the ways, which they want to use, in the course of pre-validation on-site audit, and to provide them consultation opportunities in the interest of surmounting the difficulties and of the smoothest possible implementation of the validation processes at the same time. And in the course of the approval process and the further reviews, following the proportionality principle it considers the special features of the certain institutions: above all their size, subtlety and complexity of their business activity.[16]

It keeps continuous exchange of information with the foreign supervisory authorities for the sake of smoothing the cooperation, striving for agreement relating to understanding of the directives, and for harmonizing the supervisory practices.

As a consequence of the regulatory and structural transformations, intensive modifications have become necessary regarding the practices of the certain national authorities as well, which referred to rather numerous laws (the Hpt. and the Hkr. among others) and many relating decrees in Hungary as well.

Implementing the new prescriptions into the national regulations and putting into practice meant notable challenge for the legislators, the Supervisor and also the market participants, since the adaptation into the national laws and orders and the harmonization had to be done already in 2010, considering that the institutions are obliged to apply the CRD I-II-III alterations from 1st January 2011. Moreover the implementation of CRD IV will have to be started after its endorsement and coming into force, as the time period available until 31st December 2012 is notably short.

[16] However, the so-called minimum requirements are obliged uniformly for all institutes.

From the point of view of the present dissertation, one of the most important regulatory elements is the *Act CLXII of 2009 on the credit to customers*, which has to be applied not only for the customer credit loans, but also for the mortgage credits and the financial leasing contracts.[17] The provisions of the law came into force in two steps, one part of them has been effective since 1st March 2010, another part since 11th June 2010.

The rules relating to the early repayment (*23-25. § and 35. §*) were initiated from 1st March 2010. In accordance with them the customers have the right to partial or full early repayment of the credit, but the creditor is entitled to compensation for its costs, if the early repayment falls within an interest period or within a phase when the borrowing rate is fixed, and in case of full repayment the amount is above one million HUF[18] or the customer already performed a partial early repayment in the previous 12 months. The *Paragraph 25* appoints the maximum early repayment fee which may be used for the mortgage credits.

Another part of the prescriptions has to be applied from 11th June 2010, concerning for example the customer information, the advertising communication (*4. §*), the pre-contractual information (*5-13. §*), the assessment of creditworthiness (*14. §*) and the information relating to changing the credit costs (*18. §*). Their most important aims are clearly advancing the responsible lending and borrowing, and enhancing the consumer confidence.

The *Government Decree No. 361/2009 on prudent retail lending and the assessment of credit eligibility* contains closely relating rules, which came into force in 1st March 2010 and prescribes that the loan amount is not allowed to exceed 75% of the market value of the residential property at the date of the credit approval in case of HUF loans secured by property collateral, 60% of it in case of EUR loans and 45% of it in case of other foreign currency loans.[19]

This Government Decree also introduced further aggravations regarding the foreign currency loans, so for example the monthly repayment amount is not permitted to be above 80% of the credit eligibility limit at the date of credit approval in case of EUR loans, and 60% of it for other foreign currencies. However, the method of calculating the limit, the income categories which can be considered and the way of their

[17] Considering that this dissertation focuses on the mortgage credit loans, so I mention only the prescriptions relating to them.

[18] In case of partial early repayment the creditor is entitled to the early repayment fee independently from the amount.

[19] This Government Decree contains prescriptions relating to the financial leasing and the vehicle financing loans as well, but I do not present them, since they do not belong to the topic of the dissertation.

verification have not been made concrete, so their defining remained in the own authority of the credit institutions on condition that the institutions have to put down these details in writing in the framework of their internal regulation, and the Supervisor is checking the suitability and prudence of the procedures being used.

The same regulation contains the general provisions regarding the prudent retail lending, which has to be applied by the credit institutions from 11th June 2010. In accordance with that the check of the customers' credit eligibility has to be carried out on a mandatory basis in every case of credit approvals, the check of the credit eligibility has to be based on the credit eligibility limit derived from the income position of the customers, so the opportunity of the purely collateral-based lending has come to an end. In principle the new decree also promotes introducing the positive debtor list, since it prescribes that also all the known existing borrowings of the customers have also be taken into account for calculating the credit eligibility limit.

Unquestionably that the regulations regarding the mortgage lending (and the lending in general) became significantly stricter in the past period. In this course a large role can be assigned among others to the fact that the CRD-modifications, which aim to prepare the uniform prudent regulation, are lifting a considerable part of the discretions, but it is not a negligible factor either that certain national regulations have also introduced significant aggravations, learning the lessons from the crisis.

2.5. Conclusion

During the taking over of the changes of CRD, it is also an important aspect to correct the differences and inaccuracies arisen at the implementation of the original CRD, to unify the national regulation relating to the credit institutions and the investment firms (reducing the differences between the sectors), and to carry out the complete harmonization. At the same time it is also possible to modify and supplement the statutes concerned on the basis of the experiences from the supervisory reviews and controls, as well as the questions of interpretation, which came up in the course of practice.

However, it also has to be emphasised that *"… the change of the regulatory context is basically a reaction to the past crises. In addition to the regulatory adequacy, the conscious improving the quality of the banks' risk management is essential to prepare for the future crises" (KPMG [2010], pp. 14.).* The improvement of the efficiency of the risk disclosure may be an important result of the regulation reform, in consequence of which the losses and the reserve obligation can be reduced, so both the profitability and the liquidity of the credit institutions can be improved.

The aggravation of the rules, particularly the CRD IV, increases the capital requirements of the financial institutions significantly, thus the risk-conscious operation is also becoming more valuable. The implementation of the advanced methods under Basel II brings a considerable demand for both financial and other resources, but it will be a profitable investment in the long run by all means, deriving from the reduction of the regulatory capital requirements, so the present CRD-modifications may operate also as a catalyst regarding the adaptation of the advanced methods and increasing the importance of the hybrid capital elements.

At the beginning strong duality characterized both the CRD and the Hungarian regulation, which implemented it, regarding that on the one hand they contained strict prescriptions, but on the other hand they gave the institutions considerably free hand in certain questions for developing their own practice; but this discrepancy dissolved gradually parallel to declining the number of discretions during the further modifications. This is favourable on the one hand, because it results in larger predictability, but the increasingly universal nature of the developing rules can be also expressly disadvantageous, if it influences by uniform quantitative and qualitative limits the operational logic of those institutions as well, which would be able to act in a prudent way even without these prescriptions. There is also a related question, which arises more and more often regarding Basel III: the dilemma between the importance of aggravation (prevention of crisis) and growth.

In the immediate future we can expect further initiatives directed towards the standardization of the regulation by all means, while important objectives remain to surmount the competitive disadvantages, improving and preserving the competitiveness, keeping up with market development and promoting the responsible and prudent behaviour in all areas of the financial activity.

I study in the following chapter, what kind of possible methodologies are available in the area of LGD calculation, and which of them enable to be used on the basis of both the Basel principles and the Hungarian conditions.

3. Theoretical models for LGD calculation, data models and opportunities of their application

While the literature of the credit risk has been paying notable attention to estimating Probability of Default (PD) for a long time, the quantification of Loss Given Default rate (LGD) has hardly gained emphasis.

Only in the latest few years came modelling the LGD and the recovery rate into the limelight, basically because in the period between 1999 and 2002 the recovery rates showed a declining tendency simultaneously with increase of the default rates. I provide a detailed review of the literature in Chapter 5, therefore I make only general indication here to the observation of the researchers, on the basis of which the collateral values and recovery rates can be rather unstable as a result of different external and internal circumstances.

In the past years considerable theoretical and operational achievements arose in the field of LGD calculation, but the revolution has not occurred for the retail loans yet. Although the corporate theoretical LGD models concerning the retail loans do not stand their ground in all respects *(Bellotti – Crook [2008])*, however, it is worth going through them briefly as possible starting points as well.

3.1. Theoretical LGD models

In the present subsection I describe the theories of LGD calculation which appeared in the international literature. Considering that actually the application of the workout LGD methodology presents the only real opportunity in Hungary, so I give only the broad outline of cross tables, implicit historical LGD, market and implied market LGD methodology, emphasizing more the characterization of workout LGD model.

3.1.1. Cross tables: the "primitive" methodology

The use of cross tables is the simplest manner of quantifying the LGD. These tables forming matrices contain the average LGD values classified by different factors. Both their establishment and their usage are simple, namely they do not require complex modelling techniques *(Paulovics [2005])*.

As notable disadvantage is to be mentioned that finding golden mean of specifying the "optimal" number of classifying factors is difficult, since it can happen that in case of enhancing the number of dimensions a certain cell contains so few deals that the average LGD is misleading because of the small number of elements or the lack of representativity; moreover in extreme case it is also possible that certain cells remain completely empty. At the same time the excessive contraction, the classifying of deals based on too few factors is unfavourable as well, since in this case the divergence of the LGD values within the subcategories remain hidden.

Distortions can also arise when averaging the deal level loss rates is not done in appropriate manner. Regarding the weighting there are numerous alternatives,[20] all of which have their particular advantages and disadvantages, so the characteristics of the portfolio, micro- and macroeconomic circumstances and other factors have to be considered in the course of the decision making alike.

In addition to this problem it means another difficulty or disadvantage in using cross tables that the dynamic approach does not prevail in it. *Gupton and Stein [2005]* pointed out as well that using the product type, seniority, collateral features, loan target and industry as classifying factors are almost sole. Expansion, narrowing and any other modification on the scope of table dimensions are only occasional, so the changes of LGD generally occur after a fairly long delay in case of using this model.

Neither the Basel recommendations nor the CRD forbids explicitly the application of this methodology, but do not even mention among the possible procedures, furthermore the drawbacks referred to earlier contradict using this methodology as well.

3.1.2. Implicit historical LGD

The implicit historical LGD method performs the calculation based on realized losses, derived from a group of loans or loans which belong to the same rating grade, and the estimated PD.

It has to be mentioned that contrary to the Basel prescriptions relating to the LGD *(EC [2011c] Article 160; Hkr. 74. § (1) Paragraph)*, this method essentially provides a weighted average, but considering that according to the regulation this is allowed to be used only for retail portfolio and purchased receivables *(EC [2011c] Article 177 (2); Hkr. 76. § (1) Paragraph)*, and these portfolios are characterized with appropriate

[20] I demonstrate the different weighting methods in Chapter 3.2.2.

granularity, so the weighted and non-weighted averages do not deviate from each other in a measure, which could result in significant distortion.

Putting to use the relation that the ratio of expected loss and exposure (EL/EAD) is the product of PD and LGD, the LGD can be quantified by the under-mentioned formula, knowing the loss rate and the PD:

$$\mathrm{LGD}_{(\%)} = \frac{\mathrm{EL}/\mathrm{EAD}}{\mathrm{PD}_{(\%)}} = \frac{\mathrm{EL}}{\mathrm{PD}_{(\%)} * \mathrm{EAD}} \tag{3.1}$$

Using this seemingly simple methodology meets with remarkable difficulties, because the Supervisor allows its application only in case when both the estimate of PD and expected loss is properly accurate. Concerning that no data series are available in Hungary yet whose number of elements is high enough and whose length is enough, so this methodology does not present a real alternative.

3.1.3. Market and implied market LGD

The market LGD method quantifies the LGD value based on market prices (or credit margins) of traded non-performing bonds and/or debts evolving after the default event, so it can be used only in those countries where the prices are able to reflect properly the risk belonging to bond issuers, borrowers, in other words where the secondary market of bonds and debts is developed enough *(Gupton et al. [2000]; Gupton – Stein [2005])*.

Numerous aspects have to be considered in the course of specifying which time relating market prices shall the modelling be based on, I mean how long shall the lag (period of delay) take from the time of non-performing event.[21] It is important that enough time should be at the market participants' disposal to obtain all the necessary information to establish the prices, but on the other hand the reactions on the default event should still make their impacts felt.

The recovery and loss models of rating agencies are usually founded on this approach, according to which the recovery rate can be calculated based on the market prices of non-performing bonds and debts, and the LGD can be also quantified on the grounds of it. In accordance with the logic backgrounds the methodology, in case of existing properly efficient secondary market, the price valid on the market after the default event incorporates the current market expectations relating to the recoveries, considering the costs, uncertainty, time-consumption etc. of possible reorganizational process as well.

[21] For Moody's recovery model this was a 1-month period (*Gupton – Stein [2005]*).

The implied market LGD method differs from the former one on that respect that this is founded on prices or margins of risky, but not defaulted bonds and/or debts, and quantifies the LGD on the basis of them by using a theoretical pricing model. It has the reason for the existence especially in that case, if the market of bonds and debts is developed properly, but the non-performing rate is rather low. Since in such a case despite the availability of prices of debts and bonds in a considerable number the market LGD method can not be used, because there are only very few non-performing deals among them (*Bakshi et al. [2001]*).

This methodology contains a notable model risk, and in spite of the fact that it is considerably sophisticated and complicated, it is not able to make reliable forecast in many cases.

The implied market LGD methodology is neither in the theory nor in the practice universal yet,[22] moreover it is actually linked almost exclusively with the name of the large international credit rating firms, contrary to the market LGD methodology which possesses notable literary and practical background,[23] on which detailed survey can be read for example in the study of Altman, Resti and Sironi *(Altman et al. [2005b])*, Crouhy, Galai and Mark *(Crouhy et al. [2000])*, as well as *Gordy [2000].*

In Hungary there is no possibility of using either the market LGD or the implied market LGD methodology, because the properly developed secondary market of debts and bonds required as primary condition is not available. Furthermore it is not a minor matter that these methods do not explicitly take into consideration the role of collaterals acting in the recovery of non-performing exposures, although considerable part of the retail credit portfolio of the Hungarian credit institutions is made up of secured deals.

Concerning the Basel regulation it has to be mentioned as well that, considering the notable model risk, the use of these models is not allowed for establishing the capital requirement under Pillar I, but where a properly liquid and developed secondary market

---

[22] For example the study of Unal, Madan and Güntay (*Unal et al. [2001]*), *Duffie and Singleton [1999]*, Bakshi, Madan and Zhang (*Bakshi et al. [2001; 2006a; 2006b]*), Collin-Dufresne, Goldstein and Hugonnier (*Collin-Dufresne et al. [2004]*), as well as *Das and Hanouna [2009]* can be mentioned as application of it.

[23] Some important examples: *Merton [1974]*, Kim, Ramaswamy and Sundaresan (*Kim et al. [1993]*), Nielsen, Saà-Requejo and Santa-Clara (*Nielsen et al. [1993]*), *Hull and White [1995]*; *Longstaff and Schwartz [1995]*, *Duffie and Singleton [1999]*, *Jarrow and Turnbull [1995]*, *Duffie [1999]*, Jarrow, Lando and Turnbull (*Jarrow et al. [1997]*), Gupton, Finger and Bathia (*Gupton et al [1997]*), *Wilson [1998].*

of debts and bonds is available, an opportunity of their use presents itself for calculating capital requirement under Pillar II.

3.1.4. Workout LGD

The workout LGD methodology calculates with properly discounting the expected net cash flows deriving from each of the loans after the default event, and is mostly founded on the past workout observations.

In the interpretation of CRD the loss: *"means economic loss, including material discount effects, and material direct and indirect costs associated with collecting on the instrument" (EC [2011c] Article 4 (28))*.

The notion of economic loss is not the same as the one understood by accounting, but it forms a broader category, since numerous cost elements are part of it which are not there (or perhaps not at all) handled by the accounting.

In the course of calculating economic loss the discounted value of the collectable amounts from the deal reduced by the costs has to be compared with the exposure at the time of default event. This is actually the basis of the workout LGD methodology.

The literature distinguishes two typical subtypes within this model class *(Paulovics [2005])*:

a) the contract model and
b) the collateral model.

(a) Contract model

The contract model emphasizes not the source of recoveries, but the measure of those, and searches for those explanatory factors (risk drivers), on the basis of which the prediction of expected recovery for the non-performing deals with similar features can be carried out, having discounted to the present the recovery amounts and the costs in connection with the workout.

The contract model charts the scope of relevant explanatory factors, on which it usually builds up a sort of model, and using it prepares the LGD estimate for the groups, categories or pools of deals which can be characterized by different values of the explanatory factors and can be distinguished relatively well from each other.

In the course of its application it means a problem to what extent the recoveries and incidental costs from the non-closed deals can and have to be considered. If this deal category possesses entirely different features from the deals already closed, then the manner of handling them has a significant effect on the model and on the results of calculation made with it.

Paulovics [2005] published a model constructed with using Hungarian data, but the database examined by him consisted of only 109 recovery data from 23 enterprises' 51 deals, so we can draw conclusions from his results only with doubts. He selected the potential influential factors based on the correlation with the recoveries from each deals, then he filtered them on the basis of their economic sense, and following that he prepared a linear regression on the grounds of the selected explanatory factors for predicting recovery rate.

On the basis of the two-factor regressions he found the gross margin, the return on income and the solvency rate significant. According to the calculations carried out on the examined data the exposure at default and the size of the firm, which are considered dominant by the literature, did not prove to be significant influential factors.

(b) Collateral model

The collateral LGD model actually focuses on the source of cash flows in contrast with all the models mentioned earlier, so it puts the collaterals and guarantees into the focus of the study. However, the measure of all the net recoveries deriving from the non-performing deals gets important role in this model as well.

The difference can be formulated mainly in a manner that in the framework of this model the calculation of expected recoveries from the collaterals and from the remainder unsecured claims are also separated explicitly from each other, while in case of the collateral model and in case of all the other methodologies presented earlier the emphasis is positioned rather only on the measure and the in-time distribution of the cash flows. It is unquestionable that the liquidity and collectability of the collaterals of the deals as well as the required loan-collateral rate influences the recovery rate and the LGD for secured loans, so in the case of those the collateral model seems to be a preferable solution.

Paulovics [2005] prepared a model using also this approach, on the basis of recovery data from 56 deals, with closed workout process, of a Hungarian commercial bank. Although the low number of observations would have been extendable by considering the data relating to the deals which were not yet closed, but this would have raised further problems already mentioned earlier.

The author observed that the efficiency of reorganization intensely influenced the measure of recoveries: for those deals which were ended in liquidation the recovery rate proved to be significantly lower, than in case of any other outcomes of the reorganization process. However, considering that he did not succeed in finding at least one factor using regression calculation, which would have notably contributed to predicting the outcome of reorganization, he fixed the probability of liquidation at 60% on the grounds of observations.

He considered basically three types of costs: the expenses pertaining to the enforcement of collaterals, the administrative and the financing costs. This latter one is actually the cost of funds of the bank, which the author quantified as a product of the duration needed for the workout process and the expected return from the capital.[24]

Similarly to the length of the period needed for the workout process he calculated the expected recovery rates from the collaterals one by one for the collateral types, as a proportion of the market value of the collaterals, as an amount weighted average of the observed data. However, he considered the recoveries from the remainder unsecured claims as zero, because he experienced it marginal.

It can be said of both the gross recoveries and the costs in connection with the workout procedure that their measure and in-time distribution depends on among others the type of the examined product in an important manner. Deriving from that this methodology can be better used for the typical mass products, compared to the special small loan portfolios.

3.2. Special issues of the workout LGD methodology

In the present chapter I review some important issues of the workout LGD methodology in more details, considering that only the application of this methodology has actually the reason for existence in Hungary recently, since according to the statements of the previous subsection the lack of dynamic approach can primarily be mentioned as a counter-argument against using the cross tables, no data series are available whose number of elements is high enough and whose length is enough to apply the implicit

[24] The expected return from the capital is the average cost of funds added by margin.

historical LGD method, while the market and implied market LGD methodology is not appropriate because of the backwardness of the secondary market of bonds and debts.

3.2.1. Cost allocation and consideration of discount effects

In the course of using the workout LGD methodology the basis of the calculation is discounting the cash inflows and outflows relating to the non-performing deals during the workout process by the time of default, so both the allocation of costs and the correct consideration of discount effect mean a notable challenge for the credit institutions.

The quantification of LGD has to be founded on calculating economic loss, which is not the same as the loss understood by the accounting. In the course of LGD calculation all substantial direct and indirect costs relating to the workout process have to be considered in addition to the "classical" accounting loss, and for the sake of allocating the latter to each deal,[25] it is essential to have a precisely detailed and documented dividing logic, which is in line with the calculations of the controlling of the bank. In the course of cost allocation the total exposure, the total recovery and the number of deals in non-performing status in the certain period are the most often used base of comparison.

The discount effect, which is the cost of "keeping" the non-performing deals during the workout process, means a special cost category. Concerning that the recoveries and the different direct and indirect costs appear keeping on during the period from the default event to the ending of workout process, so it is necessary to specify a discount rate, with which the present value of these cash flows relevant to a common time (the time of becoming non-performing) can be quantified. Through this is the consideration of time value of the money in the course of LGD estimation carried out, which has especially great importance in case of high interests or notable lasting of the workout period. For example *Moral and Oroz [2002]* published a thorough analysis on this issue.

Actually neither the CRD nor the national regulation contains particular prescriptions regarding what kind of method should the discount rate be defined with, but basically two solutions are possible *(Info-Datax [2006])*:

[25] For example the sales costs of the real estate, the fees of legal and other professional extra works, as well as the charges of notification letters and any other costs.

- Fixed discount rate (historical discount rate): Some kind of interest rate relating to the deal at a certain time or a point of the yield curve is applied for each deal, for example the current interest rate at disbursement or at default event, the refinancing interest rate of the exposure or the risk-free rate added by a certain risk premium.
- Current discount rate: This type means the use of different discount rate for each period (for example for each month), namely it associates to each period the appropriate point of yield curve adjusted by the current risk premium.

The use of fixed interest rate (historical discount rate) can be explained with the fact that the creditor actually expected this return on the certain deal originally, so this embodies the time value for it. On the other hand applying current discount rate means that the institute lays in each period the alternative return expectation which is typical for the certain period towards the deal. However, if the rates are relatively stable in the examined term, then only a fairly small difference derives among the results calculated with using each of the methods.

Over and above that, the decision making between the fixed (historical) and the current discount rate is only the first step in labelling the rate, which strives at consideration of the time value, since there is not a complete agreement even in the literature in that respect how the factor being used for discounting should be selected.

The approaches mentioned the most often are the followings *(Maclachlan [2005])*:

- Discounting with the interest of the exposure or with its adjusted value increased by a proper premium: This grounds on the concept that this embodies the alternative cost of the loans with similar level of risk. As a considerable problem appears in this case the assignment of the "appropriate" premium, which is guided by the risk level of the certain deal.
- Calculating present value by the cost of the capital of the bank: This is a quite problematic solution, because on the one hand it does not consider the differences between the deals, and on the other hand it focuses on the risk of possible losses of the credit institution instead of the risk of recoveries.
- The use of risk-free rate: This means an alternative issuing from its unquestionable simplicity, but its adequacy is strongly disputable. An argument for it is that both the probable recovery and the risk-free rate are in connection with the current economic circumstances, so an indirect relation exists among them. Then again it is another fact that the risk-free rate does not reflect the risk relating to the exposure. As another argument against it can be mentioned that

occasionally the Central Bank's interventions are moving the risk-free rate in the opposite direction to the economic automatisms.

- Discounting with the expected rate of the non-revolving bonds: This alternative can be applied only in that case, if the secondary market of the exposures is properly developed and efficient, so the necessary discount rates are available. In addition to this it arises also as a question that to which period relating rates have to be used.

Using of the varying discount rate in compliance with the possible source and risk of recovery can be mentioned as a further opportunity, especially in that case, if the recovery depends primarily on the value of collateral behind the exposure, and the risk relating to the collateral can be correctly quantified based on the prices on the efficient secondary market, enabling the regular actualization of the discount rate.

Considering that this question is still open, my 4th Hypothesis is directed towards examining the importance of the differences deriving from using different discount rates.

3.2.2. Establishment of the long-term average

In accordance with the Basel regulation a long-term average has to be applied for measuring the LGD on portfolio level. However, on the one hand the averaging techniques are considerably diverse on basis of averaging type, and on the other hand depending on, whether those deals' recovery rates are used also in the course of calculation, which are not yet closed.

The deals can be arranged into cohorts[26] according to the date of non-performing event. If there are enough data available, then it is practical to use monthly division, so those deals have to be categorized into the same cohort, which became non-performing in the same month. In case of smaller number of deals it can be reasoned to merge more then one month, so for example a division into quarters can be applied.

The long-term averages performing in the under-mentioned table are applicable easily in practice because of their simplicity, if the necessary data are available.

[26] *"Cohort: Group whose members share a significant experience at a certain period of time or have one or more similar characteristics"* (Source: http://www.businessdictionary.com/definition/cohort.html). According to this definition I refer to the group of deals whose default date falls into the same period (month) as cohort.

Table 1: Alternative calculation methods of the long-term average

UNWEIGHTED AVERAGE	$LGD = \frac{\sum_{i=1}^{M}(1-CRM_i)}{M}$	*(3.2)*
AVERAGE WEIGHTED BY NUMBER OF NON-PERFORMING DEALS	$LGD = \frac{\sum_{i=1}^{M}[(1-CRM_i)*N_i]}{\sum_{i=1}^{M}N_i}$	*(3.3)*
AVERAGE WEIGHTED BY TIME	$LGD = \frac{\sum_{i=1}^{M}[(1-CRM_i)*w_i]}{\sum_{i=1}^{M}w_i}$	*(3.4)*
AVERAGE WEIGHTED BY NUMBER OF NON-PERFORMING DEALS AND TIME	$LGD = \frac{\sum_{i=1}^{M}[(1-CRM_i)*N_i*w_i]}{\sum_{i=1}^{M}(N_i*w_i)}$	*(3.5)*
AVERAGE WEIGHTED BY EXPOSURE	$LGD = \frac{\sum_{i=1}^{M}[(1-CRM_i)*EAD_i]}{\sum_{i=1}^{M}EAD_i}$	*(3.6)*
AVERAGE WEIGHTED BY EXPOSURE AND TIME	$LGD = \frac{\sum_{i=1}^{M}[(1-CRM_i)*EAD_i*w_i]}{\sum_{i=1}^{M}(EAD_i*w_i)}$	*(3.7)*

where: CRM_i: cumulative recovery rate until the i^{th} period,
M: number of periods,
N_i: number of non-performing deals in the i^{th} period,
EAD_i: total exposure of non-performing deals in the i^{th} period,
w_i: weight associated to the i^{th} period.

According to *Chalupka and Kopecsni [2009]* the averaging processes using the number of non-performing deals or the exposure as weights can be considered universal concerning the retail sector, moreover the CRD also prescribes applying averaging weighted by the number of defaulted deals.

Concerning that the weighting by time usually assigns larger weights to those periods which are closer to the time of calculation, the literature often advances the argument against it that it excessively smoothes the low and high LGD figures appearing in the certain periods, and through this results in underestimating the LGD.

3.2.3. Length of the recovery period, scope of deals covered by the analysis

The workout process starts actually when the default event happens, or when the treatment of the deal concerned is transferred to the workout department of the bank, and ends when the deal is not in non-performing status any more, I mean it is "cured" or has been sold or maybe written-off.

On the basis of Basel principles those non-performing deals have to be also considered in the course of estimating the LGD, whose workout process is not yet closed. The only exception for this rule is if the institute is able to verify that those are irrelevant or leaving them out does not result in underestimating the capital requirements. In the view of HFSA only those open non-performing deals are allowed to be brought into the calculation, whose future cash flows can be estimated quite properly, because otherwise a remarkable estimation error would burden the result of the calculation.

It is worth mentioning the study from *Chalupka and Kopecsni [2009]*, in which the authors considered the recovery period as closed, I mean they took into consideration the deal in the course of the LGD calculation, in that case as well if one of the under-mentioned conditions were fulfilled:

- the amount not having been recovered was less then 5% of the exposure at default event,
- at least 1 year has passed since the deal became into non-performing status,
- longer duration has passed since the default of the certain deal, then the period which corresponds to the upper 25 percentage of the distribution of the workout processes' length,
- longer duration has passed since the default of the certain deal, then the effective recovery period, so no further remarkable recoveries are expected yet.

Specifying the effective length of the workout process and examining the in-time distribution of the recoveries are important tasks from the regulatory and modelling aspect as well, especially because the deals not yet closed represent now a considerable proportion within the portfolio of the Central-European credit institutions. With establishing the concept of "effective length of the workout process" the Basel recommendations enabled the credit institutions to consider a part of these deals quasi closed in the course of LGD calculation, moreover the institutions are even allowed to leave out all the non-performing deals from the calculation which are not closed, if they can give appropriate reasons for their decision.

I consider it necessary to emphasise two important aspects in connection with this:

- The cash flows from the deals which are not yet closed are uncertain, and the measure of the estimation error is basically determined by the degree of the uncertainty, so it is an important aspect in the course of appointing the scope of deals which are to be involved in the analysis.
- On the other hand the credit institutions have a considerable freedom in this respect as well, since although the approval of the Supervisor is necessary, it depends basically on the judgement of the certain institutions what is considered as "well predictable".

So in principle it can be an object of consideration for the credit institutions to decide, whether they use for the calculation only the data of deals already closed, or those deals are involved into the LGD quantification as well, which are in on-going workout process at the time of calculation.

The credit institutions worked out numerous hybrid methods for surmounting these problems in practice. A good example is the application of different extrapolation techniques, which calculate the future recovery rates of the deals whose workout process is not yet closed on the basis of the recovery rates of former periods. The differences between the certain extrapolation techniques lie in what kind of assumptions they use regarding the changes of cumulated recovery rates from period to period, or how many former periods' data are considered in the course of calculation.

An assumption lies behind the concept of additive extrapolation that the cumulative recovery rates are increasing from period to period in the same degree, than the average growth in the former ***j*** periods.

$$RR_{i,k}^{(additive)} = \min\left(\frac{\sum_{j=1}^{t} N_{i-j} * \left(RR_{i-j,k} - RR_{i-j,k-1}\right)}{\sum_{j=1}^{t} N_{i-j}} + RR_{i,k-1}; 1 \right) \qquad (3.8)$$

where: ***i***: serial number of that period, relating to which the estimation of recovery rate is made,

k: length of the duration between the default event and the given period,

$\boldsymbol{RR_{i,k}}$: recovery rate in the $\boldsymbol{k}^{th}$ period from the default event of those deals which became non-performing in the $\boldsymbol{i}^{th}$ period,

j: number of periods used for extrapolation,

N: number of non-performing loans.

In contrast with the additive extrapolation, the multiplicative extrapolation technique considers the pace of growth, not the degree of growth, as stable: [27]

$$RR_{i,k}^{(multiplicative)} = \min\left(\frac{\sum_{j=1}^{t} N_{i-j} * RR_{i-j,k}}{\sum_{j=1}^{t} N_{i-j} * RR_{i-j,k-1}} * RR_{i,k-1} ; 1 \right) \qquad (3.9)$$

In addition to these simplest extrapolation proceedings, many further combined variant can also be formed by modifying the assumptions behind them.

A similar logic lies behind the "mortality based" approach as well, which served as a basis for numerous studies since *Altman [1989]*. A detailed description can be read about this methodology for example in the publication of *Dermine and Neto de Carvalho [2005]*, which illustrates the application of the proceeding with a numerical example as well.

3.3. Some further aspects

In accordance with the Basel recommendations the credit risk parameters, among them the LGD as well, has to be quantified by exposure categories, pools, but there are particular prescriptions regarding neither the number of categorization levels nor the factors for grouping, so the manner of the pooling falls within the competence of the individual credit institutions.

However, considering the shortage of the available data the credit institutions have rather narrow latitude, and the detailed categorization of deals occurs very rarely in practice, although for example in case of the secured loans the pooling into Loan-to-Value (LTV) bands would be beneficial without doubt.

Further question is what kind of expectations we can use concerning the future, because these assumptions may divert the estimates significantly from the realized LGD values. The historical long-time average recovery rates can be quantified fairly exactly in case of availability of the proper data even on deal level, but they constitute only the starting point to estimations concerning the future, because both the present and the future

[27] The marks are the same as the ones in the formula of additive extrapolation.

economic circumstances, conditions have be to considered. This means in practice that the credit institutions have to examine the connection between the economic circumstances and the LGD.

The CRD prescribes the use of the so-called downturn LGD in order to calculate the risk weighted assets, in the course of which also the changes arising from the cyclicality of the economic conditions have to be taken into account. This serves actually as an adjustment for the deficiency of the quantification formula of the capital requirements that it does not treat appropriately the correlation between PD and LGD, consequently it underestimates the measure of the needed capital requirements.

One of the most important provisions relating to the downturn LGD is that it is not permitted to be lower than the long-time average weighted by the number of non-performing deals.[28] One of the most obvious manners of quantifying downturn LGD would be, if following the defining of the criteria of the economic crisis and mapping the connection between the macroeconomic conditions and the LGD the credit institutions forecasted the influencing factors of the recovery rates by using some econometric methodology *(Info-Datax [2006])*.

Considering that the Central-European commercial banks are not able to suit this condition, among others deriving from the lack of data of the appropriate quantity and accuracy, so they can use one of the under-mentioned alternatives for quantifying downturn LGD *(Chalupka – Kopecsni [2009])*:

- use of a higher discount factor which reflects also the downturn effects,
- use of the long-time average weighted by the number of the non-performing deals,
- applying stress scenarios which ground on macroeconomic factors,
- considering also those deals whose workout process is not yet closed in estimating the LGD.

In the course of the present thesis I do not dwell on quantifying those adjustments which refer to express the impact of the economic crisis, however, I mention the study of *Sabato and Schmid [2009]*, in which the authors calculated downturn adjustment factors regarding secured and unsecured deals, by using stress scenarios on the data from the period 2002-2007.

[28] *Barco [2007]* formalized also an analytic relation between the long-time average and the downturn LGD on the basis of Merton’s principles.

The considerable claim to data causes a significant problem in case of using the workout LGD methodology, however, it presents unquestionably the largest opportunity to consider the credit risk mitigating and recovery increasing role of the collaterals exactly. It has to be emphasized that neither the CRD nor the Hkr. prescribes concretely for the institutions which procedure they have to use, but it is conspicuous from the circumstances mentioned before that the workout LGD methodology has not got any real alternative in the Hungarian practice currently.

On the other hand it is indisputable that the majority of the Hungarian banks does not have long enough historical data series to perform reliable estimations, or the quantity of available default data is inadequate in many cases, so calculating own LGD values came up against numerous difficulties. The shortage of data makes the model building impossible or clips the wings of the opportunities for preparing appropriately accurate estimations. The use of some external database can be a solution to this problem, but the importance of appropriate caution has to be emphasized, since for example the applied workout processes, which can differ significantly from each other institute by institute, influence strongly the recovery rates *(Thomas et al. [2007a]; Moral – Oroz [2002])*.

3.4. Systems for collecting data about recoveries from real estates on the international scene

Mostly in those countries means establishment of LGD data models an unsolved problem, where there are not any potentialities of applying the market LGD or the implied market LGD methodology in a reliable manner, because of underdevelopment and inefficiency of the subsidiary bond market or rather the money market. Since in this case, in connection with the circumstances mentioned in the former chapter, the only opportunity is presented by the workout LGD methodology which requires historical recovery data.

As a matter of fact, development of LGD data models means a really relevant problem only in the continental Europe, so foreign experiences can be derived exclusively from the practice and published studies of those countries.

3.4.1. Foreign studies serving as models

Development of the Hungarian Interbank Retail Mortgage Database[29] could be basically based on three publications which presented the international observations *(Info-Datax [2006])*.

(a) European Loss Given Default Study, Summary Information Package *(ISDA [2003])*

The common study of the British Bankers' Association, the European Banking Federation, the International Swaps and Derivatives Association, and the Risk Management Association (RMA) include the description of a data model plan.

Under the direction of RMA, the preparations were started in 2002 with the intension of supporting banks' credit risk management. However, the implementation (uploading of the database) has not been carried out yet. The data model concerned a considerably limited scale of debts: it did not contain household ones at all. Data pertaining to clients, deals and collaterals were not divided into separated tables they were rather installed into a uniform integral data structure.

With regard to the fact that the authors did not intend to oblige the participation in the database system for the European banks, only the voluntary collectors would have got the right of access. From technical point of view the data uploading would have been made in a way that each bank would have put their own gathered data in Microsoft Access format into the system, and RMA would have created an analysis based on those. Numerous arguments and counter-arguments could be mentioned in connection with this manner, since it would not have made the access to the specific data for the banks possible, but only to a structured "extract" which could have been applied directly for analytical purpose, it would have let a considerable part of involved opportunities[30] unused. However, it is indisputable that the comprehensive analysis created by RMA would have produced notable professional value added.

This study proved to be useful for the editors of Hungarian LGD Database, who drew many "ideas" from it. The most important outcome is that it made possible a consensus

[29] Henceforth the Hungarian Interbank Retail Mortgage LGD Database will be abbreviated as "LGD Database".

[30] If the banks had access to the rough data, they would have had the ability to create more customized analyses having taken their own portfolio's specific characteristics into consideration as well.

interpretation of the concepts, and it provided appropriate organized frameworks for gathering and sharing experience.

(b) Guidelines on Credit Risk Management – Rating Models and Validation *(OeNB [2004])*

The joint study of the Austrian National Bank and the Financial Market Authority provides practical support for the Austrian banks' data collection and appraisal methodology in a way that it presents the factors classified into client-related, deal-related and collateral-related section, which can be and have to be taken into consideration for calculating the LGD, and make the compliance for the requirements of Basel II regulation, the application of workout LGD methodology possible.

In spite of the fact that it does not contain presentation of the particular data model, this piece of work served as one of the most considerable starting points in the course of development of the Hungarian LGD Database.

(c) Italian LGD study *(ABI [2002])*

An interbank work team under the direction of the Italian Bank Association constructed a complete LGD study which presents the data structure, the particular data model and the calculation methodology as well. The data structure also consists of three parts, but its dimensions are constituted by the client/deal, the collateral and the cash flows, contrary to the Austrian one.

The fact that it also contains the data regarding not-closed deals can be mentioned as the model's speciality.

3.4.2. General features

Speaking in general terms, the foreign studies and the already applied databases collect historic data which are necessary for calculating the LGD in thematically organized data tables. The data related to clients, deals or collaterals are usually uploaded and stored as structured into separate sets, moreover some databases keep a record of cash flows in another data table.

The database of RMA, mentioned in the previous subsection, would have been an exception if the uploading had been carried out in practice.

This type of handling seemed to be rational from the model authors because in this case only the access to the ready analyses would have had to be provided for the users, specific queries would not have been possible. Moreover, the RMA would have been able to create its own previously specified analyses based on the only unified data table, which would not have contained the data structure connections in an explicit manner.
However, the specification of the scope for data collection has an extraordinary importance apart from the structure, since when making decision about this question the authors of the databases predefine, which ones could be considered as potential influential factors for calculating the LGD, which ones could be inputs for model building.

It can be considered as a typical feature that the databases, into which different client types and deal types are uploaded, contain detailed classification regarding this matter, considering that these characteristics are proved to be significant influential factors of the LGD by numerous preceding studies which are generally presented in Chapter 5.
Similarly, it is widely used to classify the clients based on the country[31] and the estimated risk, or the risk founded on any external rating methods, moreover (naturally only for corporate loans) the values and indices[32] derived from the annual reports.
The type of the deal and the loan purpose, the starting date, the length of term period or the maturity date, the exposure at default event and the possible later lent sum of money are usually present among the deal-related information. In addition to these basic data, each database collects and keeps a record of further different information. In case of corporate loans the appearance of the seniority and the rating (or for lack of rating, any other figure on risk level) among the data is general as well.
In contrast with the relative uniformity of the scope of data referring to clients and deals, the range of recorded data relating to collaterals is considerably diverse. There is no way to draw up any features which characterize the databases unanimously in this regard; only the type of collaterals can be found in each database structure.
The diversity presumably also stems from the differences between dissimilar regulatory and legal environment of the countries. Recurrent elements are among the recorded pieces of information the rating category, the collateral value at the time of default event or the frequency and method of possible revaluations, but these are not in general use, moreover each database contains differently deep and detailed records. However,

[31] Of course only in the databases which ones include international data also.
[32] The total balance, the number of employees, the annual income, the organizational form, the scale of activities and the sector category could be mentioned as the most frequent examples.

numerous studies (for example *Qi and Yang [2007; 2009]*) back up the importance of them.

The variety mentioned in connection with registering of collaterals is valid for cash flows to a larger extent: although some studies and databases involve the particular itemized cash flows of each deal, in case of the majority of databases there is no data collection and register at all, perhaps only the frequency of instalments are put down.

So the scope of data which serve as input is considerably varied.

A similar statement is also valid for the output. From this point of view, the database plan which is presented in the study of RMA may be considered extremist, since in case of its implementation it would not enable for the banks to carry out specific queries, as the participant institutions would get access only to the regular analyses being made by RMA. Although it would result in support for users in some respects, but it would not make possible the exploitation of the opportunities which are in the database, as calculating uniformly with the same LGD values the banks would not be able to consider the characteristics of their own portfolio.[33] This would lead to the circumstances that the banks which obey IRB methodology would not be able to make the best of the larger liberty offered by Basel regulation.

The databases which provide the possibility of specific queries, or perhaps the hybrid implementations which adopt the combination of these variants could be considered without doubt more advantageous from numerous aspects.

3.5. The Hungarian Interbank LGD Database

The Hungarian Financial Supervisory Authority announced a competition in 2004 with the title *"Development conception for the classification of household and corporate risk, with regard to the requirements of Basel II regulation"*, on the basis of that the conception of the common LGD Database model of the banks came into existence in July 2006 *(Info-Datax [2006])*. On the one hand this study gave a survey of the literature and the international experience, and on the other hand it sketches a database structure which is considered fit for calculating own LGD values for the Hungarian banks.

[33] The covering ratios which are different from the competitors' ones, the diversity in workout processes, its efficiency and term extension, or even the dissimilar clientele could be mentioned as examples for these characteristics.

Afterwards in 2007 the LGD project started being coordinated by the Hungarian Mortgage Association (HMA) and with the participation of five Hungarian banks, with the aim of supporting modelling of the expected losses of mortgage lending based on real loss data. For the sake of success of the initiative, the Hungarian Mortgage Association closely cooperated with the Federation of the German Mortgage Bond Issuer Banks (Verband Deutscher Pfandbriefbanken).

The Hungarian Interbank Retail Mortgage LGD Database, which is the first one founded in Europe, collects the anonymous data about defaulted mortgage deals with the purpose of enabling the participant banks to carry out better-established estimations regarding the mortgage LGD parameter, for the sake of meeting the requirements of the Hungarian regulation in connection with credit risk.

It is the jobs of Hungarian Mortgage Association which coordinates this project to gather in structured form in a common database and making possible the access for the institutions, which place their own data anonymously at the other participants' disposal, by meeting the regular data uploading requirements, which they undertook by signing the joining contract (*HMA [2008]*).

Through this database and the underlying system, credit risk analysts of the participant banks could base their calculations of the LGD regarding the retail segment on a larger and more representative sample, as they can utilize for analyses not only their own data, but the ones of the other provider banks also. These circumstances could promote the improvement of the accuracy and reliability of the calculations, compared to the estimations prepared on the bases of each bank's own data exclusively. Considering that LGD is one of the most important influential factors in quantifying capital requirements of the banks, it certainly deserves increased attention.

The established LGD Database helps the credit risk analysts by offering numerous functions, among others:

- receives and registers the data uploaded in appropriate form,
- keeps a historical record of the data uploadings of the participant institutions in order to make it unambiguous for the other members, which ones are the most recent data in the system,
- makes the downloading of data files in specified structures possible,
- keeps a historical record of these downloadings in the same way as of the uploadings as well,

- provides the opportunity of filtering and querying of the deals and real estates based on various criteria, which are registered in the interbank database anonymously, and
- enables to prepare predefined and ad hoc statistics as well (*HMA [2008]*).

If a new bank would like to join the central data provider system, it has to contract with the Hungarian Mortgage Association in the same way as the present participants, promising the compliance with the means for data providing, and it has to pay the membership fee.

The resignation from the system can be set going by the particular bank itself, or the exclusion can be carried by the consultative council, being composed of the representatives of participant banks, with a majority of 2/3 of the delegates, if the particular bank seriously and systematically offends the obligations having laid down and undertaken by the contract.

3.5.1. Uploading the LGD Database

For each participant bank, the clients' data, the real estates' data and the others pertaining to the defaulted mortgage loans are usually collected by credit risk analyst staffs or by those departments which serve data for them. In order to have larger and more representative data mart at their disposal for the analyses, by means of cooperation of the participant banks and support of the implemented system, the staff of the banks which are in possession of appropriate rights, record the gathered data at least quarterly by uploading data files in a format which comply with detail specified strict requirements.

Although the Database was established only in 2008, it contains the data related to 2005, 2006 and 2007 as well, because participant banks undertook to carry out the "primordial uploading", by providing historical data with reference to those three years retroactively.

The LGD Database is able to admit data in the appropriate structure, pertaining to the participant banks' defaulted mortgage loans. From the constitutional aspect the dataset consists of three parts:

- specific data of the defaulted mortgage loans,
- specific data of the real estates referring to these deals,

- data which enable to join deals and related real estates together (relations of the type 1:1, 1:n and m:1 may occur alike).

The data uploading becomes possible by being logged in the central server, but only for those who are in possession of the right for recording and required password too. The data files which are to be uploaded may contain new records and modifications[34] as well; recognition and appropriate treatment of them is automatically provided by the system functions. The system does not receive data files which are faulty that is to say do not meet perfectly the prescriptions in some respects, for example contains incomplete fields or invalid dates.

The users being in possession of the right for recording may upload data only under the name of the bank, which they work for, and they can not see all data uploaded by the other participants, but only those which are already approved by the member of the particular bank, who is in possession of the right for approval and secured password. The approver is supposed to check the data previously, that is to do a further error screening following the thorough control performed by the automatisms of the system.

Following the end of each calendar quarter the system appraises the uploadings automatically, and if a participant bank did not record data in the particular period, it does not have the opportunity in the forthcoming quarter to query the data deriving from the other banks. Possibilities of "empty uploading" are given as well, with which the bank indicates that there were not any deals in the particular quarter which can be uploaded into the LGD Database. However, this alternative is allowed only with good reason, since the appropriate data supply is a contractual obligation.

For the banks which comply with the obligation to provide data, the credit risk analysts or the other staff-members who are charged with this function and are in possession of the right for downloading and password, can download them as anonymous data in pre-specified format whenever they wish. There are not any restrictions in connection with the frequency of downloadings, but only those data can be accessed which relate to the already closed periods, therefore the quantity of downloadable data from the system do not change during a particular quarter.

3.5.2. Deal-related data

Deal-related data can be classified into three groups basically:

[34] If a particular identification number already exists in the system, then it will be treated not as a new data, but a modification.

- basic data,
- data in connection with the claims derived from the deals, and
- recovery and loss data.

The data which constitute the first group (Table 2) are already known at the start of the deal, so these are to be recorded at that time in the framework of data registering of the bank, and do not change regularly later on, except for some special cases, for example restructuring.

These basic data are historically available for each bank, and recording of them in the internal bank systems usually possess long foregoing. Perhaps the statement of client type means exception to this characteristic, so the filling in of this field is not compulsory because of this reason.[35]

Table 2: Basic data of deals

	CONTENTS OF THE FIELD
Central deal identification number	Hidden random individual code for identifying the deal, which is generated by the central database manager system.
The bank's deal identification number	Individual deal identification number which is generated and applied by the bank.
Product type	Product type of the deal depending on the loan purpose of the client. Possible code values: 1: Home purpose – Buying, 2: Home purpose – Building, 3: Home purpose – Renovation, enlargement, other, 4: Free use.
Currency	3-character length ISO code which identifies currency of the deal at disbursement.
Client type*	Client type of the borrower. Possible code values: 1: Private individual, 2: Individual entrepreneur.
Date of disbursement	If the disbursement occurs in details than the date of the first subpayment, else the only date of disbursement.
Disbursed amount in original currency	Disbursed amount for the client, given in the original currency of the deal. If the disbursement occurs in details than the sum of disbursed subpayments.

*: Filling in is not compulsory. (Self-made table)

The following group of data (Table 3) includes the capital, interest and other claims in connection with the deals at default event and at denunciation. These pieces of information about exposures are applied in the course of calculating recovery rate directly.

[35] The basic idea in the course of development of the Database was that in order to achieve the largest possible quantity of deals the filling in of every field has to be optional, which can not probably be served on the basis of the internal bank systems which reflect the historical data collecting culture.

Table 3: Data in connection with the claims derived from the deals

	CONTENTS OF THE FIELD
Date of default event	Date of getting into default status.
Capital claim at default event	Capital claim at default event, given in domestic currency (HUF).
Interest claim at default event	Interest claim at default event, given in domestic currency (HUF).
Other claim at default event	Late interest, fee, charge and other claim at default event, given in domestic currency (HUF).
Date of denunciation	Date of getting into denunciated status.
Capital claim at denunciation	Capital claim at denunciation, given in domestic currency (HUF).
Interest claim at denunciation	Interest claim at late interest, fee, charge and other due at default event, given in domestic currency (HUF), given in domestic currency (HUF).
Other claim at denunciation	Late interest, fee, charge and other claim at denunciation, given in domestic currency (HUF).

(Self-made table)

In the practice of banks, there can be notable differences between putting the deals into defaulted or denunciated status,[36] furthermore small-scale temporal changes may occur for particular banks as well.

Considering the fact that data uploading passes anonymously, there is no way to retrieve information from the database directly concerning the proper reason behind the changing of average duration between date of default event and denunciation:

- the particular banks, in consequence of changing in their workout policy, denunciate their defaulted loans in shorter / longer time than previously, and/or
- the compound of the portfolio performed in the Database changed in a way that the banks which apply generally faster / slower denunciation period represent larger proportion lately.

The third group of deal-related data (Table 4) serves as particular input for calculating the recovery rate as well, since recovery and cost data are performed in it.[37] Although the emergence date of these latter ones do not appear explicitly, according to the practical experiences this is not significantly different from the date of recoveries, as the most considerable costs arise usually simultaneously with the recoveries or shortly before.

[36] Since this is not prescribed by regulations particularly, it serves only as indication.
[37] The amounts are to be recorded in Hungarian Forint in this case also, irrespectively of the original currency of the deal.

Table 4: Recovery and loss data in connection with the deals

	CONTENTS OF THE FIELD
Selling price of the debt*	Real selling price of the debt in case of factoring, given in domestic currency (HUF).
Amount of other recovery**	Sum of other recoveries from a deal which came in following the denunciation, given in domestic currency (HUF), excluding recoveries from selling the collateral.
Date of other recovery**	Date of the coming in of the latest other recovery.
Collection cost	Sum of the direct costs pertaining to the deal following the denunciation, for example distraint costs, excluding the wage costs of the bank's staff and the allocated other general costs.
Own factor*	The debt is sold for a factor firm which is member of the same company group or not. Possible code values: 0: False, 1: True.
Type of cut-off	The reason behind striking out the deal that result in expiration of the bank's claim on the client. Possible code values: 1: Auction of the real estate, collective selling, 2: Factoring (selling of the debt).
Date of cut-off	Date of striking out the deal and the expiration of the bank's claim on the client.
Written-off principal amount	The amount of written-off principal which is booked as a loss by the bank, given in domestic currency (HUF).
Written-off interest amount	The amount of written-off interest which is booked as a loss by the bank, given in domestic currency (HUF).
Written-off other claim	The amount of written-off late interest, fee, commission and other claim which is booked as a loss by the bank, given in domestic currency (HUF).

*: Filling in is compulsory only for factoring. (Self-made table)
**: Filling in is not compulsory.

The Database does not contain client-related data and does not enable to link the deals of a particular client with each other. In principle this is not necessary for household exposures, since the Basel regulation allows the treatment on client level *(Hkr. 68. § (1) paragraph)*, however, managing the so called "cross-default" event attention has to be taken in the course of model building for estimating the LGD, since when a loan becomes defaulted than the other deals of the particular client will carry higher risk as well. This problem has to be treated in the course of attaching default status to deals in the framework of the internal processes of the bank.

3.5.3. Real estate related data

Similarly to the deal-related data the first block (Table 5) contains only general pieces of information which are theoretically at disposal already at the start of the deal. These descriptive data as potential determinative factors of the recovery rate ground properly for establishing the regression. However, considering the fact that the filling in of the majority of them is optional, the opportunity of their application is considerably limited in consequence of the notable lack of data.

Table 5: Basic data of real estates

	CONTENTS OF THE FIELD
Central real estate identification number	Hidden random individual code for identifying the real estate, which is generated by the central database manager system.
The bank's real estate identification number	Individual deal identity number which is generated by the bank and applied in the Real Estate Register System.
Real estate type	The type of the real estate. Possible code values: 1: House (detached house, owner-occupied block, terraced house, part of a house, semi-detached house), 2: Holiday home, 3: Building site, 4: Garage and storing, 5: Land (other parcel, agricultural area, pasture, plough-land, forest).
Detailed real estate type*	For houses, more detailed specification of the type of the house. Possible code values: 1: Owner-occupied block, 2: Detached house, 3: Terraced house, part of a house, semi-detached house.
Basic area*	Area of the building (m^2) without the related ground.
Ground area*	Area of the related group of the building (m^2).
Building type of the real estate*	Building type and material of the real estate. Possible code values: 1: Prefabricated flat, 2: Brick, stone, 3: Light construction, wood, 4: Adobe, other, 5: Mixed.
Building year*	Date of building the real estate (not the age of the real estate).
Number of rooms*	Number of rooms in the building: two half-rooms are to be considered as one room.
Heating type of the real estate*	Primary heating type of the building. Possible code values: 1: Individual heating (convector etc.), 2: Gas-fired heating, 3: Central heating, 4: District-heating, 5: Other.
Location type*	For houses, the floor. Possible code values: 1: Ground floor, 2: Upper floor, 3: Loft.
Zip code	Zip code of the real estate.
Settlement name	Full settlement name of the real estate.
Generated name of settlement	Full settlement name of the real estate on the basis of zip code, generated automatically by the central database manager system.
Renovation year*	Date of the latest renovation of the building (not the elapsed time since then).

*: Filling in is not compulsory. (Self-made table)
**: Filling in is allowed only for houses, but not compulsory at all.

The aim of applying "generated name of settlement" is to check the data identity: if for example the given zip code is wrong, then this error can be immediately recognized in the course of comparing the system-generated settlement name to the one which is filled in by the bank. However, in the course of the data processing the use of system-generated name is appropriate for example for classifying the deals by settlements or by larger geographical units in order to eliminate the possible misprints of the name of the settlements.

Table 6: Data of realization of the real estate value

	CONTENTS OF THE FIELD
Selling price*	The real selling price of the real estate in case of successful auction or collective selling (HUF).
Cut-off type of the collateral*	The type of closing the deal in case of selling the real estate. Possible code values: 1: Went off in the course of first or second auction, 2: Went off in the course of third or subsequent auction, 3: The borrower found a buyer without judicial execution, 4: Sold by the bank based on Ptk. 257. § (2).
Start date of the execution *	Start date of the non-bank-specific part of the workout process: date of becoming legally binding of the judicial codicil, decree, or date of handover to the bidder.

*: Filling in is not allowed for factoring. (Self-made table)

It is apparent that in case of factoring recoveries from the selling of debt and the pertaining information are included in the deal-related data, whereas in case of action of the real estate and collective selling recoveries and other information appear among the real estate-related data. From the logical view this is a perfectly correct resolution, moreover it does not basically influence the methodological analysis.

3.5.4. Joining data of deals and related real estates

There can be different relational connections between deals and real estates. In most cases only one particular real estate pertains to each deal (1:1 relation), but occasionally there are more than one real estate collateral behind a particular deal (1:n), or the same real estate serves as collateral for more than one deal (m:1).[38]

These relations appear in the system in a way that each deal or each real estate occurs only once in the table of deals or real estates, but that table which contains the connective data represents each link as separate record, so if two real estates serve as collateral of a particular deal, then these result in two records in the table of

[38] Theoretically the occurrence of m:n relation is also possible.

connections, and it can be recognized from the deal and real estate identification numbers which deal they pertain to.

Table 7 contains the basic data which make the connections. It is apparent that these recovery amounts are not the same as the selling prices which are registered among the real estate related data, since the recovery amount is a net value which is the residual from the selling price minus the various costs connected with selling (for example costs of auction, commission of selling). Furthermore the dates are not the same as well, since in this case the date of the real coming-in is registered, while the date of the expiration occurs among the real estate related data.

Table 7: Basic data of connection between deal and real estate

	CONTENTS OF THE FIELD
Central deal identification number	Hidden random individual code for identifying the deal, which is generated by the central database manager system.
Central real estate identification number	Hidden random individual code for identifying the real estate, which is generated by the central database manager system.
Charging*	Part of the value of real estate which is charged by the deal, id est the amount of claim which is enregistered to the real estate on the basis of possession letter, given in domestic currency (HUF).
Recovery amount**	The real paid-up recovery amount derived from auction or collective selling of the real estate, given in domestic currency (HUF), minus for example the occurring commission.
Date of recovery**	Date of the coming-in of the recovery derived from auction or collective selling of the real estate.
Existing chargings	Sum of possible chargings which are enregistered into the possession letter before the claim of the particular bank, given in domestic currency (HUF).

*: Filling in is not compulsory. (Self-made table)
**: Filling in is not allowed for factoring.

The last data block (Table 8), which is also partitioned only in logical respect, consists of the value data pertaining to the deal.

Table 8: Value data pertaining to the deal

	CONTENTS OF THE FIELD
Valuation method for loan collateral at disbursement*	Assessment method for realization value which is current at disbursement. Possible code values: 1: Based on appraisal, 2: Based on contract of sale.
Traffic value at disbursement*	Current traffic value at disbursement, given in domestic currency (HUF).
Realization value at disbursement*	Current realization value at disbursement, given in domestic currency (HUF).
Date of assessment of the values for disbursement*	Date of assessment of the traffic value and realization value which are current at disbursement.
Realization value at default event*	Current realization value at the default event, given in domestic currency (HUF).
Traffic value at default event*	Traffic value at the default event, given in domestic currency (HUF).
Date of assessment of the values for default event*	Date of assessment of the traffic value and realization value which are current at the default event.
Collateral value at closing the deal*	Current realization value at closing the deal, given in domestic currency (HUF).
Traffic value at closing the deal	Traffic value at closing the deal, given in domestic currency (HUF).
Date of assessment of the values for closing the deal*	Date of assessment of the traffic value and realization value which are current at closing the deal.

*: Filling in is not compulsory. (Self-made table)

In this last data group only the "traffic value at closing the deal" is compulsory to fill in. These enrolled data are usually not collected historically by the banks, therefore a decision was taken in the planning phase of the LGD Database, according to which the filling-in of these fields are not compulsory, since otherwise the banks would hardly be able to make these pieces of information available and to comply with their obligations of data providing which is quarterly due.

However, the examination of the effect of LTV (Loan-to-Value) and CLTV (Current Loan-to-Value) on the loss rate regarding the mortgage deals performs a stressed theme in the literature, but the Hungarian Interbank LGD Database can provide input for that only to a limited degree, as a result of what was mentioned above. I touch upon these questions in more details in Chapter 5.3.

3.5.5. The actual state of the database

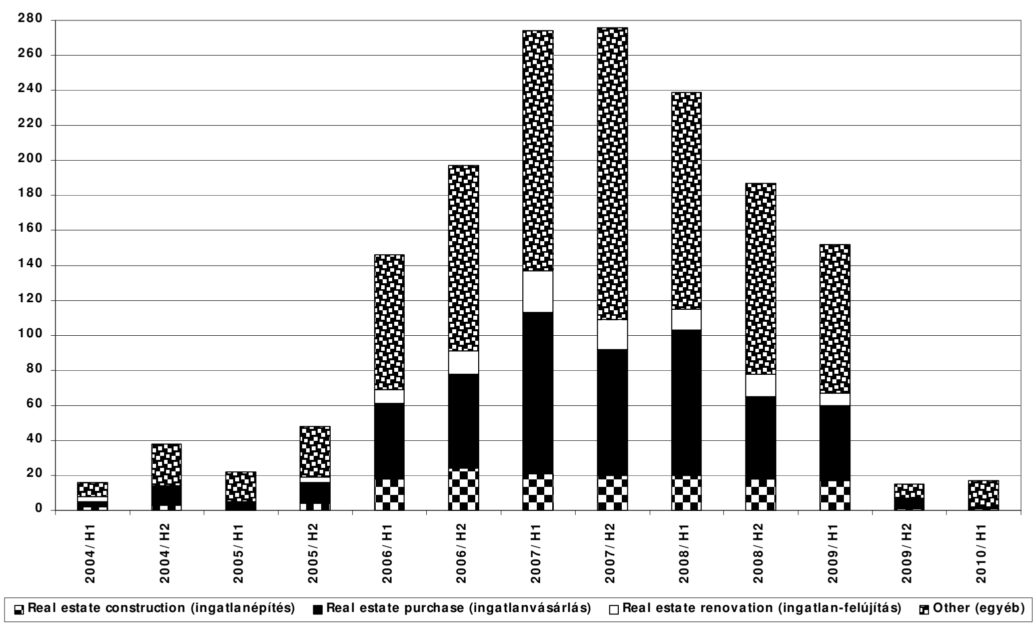
On 30th June 2011 the Hungarian Interbank LGD Database contained 1770 deals and 1719 real estates, which constituted 1881 records because of the 1:n and m:1 relationships between the deals and the real estates.[39]

In the course of my empirical researches I had respect only to those deals, in case of which the default event occurred after 31st December 2003, because the bank database (Chapter 6.1), on which the most considerable part of my analyses rested, also contains only the deals whose default event occurred after December of 2003. In addition to that I picked out from the database the deals as well, in case of which not residential real estate (or not only residential real estate) serves as collateral.

The reason behind these adjusting steps was that only those data should be applied in the course of the empirical analyses, which are directly comparable with the data of the bank database. Namely according to the *71. § (1) Paragraph of the Hkr.* one of the important base conditions of using the common database is that it has to reflect the portfolio representatively, relating to which the common data are applied.

The following Figure shows the distribution of the deals in the filtered database according to the date of the default on half yearly basis.

Figure 2: The distribution of the deals in the LGD Database according to the date of the default on half yearly basis



(Self-made figure: own calculation results)

[39] 30th June 2011 was the closing date of the last quarter which preceded the carrying out my empirical analyses, thus I consider as actual the state of the database at this time.

It is visible that the majority of the default events happened before the crisis stroke in the second half of 2008 or shortly afterwards, and the quantity of the deals which became non-performing in the last two years is rather small. Its reason is supposedly not the decrease of the number of the default events, but that in case of them the closing of the deals has not occurred yet. In addition to the shortness of time the trouble of the real estate market also delays the closing of the deals as selling the deals or the underlying collaterals. This aspect can not be left out of account even in the course of defining the length of the "effective recovery period" (Chapter 6.2.3).

In the following chapter I present a review about the econometric methodological bases of LGD calculation, then I make known the results published in the literature as well as my own empirical researches.

4. Econometric methodological bases of LGD calculation

None of the methodologies of the LGD calculation presented in Chapter 3 can dispense with using the repertoire of the econometric techniques. This statement is especially true for the market and implied market LGD described in Chapter 3.1.3 and for the workout LGD methodology made known in Chapter 3.1.4. For the sake of establishing my research, it is necessary to clarify some econometric concepts and to review certain proceedings.

Hereinafter I outline the problems of data samples which characterize the databases serving as a basis for LGD calculation, and the possible manners of solving them, the most frequently used parameter estimating methodologies, the aspects of model selection and testing, then I describe briefly those model types which are relevant from the viewpoint of my thesis.
In the course of that at first I deal with those regression models which contain a dummy as dependent variable, because these enable the classification during the LGD calculation for example on the basis of closing type of the workout process. I describe the logistic regression in more details, concerning that this methodology gets a significant role in the course of the empirical research. Following that I mention that the logistic regression is actually a special case of Generalised Linear Model (GLIM), which is destined for the linear modelling of explanatory variables which have blended measuring scales.

4.1. Problems with data samples and its treatments

The deficiencies and errors of the databases containing the historical data which are used for the statistical computations, the reductive assumptions applied for the modelling and the errors of the model can lead to biased estimates. The adjusting factors which are prescribed by the CRD and are required to use, aim to solve these problems or to mitigate their impacts, however, we have to be aware of all these sources of danger in the course of modelling, and have to arrange for treatment of their damaging consequences. Hereinafter I make known the most important aspects of this scope of problems, without intending to be exhaustive.

Concerning the LGD estimation several types of data efficiency arise, which require different treatments. Data deficiency can derive for example as a consequence of incomplete or faulty data recording. This is a rather general problem, so it is not surprising that on the one hand it has a considerable literary background, and on the other hand numerous methodologies exist which are already used "as a routine" for its treatment. I describe the most important of them in the framework of this chapter.

The latent variates, because they lack observability, can also be interpreted as data deficiency, but these are special from that point of view that in case of them none of the values relating to the observations is known, I mean they present themselves as generic deficiency. The treatment of this problem can basically happen in two ways: either by collecting additional data[40] or by using multivariate statistical analysing methods, for example Principal Component Analysis. This field of questions hangs out from the scope of my thesis, so I make only allusion to that. One of the most high-standard, detailed overviews can be read in Hungarian in the book of Füstös, Kovács, Meszéna and Simonné *(Füstös et al. [2004])*.

Regarding the LGD estimation the lack of recovery data of the deals, whose workout process is not yet closed, is a more relevant problem. Concerning that it represents a considerable proportion in the portfolio, this problem can not be left out of consideration.

Little and Rubin [2002] named the connection between the data deficiency and the values of the variables in the database as data deficiency mechanism, and they distinguished three types:

- Missing Completely at Random (MCaR),
- Missing at Random (MaR),
- Non-ignorable / Not Missing at Random (NMaR).

While in case of MCaR data deficiency the deficient and the complete observations derive from the same distribution, and so the problem can be solved quite simply, the negative impact of the other two on the data quality and on the applicability of the models which are made by using them, occasionally can not be eliminated even with sophisticated methodologies.

There is a common feature of the two latter types (MaR and MCaR) that the deficient and the complete observations do not derive from the same distribution. However, it is a

[40] The implementation of that would cause irrationally high costs in comparison with improvement of the predictive power of the model in case of historical databases of banks.

fundamental difference between these two data deficiency mechanisms, whether the lack is in connection with other variables, namely whether it is possible to predict the characteristics of the data deficiency by using other known variables. Since while this can be carried out in case of MaR data deficiency, the exact data replacement can not be executed by using the fundamental methodologies in case of NMaR, because the lack is in connection with the variable itself which contains the data deficiency.
Regarding the LGD estimation only the MCaR data deficiency mechanism means actually a problem which has to be treated, so I mention these most important methodologies hereinafter.

4.1.1. Methods for treatment of MCaR data deficiencies

Numerous methods for treating different data deficiencies are known in the literature *(Little – Rubin [2002])*, and just in the case of this simplest type is the storehouse of opportunities the broadest. Hereinafter I describe briefly some simple methodologies, whose application can be carried out easily.

(a) Listwise/casewise data deletion

The listwise/casewise data deletion is “the most trivial” type of treating data deficiency, moreover this is the default setting technique in numerous statistical program packages. Using this methodology means that all elements come to be deleted from the database, where at least one data field is incomplete *(Acock [2008])*.
This is a simple method which can be applied very well for treating the MCaR data deficiencies, and provides the comparability of the univariate statistics. Considering that the deficient and the complete observations derive from the same distribution, the deletion of the cases which contain data deficiency does not cause biases. Notwithstanding the use of this methodology can be strongly criticized in case of databases, where the proportion of the deficient observations is notable.

(b) Analysis based on available data

This procedure is the most fundamental alternative method for treating data deficient cases, in the course of which all data come to be used in the analysis of each variable, which are known for this variable, independently of the fact whether there is a data deficiency in other variables relating to the same element.

Its application enables that all the information which are available from the data collection come to be built in the computations, but on the other hand it can be mentioned as its unquestionable disadvantage that in this case for example the regression models built from different variables are carried out with using different databases, so the comparison can become problematic.

According to the analysis made by *Kim and Curry [1977]*, in case of databases which contain only MCaR data deficiencies and weak correlation, this methodology enables more efficient data management in comparison with the listwise/casewise data deletion. Nevertheless *Azen and Van Guilder [1981]* experienced just the very opposite of that in case of strong correlation.

For example the study of *Bellotti and Crook [2008]*, which will be presented in Subsection 5.2, used this proceeding as well. The cited authors filled in the deficient or faulty data fields with zeros, and introduced a dummy variable for the sake of indicating this correction.

(c) Re-weighting

The re-weighting methods classify the observations into quite homogeneous groups (categories) in the first step, then supposing that where a data deficiency occurs there the existing data represent proportionally more elements, they assign higher weights to these categories in the course of the data processing. Several subvariants of these methods are known.

The simplest re-weighting procedure can be formalized in the following manner: if a data deficiency with $\boldsymbol{n_{j(k)}}^{*}$ proportion occurs in the $\boldsymbol{k}^{\text{th}}$ category, then this group gets a weight of $p_k^* = \dfrac{1}{1 - n_{j(k)}^*}$ as a result of re-weighting *(Hunyadi – Vita [2004])*.

This proceeding can actually be identified with the imputation by part average, being mentioned hereinafter.

Deriving from its simplicity and cost-efficiency, it is rather widely used in practice, but its disadvantageous feature is that in case of considerable data deficiency it can result in biased estimates, and requires additional assumptions regarding the distribution. If the data deficiency is not MCaR (not completely at random) within the certain categories, then its applicability is strongly controversial.

It can be mentioned as a further problem that the statistical program packages do not offer at all or only a fairly limited opportunity for treating the different weights from variable to variable.

(d) Basic imputation procedures

The imputation means the subsequent artificial replacement of an originally missing data with the most similar value. As a special case the deductive (logical) imputation can also be ranked here, which deduces the missing data from the values of other variables of the certain element. The fundamental difference between the imputation procedures derives from how they understand the term similarity, I mean which criterion is considered as the most important one *(Hunyadi – Vita [2004])*.
The most frequently occurring method is the use of some sorts of mean values (for example average, mode or median) which can be calculated on the basis of the existing data relating to the certain variant.[41] The downward bias of standard error, which is otherwise the largest disadvantage of this methodology, can be mitigated by using mean values relating to relatively homogeneous groups. After all it has to be considered anyway in the course of the analyses that the imputed mean values are not independent from the other observed values, and this makes the use of numerous statistics problematic *(Little – Rubin [2002])*.
It can be generally said about all the imputation procedures that they do not leave the correlations between the variants untouched, and this causes worry during the building of the regression models, especially in case of considerable data deficiency. This symptom is a relevant source of problem in the course of LGD modelling as well: for example the study of *Bastos [2009]*, which I make known in Subsection 5.2.3, was carried out on the basis of a loan portfolio, in which the rating grade was missing in case of approximately 50% of the deals, and the author replaced it with the average value.
The literature often arranges the imputation proceedings on the ground of what the source of replacing the data deficiency is. According to that hot-deck and cold-deck procedures can be distinguished *(Little – Rubin [2002])*.

The hot-deck methods perform the replacement of the data deficiencies based on the other observations which are available. Their simplest version is the random imputation

[41] For example the study of *Chalupka and Kopecsni [2009]* reports in details on the imputation methods being used in the course of the LGD calculation.

carried out on the whole sample or on its rather homogeneous subsample, which means the substitution of the missing data with a randomly selected observed "donor". The sequential method is a little bit more subtle version of it, which makes the replacement with the first element of the database, which belongs to the same imputation group as the missing one. For this reason the result of the sequential imputation is not independent from the order of the elements within the database.

Searching for donor on the basis of distance function is a further proceeding, which selects the element within the database, which is "the nearest" to it according to one or more considerably important variants relating to the deficient element,[42] then it replaces the lack with the data relating to it. In case of using these methods it has to be considered in the course of the regression building that if a variable appears in the regression, which was left out of the distance calculation, then the parameters of the regression relating to the other variables will be biased.

Similarly to the imputation with mean values, these procedures also bias the estimates of standard errors, so they require additional assumptions in the course of making analyses *(Roth – Switzer [1995])*.

Searching for donor on the basis of distance function means a transition to the cold-deck methods, which handle the data deficiency by using external sources. The most wide-spread subtype of the cold-deck procedures is the regression imputation, which replaces the data deficiency on the basis of multivariate regression, made on the grounds of complete observations. Its special version is the stochastic regression imputation, which also includes a random error factor in the regression destined for handling the data deficiency *(Little – Rubin [2002])*.

The multiple imputation can be mentioned as the developed version of the imputation techniques *(Rubin [1976])*, which treats the uncertainty deriving from the data deficiency in a way that it multiplies the database by preparing several possible imputations, thus enabling the quantification of the error caused by the imputation, on the basis of which the standard error of the theoretically complete database can be calculated *(Schlafer [1997]; Barnard – Rubin [1999])*.

This latter method can be interpreted as a special version of the subsidiary sampling procedures described hereinafter.

[42] Numerous sophisticated methodologies are available in the literature for defining "the shortest distance". Detailed description can be read for example in the work of *Füstös et al. [2004]*.

(e) Model based procedures for treating data deficiency

The model based procedures for treating data deficiency specify a model according to the observed data, based on which estimations can be carried out on the grounds of probability or likelihood principles. These methods can be applied extremely flexibly, they enable the handling of numerous problems of other procedures for treating data deficiency, but they require the application of rather complicated mathematical-statistical methods in many cases.

A very significant element of this family of methods is the Maximum Likelihood (ML) procedure described in Subsection 4.3.1, on the basis of which the Expectation Maximization (EM) methodology spread most widely *(Dempster et al. [1977])*. This procedure was used in the field of the LGD calculation for example by *Hlawatsch and Ostrowski [2010]*. This is a multi-step iterative algorithm, which carries out ML estimation at first on the basis of those elements of the database which do not contain data deficiency, then it replaces the faulty data on the grounds of its results. Following that it makes a new ML estimation on the basis of the database which does not have any missing elements yet, then it replaces the previously supplemented values with the newly received results, and goes on with the iteration until as a result of the ML estimates' convergence the changes in the estimated values between two steps are not significant any more.

However, these model based procedures also have some disadvantages beside their numerous advantages. Concerning that handling the random error factor, which denotes the uncertainty, is an unsolved problem in the framework of these methods as well, the standard errors and the test statistics which are made with using them are not accurate *(Little – Rubin [2002])*.

Techniques arise as alternative solutions, as for example the multiple imputation mentioned previously, the Monte Carlo simulation *(Roth – Switzer [1995])* or the application of the methods made by generating artificial samples. I describe briefly these latter ones in the following subsection.

4.1.2. Artificial (repeated or subsidiary) samples

Concerning that rather little data are available for estimating the credit risk parameters, especially the LGD, it is true in a larger extent that the accuracy of estimates can be significantly increased by using artificial samples. In this chapter I mention the most important artificial sampling procedures which are able to provide appropriate data for non-parametric estimations as well.

The repeated or subsidiary sampling methods prepare new samples with artificial procedures from the existing observations mapping their structure. Their importance derives from the fact, that conclusions, relating to the distribution or some parameters of the whole population, can be drawn from the features of the statistics estimated on the basis of the got random samples *(Kröpfl et al. [2000])*.

(a) Method of independent subsamples

Using the method of independent subsamples enables the testing of sampling errors in those cases as well, when there are not enough proper data for performing the fundamental testing procedures. It prepares new samples from the existing data with independent and random "cutting up", then the sampling error can be estimated better from the available samples *(Hunyadi – Vita [2004])*.

The first step of the procedure is to construct ***k*** samples, each containing ***m*** elements, which can happen in two ways:

- by the selection of ***k*** samples, each contain ***m*** observations, or
- by randomly cutting up a dataset with k*m=n elements into ***k*** parts (which contain ***m*** elements each).

The average of the appropriate mean values of the ***k*** samples provides the estimate of the mean values of the whole dataset, whereas their variance results the estimate of the variance of the certain mean value relating to the whole dataset. Moreover in case of sufficiently large ***k*** the estimator function of the independent subsamples can be well approximated with normal distribution, which enables simple calculation of the confidence intervals.

Unquestionable advantage of this method is the general applicability, since no kind of assumptions are needed regarding the distribution of the variant. It is practical to rely on professional considerations for appointing the measure of ***m*** and ***k***, nevertheless for example according to the study of *Witten and Frank [2005]* the smallest estimation error derives in case of ***k***=10.

The study of *Bastos [2009]* and of *Bellotti and Crook [2008]* can be cited as examples from the literature of LGD calculation. Their common trait is that they divided the available data randomly into 10 parts, which contained the same quantity of deals, leaving out one part of them in each case, they built models on the basis of data from the other 9 parts, then they tested these models on the data from the 10^{th} part which was left out. They measured the average and the standard error in each case, so finally they got 10-10 averages and standard errors as a result for each portfolio, which they analysed one by one and together as well.

(b) Method of balanced repetitions

The method of balanced repetitions is one of the most popular types of the artificial sample repetitions. Similarly to the other versions, it prepares new artificial samples with numerous repetitions, then deduces the characteristics of the whole dataset from the estimates based on them.

Based on prearranged schemes it selects pairs from the samples which are made by bisection in every possible way, so that these pairings do not cause systematic biases *(Kröpfl et al. [2000]).*

The name of the method is derived from the prearranged scheme providing for each element to get involved in the certain subsamples with the same probability. Although using all divisions is possible in theory, but it is not typical in practice, since their quantity (in case of large ***n***) can be considerably high: $m = \binom{n}{n/2}$.

The course of the estimation is the same as using the method of independent subsamples, namely the average of appropriate mean values of the samples provides the estimate of the mean value relating to the whole dataset, whereas their variance results the estimate of the variance relating to the mean value of the whole dataset. Furthermore it is true in this case as well that the quantification of the confidence intervals is supported by the characteristic that it can be approximated well with normal distribution in case of large samples *(Hunyadi – Vita [2004]).*

(c) Jackknife method

The main point of the Jackknife method is preparing new samples from the original observations by omitting one (always another) element, as a result of which ***n*** artificial samples containing ***n-1*** elements derive from the dataset which consists of ***n*** elements, then performing the estimations on the basis of each of them, the combined Jackknife estimate can be made from these results *(Hunyadi – Vita [2004]).*

In case of the Jackknife method, similarly to the procedures described previously, the mean value and variance calculated in this way provide the estimate of the mean value and variance relating to the whole dataset. It can be mentioned as an advantage that in case of the availability of sufficiently large sample, the confidence intervals can also be simply quantified by approximation with standard normal distribution.

(d) Bootstrap method

The Bootstrap method generates with replacement new samples, containing ***n*** elements, from the existing set of observations which has ***n*** elements. Its logical basis is that if the empirical cumulative distribution function, estimated on the grounds of the Bootstrap-samples which contain ***n*** elements, approximates the distribution of the whole dataset well, then also the estimation of the parameters and of the variance can be carried out on the basis of the samples *(Hunyadi – Vita [2004])*.
Concerning that the number of the possible samples is rather large (n^n), the estimations are generally made only on the grounds of randomly selected artificial samples. Nevertheless in addition to the accuracy, it can be mentioned as a significant advantage that because of the numerous calculations it also enables the calculation of standard errors of statistics whose estimation is impossible or difficult by other procedures *(Füstös et al. [2004])*.

With reference to the LGD calculation it is worth emphasising that more accurate estimates can be prepared by using artificial sampling techniques, especially in that case if few data are available. According to this realization numerous studies have been already born in which one of these procedures was used, moreover a detailed description can be read about using the artificial sampling methods in the course of LGD estimation for example in the study of *Bellotti and Crook [2008]*.
Following the review of the possible techniques for treating data deficiencies, I make known the bases of the hypothesis testing and the tests which are relevant from the viewpoint of my thesis hereinafter.

4.2. Hypothesis testing procedures

In the statistical terminology the hypotheses are various assumptions regarding the whole population (the type of their distribution or certain parameters of them), and the

hypothesis testing is the examination of their adequacy on the basis of the results of sampling. [43] The tests deliberate to what extent the statement referring to the population is believable knowing the result of the sampling, considering the sample deviation as well *(Hunyadi – Vita [2004])*.
In the framework of the present subsection I briefly demonstrate the most important conceptions in connection with the hypothesis testing, then I make known the statistical tests which get part in the framework of my dissertation.

4.2.1. The basis of the hypothesis testing

All the hypothesis tests which are used in present dissertation are directed towards checking whether there is a difference between the sample estimated value and the population value composed in the hypothesis. After *Jerzy Neyman* and *E. S. Pearson* the literature calls the hypothesis which is to be tested as null hypothesis *(Maddala [2004])*, thus I also use this terminology hereinafter.
The very first step of the hypothesis testing is the composition of the null hypothesis ($\boldsymbol{H_0}$) which is to be examined and the alternative hypothesis ($\boldsymbol{H_1}$) which is opposed to it. As the result of the hypothesis testing that one of the two can be considered as true which seems to be more believable on the basis of the sampling. Since the null hypothesis and the alternative hypothesis exclude each other, the decision concerning $\boldsymbol{H_0}$ hypothesis means also a decision concerning $\boldsymbol{H_1}$ at once: the acceptance of $\boldsymbol{H_0}$ infers the rejection of $\boldsymbol{H_1}$, and the rejection of $\boldsymbol{H_0}$ infers the acceptance of $\boldsymbol{H_1}$ as well *(Hunyadi – Vita [2004])*.
Respecting its value range the statistically testable hypothesis can be simple and composite: in the first case it refers to one fixed numerical value, whereas in the second case it refers to some range of the values. The composite hypothesis is always the aggregate of simple hypotheses, and its examination can be reduced to testing the simple hypothesis *(Hajdu [2004])*.

The means of the hypothesis testing are the test statistics (test functions), towards which there is a requirement that their sampling distributions are known and mathematically treatable *(Maddala [2004])*.

[43] The result of the hypothesis testing is not an evidence, it serves only to approve or weaken the researcher's conviction in the adequacy of the hypothesis.

The test function ($T(y_1,y_2,\ldots,y_n)$) is a variate, and following the sampling it is a concrete realization of the given variate. In the course of the hypothesis testing its possible value range has to be cut into two parts: a region of acceptance (***E***) and a critical region (***K***). An alternative term for the critical region is the region of rejection. The limits of the ranges have to be defined in a way that in case of the validity of the null hypothesis the test function falls into the critical region only with a small probability (α), and the probability of falling into the region of acceptance is large ($1-\alpha$) *(Hunyadi – Vita [2004])*:

$$P(T(y_1, y_2, \ldots, y_n) \in E) = 1 - \alpha \qquad (4.1)$$

$$P(T(y_1, y_2, \ldots, y_n) \in K) = 1 - P(T(y_1, y_2, \ldots, y_n) \in E) = \alpha \qquad (4.2)$$

To carry out the hypothesis testing we have to take a concrete sample of the population, then on the basis of the position of the test function's given sampling value compared to the critical region we have to decide on the acceptance or rejection of the null hypothesis *(Hajdu [2004])*. If the value of the test function falls into the region of acceptance, it supports the justness of H_0 (H_1 alternative hypothesis has to be rejected), or else H_1 gets to be accepted (H_0 null hypothesis has to be rejected).

The literature refers to the α values as significance level, and regarding it *Sir R. A. Fisher* (1890-1962), who is reckoned as the father of the modern statistical methods, suggested the use of 5% or 1% as α, which values became generally admitted since then *(Maddala [2004])*. On the other hand it has to be emphasized that the selection of the significance level is subjective to a certain degree, its modification enables the aggravation or the easing of the examined hypothesis's acceptance, since it decrease or increase the extent of the critical region. In compliance with it the complement of the significance level of α, namely the probability of $1-\alpha$ of the acceptance of the right null hypothesis, can be interpreted as the level of reliability of the test *(Hajdu [2004])*.

The placing of the region of acceptance and the critical region compared to each other is determined by the circumstance, what kind of direction has the deviation of the assumption composed in the alternative hypothesis from the condition which is formulated in H_0 *(Hunyadi – Vita [2004])*.

The particular directional deviations of the fact from the state which is defined in the null hypothesis can be formalized with one-tailed (left-tailed or right-tailed) alternative hypotheses. Under such circumstances the given directional deviations from the assumption composed in H_0 hypothesis do the test function either relatively low or relatively high in comparison with the value in case of H_0, thus in these cases the whole

critical region has to be placed either only on the left side of the test function's distribution or only on its right side. The critical value is the $p=\alpha$ quantile of the distribution of the test function (c_a) in case of left-sided critical region, and the $p=1-\alpha$ quantile of the distribution of the test function (c_f) in case of right-sided critical region.
On the contrary, two-sided critical region is necessary to be assigned if only the fact of the deviation from the statement composed in the null hypothesis has importance, but the direction of the deviation is indifferent. In such cases the whole probability (α) of falling into the critical region has to be cut into two parts: the lower critical value (c_a) is the $p = \alpha/2$ quantile of the test function's distribution, and the upper critical value (c_f) is the $p = 1 - \alpha/2$ quantile of the distribution.

In the course of the hypothesis testing one or more samplings serve as a basis for the decision, thus making mistakes is possible. If H_0 hypothesis is true, but the value of the test function calculated from the given sample falls into the critical region, the researcher rejects H_0 hypothesis in spite of the fact that it is in reality true. The mistake which is made in case of rejecting the true null hypothesis is the Type I Error, whose probability is α *(Hunyadi – Vita [2004])*:

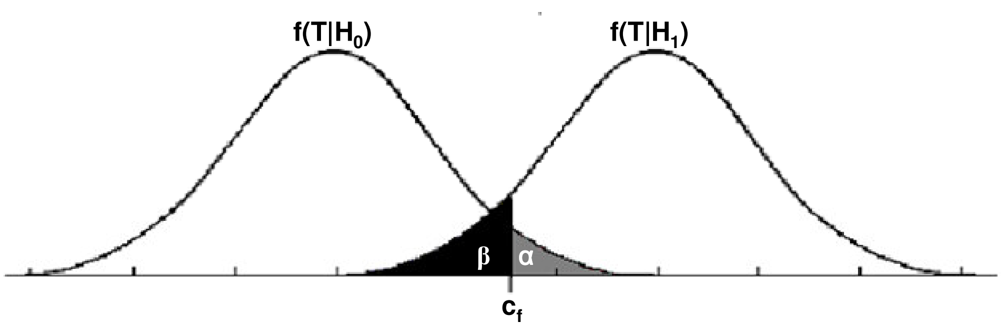
$$P\left(T(y_1, y_2,\ldots, y_n)\in K \middle| H_0\right) = \alpha \qquad (4.3)$$

On the other hand a mistake can derive not only from rejecting the true null hypothesis, but also from the researcher's accepting the false null hypothesis. This is the Type II Error which happens if H_0 is not true, but the value of the test function falls into the region of acceptance. The probability of the Type II Error is β *(Hunyadi – Vita [2004])*:

$$P\left(T(y_1, y_2,\ldots, y_n)\in E \middle| H_1\right) = \beta \qquad (4.4)$$

The following Figure illustrates the probabilities of the Type I Error and Type II Error.

Figure 3: Type I Error and Type II Error in the course of the hypothesis testing



(Self-made figure on the basis of *Maddala [2004], pp. 62.*)

The statistical literature refers to the complement of the probability of the Type II Error, namely to the ***1-β*** probability of the rejection of the false null hypothesis, as the power of the test. The probability of the Type I Error can be limited with choosing lower ***α***, but its reduction increases the probability of the Type II Error, since the region of acceptance enlarges because of the narrowing of the critical region.
Issuing from it we have to deliberate at choosing the significance level, which type of wrong decision carries more damaging consequences *(Hajdu [2004])*.
Considering that in case of given sample size and simple alternative hypothesis the probability of the Type I Error and Type II Error can be reduced only at the expense of each other, according to the Neyman–Pearson-approach ***β*** can be minimized with increasing the sample size or using the strongest test statistic in case of given ***α*** *(Maddala [2004])*.

Table 9: The mistakes which can occur during the hypothesis testing and their possibilities

	H_0 **IS IN FACT TRUE**	H_0 **IS IN FACT FALSE**
Rejection of H_0	Type I Error (its probability: α)	Right decision (its probability: $1-\beta$)
Acceptance of H_0	Right decision (its probability: $1-\alpha$)	Type II Error (its probability: β)

(Self-made table on the basis of *Hunyadi – Vita [2004], pp. 421.*)

Some statisticians (for example *Kalblfeisch – Sprott [1976]; Lindley [1957])* do not agree with using the Neyman–Pearson Theory, and they consider it as a strong simplification to regard the significance level as a decision rule. Numerous authors own the principle that the significance level has to depend on the sample size *(Maddala [2004])*.
An alternative opportunity (or parallel applicable method) is the ***p***-value approach. The ***p***-value is the probability that in any random sample the test function has a value which is less believable than the observed concrete value in this particular case, if the null hypothesis is valid.
The ***p***-value is often called empirical significance level as well, since this is the lowest significance level at which H_0 can already be rejected against H_1. In case of one-tailed alternative hypothesis the given concrete realization of the test function has to be considered lower or upper critical value depending on the direction of H_1 alternative hypothesis from H_0, and on the basis of that can the relating significance level be defined. In case of two-tailed alternative hypothesis the ***p***-value can be quantified as the

double of the ***p***-value which relates to the lower or upper critical value *(Hunyadi – Vita [2004])*.

4.2.2. The relevant tests from the viewpoint of the dissertation

The previously presented elements of the hypothesis testing can be applied in the course of numerous different statistical tests. While the one-sample tests examine the population against some condition assumed by the researcher, the two- or more-sample tests serve for comparing the populations with each other *(Hunyadi – Vita [2004])*, in other words they search for the answer whether the examined populations differ from each other from a particular point of view.

In the following I make known only the test statistics which were applied in the framework of my dissertation during the empirical research (Chapter 6.3). I briefly present the relevant two-sample (paired two-sample and two independent samples) tests for expected values as well as the Homogeneity Analysis which verifies the equivalence of the distributions.

(a) Paired two-sample tests

To carry out the paired two-sample tests one sample is necessary from each of the two different populations, in which:

- the n_Y elements of the sample from the ***Y***-population are: $y_1, y_2, \ldots, y_{nY}$, and
- the n_X elements of the sample from the ***X***-population are: $x_1, x_2, \ldots, x_{nX}$.

The examined variable is symbolized in the first population (***Y***-population) with ***Y***, and in the second population (***X***-population) with ***X***. The sample characteristics are the unbiased estimators of the appropriate population characteristic in all cases *(Hunyadi – Vita [2004])*.

The paired two-sample tests are special ones in the group of the two-sample tests: in case of them the elements of the two samples can not be considered independent from each other, since the selection of the elements of one sample entails the selection of the elements of the other sample. The sizes of the paired samples are always equal, that is $n_Y = n_X$.

In case of the tests directed towards examining the deviation of the expected values, one of the most obvious manners of their handling is to calculate the difference

$(d_i = y_i - x_i)$[44] of the elements belonging together (pairs), which can be considered the elements of a sample with ***n*** elements hereinafter *(Hunyadi – Vita [2004])*.

In this case the null hypothesis can be formalized in the following way:

$$H_0 : \mu_d = \delta_0 \qquad (4.5)$$

where: μ_d is the assumed expected value of the differences relating to the pairs of elements.

If the distribution of the $\boldsymbol{d_i}$ differences is normal or a large sample is available, the justness of the null hypothesis can be tested against the proper left-tailed $(H_1 : \mu_d < \delta_0)$, two-tailed $(H_1 : \mu_d = \delta_0)$ or right-tailed $(H_1 : \mu_d > \delta_0)$ alternative hypothesis with the one-sample expected value tests. These tests differ from each other regarding the conditions of use *(Hunyadi – Vita [2004])*:

- If the available random sample derives from a normal distribution with known deviation ($\boldsymbol{\sigma_0}$), the *z*-test can be applied for examining the justness of the null hypothesis. In this case the ***Z*** test function has standard normal distribution (N(0,1))[45], independently of the sample size:

$$Z = \frac{\bar{y} - \mu_d}{\frac{\sigma_d}{\sqrt{n}}} \qquad (4.6)$$

- If the distribution is normal, but the population deviation is unknown, the ***t***-test can be applied, namely the hypothesis can be verified using the estimated population deviation of $\boldsymbol{s_d}$ with the ***T*** test function. If the null hypothesis is true and the population distribution is normal, the ***T*** test function follows Student's ***t***-distribution with ***n-1*** degree of freedom[46]:

$$T = \frac{\bar{y} - \mu_d}{\frac{s_d}{\sqrt{n}}} \qquad (4.7)$$

- If the conditions of use of the two previous tests do not exist, but the available sample of the $\boldsymbol{d_i}$ differences is large[47] and its deviation is finite (its estimated

---

[44] The other most frequent procedure is the calculation of quotient, but I do not dwell on it because I used only the tests regarding the differences in the course of the empirical research.

[45] The standard normal distribution is a special case of the normal distribution: its expected value is 0 and its deviation is 1.

[46] Student's ***t***-distribution with ***n*** degree of freedom occurs as the distribution of a variate defined by independent variates ($\boldsymbol{\eta, \xi_1, \xi_2, \ldots, \xi_n}$) with standard normal distribution *(Medvegyev [2002])*, and in case of fairly large ***n*** it can be described with standard normal distribution. The formula of the ***t***-distribution is:

$$t = \frac{\eta}{\sqrt{\frac{\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2}{n}}} = \frac{\sqrt{n} * \eta}{\chi}$$

[47] The less the population distribution differs from the normal distribution, the smaller sample size is enough for carrying out the test.

deviation is: s_d), an opportunity presents itself to use the asymptotic *z*-test, since in this case the ***Z*** test function has asymptotically standard normal distribution[48]:

$$Z = \frac{\bar{y} - \mu_d}{\frac{s_d}{\sqrt{n}}} \tag{4.8}$$

On the basis of the sample the estimated value (s_d^2) of the population variance (σ_d^2) can be quantified in case of both the ***t***-test and the asymptotic *z*-test in the following manner:

$$s_d^2 = \frac{\sum_{i=1}^{n}\left(d_i - \frac{1}{n}\sum_{i=1}^{n} d_i\right)^2}{n-1} = \frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1} \tag{4.9}$$

I summarized the lower and upper critical values of the different alternative hypotheses in the under-mentioned table:

Table 10: The lower and upper critical values of the hypothesis testing

	THE CRITICAL VALUES OF THE Z-TEST	THE CRITICAL VALUES OF *T*-TEST
Left-tailed alternative: $H_1: \mu < \mu_0$	$c_a = -z_{1-\alpha}$	$c_a = -t_{1-\alpha}(\nu)$
Two-tailed alternative: $H_1: \mu \neq \mu_0$	$c_a = -z_{1-\frac{\alpha}{2}}$ and $c_f = z_{1-\frac{\alpha}{2}}$	$c_a = -t_{1-\frac{\alpha}{2}}(\nu)$ and $c_f = t_{1-\frac{\alpha}{2}}(\nu)$
Right-tailed alternative: $H_1: \mu > \mu_0$	$c_f = z_{1-\alpha}$	$c_f = t_{1-\alpha}(\nu)$
where: z_p: the ***p***-quantile of the *z*-distribution, $z_p(\nu)$: the ***p***-quantile of the ***t***-distribution with ν degree of freedom.		

(Self-made table on the basis of *Hunyadi – Vita [2004] pp. 439.*)

If the value of the test function falls into the region of acceptance, it confirms the researcher's null hypothesis regarding the equality of the expected values against the proper alternative hypothesis, otherwise it supports the justness of the statement composed in H_1 alternative hypothesis.

(b) Tests for independent samples

If, in contrast to the paired samples, the samples are independent from each other, those have to be handled separately indeed in the course of the hypothesis testing. In this case the elements of the samples can not be paired, in many cases neither the numbers of elements of the certain samples (n_Y and n_X) are equal.

[48] Fisher's *z*-distribution can also be traced back indirectly to the standard normal distribution, it is a special case of the Student's ***t***-distribution where ***n*** is large enough *(Medvegyev [2002])*. The formula of the ***t***-distribution is the following:

$$z = \frac{1}{2}\ln\frac{\frac{1}{m}\sum_{i=1}^{m}\xi_i^2}{\frac{1}{n}\sum_{i=1}^{n}\eta_i^2}$$

In the framework of the tests directed at the expected value, the justness of the following null hypothesis can be examined on the basis of the samples chosen separately from the two samples and independently from each other *(Hunyadi – Vita [2004])*:

$$H_0 : \mu_Y - \mu_X = \delta_0 \tag{4.10}$$

Likewise the paired sample tests, this examination can also be carried out against left-tailed ($H_1 : \mu_Y - \mu_X < \delta_0$), two-tailed ($H_1 : \mu_Y - \mu_X \neq \delta_0$) and right-tailed ($H_1 : \mu_Y - \mu_X > \delta_0$) alternative hypothesis alike. The δ_0 can be any value, in case of δ_0=0 the null hypothesis formulates the equality of the expected values.

The test which examines the difference between the expected values of the two independent samples can be carried out with one of the following test functions depending on what kinds of information are available about the two populations *(Hunyadi – Vita [2004])*:

- If the distribution is normal and the deviation is known in case of both populations, then the ***Z*** test function has standard normal distribution (N(0,1)) independently from the sample sizes:

$$Z = \frac{(\bar{y} - \bar{x}) - \delta_0}{\sqrt{\frac{\sigma_Y^2}{n_Y} + \frac{\sigma_X^2}{n_X}}} \tag{4.11}$$

- If the distributions of the two populations are normal, but the deviations are unknown[49], then in case of the justness of $\boldsymbol{H_0}$ and the validity of the conditions of use the ***T*** test function follows ***t***-distribution with $\nu = n_Y + n_X - 2$ degree of freedom:

$$T = \frac{(\bar{y} - \bar{x}) - \delta_0}{s_c \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}}} \tag{4.12}$$

  where: $s_c^2 = \frac{(n_Y - 1)s_Y^2 + (n_X - 1)s_X^2}{n_Y + n_X - 2}$ is the combined estimate of the same variances of the two populations, calculated with using both samples together.

- If none of the conditions of use of the two previous tests holds, but the deviations of both samples are finite and their sample sizes are large enough[50],

[49] In case of small samples there is a condition that the deviations are equal.
[50] The more the distributions of the two populations differ from the normal one, the larger samples are necessary.

the $\boldsymbol{Z}$ test function has approximately standard normal distribution (N(0,1)), likewise the asymptotic *z*-test made known among the one-sample tests:

$$Z = \frac{(\bar{y} - \bar{x}) - \delta_0}{\sqrt{\frac{s_Y^2}{n_Y} + \frac{s_X^2}{n_X}}} \tag{4.13}$$

I consider it as important to note that the result of this latter test does not serve any kind of information about the type and the deviation of the distributions, the acceptance of $\boldsymbol{H_0}$ null hypothesis supports only the justness of the statement regarding the defined difference between the expected values of the two distributions (in case of δ_0=0 the equality of them).

(c) Homogeneity Analysis

The equivalence of the two distributions can be tested with Homogeneity Analysis whose null hypothesis formalizes that the distribution of a variate is the same in the two populations (***Y***-population, ***X***-population), and its alternative hypothesis states that the two distributions differ from each other. Issuing from the special feature of the test function this test can be carried out with critical region only on the right side.
To carry out the Homogeneity Analysis of large samples, both the samples have to be divided up to the same classes on the basis of certain variable in the manner which can be seen in the following table:

Table 11: The work table of the hypothesis testing

CLASS	FREQUENCIES IN THE SAMPLE OF THE Y-POPULATION	FREQUENCIES IN THE SAMPLE OF THE X-POPULATION	TOTAL
C_1	n_{Y1}	n_{X1}	$n_{Y1}+n_{X1}$
C_2	n_{Y2}	n_{X2}	$n_{Y2}+n_{X2}$
…	…	…	…
C_i	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$
…	…	…	…
C_k	n_{Yk}	n_{Xk}	$n_{Yk}+n_{Xk}$
Σ	n_Y	n_X	n_Y+n_X

(Self-made table on the basis of *Hunyadi – Vita [2004] pp. 475.*)

If the distribution of the given variable is the same in the two distributions ($\boldsymbol{H_0}$ is true), and both samples are large enough, the χ^2 test function follows approximately χ^2-distribution[51] with $\nu = k - 1$ degree of freedom:

[51] The χ^2-distribution is the distribution of a variate, which derives as the sum of squares of ***n*** independent variates (ξ_1, ξ_2, …, ξ_n) which follow standard normal distribution *(Csernyák [1998])*: $\chi^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2$

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Yi} + n_{Xi}} \left(\frac{n_{Yi}}{n_Y} - \frac{n_{Xi}}{n_X} \right)^2 \qquad (4.14)$$

The null hypothesis states only the equivalence of the distributions, but it does not say anything about the type and certain characteristics of the distributions, thus in some respect it can be applied as a completion of the two-sample tests presented previously. For that very reason during my empirical analyses I also used the tests regarding the equality of the expected values and the Homogeneity Analysis simultaneously in the course of testing the 1st, 2nd, 3rd and 4th Hypotheses.

Following the brief review of the hypothesis testing procedures, I present the most important methodologies which serve for quantifying the regression parameters thus they played an important role during my empirical research.

4.3. The bases of the regression methodology

In the course of predicting credit risk parameters, one or more appropriately fitting models serve as a basis for numerous methodological procedures, and using some kind of regression is an obvious methodology for establishing them. The goal is to define a multivariate equation, which enables to predict the LGD or the recovery rate, on the basis of the influencing factors.

4.3.1. The methods for estimating regression parameters

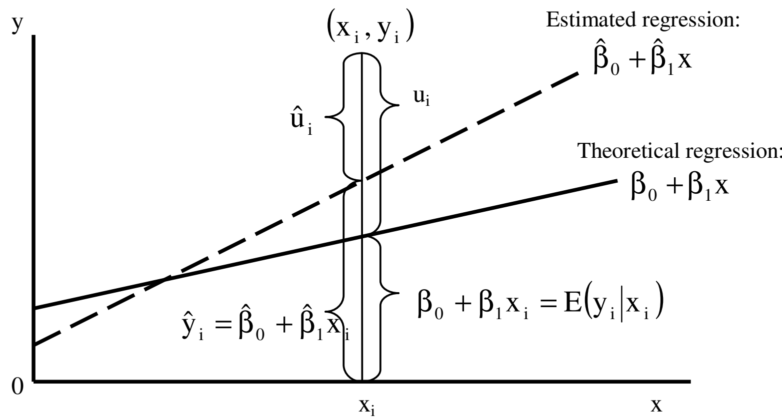
The purpose of using the three simple proceedings made known in this chapter – the Ordinal Least Squares method (OLS), the Method of Moments and the Maximum Likelihood (ML) method – is to quantify the regression parameters.

(a) The Ordinary Least Squares method (OLS)

It is not a required condition of using the Ordinary Least Squares (OLS) method to know the population distribution. Its purpose is to develop a model, in case of which the sum of squares of the deviances between the models based on the real and the estimated parameters is the lowest.

Figure 4 illustrates the deviance between the theoretical and the estimated regression function, the measure of residuum, regarding the model which contains one explanatory variable.

Figure 4: Difference between the theoretical and the estimated regression function



(Self-made figure on the basis of *Hunyadi – Vita [2004], pp. 581.*)

In the Euclidean space the distance can be defined as the sum of squares of the deviances or as its square root, so the Ordinary Least Squares (OLS) method minimizes the distance between the real observations and the values estimated on the basis of the sample, by using the methodology of calculating extreme values *(Maddala [2004])*.

The purpose of raising the deviances to the second power is double:

- on the one hand it enlarges the deviances between the real and the estimated values, thus it weights the considerable deviances more strongly in comparison with the small ones, and
- on the other hand it eliminates the problem that without squaring the deviances with opposite signs would neutralize each other.

The main point of the procedure is that it looks all the observations $\boldsymbol{x_i}$ (i=1,2,…,n) as an estimate for sample mean $\boldsymbol{\mu}$, considering that $E(x_i)=\mu$. According to that, the error of the estimate is $u_i=x_i-\mu$, so the total sum of squares of the deviances measured on the sample is the following *(Hunyadi – Vita [2004])*:

$$SSE = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \qquad (4.15)$$

The Ordinary Least Squares method serves for quantifying the $\boldsymbol{\mu}$ value, in case of which the total sum of squares of the deviances is the lowest. In the viewpoint of this methodology, the best estimate of $\boldsymbol{\mu}$ is the sample mean. The $SSE(\hat{\mu})$ described by formula *(Ramanathan [2003])*:

$$SSE(\hat{\mu}) = \sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + \sum_{i=1}^{n}(\overline{x} - \hat{\mu})^2 + 2 * \sum_{i=1}^{n}(\overline{x} - \hat{\mu})(x_i - \overline{x}) \qquad (4.16)$$

And this expression is minimal, when: $\hat{\mu} = \overline{x}$.

The Ordinary Least Squares method is the most widely used procedure regarding the quantification of the parameters of regression and other descriptive models. Its comprehensive application can be explained with its robustness, it does not require knowing the population distribution, it is applicable irrespective of the type of distribution.

The calculated value of the residuum in the linear regression models which contain ***k*** explanatory variables *(Ramanathan [2003])*:

$$\hat{u}_i = \hat{y}_i - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_j x_{ij} \qquad (4.17)$$

In consequence the optimization criterion of the Ordinary Least Squares method is the minimization of the following expression:

$$SSE(\hat{\beta}) = \sum_{i=1}^{n}\hat{u} = \sum_{i=1}^{n}(y_t - \hat{\beta}_0 - \sum_{j=1}^{k}\hat{\beta}_j x_{ij} \qquad (4.18)$$

where: i=1,2,…,n: number of the observations,

j=1,2,…,k: number of the explanatory variables.

So the Ordinary Least Squares method is searching for the best fitting regression, namely in the case of which the SSE is the lowest.

One of the alternatives of the Ordinary Least Squares method is the Least Absolute Value (LAV) or Least Absolute Deviation (LAD) method, which minimizes the sum of the absolute values of the deviances instead of the sum of squares. An example for its use is the study of Bellotti and Crook *(Bellotti – Crook [2008])*, which is made known in Subsection 5.2.

(b) Method of Moments

The Method of Moments intends basically to estimate parameters of distributions on the basis that there is function-like relation between the parameters and the moments of the empirical distribution, whose type is known. Equalizing the moments calculated from the sample with the moments of the population which are defined by parameters, it deduces the particular parameter values of the population, so it searches for parameters of the population, in case of which the adequate moments of the population and of the sample are the same *(Maddala [2004]; Hunyadi – Vita [2004]).*

In case of a distribution which can be characterized with k unknown parameters, the Method of Moments use the first k empirical moments as estimates for their adequate theoretical moments *(Ramanathan [2003]).*

The literature calls the moments, in case of which the deviation from the average (μ) performs in the formula, as central moments. The general formula for their calculation is the following:

$$M(k) = E\left[(x_i - \mu)^k\right] \qquad (4.19)$$

For characterizing the simplest and most frequently used distributions (for example the normal distribution) it is enough to quantify the first two moments *(Ramanathan [2003])*:

- the first moment is the theoretical average, in other words the expected value (estimated μ), which can be quantified by calculating the average from the x_i values weighted by probabilities,
- the second central moment is the variance of the variable:

$$\sigma^2 = E\left[(x_i - \mu)^2\right] \qquad (4.20)$$

However, the larger the number of the unknown parameters of the distribution, the more moments are needed to characterize the variable appropriately.

(c) The Maximum Likelihood (ML) method

The Maximum Likelihood method assumes known population distribution, and it is directed towards the quantification of the unknown parameters, moreover it can be widely used in the field of various statistical tests as well *(Hunyadi – Vita [2004]).*

The so-called likelihood function depends on the unknown parameters of the distribution, and quantifies the probability, that in case of the certain distribution and different parameter values the very given sample derives as a result of the sampling. In other words the likelihood shows the probability of the certain observed elements *(Maddala [2004]).*

The parameter values are quantified by maximizing the likelihood function.

The use of the Maximum Likelihood method assumes that a random sample with n elements ($x_1, x_2, \ldots, x_n$) from independent observations is given relating to the variant x, where the probability distribution of x depends on an unknown parameter θ. So the density function of the variant x is: $f(x|\theta)$.

The x_i values are independent from each other, thus the joint density function is equal to the product of the density functions of the elements, namely to the probability, which is defined by the likelihood function *(Ramanathan [2003])*:

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta) = f(x_1|\theta) * f(x_2|\theta) * \ldots * f(x_n|\theta) \qquad (4.21)$$

The Maximum Likelihood method has to be applied in two different ways depending on whether the parameter θ is discrete or continuous:

- In case of discrete parameter, the likelihood function $L(\theta|x)$ has to be calculated for each possible value, than the highest of them has to be selected.
- But if the parameter θ is continuous and the likelihood function $L(\theta|x)$ is differentiable, then the maximization of the function can be carried out with derivation. The likelihood is maximal, when the first derivative is 0, and the second one is negative ($\frac{dL}{d\theta} = 0, \quad \frac{d^2L}{d\theta^2} < 0$).

Considering that the logarithm is a strictly monotonic transformation,[52] thus the likelihood function and its natural logarithm ($\ln L(\theta|x)$) are maximal at the same point, so it is a general practice to maximize the log-likelihood function instead of $L(\theta|x)$. The log-likelihood can be formalized in the following way:

$$\ln L(\theta|x) = \sum \ln[f(x|\theta)] \qquad (4.22)$$

One of the most notable disadvantages of the Maximum Likelihood method is that it assumes knowing the population distribution. A further problem is that the quantification of the conditional probabilities is very difficult in many cases, and it can occur as well that the likelihood function does not have a maximum.

However, it has the advantageous attributes that its estimates are:

- consistent, namely the estimators are unbiased, and in case of large n the variance tends to 0,
- asymptotically efficient that is in case of large n a consistent estimate, whose variance is lower, does not exist,

[52] The strict monotonicity described with schematic formula: if $x_1 > x_2$, then $f(x_1) > f(x_2)$.

- asymptotically normal, so in case of large ***n*** independently from the type of the examined distribution they follow approximately normal distribution, namely their limiting distribution is normal *(Hunyadi – Vita [2004])*.

It is worth mentioning that during the estimation of the parameters $\boldsymbol{\beta}$ in the linear regression model the maximization of likelihood function is equal to the minimization of the SSE, so the Maximum Likelihood procedure provides results, regarding the parameters $\boldsymbol{\beta}$, which are appropriate for the Ordinal Least Squares method as well, if the residuums are independent from each other *(Ramanathan [2003])*.

4.3.2. Basic model types (function forms)

Concerning the regression models there are not any generally "best" function forms, the most appropriate type has to be selected case by case on the basis of the subject of modelling. In Table 12 I present the simplest cases of the function forms, which occur most frequently in the literature:

Table 12: Regression function forms

NAME	FORMULA OF THE FUNCION
Linear	$y = \beta_0 + \beta_1 * x$
Lin-log	$y = \beta_0 + \beta_1 * \ln x$
Hyperbolic / Reciprocal	$y = \beta_0 + \beta_1 * \frac{1}{x}$
Quadratic	$y = \beta_0 + \beta_1 * x + \beta_2 * x^2$
Cross-effect	$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_1 x_2$
Log-lin	$\ln y = \beta_0 + \beta_1 * x$
Log-reciproc	$\ln y = \beta_0 + \beta_1 * \frac{1}{x}$
Log-quadratic	$\ln y = \beta_0 + \beta_1 * x + \beta_2 * x^2$
Loglinear (log-log)	$\ln y = \beta_0 + \beta_1 * \ln x$
Logistic	$\ln\left(\frac{y}{1-y}\right) = \beta_0 + \beta_1 * x$

(Self-made table on the basis of *Ramanathan [2003], pp. 258.*)

In the course of designing the regression model, appointing the scope of the adequate explanatory variables raises further questions in addition to selecting the function form. The literature offers numerous techniques for treating them, from which I mention two basic methods here:

1. In one case the best model is selected from a prearranged scope of models based on several indices. The largest disadvantage of this method is that it does not tend to find the optimal model, it is rather limited to rank the "previously appointed" models.
2. The other method is using a kind of (generally automated) stepwise procedure which develops the wanted model with iteration.

In the following chapter I demonstrate the most fundamental techniques.

4.3.3. Selection criteria and procedures

Hereinafter I present the most typical model types as well as the most important criteria for selecting the model and the explanatory variables. The quantification of the parameters of these models can be usually carried out by using one of the estimation methods reviewed in Chapter 4.3.1.

(a) Comparative indices

The coefficient of determination $\boldsymbol{R^2}$ and the sum of squares of the residuums are frequently used indices of the model fitting. Nevertheless it means a notable problem that both indices give preference to the models which contain more variables ceteris paribus, in such case as well, if the newly introduced variable hardly contributes to the predictive power of the model *(Hunyadi – Vita [2004])*. This raises basically two problems:

- on the one hand the hazard of multicollinearity is intensifying in line with increasing the number of variables in the model, which results in decline of the estimates' accuracy, and
- on the other hand the degree of freedom is determined by the number of explanatory variables, so in case of the models, which contain too many explanatory variables, the number of estimated parameters of the estimators is too high, and this prevents the statistical characteristics of estimates from enforcement *(Ramanathan [2003])*.

For the sake of eliminating these problems, numerous new model selecting criteria have been developed, from which I describe the Theil's coefficient of determination adjusted by degree of freedom and some other indices, which use the correction of the sum of squares of the residuums by some “penalty factors”.[53]

In contrast with the raw $\boldsymbol{R^2}$ index, the Theil’s coefficient of determination adjusted by the degree of freedom ($\overline{R}^2$ or R^2_{adj}) considers the number of explanatory variables appearing in the model, and applies a correction according to it. During calculating the correction factor it considers both the quantity of the observations (***n***) and the number of explanatory variables (***k***).

Its formula is the following *(Hunyadi – Vita [2004])*:

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right) \tag{4.23}$$

The best model can be selected from a certain scope of them by maximizing the $\overline{R}^2$.

The logic serves as a basis for using the adjusted index that if a new variable does not improve the explanatory power of the model significantly, then, although the value of the raw $\boldsymbol{R^2}$ slightly increases, the adjusted $\boldsymbol{R^2}$ shows decrease, indicating that it is not a good decision to extend the model with the new variable.

In addition to that, it can be diagnose according to the ***t***-test statistic of the explanatory variables, whether it is worth involving them into the model. Since if the absolute value of the t-statistic of the explanatory variable under discussion is higher than 1 ($t>1$ or $t<-1$), then introducing the variable will result in increasing the $\overline{R}^2$, which indicates the rise of the explanatory power of the model. Otherwise, that is if $-1<t<1$, it is not worth extending the model with the new variable *(Ramanathan [2003])*.

The residuum-square indices adjusted by “penalty factors” actually measure to what extent the pieces of information existing in the observations are built into the model. The model, whose information utilization is the best, can be selected by minimizing some of these model selecting criteria. These indices normally rank the models into the same order, but exceptions can occur.

Table 13 summarizes the most widely used model selecting criteria of the literature.

[53] Detailed review can be read in the study of *Engle and Brown [1985]*.

Table 13: Criteria for model selecting

INDEX[54]	NAME
$SGMASQ = \left(\frac{ESS}{n}\right)\left[1-\left(\frac{k}{n}\right)\right]^{-1}$	Sigma Square
$AIC = \frac{SSE}{n} e^{2k/n}$	Akaike Information Criterion
$FPE = \left(\frac{ESS}{n}\right)\frac{n+k}{n-k}$	Finite Prediction Error (Akaike)
$GCV = \left(\frac{ESS}{n}\right)\left[1-\left(\frac{k}{n}\right)\right]^{-2}$	Generalized Cross Validation (Craven – Wahba)
$HQ = \left(\frac{ESS}{n}\right)(\ln n)^{2k/n}$	Hannan and Quinn Criterion
$RICE = \left(\frac{ESS}{n}\right)\left[1-\left(\frac{2k}{n}\right)\right]^{-1}$	Rice Criterion
$SCHWARZ = \left(\frac{ESS}{n}\right) n^{k/n}$	Schwarz Criterion
$SHIBATA = \left(\frac{ESS}{n}\right)\frac{n+2k}{n}$	Shibata Criterion

(Self-made table on the basis of *Ramanathan [2003], pp. 173.*)

These comparative indices are not suitable for qualifying the regression models involved in the examination in absolute sense, they are only able to rank them.

(b) Stepwise procedures for model selection

During the selection of explanatory variables the automated model selection systems generally use one of the two fundamental stepwise procedures, the backward elimination and the forward strategy, or the combination of them *(Draper – Smith [1981]; Derksen – Keselman [1992]).*

- The backward elimination narrows down the model step by step from the one containing the most variables, and makes the decision about continuing the process on the basis of results of the ***t***- and ***F***-tests, as well as the values of model selecting indices.
- The forward strategy follows just the opposite logic than the previous one, it starts with that model which contains only the most correlating explanatory

[54] ***k*** indicates the number of estimated parameters, *n* the quantity of observations in all formulae.

variable, then enlarges the model step by step, applying the ***t***- and ***F***-tests as well as the model selecting indices ($\overline{R}^2$, AIC, SBC) similarly.

Concerning that the relevance of each variable also depends on that in which step it is involved in the model, so stepping-backs generally occur in the course of using both strategies:

- This means in case of the backward proceeding that variables can be put back into the model, which previously fell out exactly in the course of this process.
- In case of the forward strategy just on the contrary, some of the variables having been involved previously can fall out from the model again, if they become redundant in consequence of the new variable *(Hajdu [2004])*.

In most cases neither the backward elimination nor the forward strategy is used exclusively during the model selection, applying some kind of combination of the procedures is more widely used, where the professional aspects take an important role as well.

4.3.4. Testing the model specification

Following the brief review of the model selection, I present the most important procedures directed towards testing the model specification. In this framework I mention only those tests, which deal with embedded hypotheses, namely in case of which the limited model is a "restricted" version of the unlimited one.

(a) Tests of single coefficients (*t*-test, *p*-value approach)

Using the ***t***-test and the ***p***-value approach are universal tools for testing the regression coefficients one by one.

These proceedings are based on the assumption that the distributions of the estimated coefficients are normal, whereas $\frac{SSE}{\sigma^2}$ follows a χ^2-distribution whose degree of freedom is n-k-1, where n is the number of observations, and k is the quantity of explanatory variables in the model. The null hypothesis (H_0: β_i=0) of the ***t***-test draws up that one of the explanatory variables (x_i) in the model is not relevant, so its coefficient β_i does not differ from 0 significantly *(Ramanathan [2003])*.

The comparison of the ***t***-test statistic ($t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$), which follows a t-distribution with ***n-k-1*** degree of freedom, and the critical value relating to significance level of α shows, whether the explanatory value under discussion contributes to the predictive power of the model *(Hunyadi – Vita [2004])*. Accepting H_0 hypothesis means that the explanatory variable has to be fallen out of the model, while rejecting of the null hypothesis indicates the relevance of the variable.
With this methodology linear combination of regression coefficients can be tested as well, analogously with the above.

The ***p***-value approach offers an alternative opportunity to making decision about accepting the hypothesis. According to it, the significance level (***p***) relating to the ***t***-test statistic has to be compared to the predefined α. If $p \geq \alpha$, then the null hypothesis has to be accepted, otherwise has to be rejected. The ***p***-value is actually the first kind error, namely the probability that the true H_0 hypothesis will be rejected *(Hunyadi – Vita [2004])*.

(b) Wald test

The Wald test is frequently mentioned in the literature as multiple/grouped ***F***-test (for example *Spanos [1999]*), considering that this procedure enables testing which relates to more than one explanatory variable, in contrast with the ***t***-test. The term ML-(Maximum Likelihood) test is also in general use *(Hajdu [2003])*.
The Information Matrix (***I***) is an "auxiliary function" of these testing proceeding, which is the expected value of the second derivative of the log-likelihood function with respect to β, in other words the concavity of the log-likelihood function *(Ramanathan [2003])*:

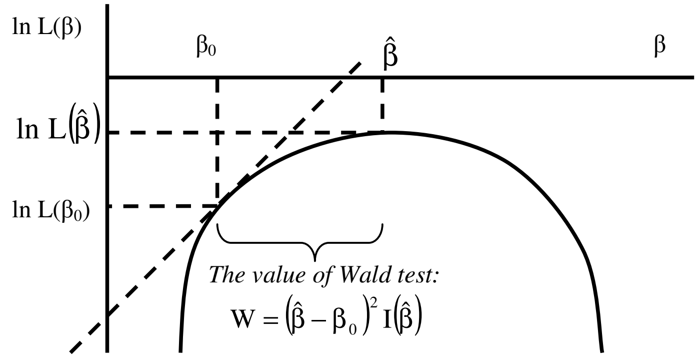
$$I(\beta) = -E\left[\frac{\partial^2 \ln L}{\partial \beta^2}\right] \tag{4.24}$$

Quantification of the Wald test using the ***I*** Information Matrix:

$$W = \left(\hat{\beta} - \beta_0\right)^2 I\left(\hat{\beta}\right) \tag{4.25}$$

Figure 5 shows for the case of the model which contains only one explanatory variable that the Wald test bases on the deviation between the estimated and the real parameter values.

Figure 5: Demonstration of the Wald test in case of the model which contains one parameter



(Self-made figure on the basis of *Ramanathan [2003], pp. 306.*)

In the course of using the Wald test, in the first step an unrestricted (***U***) and a restricted (***R***) model have to be defined, which differ from each other in that respect that the latter one contains ***k-m*** less variables (***x_{m+1}, x_{m+2}, …, x_k***) *(Ramanathan [2003]).*

$$U:\quad y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u \qquad (4.26)$$

$$R:\quad y = \beta_0 + \sum_{j=1}^{m} \beta_j x_j + v \qquad (4.27)$$

The testing is directed towards the fact, whether leaving out the examined ***k-m*** variables (***x_{m+1}, x_{m+2}, …, x_k***) does not damage significantly the regression fitting. The purpose of the test is to decide, whether it is worth replacing the unrestricted model with the restricted one.

According to the null hypothesis of the test the regression coefficients of the ***k-m*** variables left out are 0 (***H_0***: ***$\beta_{m+1}=\beta_{m+2}=\ldots=\beta_k=0$***), so the rejection of the null hypothesis means that at least one of the explanatory variables is significant in the model.

The Wald's ***F***-test statistic can be formalized in the following manner:

$$W = F = \frac{(SSE_R - SSE_U)/(k-m)}{SSE_U/(n-k-1)} = \frac{(R_u^2 - R_R^2)/(k-m)}{(1-R_u^2)/(n-k-1)} \qquad (4.28)$$

The sum of squares of variates whose distribution is independent standard normal, follows χ^2-distribution, thus SSE_U/σ^2 and $(SSE_R - SSE_U)/\sigma^2$ also have χ^2-distribution with degree of freedom ***n-k-1*** and ***k-m***. So ***F***-test statistic is a quotient of two χ^2, in the case of that model which contains only one explanatory variable its formula is: $W = \frac{nR^2}{1-R^2}$, and its distribution is also χ^2, in case of large sample *(Maddala [2004]).*

In the course of using the Wald test the comparison with the critical value relating to significance level of $\boldsymbol{\alpha}$ or the $\boldsymbol{p}$-value approach can serve as a basis for the decision whether to accept or reject the null hypothesis. Accepting the null hypothesis means in this case that none of the $\boldsymbol{k\text{-}m}$ explanatory variables ($\boldsymbol{x_{m+1}, x_{m+2}, \ldots, x_k}$) contributes considerably to the predictive power of the regression model, thus it is worth leaving them out of the model *(Engle [1984])*.

The test which examines the general goodness of the fitting can be considered as a special case of the Wald test. The special nature issues from the fact that the unrestricted model (***U***) has to be compared with the so-called super-restricted model (***SR***), which can be formalized in the following manner *(Ramanathan [2003])*:

$$U: \quad y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + u \tag{4.29}$$

$$SR: \quad y = \beta_0 + w \tag{4.30}$$

The test ($\boldsymbol{H_0}$: $\boldsymbol{\beta_1}=\boldsymbol{\beta_2}=\ldots=\boldsymbol{\beta_k}=0$) examines in this case, whether the hypothesis is true that apart from the constant none of the explanatory variables is significant in the model.

Having used that $\hat{\beta}_0 = \bar{y}$ in the super-restricted model, the formula of the Wald's F-statistic also differs from the general case:

$$W = F = \frac{\left(SST_U - SSE_U\right)/k}{SSE_U/(n-k-1)} = \frac{SSR_U/k}{SSE_U/(n-k-1)} = \frac{R^2/k}{\left(1-R^2\right)/(n-k-1)} \tag{4.31}$$

In case of accepting the null hypothesis the whole model has to be re-specified, because none of its explanatory variables is able to contribute significantly to the explanatory power of the model.

(c) Lagrange Multiplier (LM) test

In contrast with the Wald test the Lagrange Multiplier test *(Silvey [1959])* is used for deciding, whether it is worth putting one or more explanatory variables into the model, so the LM test carries out a test which is directed towards expanding the model. It is often mentioned in the literature as Rao's score test.

It grounds on the Lagrange Multiplier procedure which is widely used in conditional optimization: the restrictions relating to the parameter values which are formulated in the null hypothesis mean the conditions, and the goal is to maximize the log-likelihood function under these conditions *(Ramanathan [2003])*.

The sign function which is the essence of the proceeding is the Lagrange Multiplier itself:

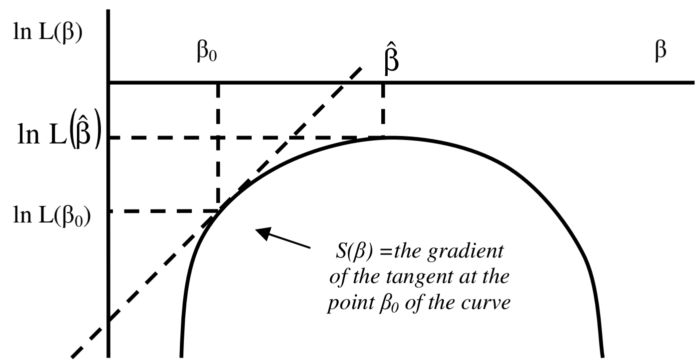
$$S(\beta) = \frac{\partial \ln L}{\partial \beta} \qquad (4.32)$$

According to that the test statistic can be defined with the following formula:

$$LM = S^2(\beta_0) * I(\beta_0)^{-1} \qquad (4.33)$$

The value of the sign function is the partial derivative of the log-likelihood function, namely the gradient of the tangent at the point $\boldsymbol{\beta_0}$ of the function. Figure 6 illustrates that in case of the model which contains only one parameter.

Figure 6: Demonstration of the LM test in case of the model which contains one parameter



(Self-made figure on the basis of *Ramanathan [2003], pp. 306.*)

The Lagrange Multiplier principle is directed towards examining, whether the partial derivative of the log-likelihood relating to the value estimated by the Maximum Likelihood procedure is 0 in the unrestricted (***U***) model, whereas it is differing in the restricted (***R***) model *(Hajdu [2003])*.

The hypothesis ***S(β)***=0 issues from the fact that the value of the test is 0, if the estimated parameter value corresponds to the real one (the tangent is horizontal).

In contrast with the Wald test the LM test compares the unrestricted (***U***) model to the restricted one (***R***). In this case the unrestricted model means the alternative to the restricted one, which contains ***k-m*** less variables ($x_{m+1}, x_{m+2}, \ldots, x_k$) *(Ramanathan [2003])*.

$$R: \quad y = \beta_0 + \sum_{j=1}^{m} \beta_j x_j + u \qquad (4.34)$$

$$U: \quad y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + v \qquad (4.35)$$

The null hypothesis of the test states here likewise that the regression coefficients of the ***k-m*** variables are zero (H_0: $\beta_{m+1}=\beta_{m+2}=\ldots=\beta_k=0$). So if there is at least one explanatory variable among the ***k-m*** pieces, which contributes significantly to the explanatory power of the model, then the null hypothesis has to be rejected.

The residuum of the restricted model can be calculated by the following formula:

$$\hat{u}_R = y - \hat{\beta}_0 - \sum_{j=1}^{m} \hat{\beta}_j x_j \qquad (4.36)$$

This residuum has to be explained with the explanatory variables which are left out, using an auxiliary regression.

In case of large samples the value nR^2 of the auxiliary regression follows χ^2-distribution with the ***k-m*** degree of freedom *(Engle [1982]; Maddala [2004])*, and if its fitting is appropriate at the significance level of α ($nR^2 > \chi^2_{k-m}(\alpha)$), then the null hypothesis has to be rejected, because it is worth putting in at least one of the explanatory variables left out. Selecting the relevant variable requires further examinations.

(d) Ramsey's RESET test

The Ramsey's RESET test (Regression Specification Error Test) is a simple procedure which is directed towards examining the specification error of the regression *(Ramsey [1969])*.

As a first step the regression ($\hat{y}_t$) of the original model has to be constructed according to the Ordinal Least Squares method, then further estimation has to be made complemented with the variables $\hat{y}_t^2$, $\hat{y}_t^3$ and $\hat{y}_t^4$. Having carried out the Wald's ***F***-test relating to rejecting these three variables, accepting the null hypothesis refers to the not appropriate specification of the model.

Although this test does not give any information about the type of the specification error, but it can be used very well as a diagnostic instrument *(Wooldridge [2009])*.

(e) Likelihood Ratio (LR) test

The Likelihood Ratio (LR) test is a classical testing procedure. As regards its calculation, it bases on a quotient of the likelihood functions, where the value of the likelihood function according to the null hypothesis appears in the numerator, and the maximum value of the same function without restrictions in the denominator *(Maddala [2004])*.[55]

Respecting whether an explanatory variable contributes to the predictive power of the model, the following Likelihood Ratio can be defined *(Ramanathan [2003])*:

[55] Some authors, for example *Spanos [1999]* understand the Likelihood Ratio inversely.

$$\lambda = \frac{L(\beta_0)}{L(\hat{\beta})} \quad (4.37)$$

where: $L(\beta_0)$ is the Likelihood function in case of the null hypothesis $\beta=\beta_0=0$.
Considering that the value of the Likelihood function relating to the unrestricted model (denominator) can never be less, than the value according to the null hypothesis (nominator), so the quotient falls between 0 and 1, and the H_0 null hypothesis has to be rejected, if $\lambda \leq K$, where K, relating to a significance level of α, is the following:

$$P(0 \leq \lambda \leq K | \beta = \beta_0) = \alpha \quad (4.38)$$

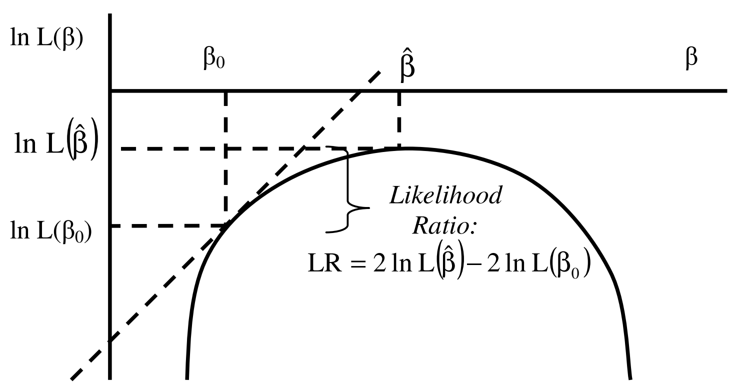
Usually it can be converted into some simple test statistic (for example ***t***-, ***F***- or χ^2-test) *(Mood et al. [1974]*), otherwise large-sample tests can be used. The following statistic follows χ^2-distribution with the same degree of freedom as the number of the restrictions *(Hajdu [2003])*:

$$LR = -2\ln\lambda = 2\ln L(\hat{\beta}) - 2\ln(\beta_0) \quad (4.39)$$

It is worth mentioning that in case of the model which contains only one explanatory variable, the test statistic takes a simpler form: $LR = -\ln(1-R^2)$, where R^2 is the raw coefficient of determination of the unrestricted model *(Maddala [2004])*.

Figure 7 illustrates the underlying logic of the LR test relating to the model which contains only one explanatory variable.

Figure 7: Demonstration of the LR test in case of the model which contains one parameter



(Self-made figure on the basis of *Ramanathan [2003], pp. 306.*)

The comparison of the test procedures

In the course of testing the significance of the $\boldsymbol{\beta}$ parameters the Wald, the Lagrange Multiplier (LM) and the Likelihood Ratio (LR) tests give the same result, if the log-likelihood function is quadratic *(Engle [1982]; Buse [1982])*.

Otherwise, if this condition is not valid, namely the degree of the log-likelihood function is more then two, then the equality does not prevail, but the following relation is in existence in this case as well *(Maddala [2004])*:

$$W \geq LR \geq LM \qquad (4.40)$$

It follows from the foregoing that the Wald test is the severest and the Lagrange Multiplier is the least severe testing procedure in all cases. Considering that all three tests lead to asymptotically equivalent and consistent results, in the course of deciding which of them should be used by the analyst, basically the characteristics of the certain problem have to be kept in view *(Hajdu [2003])*.

4.4. Models with dummy dependent variable

Following the review of the aspects in connection with the model specification, hereinafter I deal with some special models whose dependent variable is a dummy.

4.4.1. The most important model types

In Chapter 4.3.2 I have already mentioned the most important function types, nevertheless it is worth emphasising the models with dummy dependent variable, because these also appear frequently in the literature of LGD estimation, in spite of their special nature. Their widespread application is typical in econometric analyses, where the goal is to explain some qualitative (nominal, discrete) variables.

The value set of the dummy variable may consist of two or more elements *(Maddala [1983]; Amemiya [1981]; Cox [1970])*, but from the viewpoint of the present dissertation basically the treatment of the dichotomous case is reasoned, thus the presentation of the polychotomous models gets only a marginal part.

(a) Linear probability model

The linear probability model is a special kind of the models with discrete dependent variable: it is a linear regression containing ***k*** explanatory variables ($\boldsymbol{x_i}$), in which the value of the dependent variable is 1 in case of supervention of a predefined event, and otherwise it is 0 *(Ramanathan [2003])*.

The linear probability model can be formalized by the following equation:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \tag{4.41}$$

where: $\boldsymbol{u_i}$ is the random variable (error factor), whose expected value is: $E(u_i)=0$.

The coefficients $\boldsymbol{\beta}$ show, how the changing of some explanatory variable influences the probability of the supervention of the event.

The conditional expected value of the dependent variable means the probability of the event under discussion in case the value of the explanatory variable is $\boldsymbol{x_i}$, so ***y***, calculated from the regression equation, can be interpreted as an estimate relating to the probability of the event in case of some certain values of ***x*** *(Maddala [2004])*.

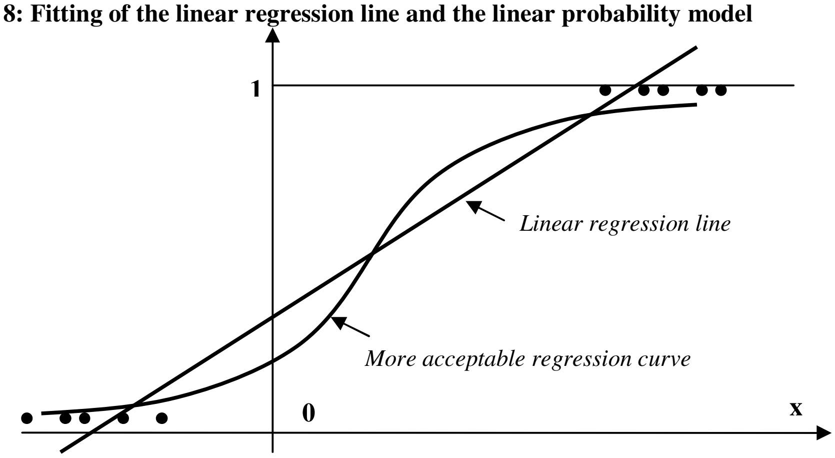
Resulting from that in the course of LGD calculation by estimating ***y*** it can be quantified, how much its probability is that a deal, which may be characterized by some given ***x*** features will "be cured", namely taking the advantage of a little simplification 100% recovery can be realized from it. In this case the definition of the dummy variable:

$$y = \begin{cases} 1, & \text{if the deal 'is curing'} \\ 0, & \text{otherwise} \end{cases} \tag{4.42}$$

Nevertheless the practical use of the linear probability model is rather scarce, because the value of ***y*** can be only 0 and 1 *(Nerlove – Press [1973])*, and this model does not guarantee that the estimated probabilities fall between 0 and 1 in all cases *(McGilvray [1970])*.

Figure 8 shows that in the linear probability model it is possible to construct a curve, which fits better than the linear regression line.

Figure 8: Fitting of the linear regression line and the linear probability model



(Self-made figure on the basis of *Maddala [2004], pp. 372.*)

The heteroscedasticity of the residuums can be mentioned as its notable disadvantage, in consequence of which using the Ordinary Least Squares (OLS) method results in inefficient estimates *(Ramanathan [2003]; Maddala [2004])*.

The heteroscedasticity derives from the fact that the distribution of the error factors u_i is not normal, but binomial, and the variance is not stable, but depends on i:

$$\operatorname{var}(u_i) = \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\left(1 - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right) = E(y_i)[1 - E(y_i)] \qquad (4.43)$$

Considering that the OLS-estimates are linear combinations of random variables which are the same (not definitely normal) distribution, so the assumption of asymptotic normal distribution is affirmable under the central limit theorem. However, the validity of the statistical tests is questionable because of the heteroscedasticity even in this case. The use of the Weighted Least Squares (WLS) method can be an alternative solution, but only the asymptotic efficiency is accessible by this procedure, it remains an unsolved problem that the estimated probabilities do not always fall between 0 and 1 *(Ramanathan [2003])*.

(b) Linear discriminant function

The linear discriminant function serves for classifying the observations into two categories (***π₁*** and ***π₂***), if we suppose that from the ***n*** quantity of observations $\boldsymbol{n_1}$ quantity belong to the $\boldsymbol{\pi_1}$ group and $\boldsymbol{n_2}$ quantity to the $\boldsymbol{\pi_2}$ group. The function is a linear function containing ***k*** explanatory variables, which can be formalised by the following formula *(Maddala [2004])*:

$$Z = \lambda_0 + \sum_{i=0}^{k} \lambda_i x_i \qquad (4.44)$$

For the sake of developing the model whose discriminatory power is the very best, the goal is to quantify $\boldsymbol{\lambda_i}$ values, in case of which the deviation within the groups is negligible for the deviation between the groups. By formula:

$$\frac{\sigma^2_{Z_{\text{csoportok közötti}}}}{\sigma^2_{Z_{\text{csoportokonbelülii}}}} \rightarrow \max \qquad (4.45)$$

Fisher [1936] also showed that there is a close connection between the linear discriminant function and the multivariate regression:

$$y = \begin{cases} \dfrac{n_2}{n_1 + n_2}, & \textit{if belongs to the } \pi_1 \textit{ group} \\ -\dfrac{n_1}{n_1 + n_2}, & \textit{if belongs to the } \pi_2 \textit{ group} \end{cases} \qquad (4.46)$$

He deduced that because the linear discriminant function is actually the transformed version of the linear regression model by adding $-\dfrac{n_1}{n_1 + n_2}$ to it, these two models differ from each other only in terms of the estimated value of the constant.

(c) Probit and logit models

It is an alternative for the linear regression model to carry out a monotonic transformation, which limits the possible value range of the result variables *(Ramanathan [2003])*, so ensures the falling of the estimated probabilities between 0 and 1.

A linear regression constitutes the starting point in this case as well:

$$y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \qquad (4.47)$$

where: y_i^* is an underlying latent dummy variable, whose observation is impossible, but it can be defined in the following manner *(Maddala [2004])*:

$$y_i = \begin{cases} 1, & \textit{if } y_i^* > 0 \\ 0, & \textit{otherwise} \end{cases} \qquad (4.48)$$

The quantification of the $\boldsymbol{\beta}$ regression parameters basically can be carried out by using ML procedure based on the under-mentioned function:

$$L = \prod_{p_i=0} F\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) * \prod_{p_i=1} \left[1 - F\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)\right] \qquad (4.49)$$

where: F is the cumulative distribution function for the $\boldsymbol{u}$ error factor.

The types of the models differ from each other in that respect, what kind of transformation they apply and what kind of assumptions they have relating to the distribution of the ***u*** error factor. The best-known types are the probit and the logit models *(Maddala [2004])*.

In case of the probit model the standard normal distribution describes the estimated probability *(Greene [2003])*:

$$\hat{p}_i = \Phi\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) = \int_{-\infty}^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}} \varphi(z) dz \qquad (4.50)$$

where: $\Phi(.)$: the standard normal cumulative distribution function,

$\varphi(.)$: the standard normal density function.

The logit model differs from it in that respect that it uses the logistic cumulative distribution function for characterising the estimated probability *(Ramanathan [2003])*:

$$\hat{p}_i = F\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right) = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}} = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}\right)}} \qquad (4.51)$$

where: $F(.)$: the logistic cumulative distribution function.

In case of the logit model the cumulative distribution function can be written in a close formula, so the natural logarithm of the estimated probability's quotient (odds = $\frac{\hat{p}_i}{1-\hat{p}_i}$) is a linear function of the explanatory variables *(Ramanathan [2003]; Everitt [2002])*:

$$\ln(\text{odds}) = \text{logit} = \ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \qquad (4.52)$$

Cox [1970] published the use of the logit transformation in the area of analysing the logistic qualitative dependent variables at first.

Considering that the logistic regression has an important role also in the course of estimating the risk parameters, so I present it in details in Chapter 4.4.2.

(d) Censored normal regression model

The censored normal regression model, which is also one of the relevant special regression methodologies referring to the estimating the LGD, was applied by *Tobin [1958]* at first, so it is known in the literature as Tobit model ("Tobin's probit").

The object of the observation is also the $\boldsymbol{y_i^*}$ latent variable, similarly to the probit and logit models. The difference arises from the fact that whereas the result variable is a dummy for the probit and logit models, in case of the Tobit model $\boldsymbol{y_i}$ is to be observed if $\boldsymbol{y_i^*}>0$.[56]

Maddala [2004] formalised this model in the following manner:

$$y_i = \begin{cases} y_i^* = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}, & if\ y_i^* > 0 \\ 0, & otherwise \end{cases} \qquad (4.53)$$

Because the values of $\boldsymbol{y_i^*}$ are censored in case of $\boldsymbol{y_i^*}\leq 0$, that is not their real values are observed, so the distribution of the error factor ($\boldsymbol{u_i}$) is a truncated normal distribution, its expected value depends on $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{x_i}$, consequently differs observation by observation. Arising from that the Ordinal Least Squares method (OLS) is not proper for estimating the parameters, but the Maximum Likelihood procedure (ML) has to be followed, since the estimates are not unbiased.

The likelihood function of the Tobit model can be calculated by the following formula:

$$L = \prod_{y_i > 0} \frac{1}{\sigma} f\left(\frac{y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right)}{\sigma} \right) * \prod_{y_i \leq 0} F\left(-\frac{\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \right)}{\sigma} \right) \qquad (4.54)$$

where: $\Phi(.)$: the standard normal cumulative distribution function,

$\varphi(.)$: the standard normal density function.

The desired estimated parameters arise from maximizing the likelihood function by $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ *(Maddala [2004])*.

In several cases maximizing the log-likelihood function (the logarithm of the likelihood function) instead of the likelihood function is simpler. So for example *Bellotti and Crook [2008]* in their study presented later, used the following log-likelihood function for estimating the $\boldsymbol{\beta}$ coefficients and the variance of the residuals:

$$\ln L(\beta, \sigma) = \sum_{y_i > 0} -\frac{1}{2} \left[\ln(2\pi) + \ln \sigma^2 + \frac{(y_i - \beta * x_i)^2}{\sigma^2} \right] + \sum_{y_i = 0} \ln \left[1 - N\left(\frac{\beta * x_i}{\sigma} \right) \right] \qquad (4.55)$$

Although the archetype of the Tobit model contains only a one-tailed truncation, it can be extended to both sides *(Greene [2003])*, so it enables the censoring of the result

[56] Another value can also be defined as a limit (not only 0), moreover it is also possible that not the lower limit is defined in the model, but the upper one.

variable into the desired limits (for example 0 and 1 in case of the LGD) without applying any other transformation procedures.

4.4.2. Logistic regression: use of the logit model

"The logistic regression is predestined to quantify the probability of the emergence of an uncertain categorical result variable's certain category, while providing known outcomes of other explanatory variables" (Hajdu [2004], pp. 279.).

On the basis of the conditional probability the function can be defined, according to which the certain observation units can be classified into one of the predetermined categories.

This model does not have any assumptions regarding the distribution of the explanatory variables, so it is particularly suitable for classifying the result variables whose distribution is discrete, since in this case the discriminant analysis is not applicable because of the unfulfilment of the multivariate normality of the explanatory variables.

The dichotomous logistic regression model carries out the categorisation of the observations based on the $\boldsymbol{\beta}$ parameters in a way that it defines the critical value (cut-off) of the certain event's emergence, and it classes the observations, in case of which the conditional probability exceeds this value, into the given category, and the other observations into the complementary one *(Hajdu [2004])*.

Quantifying the cut-off is generally done by minimizing the function which characterises the loss deriving from the wrong classification.

The polychotomous (multinomial) model differs from it in that respect that the observations are classified not into two, but more groups, so the polychotomous model can be originated from several dichotomous logistic regressions. If the order of the categories does not contain any information, then the odds-ratios of all the category pairs can be interpreted, however, it is practical to compare the chance of the certain categories to the probability of the emergence of an arbitrarily chosen "base" category *(Amemiya [1981])*. In this so-called nominal polychotomous model separate linear regressions explain the logits of each pair of categories.[57]

[57] The ordinal polychotomous models mean the other type of the polychotomous models, in case of which the order of the categories also contains information. A detailed description can be read for example in the work of *Amemiya [1981]*.

(a) Estimating and explaining parameters

In case of ***n*** quantity of independent, conditional observations of the result variable the outcomes derive from a Bernoulli process, so estimating the parameters is carried out by the Maximum Likelihood procedure. Providing the certain range of the explanatory variables those parameter values have to be quantified, in case of which the likelihood function and its logarithm is maximal *(Hajdu [2003])*.
The weighted form of the likelihood function:

$$L = \prod_{x_i} P_{xb}^{f_x} \left(1 - P_{xb}\right)^{n_x - f_x} \qquad (4.56)$$

where: x_i: covariant (a certain combination of the explanatory variables' emergences),

P_{xb}: the estimated value of the probability of the event's emergence providing the **x** covariant and the **b** parameters,

n_x: the incidence rate of the **x** covariant,

f_x: the observed quantity of the event's emergences providing the **x** covariant.

The maximizing of the log-likelihood function is deducible from using the non-linear least squares method. *Hajdu [2003]* proposed the iteratively re-weighted Gauss-Newton non-linear least squares method for estimating the parameters. On the basis of that the object is to maximize the weighted sum of squares of all the covariants:

$$\sum_{x} \frac{\left(f_x - n_x P_{xb}\right)^2}{n_x P_{xb}\left(1 - P_{xb}\right)} \qquad (4.57)$$

Proceeding from the theorem that a significant relation exists between some explanatory variable and the result variable, if the partial coefficient of the regression is not 0 at a given confidence level, the testing of the significance of the parameters can be done inversely by two types of methods *(Hajdu [2004])*:

- The $z = \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)}$ statistic is suitable for both one-tailed and two-tailed testing,[58] whose distribution is asymptotically standard normal for large samples in case of the validity of the $\boldsymbol{H_0}$: $\boldsymbol{\beta_j}$=0 null hypothesis.
- The testing against the two-tailed $\boldsymbol{H_1}$: $\boldsymbol{\beta_j}$<>0 alternative hypothesis also can be carried out by using the $\chi_1^2 \cong \left(\frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)}\right)^2$ Wald statistic, which also follows roughly

[58] The $se\left(\hat{\beta}_j\right)$ is the estimated, asymptotic standard error.

χ^2-distribution with the degree of freedom ***df=1*** for large samples *(Wooldridge [2009])*.

The handling of the nominal variables is special in this model as well. If the quantity of the possible categories is ***K***, then ***K-1*** dummy variables are needed.
The impact of the nominal variables can be formalised in the following manner:

$$\beta_{j_1} D_{j_1} + \beta_{j_2} D_{j_2} + \ldots + \beta_{j_{K-1}} D_{j_{K-1}} \qquad (4.58)$$

Interactions of the explanatory variables can also be placed in the logistic regression, so the quantity of the model's parameters can be reduced by coding the interaction x_j*x_k*... of the nominal variables with a product of well-chosen dummy variables.

The results of the logit model can be easily interpreted, since the $e^{\beta j}$ factor directly shows what partial multiplicative effect is made by a unity absolute increment of the x_j explanatory variable ceteris paribus on the odds, namely on the quotient of the emergence's probability of the certain event and the complementary event *(Maddala [2004])*.
Considering that

$$\text{odds}^{x_j+1} = e^{\beta_j} * e^{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}} = e^{\beta_j} * \text{odds}^{x_j}, \qquad (4.59)$$

and the partial derivative of the emergence's probability of the certain event by the x_j explanatory variable is $\beta_j * \hat{p} * (1-\hat{p})$, so the unity absolute change of the x_j has a $\beta_j * \hat{p} * (1-\hat{p})$ additive effect on the probability of the certain event *(Ramanathan [2003])*.
The β_j indicates the partial additive impact of a unity change of the x_j explanatory variable on the logit. However, if x_j and x_t are in interaction with each other and β_j indicates the effect of this interaction on the logit, then the impact of a unity change of the x_j on the logit also depends on the current value of x_t *(Hajdu [2004])*:

$$e^{\beta_j + \beta_{jt} x_t} = e^{\beta_{jt}} e^{\beta_{jt} x_t} \qquad (4.60)$$

(b) Fitting of the logistic model

The fundamental manner of testing the hypotheses which examine the fitting of the logistic model is the use of the Maximum Likelihood ratio test,[59] whose degree of freedom is ***b***, and takes the following form *(Hajdu [2004]; Wooldridge [2009])*:

[59] Other tests also occur in the literature. For example *Maddala [2004]* mentions the Effron's and the Amemiya'a measure which are based on the sum of residual squares, the Cragg-Uhler's pseudo R^2 and

$$\chi^2_{(b)} \cong -2\ln\lambda = -2\ln\frac{L(\hat{\beta}_{H_0})}{L(\hat{\beta}_{H_1})} = -2(l_0 - l_1) \qquad (4.61)$$

where: $\hat{\beta}_{H_0}$: the restricted Maximum Likelihood estimate of the parameter vector providing the $\boldsymbol{H_0}$,

$\hat{\beta}_{H_1}$: the unrestricted Maximum Likelihood estimate of the parameter vector,

b: the quantity of the parameters, which have to be estimated in the $\boldsymbol{H_1}$ model.

Hereunder I present certain variants of this testing procedure referring to selecting the explanatory variables and to examining the goodness-of-fit.

The most frequently used method of developing the model with maximal explanatory power is the stepwise algorithm (Chapter 4.3.3), which puts in or leaves out of the model only one variable step by step, and in the meantime it examines the run of the likelihood function's value. The test statistic, whose degree of freedom is ***b***, can be written down in the following form *(Hajdu [2003])*:

$$\chi^2_{(b)} \cong -2\ln\left(\frac{L_{b_0}}{L_{b_1}}\right) = 2\left|\ln L_{base} - \ln L_{current}\right| \qquad (4.62)$$

where: b_0: the vector of the estimated parameters of the restricted or enlarged model,

b_1: the vector of the estimated parameters of the model in the current step,

L_{base}, $L_{current}$: the likelihood function of the base and the current model,

b: the quantity of the parameter restrictions composing the difference between the models.

In the given step always the significant explanatory variable has to be put into the model, whose test statistic value is the highest, or the one has to be left out of the model whose test statistic value is the lowest. So using the stepwise algorithm in the course of a multistep iteration we can develop the "optimal" model, which has the maximal explanatory power.

This procedure can also be used jointly for groups of variables for the sake of examining whether their simultaneous leaving-out of the model or their simultaneous putting-in into the model considerably influences the fitting of the tested model.

---

the McFadden's R^2, but he notes that they are not equivalent in case of the models which contain qualitative dependent variable.

The Pearson's statistic whose degree of freedom is $df=n_l-n_b$ and which grounds on the Lagrange Multiplicator theorem is a widely used alternative test for characterising the model fitting comprehensively *(Hajdu [2003])*:

$$\chi^2_{\mathrm{Pearson}} = \sum_x \left(\frac{(f_x - n_x P_{xb})^2}{n_x P_{xb}} + \frac{((n_x - f_x) - n_x(1-P_{xb}))^2}{n_x(1-P_{xb})} \right) = \sum_x \frac{(f_x - n_x P_{xb})^2}{n_x P_{xb}(1-P_{xb})} \qquad (4.63)$$

Furthermore the Wald theorem is also suitable for preparing a test function, whose degree of freedom is $df=n_l-n_b$ as well, and which extensively characterises the fitting of the model:

$$\chi^2_{\mathrm{Wald}} = \sum_x \left(\frac{(f_x - n_x P_{xb})^2}{f_x} + \frac{((n_x - f_x) - n_x(1-P_{xb}))^2}{n_x - f_x} \right) = \sum_x \frac{(f_x - n_x P_{xb})^2}{p_x(n_x - f_x)} \qquad (4.64)$$

Considering that the extreme values intensively influence the values of the fit indices presented previously, the quantification of the standardised residuums also helps to select the proper model by supporting the indication of the extreme values. The most frequently used type is the Pearson's standardised residuum:

$$e_x^P = \frac{p_x - P_{xb}}{\sqrt{\dfrac{P_{xb}(1-P_{xb})}{n_x}}} \qquad (4.65)$$

In case of large n_x the distribution of the Pearson's residuum is approximately standard normal, thus the sum of squares is roughly χ^2-distributed with the same degree of freedom as the quantity of the sampling logits.

If the $n_x P_{xb}$ is not large enough or the frequencies of the certain covariants are 1 in many cases (a large proportion of the observations compose substantive sampling logits), then neither the $\chi^2_{\mathrm{Pearson}}$, nor the χ^2_{Wald}, nor the e_x^P is relevant. In such cases the Hosmer-Lemershow test can be applied for examining the fitting, which is actually a special Pearson's χ_p^2 test. It has to be carried out on the groups defined by the quantiles of the non-decreasing order on the basis of the estimated probabilities of the sampling logits, and its degree of freedom is $df=g-2$, where g is the quantity of the groups generated by the quantiles *(Hajdu [2004])*.

(c) The generalized linear model (GLIM)

The logistic model is actually a special case of the generalised linear model (GLIM: Generalized Linear Interactive Modelling) worked out by *Nelder and Wedderburn*

[1972], which enables the linear modelling of explanatory variables whose measuring scale is different.

The elements of the GLIM model are the followings *(Füstös et al. [2004])*:

y: the observation values of the result variable (y_i, where i=1,…,n),

μ: expected value = E(y), the systematic component of the result variable (μ_i, where i=1,…,n),

e: error factor, the random component of the result variable ($y_i = \mu_i + e_i$),

η: linear estimate,

$x_1, \ldots, x_k$: explanatory variables (x_{ij}, where i=1,…,n; j=1,…,k).

This general model assumes that the distribution of the result variable's probability is normal, binomial, Poisson or gamma with expected value of μ. It is conceivable on the basis of Chapter 4.4.1 about the distribution of the logit and probit models' result variables that the assumption of the binomial distribution is valid.

In case of the GLIM the linear structure can be formalised as the function of the unknown β_j parameters in the following manner:

$$\eta_i = \sum_{j=1}^{k} \beta_j x_{ij} \qquad (4.66)$$

The interpretation of the β_j parameters depends on the measuring scale of the x_{ij}:

- if x_{ij} is dichotomous (namely it means the inherence of the explanatory variable into the given category), then β_j expresses the impact of the given category,
- but if x_{ij} is an observed value of a quantitative variable, then β_j is the weight of the j-th variable.

The function which describes the connection between the systematic component of the dependent variable (η) and the linear estimate (μ) can be sevenfold *(Füstös et al. [2004])*:

- Identity: $\eta = \mu$,
- Logarithm: $\eta = \ln(\mu)$,
- Inverse: $\eta = 1/\mu$,
- Square root: $\eta = \mu^{0,5}$,
- Logit: $\eta = \ln(p/(1-p))$,
- Probit: $\eta = N(\mu)$,
- Complementary log-log: $\eta = \ln(-\ln(1-p))$.

It is already directly apparent from this list that the probit and logit models discussed previously can be considered as special cases of the generalised linear model.

It is unquestionable that the logistic regression is difficult to handle in many aspects, since numerous special problems derives from the discrete distribution of the result variable, however, it offers a classification method, which can be properly used in the course of calculating the LGD as well.

5. Analyses in the literature

Following the demonstration of the methodological aspects, further on I focus on the empirical area of calculating the LGD. First I make known the examinations and conclusions published in the literature,[60] then in the following chapter I outline my own research and its results.

The LGD models, presented in the literature, can be arranged basically into two types:

- the members of one of the model types prepare forecasts on the basis of historical recovery and LGD data by using analytical procedures, particularly applying regression methodologies, modelling distributions, while
- those from the other group, deal with modelling the recoveries as stochastic variants mainly on the basis of market information.

The models, belonging to the latter type, are not linked directly with the topic of my research, so I disregard the outline of the empirical works dealing with them. Moreover, concerning that the aims of the present thesis are basically to study the models and to develop a new one, instead of emphasising the values calculated and published by the researchers, I focus on methodological aspects such as for example specifying the factors, which influence the recovery rate, studying the distributions and the transformation procedures used in the course of modelling.

I do not touch upon the analysis of the LGD's impact on the procyclicality of the capital requirements. A detailed review can be read about it and other relating topics in the actualized literary general work of *Altman [2009]*.

5.1. General empirical studies

Rather few studies have been written up to now in the area of LGD modelling for the retail deals regarding both the theoretical and the empirical results, thus I think it necessary to present some important publications from the empirics carried out by examining the corporate segment. I premise however, that the achievements of the empirical studies relating to the recovery rates of corporate loans are only partially relevant from the point of view of the retail deals, because some of the factors, which proved to be significant for the corporate deals, can not be interpreted concerning the

[60] There is a detailed summary about the review of the literature in the study of *Altman et al. [2005b]*, which was notably instrumental for me in collecting the sources.

retail exposures.[61] However, these studies also give valuable experiences relating to the factors which are influencing the LGD and the recovery rates.

Since at the beginning the financial institutions generally did not have enough data, the modellers and the analysts started to make the first calculations on the market of the public bonds. Then again some banks tried to build on the ground of their own databases models, which suited the requirements of the advanced IRB method *(Altman et al. [2005b])*.

The first significant empirical study relating to the recovery rates expected in case of default event is linked with the name of Altman, Haldeman and Narayanan *(Altman et al. [1977])*. Not market price information, but results of the survey of the bank collection processes between 1971 and 1975 served as a source of the recovery data used by them.

Later on this topic was getting slowly, gradually into the horizon of the researchers. One of the most detailed analyses is linked with the name of *Altman and Ebenhart [1994]*, which is about the recovery rates of defaulted bonds from the period from 1981 to 1993, and about the price development of these bonds.

5.1.1. Expected LGD vs. loss distribution

During the following years researchers put under investigation also a new aspect of the questions relating to credit risk parameters, so numerous studies were carried out, which tried to quantify not only the expected value of the recovery rates, but their complete loss distribution as well.

The largest part of the researches were not based on recoveries from bank loans, much rather from public corporate bonds, nevertheless the questions arising here possess some relevance also from the point of view of quantifying the expected loss rate of the loans; this is the argument for discussing them.

As the first step towards studying the distribution, also the modes and/or the medians and the deviation were quantified in addition to the average LGD. So for example *Carty and Liebermann [1996]* have already mentioned the asymmetry of the LGD distribution: 29% for average and 23% for median were derived on the basis of studying

[61] The number of applicants, the total assets and the business sector can be mentioned as typical examples.

58 American senior secured syndicated loans from the period between 1989 and 1996, so those LGD values dominated which were lower than the average.

Altman and Kishore [1996], examining 696 bonds from the period between 1978 and 1995, calculated 58.3% as average LGD with considerable deviation, but they also took notice of the fact that using arrangement according to securedness and seniority the deviation within the subportfolios decreased notably, and the partial averages differed from each other: 42.11% was for the senior bonds, 65.62% for the senior subordinated bonds and 68.66% for the junior subordinated bonds.

Roche, Brennan, McGirt and Verde *(Roche et al. [1998])*, who wrote about studying the LGDs of 60 secured syndicated loans in 18 retail branches from the period between 1991 and 1997, found that while the average loss rate of all the secured bank loans was 18%, the one of the subordinated debts was 61% and the one of the senior subordinated debts 58%.

Keisman and Van de Castle [1999] analysed the recoveries of 829 bank loans in the Standard & Poor's credit loss database from the period between 1987 and 1997. The authors emphasized the quantification of mean values, and considering the full portfolio they measured 15.5% as average LGD and 24.8% as deviation.

On the basis of examining recovery data of American trade credits, defaulted senior secured and subordinated corporate loans in the period between 1989 and 2000, Gupton, Gates and Carty *(Gupton et al. [2000])* noticed strong asymmetry, but in this case, in contrast with for example the achievements of *Carty and Liebermann [1996]*, the typical LGD exceeded the average. It was a further important observation that they experienced a significant difference between the LGD of the secured and the unsecured deals: while they measured 30.5% as average LGD for the senior secured loans, whereas the one of the senior unsecured loans was 47.5%.

Also strongly left-skewed distribution appears in the study of Altman, Resti and Sironi *(Altman et al. [2001])*, which presents the results of analysing the loss rate of 1000 bonds from the period between 1982 and 2000: according to the observations the average LGD (64.15%) notably exceeded the median (59.95%). The authors experienced a similar relation in that case as well, when they carried out the calculation limited to the senior secured bonds (the average LGD of them was 47.03%, while the median was 42.58%).

On the basis of studying the recovery data of 35 senior secured corporate loans and subordinated bonds in the period between 1997 and 2000, O'Shea, Bonelli and Grossman *(O'Shea et al. [2001])* pointed out that the LGD distribution of both the loans and the bonds were asymmetric, but while the former ones were characterized by low LGD (the average LGD was 37% and the median 17.01%), so the distribution skewed to the left, whereas in case of the latter ones the high LGD values dominated, which implied right-skewed distribution. Their further observation was that they experienced 88% as average LGD for the bonds contrary to the quite low LGD of the loans.[62]

Hu and Perraudin [2002] found the distribution of the LGD unimodal and also asymmetric, examining the recovery data of 958 long-term senior subordinated and secured bonds in the period between 1971 and 2002.

Analysing the recovery data of 1800 defaulted bonds, preference shares as well as senior secured and unsecured loans in the period from 1981 to 2002 on the basis of Moody's LossCalc model, *Gupton and Stein [2002]* observed that using beta distribution is much more appropriate contrary to the normal distribution in modelling. In addition they drew attention to the fact that the LGD can be just as well negative, if the cumulative recoveries altogether exceed the existing exposure at the time of the default event.

Dermine and Neto de Carvalho [2003], who used the workout LGD methodology, so I present their study in Chapter 5.2, concretely expounded their recognition that a significant part of the non-performing loans resulted in very low (near 0%) recovery or got out of the non-performing status in a short time, so the recovery was near 100%. On the basis of all that they came to the conclusion that the LGD distribution is not normal, but it can be referred to much more as bimodal, and the assumption of neither the fix recovery rate nor the beta distribution is appropriate.

Hamilton, Varma, Ou and Cantor *(Hamilton et al. [2003])*, who analysed the recovery data of 2678 non-performing bonds and loans altogether from the period between 1982 and 2002, took notice also of the fact that the median LGD (70%) exceeded the average (62.8%) looking at all the deals, so the LGD distribution skewed to the right, but when they limited the investigation to the secured deals, then just an opposite picture was outlined (the average LGD was 38.4% and the median 33%). This study is worth being underlined for the reason as well, because it emphasised strongly the mortgage credits. However, it also has to be considered that it was prepared basically on the grounds of data relating to the corporate segment, thus its conclusions can not be adapted directly to the retail mortgage deals.

---

[62] The investigations of *O'Shea et al. [2001]* covered only 35 corporate deals, so their conclusions have to be handled with reservations because of the small quantity of items.

Renault and Scaillet [2004] studied the recovery rates of defaulted bonds in the period from 1981 to 1999, and their research was directed towards defining the type of the LGD distribution. They started from the fact that the bimodality makes modelling the recoveries with parametric methods more difficult, so it requires non-parametric approach. After all in the course of estimating the density function of recovery rates by non-parametric methodology, they realized that it can not be considered beta distribution in contrast with the assumptions accepted by a significant part of the researchers.

Altman, Brady, Resti and Sironi in one of their further studies *(Altman et al. [2005a])*, in which they analysed the recoveries of 1300 defaulted corporate bonds regarding the period between 1982 and 2001, found that the LGD deviated extremely around the average of 62.8%, and the distribution was far from normal.

Similarly to the study of Gupton, Gates and Carty *(Gupton et al. [2000])*, Acharya, Bharath and Srinivasan *(Acharya et al. [2007])*, who carried out their empirical research on the basis of bank loans and corporate bonds from the S&P database regarding the period between 1982 and 1999, also found significant difference between the LGD mean values: the average LGD of the bonds was 58.04% with a deviation of 25.34% and median of 62%.

Table 14: Summary of the literary empirical results I

AUTHORS	PERIOD, DATA	TYPE OF DISTRIBUTION	LGD MEAN VALUES
Carty – Liebermann [1996]	1989-1996: 58 American senior secured syndicated loans	asymmetric, left-skewed	average: 29%, median: 23%
Altman – Kishore [1996]	1978-1995: 696 bonds		total average: 58.3%, average for the senior bonds: 42.11%, average for the senior subordinated bonds: 65.62%, average for the junior subordinated bonds: 68.66%
Roche et al. [1998]	1991-1997: 60 secured syndicated loans		total average: 18%, average for the subordinated debts: 61%, average for the senior subordinated debts: 58%
Keisman – Van de Castle [1999]	1987-1997: 829 bank loans from the Standard & Poor's credit loss database		average: 15.5%, deviation: 24.9%
Gupton et al. [2000]	1989-2000: American trade credits, senior secured and subordinated corporate loans	asymmetric, right-skewed	average for the senior secured loans: 30.5%, average for the senior unsecured loans: 47.5%
Altman et al. [2001]	1982-2000: 1000 bonds	Left-skewed	total average: 64.15%, total median: 59.95%, average for the senior secured bonds: 47.03%, median for the senior secured bonds: 42.58%

Table 14 (continuation): Summary of the literary empirical results I

AUTHORS	PERIOD, DATA	TYPE OF DISTRIBUTION	LGD MEAN VALUES
O'Shea et al. [2001]	1997-2000: 35 senior secured corporate loans, subordinated bonds	asymmetric (for the loans: left-skewed, for the bonds: right-skewed)	average for the loans: 37%, median for the loans: 17.01%; average for the bonds: 88%
Hu – Perraudin [2002]	1971-2002: 958 long-term senior subordinated and secured bonds	unimodal, asymmetric	average for the senior secured bonds: 47%
Gupton – Stein [2002]	1981-2002: 1800 bonds, preference shares, senior loans	beta distribution (negative value is also possible)	
Dermine – Neto de Carvalho [2003]		bimodal	
Hamilton et al. [2003]	1982-2002: 2678 bonds and loans	aggregately right-skewed, left-skewed for the secured deals	total average: 62.8%, total median: 70%; average for the secured deals: 38.4%, median for the secured deals: 33%
Renault – Scaillet [2004]	1981-1999: bonds	bimodal, but not beta distribution	
Altman et al. [2005a]	1982-2001: 1300 corporate bonds		average: 62.8%, excessively large deviation
Acharya et al. [2007]	1982-1999: bank loans and corporate bonds from the S&P database		average for the bonds: 58.04%, median for the bonds: 62%, deviation for the bonds: 25.34% (on an average 22% lower for the bank loans)

(Self-made table)

These analyses and their achievements contributed considerably to improvement in the area of the methodology of LGD modelling regarding both the bonds and the loans in the past years.

5.1.2. The influencing factors of LGD

Regarding the influencing factors most empirical studies investigated the run of the LGD in time, as well as its relationship with the default rates and indirectly with the state of the economy, but some publications came into being also about the influencing role of numerous other factors.

According to the experiences of *Altman and Kishore [1996]* the duration between the bond issuance and the default event did not affect considerably the measure of recoveries, but for example the unsecured and junior debts were characterized by higher LGD values compared with the others.

The study of Roche, Brennan, McGirt and Verde *(Roche et al. [1998])*, about the LGD of 60 secured syndicated loans from the period between 1991 and 1997, reported about the existence of strong significant relation between the industry and the LGD. The authors experienced negative correlation between the LGD and the stock prices (the Dow Jones sectoral average), and the seniority also proved to be an important factor, but they did not find significant for example the size of the company.

Keisman and Van de Castle [1999], analyzing the recovery data of 829 bank loans in the Standard & Poor's credit loss database from the period between 1987 and 1997 with regression methodology, experienced that the type and the amount of the debt as well as the type of the underlying collateral influenced the LGD most considerably.

Gupton, Gates and Carty *(Gupton et al. [2000])*, who performed the modelling on the basis of secondary market prices of the bank loans 1 month after the default event, found further significant influencing factors as well. They took notice for example of the fact that there was a negative correlation between the expected recovery and the duration of the collection process, and the expected recovery was also in connection with the fact whether the certain client possessed other credits in addition to the defaulted deal. This latter relation seemed much stronger in case of the unsecured deals in comparison with the secured ones, which is explainable if the cash flow is the only source of the credit instalment, than crucial importance may be attached to the circumstance, whether one or more credits have to be paid from that.[63] Contrary to that, the industry, which is mentioned as a significant influencing factor by numerous studies published in the literature, did not prove to be a significant explanatory factor.

Frye [2000a] proposed a model for investigating the relation between the PD and the recovery rate, which was founded basically upon the approach of *Finger [1999]* and *Gordy [2000]*.[64] He started from the assumption that those factors, which are increasing the probability of default in recession time, are decreasing the recovery rates at the same time. For example in recession the value of the underlying collaterals of the loans is declining, thus only smaller recoveries can be expected. So Frye pointed out that the relationship between the PD and the LGD is actually derived from the fact that both parameters depend on the same systematic factor, the current state of the economy.

[63] This relationship prevails similarly regarding the loan amount as well, since it is actually irrelevant, whether the client has only one large loan or many smaller ones.
[64] According to that the only systematic factor is the current state of the economy.

These realizations suggest that examining those factors which are also influencing the default rate can be a good starting point in the course of modelling the recovery rate.
In another study Frye *(Frye [2000c])* found strong negative correlation between the default rate and the recovery rate. He used the default and recovery data of American corporate and government bonds from the Moody's Default Risk Service database in the period from 1982 to 1997 to test his model empirically.
Hamilton, Gupton and Berthault *(Hamilton et al. [2001])* also experienced the correlation between the PD and the LGD.

O'Shea, Bonelli and Grossman *(O'Shea et al. [2001])*, similarly to the study of *Altman [2001]*, found the extensive indebtedness as a strong increasing factor of LGD. They observed also that the LGD correlated positively with the duration of bankruptcy.
Altman, Resti and Sironi *(Altman et al. [2001])*, similarly to the study of *Altman and Kishore [1996]*, Roche, Brennan, McGirt and Verde *(Roche et al. [1998])*, and Gupton, Gates and Carty *(Gupton et al. [2000])*, took the notice of the fact that the unsecured and the junior debts were characterized by higher LGD in comparison with the others. However, they also emphasized that, though positive correlation seemed between the default rate of the examined bonds and the LGD in the period from 1982 to 2000, the general economic indices did not influence significantly the recoveries in themselves.
Hu and Perraudin [2002] also experienced a relationship between the LGD and the securedness: the LGD of the senior subordinated bonds proved to be lower than the average LGD of the senior secured bonds (47%).
Frye [2003], who investigated the LGD of 859 bonds and loans from the period between 1983 and 2001, found the LGD of the senior secured loans as the lowest ones. His further observation that a decline of 20-25 percentage points of the recoveries (bond returns) could be experienced in the period of heavy economic recession (high default rate), so a strong negative correlation existed between the default rate and the recovery rate in the investigated period.

Hamilton, Varma, Ou and Cantor *(Hamilton et al. [2003])*, analyzing recovery data of 2678 non-performing bonds and loans relating to the period from 1982 to 2002, found positive correlation between the LGD and the default rate, moreover they qualified the industry also as a significant influencing factor.

Verde [2003], examining the credit risk of a broad scale of the corporate sector, reported that the recovery rates changed dramatically between 2001 and 2002 from one year to another, but they were restored to the previous level in 2003.

Carey and Gordy [2003] analysed the data of the Moody's Default Risk Service, the Society of Actuaries, the S&P and the Portfolio Management Data, relating to the period from 1970 and 1999, and their experiences contradict numerous former investigations to a certain degree, since they did not find notable relation between the default rate and the LGD, examining the period as a whole. However, when they limited the time horizon of the investigation to the period from 1988 to 1998, significant correlation was outlined indeed, from which they concluded that other factors of the economic cycles are also influencing the relationship between the two credit risk parameters.

Keisman [2004], on the basis of examining the Standard&Poor's data, pointed out that during the crisis the recovery rates fall significantly short of the recovery rates typical for the normal economic stage.

Altman and Fanjul [2004] also investigated the relationship between the default rate and the recovery rate. According to that, similarly to the studies or Altman, Brady, Resti and Sironi *(Altman et al. [2001; 2005a])*, inverse relation (negative correlation) existed between the PD and the recovery rates, since rather low recoveries[65] performed in 2001-2002 (in the period of high default rates), but they became almost doubled parallel with the strong decrease of the default rate in the following year.

Grippa, Ianotti and Leandri, who analysed the recoveries of 20,724 Italian small and medium enterprises and retail loans in their empirical study *(Grippa et al. [2005])*, found that the loan amount, the collateral and the proportion of the non-performing loans alike correlated negatively with the recovery rate. Comparing the customer segments to each other, they observed that the recovery rate of the retail sector exceeded the average typical for the small and medium enterprise sector, and within the retail sector the recovery realized on the home loans, proved to be even higher than the one of the other loans.

Schuermann [2005], as well as Altman, Bradi, Resti and Sironi *(Altman et al. [2005a])* also emphasized the important effect of the industry on the loss rate, moreover they found inverse relationship between the PD and the recovery rates, similarly to for example the study of *Altman and Fanjul [2004]*.

[65] The recovery rate was 25% on the market of the corporate bonds.

Emery, Cantor, Keisman and Ou *(Emery et al. [2007])* examining the Moody's data experienced positive correlation between the PD and the LGD.

Acharya, Bharath and Srinivasan *(Acharya et al. [2007])* also mentioned in their empirical study, which was carried out on the basis of bank loans and corporate bonds from the S&P database regarding the period between 1982 and 1999 that the industrial conditions at the time of default event influenced considerably the recovery rates. The authors pointed out significant difference between the LGD values of the bank loans and of the bonds,[66] which drew the attention to that the adaptation of the investigations has always to be done cautiously regarding both the segments (for example corporate / retail) and the deal types (for example secured / unsecured deals).

Table 15: Summary of the literary empirical results II

AUTHORS	PERIOD, DATA	INFLUENCING FACTORS
Altman – Kishore [1996]	1978-1995: 696 bonds	securedness (–), juniority (+), economic cycle
Roche et al. [1998]	1991-1997: 60 secured syndicated loans	securedness (–), juniority (+), industry, stock prices (–)
Keisman – Van de Castle [1999]	1987-1997: 829 bank loans from the Standard & Poor's credit loss database	type of the debt, loan amount (+), type of the underlying collateral
Gupton et al. [2000]	1989-2000: American trade credits, senior secured and subordinated corporate loans	securedness (–), juniority (+), duration of the collection process (+), whether the client has other credits
Frye [2000a; 2000c]	1982-1997: American corporate and government bonds from the Moody's Default Risk Service database	economic cycle, default rate (+)
Hamilton et al. [2001]		default rate (+)
O'Shea et al. [2001]	1997-2000: 35 senior secured corporate loans, subordinated bonds	indebtedness (+), duration of the bankruptcy (+)
Altman et al. [2001]	1982-2000: 1000 bonds	securedness (–), juniority (+), default rate (+)
Hu – Perraudin [2002]	1971-2002: 958 long-term senior subordinated and secured bonds	securedness (+)
Frye [2003]	1983-2001: 859 bonds and loans from the Moody's database	securedness (–), juniority (+)
Hamilton et al. [2003]	1982-2002: 2678 bonds and loans	default rate (+), industry

[66] They measured about 22% lower average LGD for the bank loans than for the senior secured bonds.

Table 15 (continuation): Summary of the literary empirical results II

AUTHORS	PERIOD, DATA	INFLUENCING FACTORS
Verde [2003]		economic cycle
Carey – Gordy [2003]	1970-1999: Moody's Default Risk Service, Society of Actuaries, S&P and Portfolio Management Data database	economic cycle
Keisman [2004]	Standard&Poor's loss database	economic cycle
Altman – Fanjul [2004]		default rate (+)
Altman et al. [2005a]	1982-2001: 1300 corporate bonds	default rate (+), industrial factor
Grippa et al. [2005]	20,724 Italian small and medium enterprises and retail loans	loan amount (+), securedness (+), proportion of the non-performing loans (+), sector (retail / corporate), purpose of the loan (home or other)
Schuermann [2005]		industrial factor
Emery et al. [2007]	Moody's data	default rate (+)
Acharya et al. [2007]	1982-1999: bank loans and corporate bonds from the S&P database	industrial conditions at the time of default, type of the debt (bank loan / bond)

(Self-made table)

As a closing of this subsection it is worth mentioning two further studies, which are about analyzing the effects of the economic state on the LGD, examining the period of the present economic crisis. Analysing the data of the first half-year in 2009, *Keisman and Marshella [2009]*, as well as *Altman and Karlin [2009]* pointed out that the recovery rates decreased to a record level parallel with the exceptional acceleration of increase of the default rate, which started in 2008 and continued in 2009. The authors drew the attention to the importance of the economic cyclicality's effect on the credit risk parameters again. In the course of my empirical research I also try to map the role of the factors which characterize the state of the economy.

5.1.3. Methodological aspects

Numerous empirical works came out in the past years, in which the methodology of the LGD estimation was also presented. Hereinafter I outline some important studies, without claiming completeness.

Hu and Perraudin [2002] studied the relationship between the recoveries and the default rates applying the Moody's historical bond market data from the period between 1971 and 2000. They analysed the effect of the negative correlation on the credit VaR measures using the Extreme Value Theory and other non-parametric procedures, and they found it statistically significant at 99% confidence level.
Pykhtin [2003] basically followed the logic of *Gordy's [2000]* one-factor model, so he presumed that the recoveries are influenced by only one systematic factor, the current state of the economy. However, he introduced an innovation in that respect that he considered the distribution of the recoveries lognormal.
Jokivuolle and Peura [2003] used a new type of approach. They built the correlation between the collateral value and the default rate into their model based on option pricing methodology, according to which the default event is the result of the changes of the company's asset value. They assumed that the recovery rate is determined by the stochastic collateral value, and the PD is an exogen factor, thus modelling the company's asset value is not required for calculating the LGD. As a result of their researches they also found inverse relationship between the PD and the recovery rates.
On the basis of the default and recovery rate time series from the period between 1982 and 1999 regarding bonds and loans in the S&P Credit Pro database, *Düllmann and Trapp [2004]* came to the conclusion that the systematic risk factor considerably influences the recovery rate.

Chabaane, Laurent and Salomon *(Chabaane et al. [2004])* presented that the calculations based on the Basel II recommendations lead to underestimated expected credit loss figures, since they leave the correlation out of consideration.
Altman, Brady, Resti and Sironi *(Altman et al. [2005a])* studied the recovery data of about 1300 American defaulted corporate bonds from the period between 1982 and 2001. In the course of the empirical analyses they also experienced negative correlation between the default rates and the recovery rates, they realized however, that other factors may have even larger explanatory power, instead of the single systematic factor describing the state of the economy. They came to the conclusion that the most important factor is the market supply of the defaulted bonds, and in addition to that numerous other variables, which characterise the market size of the risky bonds and the economic cycle, also possessed quite large explanatory power.
In one of their later studies they performed Monte Carlo simulation on the basis of a bank loan portfolio for the sake of assessing the effect of the negative correlation

between the default rates and the recovery rates on the credit risk models. They found that if they assumed that the PD and the LGD are not correlated to each other, this resulted in significant underestimation of both the expected and the unexpected loss.

In spite of the fact that it is extraordinarily important for the financial institutions to possess a model which has appropriate predictive power, the applied models were described in only very few empirical analyses in the beginning. It was particularly atypical to publish exact formulations, which would be instrumental for the institutions in preparing their own LGD models.

In this respect the Moody's KMV LossCalc™ model demonstrated by *Gupton and Stein [2005]* was an important milestone, which was directed towards modelling the run of recovery rate on the basis of about 3026 international (Asian, Australian, North- and South-American, European) observation data in a 1-year time horizon, considering numerous factors simultaneously. The secondary market prices of the bonds and of the loans served as a basis of the recovery model in this case as well, similarly to the previous study of the authors.

They classified the predictive factors basically into the following five groups:

- data referring to the collaterals,
- deal-related pieces of information,
- client-related pieces of information,
- industry characteristics,
- macroeconomic factors.

They found that the consideration of the data referring to the collaterals increased the accuracy of the model by 72%. While the indebtedness and the probability of default played considerable influencing role among the client-related pieces of information, the type (loan / bond) and the seniority proved to be significant regarding the deal-related pieces of information. They specified the industry characteristics with features such as for example the average PD values of the certain industries, and by researching the effects of the macroeconomic factors they were able to consider special characteristics as well, such as for example the legal differences in execution.

Their study means a large leap forward not only because it included a notably broader scope of the examined explanatory factors and the data used, than the previous ones, but also because it presented fairly detailed not only the theory and the modelling results, but the particular methodology of calculating the LGD as well.

Bakshi, Madan and Zhang (Bakshi et al. [2006a]) modelled the loss rate of corporate bonds rated BBB, considering also the correlation between the PD and the recovery rates. Their empirical results agreed with the former experiences, according to which there is a negative correlation between the probability of default and the recovery rates. They found that a decline of 1 percentage point of the recovery rate occurred parallel with an increase of around 4 percentage points of the PD.

Later on the simulation procedures and other complex methodologies gained ground more and more. For example *Huang and Oosterlee [2008]* modelled the run of the LGD as a stochastic random variant by regressions based on beta distribution.

Hlawatsch and Ostrowski [2010] also turned to the simulation procedures: their research was directed towards developing a model by mixing two beta distributions, which is able to attract the bimodality of the LGD, and ensures at the same time that the values are limited into the interval between [0;1]. The authors used the Expectation Maximization procedure mentioned in Chapter 4.1.1 for estimating the parameters.

However, the appearance and gradual spreading of the new methodologies do not mean at all the ceasing of the justification of the "traditional" analytic procedures; their use underlies numerous newly coming studies in the future too.

Table 16: Summary of the literary empirical results III

AUTHORS	METHODOLOGY AND RESULTS
Hu – Perraudin [2002]	Using the Extreme Value Theory and other non-parametric procedures on the basis of historical bond market data. Studying the effect of the correlation on the credit VaR measures.
Pykhtin [2003]	Considering the distribution of the recoveries lognormal. One-factor model on the basis of Gordy's study [2000] (the state of the economy is the only one systematic factor).
Jokivuolle – Peura [2003]	Assumptions: the PD is an exogen factor, the recovery rate is determined by the stochastic collateral value. Option pricing methodology with building the correlation between the collateral value and the default rate into the model.
Düllmann – Trapp [2004]	Investigating the systematic risk factor on the basis of the time series of default and recovery rates.
Chabaane et al. [2004]	Impact study about leaving the correlation out of consideration. Conclusion: it leads to underestimated expected credit loss figures.
Altman et al. [2005a]	Univariate and multivariate econometric modelling. The most important factor is the market supply of the defaulted bonds. Other factors: variables which characterise the market size of the risky bonds and the economic cycle. Monte Carlo simulation for studying the PD-LGD correlation.
Gupton – Stein [2005]	Multivariate model on the basis of the secondary market prices of the bonds and loans. The predictive factors: (1) data referring to the collaterals, (2) deal-related data, (3) client-related data, (4) industry characteristics, (5) macroeconomic factors. Applying transformed risk factors in the regression.
Bakshi et al. [2006a]	Considering the correlation between the PD and the recovery rates. In case of the corporate bonds rated BBB a decline of 1 percentage point of the recovery rate occurred parallel with an increase of 4 percentage points of the PD.

Table 16 (continuation): Summary of the literary empirical results III

AUTHORS	METHODOLOGY AND RESULTS
Huang – Oosterlee [2008]	Regressions based on beta distribution. Handling the LGD as a stochastic random variant.
Hlawatsch – Ostrowski [2010]	Mixing beta distributions for the sake of catching the bimodality and ensuring the limitation into the interval between [0;1]. Using Expectation Maximization procedure.

(Self-made table)

5.2. Analyses referring to loans on the basis of workout LGD methodology

The first significant adoption of the workout LGD methodology is the study already mentioned in the beginning of this chapter, which is linked with the names of Altman, Haldeman and Narayanan *(Altman et al. [1977])*. The researchers studied the recoveries deriving from a 3-year period of the deals, but they did not use discounting, which unquestionably distorted the results. After all this empirical study can be considered as the first presentation of using the workout LGD methodology, to which important role can be attached concerning the establishment of the methodology. Regarding however, that the authors left the time value of money out of consideration, the numerical results are hardly informative.

After that the works presenting the workout LGD methodology have not taken part at all for two decades, but more and more studies have arisen since the middle of the 1990s.

5.2.1. Definitions and assumptions

Asarnow and Edwards [1995] analysed the losses deriving from defaults of American bank loans on the basis of the Citibank's 24-year-long data series from the period between 1970 and 1994. There were 831 senior secured and unsecured commercial and industrial loans, 89 structured secured loans, large and medium corporate loans in the scope of the investigated deals, whose exposures were between 1 million USD and 190 million USD at the date of the default event.

They applied the borrowing rate as discount rate, and the definition of default used by them suited the CRD regulation as well, since they considered all deals non-performing, which the client is not likely to pay back in its entirety or is already delayed. However, the consistency with the CRD can not be said for example about the definition regarding

the closing of the deal, since the authors considered the deals closed only in that case, if the exposure decreased to zero.
According to their observations, the LGD was fairly stable during the examined 24-year-long period, but it was not characterized by normal distribution, but left-skewed[67] or bimodal.

Carty and Liebermann [1996] prepared calculations referring to the recoveries of 229 senior secured small and medium enterprise loans, whose collection process already closed, from the Loan Pricing Corporation's loss database in the period between 1990 and 1996.
They calculated the recovery rate as the discounted sum of the interest and principal payments, with notable oversimplifications leaving the different costs, charges and fees out of consideration, and they used yields as discount rate, which they quantified in a way that they added a spread, which was calculated on the basis of the deal's risk, to the LIBOR. The distribution observed by them was strongly left-skewed: contrary to the 8% median the average LGD was 21%.

The study of *Eales and Bosworth [1998]* presented the result of the empirical investigation of the Australian Westpac Banking Corporation's recovery rates in the period between 1992 and 1995. This analysis covered 5782 clients' small enterprise, consumer and home loans as well as real estate investment financing loans already closed in all, which were smaller than 6.7 million USD and the proportion of the secured deals was almost 95% among them.
As discount rate they applied the alternative cost of capital calculated according to CAPM, but they showed that if they had used the borrowing rate similarly for example to *Asarnow and Edwards [1995]*, then the LGD would have been 10% lower.
They subtracted from the realized recoveries the amounts disbursed following the default event as well as the external and internal collection costs, and they calculated the recovery rate on the basis of these net recoveries. They truncated the distribution of the LGD values at 0% and at 100%.
They also experienced different LGD values for the different types of loans: the average LGD of the commercial loans was 31% (their median: 22%), and the one of the consumer credits 27% (their median: 20%); moreover the loss rate of the unsecured

---

[67] For example the average LGD of the industrial and commercial loans was 34.79%, and the median was only 21%.

deals also proved to be higher than the one of the secured deals. They realized that the LGD distribution of the unsecured and secured deals differed significantly from each other: whereas the one of the unsecured loans was bimodal, the secured deals were characterized by left-skewed unimodal distribution. Hamilton, Varma, Ou and Cantor *(Hamilton et al. [2003])* also found difference concerning the LGD distribution of the unsecured and the secured deals in the course of their research already mentioned, but actually they thought the difference only in the direction of the skewness.

Felsovalyi and Hurt [1998] investigated the recoveries of 1149 Latin-American loans, whose amounts exceeded 100 Thousand USD, from the period between 1970 and 1996. Similarly to the work of *Eales and Bosworth [1998]*, they also observed a strongly left-skewed distribution. The average LGD measured by them was 31.8% with 28.8% deviation, but while the proportion of those LGD values which were lower than 15% seemed definitely high, there were LGD values which exceeded the 100% as well.
Their definition for default event analogized with the one applied in the study of *Asarnow and Edwards [1995]* presented previously, so they considered all deals non-performing which the client is not likely to pay back in its entirety or is already delayed.

Araten, Jacobs and Varshey *(Araten et al. [2004])* analysed the recovery data of 3761 large corporate loans deriving from JP Morgan from the period between 1982 and 1999, and they found that the modus of the LGD (5%) fell significantly under the average (39.8%), the distribution skewed to the left. On the other hand the LGD values measured by them spread more intensively (with a deviation of 35.4%) and in broader range (from -10% to 173%) in comparison with the results of other analysis, but unlike the method of *Eales and Bosworth [1998]* mentioned earlier, they did not truncate the distribution at 0% and at 100% before calculating the mean values.

Dermine and Neto de Carvalho [2003; 2005] analysed a sample of 371 defaulted loans of the Banco Comercial Portugues, which were disbursed for small and medium enterprises between June 1995 and December 2000. They used a rather strict definition for default, they considered already the 1-day delay as default as well, thus numerous multiple defaults appeared among the observations. By analysing the results we can not disregard the feature that in these cases the authors paid attention only to the very first occasion, and this resulted in overestimating the recovery rate.

Grunert and Weber [2005; 2009] examined bank data referring to German corporate loans from the period between 1992 and 2003, and their database contained the closed deals of 120 corporate. The calculation was carried out on client level, using 5% as discount rate, considering a broad range of the cash inflows and cash outflows.
The authors analysed the distribution of the recovery rates in detail, and they observed strongly skewness to the right, but they also realized that the recovery rates were negative in many cases, so the present value of the costs exceeded the present value of the gross recovery amounts for certain deals.
Similarly to the study of *Dermine and Neto de Carvalho [2003; 2005]* they experienced that the most significant costs derived from enforcing the collaterals, thus the costs proved to be relatively higher in case of the lower credit items, than for the larger loan amounts.

Brady, Chang, Miu, Ozdemir and Schwartz *(Brady et al. [2007])* aimed at quantifying the discount rate which is adequate to the risk of the recoveries following the default event. They compared in their study market prices and empirical data of a database containing recoveries from collection process as well.
Thomas, Mues, Matuszyk and Moore *(Thomas et al. [2007a; 2007b])* examined what kind of factors influenced the recoveries derived from the collection processes done by the bank's internal collection department (in-house collection) or by the external agent (3^{rd} Party).
They considered only those deals as non-performing whose delay was longer than 180 days, and they reckoned only with the recoveries from the first 24 months after the default event. However, considering that they performed the analysis by using two considerably different databases, the comparability of the numerical results is disputable.

Keisman and Marshella [2009] adapted the workout LGD methodology for the Moody's recovery data referring to bank loans and debts from the period between 1988 and 2009. Their portfolio contained about 3000 deals. They carried out their calculations both without using discount rate and with discounting by the interest rate which existed before the default event.
They experienced that the highest recovery rate characterized the senior bank loans, which were secured measurably, while the lowest recovery was realized by the junior bonds. In the course of comparing the rates deriving from the nominal and the discount

method, they found that also the difference between them was the largest in case of the senior bank loans, and the smallest for the junior bonds.

Chalupka and Kopecsni [2009] applied the data of an anonymous Czech commercial bank's database about closed and non-closed loans of Czech small and medium enterprises and corporations for modelling the LGD. They observed significant difference between the LGD values of those deals which were closed within 1 year and those ones whose collection process was longer, and because most cases of the former category were "technical default"[68], they left them out of the analysis.

They experienced on the basis of the cumulative recovery rates that considerable recoveries did not occur after the first 3 years of the collection process, thus they considered the effective length of the collection period to 3 years. So they involved even those deals in the analysis, in case of which the length of the collection period (the time from the default event) exceeded this duration.

They disregarded the direct costs, while in the course of allocating the indirect costs they considered the cumulative recoveries as a basis, and they estimated these costs as 1.8% relative to the recovered amount based on the past experience.

They grouped the collaterals into five categories, and attached a risk premium of 0, 240, 420, 600 and 990 basis points to them, then they calculated the risk premium referring to certain clients according to the risk classes of the underlying collaterals of their deals. They defined the discount rate as the sum of the risk-free rate and the risk premium different for each asset class.

Using a flat risk premium of 0%, 1%,…, 8% and 9% as a comparison they experienced that increasing the risk premium by 1% resulted in an increase of the LGD by approximately the same percentage point, and the authors explained this moderate impact with the shortness of the collection period. The systematic asset risk class approach resulted in similar LGD figures as the use of flat risk premium of 5%.

Because of the small number of observations in individual years they classified the deals into three categories (until 1994, between 1995 and 1999, between 2000 and 2005) according to the date of their origination or their default event, which periods covered different stages in the development of the Czech banking sector or different cycles of the Czech economy. The differences were also manifested in the LGD values, their effects were significant.

[68] Delay of marginal low amount or short period.

Bellotti and Crook [2008] proposed to prepare a model which is able to predict the recovery or loss rate of the retail credit card products. The data provided by a financial institution in the UK related to four types of credit card products which originated in the period between 1998 and 2004.

The authors took only a 12-month long recovery duration into account, furthermore applying a notable simplification they did not consider the costs. They included into the calculation only those deals, in case of which there was a longer duration than 12 months since their default event.

They did not dwell on defining the form and the statistical features of the distribution, because their fundamental purpose was the model building as well as to measure and to compare the models' predictive ability.

Table 17: Summary of the literary empirical results IV

AUTHORS	PERIOD, DATA	ASSUMPTIONS	TYPE OF DISTRIBUTION, LGD MEAN VALUES
Asarnow – Edwards [1995]	1970-1994: 831 senior commercial and industrial loans, 89 structured secured loans, large and medium corporate loans (Citibank), 1 million – 190 million USD	Discount rate: borrowing rate. Definition of default: consistent with the CRD. Closed deal: if its exposure is 0.	Left-skewed or bimodal distribution. Average LGD for the commercial and industrial loans: 34.79%, their median: 21%.
Carty – Liebermann [1996]	1990-1996: 229 senior secured small and medium enterprise loans, whose collection process is already closed (Loan Pricing Corporation's database)	Leaving the different costs, charges and fees out of consideration. Discount rate: LIBOR + interest spread calculated on the basis of the deal's risk.	Strongly left-skewed distribution. Average: 21%; Median: 8%.
Eales – Bosworth [1998]	1992-1995: 5,782 clients' small enterprise, consumer and home loans as well as real estate investment financing loans already closed in all, which are smaller than 6.7 million USD (the Australian Westpac Banking Corporation's database)	Discount rate: the alternative cost of capital calculated according to CAPM. Netting of the recoveries and the costs. Truncating the distribution of the LGD values at 0% and at 100%.	Bimodal distribution for the unsecured deals, right-skewed unimodal distribution for the secured deals with lower average. Average LGD for the commercial loans: 31%, their median: 22%. Average for consumer credits: 27%, their median: 20%.
Felsovalyi – Hurt [1998]	1970-1996: 1149 Latin-American loans, whose amounts exceeded 100 Thousand USD, already closed collection process	Definition of default: consistent with the CRD.	Strongly left-skewed distribution. Average: 31.8%. Deviation: 28.8%. The LGD values which are lower than 15% are frequent. There were LGD values which exceeded the 100% as well.
Araten et al. [2004]	1982-1999: 3761 large corporate loans (JP Morgan)	The LGD can exceed the interval [0;1].	Left-skewed distribution. Average: 39.8%. Modus: 5%. Deviation: 35.4%. Range: -10%-tól 173%-ig.

Table 17 (continuation): Summary of the literary empirical results IV

AUTHORS	PERIOD, DATA	ASSUMPTIONS	TYPE OF DISTRIBUTION, LGD MEAN VALUES
Dermine – Neto de Carvalho [2003; 2005]	1995-2000: 371 small and medium enterprise loans (Banco Comercial Portugues)	Definition of default: 1-day delay. Considering only the very first default even in case of each deal.	
Grunert – Weber [2005; 2009]	1992-2003: 120 German corporate loans which are already closed	Considering a broad range of the cash inflows and cash outflows. Discount rate: 5%. Calculating on client level. The most significant cost: enforcing the collaterals.	Strongly right-skewed distribution. The LGD values which are higher than 100% are frequent.
Brady et al. [2007]	Market prices and recoveries derived from collection		
Thomas et al. [2007a; 2007b]	11,000 personal loans + 70,000 deals from an external collection database	Definition of default: 181-days delay. Effective recovery period: 24 months.	
Keisman – Marshella [2009]	1988-2009: 3,000 bank loans and bonds (Moody's)	Discount rate: 0 or the interest rate which existed before the default event.	The average LGD of the junior bonds is higher than the LGD of the senior secured bank loans.
Chalupka – Kopecsni [2009]	Closed and non-closed Czech small and medium enterprise as well as corporate loan deals (anonymous commercial bank)	Disregarding the direct costs. Effective recovery period: 36 months. Cost allocation: 1.8% relative to the recovery. Grouping the collaterals into 5 categories on the basis of the risk. Discount rate: risk-free rate + risk premium.	Different distributions in case of collection processes which are closed within 1 year and in case of the longer ones. The LGD is not the same in different stages in the development of the banking sector and different cycles of the economy.
Bellotti – Crook [2008]	1998-2004: Four types of credit card products (United Kingdom)	Effective recovery period: 12 months. Leaving the costs out of consideration.	

(Self-made table)

In the following part I present that what kind of factors are influencing considerably the run of the recovery rate and the LGD, according to the empirical analysis referring to loans on the basis of workout LGD methodology.

5.2.2. Influencing factors of the recovery rate and the LGD

Asarnow and Edwards [1995], analyzing the losses deriving from defaults of American bank loans on the basis of the Citibank's 24-year-long data series in the period between

1970 and 1994, experienced that the securedness and the type of the loan considerably influenced the LGD, but for example the impact of the loan amount was not significant. A notably difference appeared between the LGD of the different deal types: while the average LGD of the commercial and industrial loans was 34.79% (their median: 21%), the one of the structured loans was only 12.75%.

Carty and Liebermann [1996], looking at the relationship between the type of the collateral and the LGD on the basis of investigating the recoveries of senior secured small and medium enterprise loans observed that the loss rate of the credits secured by operating assets was lower than the one of those, where real estate or some kind of invested tangible asset served as collateral. A possible explanation is that the operating assets are more liquid, so the bank may depend upon quicker and larger recovery from them.

Eales and Bosworth [1998] found inverse relation between the loan amount and the LGD for the small enterprise, consumer, home and real estate investment financing loans, which contradicted the results of numerous other studies.

Felsovalyi and Hurt [1998] observed positive relationship between the loan amount and the LGD, but they did not found the effect of the national macroeconomic factors determinant. It is reinforced also by their observation that the average LGD seemed to be rather stable in spite of the fact that their analysis covered a 27-year-long period.

Investigating recoveries of retail credits, *McNabb and Wynn [2000]* drew attention to the fact that also the diversity of the reasons underlying the default event also measurably influences the outcomes of the collection process, while Bos, Kelhoffer and Keisman *(Bos et al. [2002])* pointed out the correlation between the PD and the LGD.

Araten, Jacobs and Varshey *(Araten et al. [2004])*, analysing the recovery data of large corporate loans of the JP Morgan in the period between 1982 and 1999, did not experience any relation between the default rate and the loss rate, when they studied the whole period in one, but when they left out the first four years of the analysis, then a positive correlation showed up. The authors found significant relationship between the type of the collateral and the LGD in all durations.

Dermine and Neto de Carvalho [2003; 2005], using the workout LGD methodology for the small and medium enterprise loans, drew consentaneous conclusion with many former empirical studies: the results of the analysis showed that the recovery rate is referable mostly to the loan amount and the type of the collateral. They found strong negative correlation with the loan amount, whereas an expressed positive correlation appeared with the securedness, particularly in case of the physical collaterals. The sector, the year of the deal's origination as well as the age of the firm proved to be further important influencing factors.

Considering the costs occurring from the collection process they realized that the direct costs were affected notably by the way of the collection (in-house collection or not). They experienced that the proportion of the fix costs was considerably high among the cost components, and this served partly as an explanation for the negative relation observed between the loan amount and the recovery rate.

However, it can be generally said about all the macroeconomic characteristics involved into the investigation that none of them proved to be a statistically significant influencing factor for the recovery rate.

Querci [2005] analysed the data of 15,827 loan deals of an Italian commercial bank referring to the period between 1980 and 2004 searching for those factors which especially influenced the run of LGD. He experienced that the explanatory power of the models without using client characteristics was rather low, and none of the variables appeared really important in itself. Finally the model proved to be the best, which also involved client characteristics in addition to the type of the loan, the geographical location, the type of the collateral, the client segment and the length of the collection process.

Grunert and Weber [2005; 2009] researched the factors influencing the expected loss of German corporate loans on the basis of the workout LGD methodology, and they found that the closing type of the collection process notably influenced the recovery rate: the recovery rates proved to be significantly higher in case of the deals of firms which recuperated following the default event, than for those which came to be liquidated.

In terms of the expected recovery the measure of securedness, the PD, the intensity of the connection between the client and the bank, the size of the firm and the costs occurring in the course of the collection process proved to be the most important influencing factors. The authors found negative correlation between the measure of the securedness and the PD, which contradicts on the one hand the results published in numerous studies, and on the other hand the professional experience that the credit institutions generally require larger collaterals from the clients whose creditworthiness is worse (may be characterized with higher PD).

Taking into consideration the cost of capital, as a special type of costs, the recovery rates decreased considerably, but calculating also with the tax advantages deriving from the provision forming and depletion they still increased, accordantly to the experience that the collection period was rather long. Again they did not find significant relation between the closing type of the deal and the length of the collection process.

Brady, Chang, Miu, Ozdemir and Schwartz *(Brady et al. [2007])* investigated both market prices and recoveries from collection, and they identified the rating grade of the client, the type of the debt and the sectoral-economic conditions as statistically significant influencing factors. Similarly to the study of Acharya, Bharath and Srinivasan *(Acharya et al. [2007])*, *Schuermann [2005]*, as well as Altman, Brady, Resti and Sironi *(Altman et al. [2005a])* they also experienced that the sectoral conditions affected the risk of the recovery more strongly in comparison with the overall state of the economy.

They did not find notable difference between the risk of the unsecured and the secured deals, but this result of them has to be handled with reserves, because the data referring to the measure of securedness was not available for them, they had information only about the fact of securedness.

They observed that during economic crisis (high default rates) the risk of the future recoveries increases as well, so this study also pointed out the existence of the correlation between PD and LGD. At the same time they drew attention also to the fact that a considerable estimation error charges the calculation carried out with the workout LGD methodology because of the necessity of the data available.

Thomas, Mues, Matuszyk and Moore *(Thomas et al. [2007a; 2007b])* researched the determinative factors of the recoveries from collection processes done by the bank's internal collection department (in-house collection) and by the external agent (3^{rd} Party). They experienced that in case of the in-house collection the LGD correlated positively with the loan amount, but negatively with the application score, the time of the loan until default, the number of months with arrears in the whole life of the loan, and with the number of months with arrears in the last 1 year. Conversely the recovery rate resulting from the external collection was in significant positive relationship with the age of the debtor and the monthly repayment amount, moreover it also appeared as an influencing factor whether the client disposed of phone.

Bellotti and Crook [2008], in the course of their research referring to the retail credit card products, observed that the length of the relationship between the client and the bank, the income of the client, the application score and the duration from the origination of the deal to the default event influenced the recovery rate positively, while the age of the client and the exposure at the default event affected it negatively. Moreover the date of the deal's origination also proved to have significant effect, and also whether the client was the owner of the real estate in case of deals secured by real estate property. The negative correlation between the exposure at the default event and the recovery rate agreed with the former empirical results (for example *Dermine – Neto de Carvalho [2005]; Grippa et al. [2005])*, but they were not able to demonstrate the strong explanatory role of the geographical location and the behavioural score.

The authors found a statistically significant negative relation between the default rate and the recovery rate for all four credit card products, however, in the case when they placed also the other explanatory variables in the model, than this relation was not justifiable, so actually the correlation existed between the determinative factors of the PD and the LGD, not between the PD and the LGD directly.

Zhang [2009a] studied the relationship between the collaterals and the recovery rate on the basis of data referring to loan deals of American corporations from the period between 1988 and 2007, and he found a strong positive relation between the collateral values and the recovery rate.

The connection between the business-economic cycle and the collaterals drew his attention to the lagging effect of the macroeconomic circumstances on the recovery rate, which leads to easing the prescriptions referring to the collaterals and declining the recovery rates during economic boom, while it effects towards the higher recoveries by the stricter conditions of contract during recession.

This study was prepared on the basis of data referring to the corporate loans, but it can be an interesting question concerning the retail sector as well, how the severity of the contractual conditions influences the recovery rate. Conversely it also has to be considered that the quantity and the elaborateness of the domestic data available are not appropriate yet for its investigation in many cases.

Table 18: Summary of the literary empirical results V

AUTHORS	PERIOD, DATA	INFLUENCING FACTORS
Asarnow – Edwards [1995]	1970-1994: 831 senior commercial and industrial loans, 89 structured secured loans, large and medium corporate loans (Citibank), 1 million – 190 million USD	securedness, type of the loan (commercial and industrial loans / structured loans)
Carty – Liebermann [1996]	1990-1996: 229 senior secured small and medium enterprise loans, whose collection process is already closed (Loan Pricing Corporation's database)	type of the collateral (lower for the credits secured by operating assets)
Eales – Bosworth [1998]	1992-1995: 5,782 clients' small enterprise, consumer and home loans as well as real estate investment financing loans already closed in all, which are smaller than 6.7 million USD (the Australian Westpac Banking Corporation's database)	loan amount (-)
Felsovalyi – Hurt [1998]	1970-1996: 1149 Latin-American loans, whose amounts exceeded 100 Thousand USD, already closed collection process	loan amount (+)
McNabb – Wynn [2000]	Retail credits	reason underlying the default event
Bos et al. [2002]		probability of default
Araten et al. [2004]	1982-1999: 3761 large corporate loans (JP Morgan)	default rate (+, in some periods), type of the collateral
Dermine – Neto de Carvalho [2003; 2005]	1995-2000: 371 small and medium enterprise loans (Banco Comercial Portugues)	loan amount (+), type of the collateral, securedness (-), sector, the year of the deal's origination, the age of the firm, the way of the collection
Querci [2005]	1980-2004: 15,827 loan deals of an Italian commercial bank	type of the loan, geographical location, type of the collateral, client segment, the length of the collection process, client characteristics
Grunert – Weber [2005; 2009]	1992-2003: 120 German corporate loans which are already closed	the closing type of the collection process (liquidation or not), measure of securedness, probability of default, intensity of the connection between the client and the bank, the size of the firm, costs of the collection process, cost of capital, tax advantages deriving from the provision forming and depletion
Brady et al. [2007]	Market prices and recoveries derived from collection	the rating grade of the client, type of the debt, sectoral-economic conditions, probability of default

Table 18 (continuation): Summary of the literary empirical results V

AUTHORS	PERIOD, DATA	INFLUENCING FACTORS
Thomas et al. [2007a; 2007b]	11,000 personal loans + 70,000 deals from an external collection database	in case of in-house collection: loan amount (+), application score (-), the time of the loan until default (-), the number of months with arrears in the whole duration (-), the number of months with arrears in the last 1 year (-); in case of external collection: age of the debtor (+),monthly repayment amount (+), whether the client disposes of phone
Bellotti – Crook [2008]	1998-2004: Four types of credit card products (United Kingdom)	length of the relationship between the client and the bank (-), the income of the client (-), application score (-), (-), the time of the loan until default (-), the age of the client (+), the exposure at the default event (+), the date of the deal's origination, whether the client is the owner of the real estate which serves as collateral, default rate (+, indirectly)
Zhang [2009a]	1988-2007: Loan deals of American corporations	collateral values (-), severity of the contractual conditions

(Self-made table)

5.2.3. Aspects regarding the model building

Examining the facilities of the parametric and non-parametric modelling, *Polívka [2008]* experienced that the distribution of the recovery rates is bimodal or unimodal with considerably fat tails in many cases, and there are recovery rates as well which are less than 0% or larger than 100%. As explanation why the distribution is not normal, he emphasized the followings:

- Not only the delayed deals may get default status, but for example those as well, in case of which another credit of the debtor, co-debtor or guarantor became non-performing. We may generally reckon on these deals high recoveries, in many cases around 100%.
- Normally, the credit institutions write off the exposures as loss only when the proportion of their collection costs and the gross recoveries exceeded 100%, so the value of the recovery rate is negative.

- In the case when the costs deriving in the course of the collection process exceed the inflows, the net recoveries may become negative. Using the high discount rate also contributes to that, because it charges the further recoveries more strongly then the costs which are deriving dispersedly during the period.
- The situation may occur as well that the credit institution is able to sell the underlying collateral of the deal at higher price then the sum of the present value of the exposure at default event and the costs, so the net recovery exceeds 100%.
- In case of auction or factoring the length of the period from the default event to closing the deal can be very diverse, and the considerable differences may result in internal modes for large amounts.

These specialities induced the researchers to use special models, which are able to consider all these.

Peter [2006] suggested a multistep approach, in the framework of which at first the possible scenarios, referring to the post-default period, have to be defined, than the recovery rate and the LGD can be calculated by weighting with the probabilities quantified by using logistic regression or Markov-chains.

Thomas, Mues, Matuszyk and Moore *(Thomas et al. [2007a; 2007b])*, in their study mentioned previously, used decision tree methodology in the course of investigating and comparing the recoveries deriving from the collection processes done by the bank's internal collection department or by the external agent. They proceeded from the assumption that the efficiency of the two procedures are influenced by significantly different factors, thus it is necessary to prepare separate regressions for the sake of more accurate modelling.

On the basis of their observations deriving from studying the shape of the distribution, they carried out the following classification by logistic regression:

- in case of the internal collection process they grouped the deals according to the conditions: LGD=0 and LGD>0, and
- in case of the external collection they accomplished the categorization on the basis of the conditions: LGD=1 and LGD<1.

Then they prepared the model building by using the classical linear regression, beta and lognormal transformation, Box-Cox procedure and WOE (weight of evidence) approach. In the framework of the latter one they split the deals into 10 groups on the basis of each significant influencing factor they quantified the ratio of above mean to below mean in all cases, then they calculated weighted averages by using them to working out the LGD. The WOE procedure resulted in the highest $\boldsymbol{R^2}$ figure among all of the methods.

Caselli, Gatti and Querci *(Caselli et al. [2008])* prepared calculation adopting workout LGD methodology, investigating a portfolio which contained 11,649 delayed retail, small and medium enterprise loans. Following the testing of numerous models they finally carried out a linear regression model by using the Ordinal Least Squares (OLS) method, which proved to be the best predictable one among all of the examined alternatives.

In the course of modelling the LGD of four retail credit card products which originated between 1998 and 2004, *Bellotti and Crook [2008]* compared the performance of the Ordinary Least Squares (OLS) regression, the Least Absolute Value (LAV) method, the Tobit model and the Decision Tree algorithm. For this latter procedure they used two logistic regressions to the 0% and 100% recovery rates, and they built a classical multivariate linear regression referring to the recovery rates falling into the [0;1] interval.

They worked with four types of transformation procedures. They tested the reason for the existence of the logit ($G(\alpha+\beta' x)=\left(1+e^{-\alpha+\beta' x}\right)^{-1}$), log ($G(\alpha+\beta' x)=e^{-\alpha+\beta' x}$) and probit transformation in addition to the beta distribution as well. While these latter two procedures are fairly recent in this area, the beta distribution and the logit transformation can be qualified as fairly current ones in the econometric applications *(Papke – Wooldridge [1996])*, and also particularly in the area of the LGD modelling. Logit transformation was used for example by Grippa, Iannotti and Leandri *(Grippa et al. [2005]), Dermine and Neto de Carvalho [2006]*, as well as *Bastos [2009]*.

Bellotti and Crook [2008] measured the fitting of the models in a way that they compared the predicted recovery rates to the data observed on the independent test sample. They examined the value of the Mean Square Error (MSE), the Mean Absolute Error (MAE) and the Pearson’s correlation coefficient, and they experienced that the modelling algorithms’ hierarchy according to the performance considerably depended on the index used.

The author realized that none of the transformation procedures provided consequently the gain in the performance of the models, and none of the distributions carried out by the models they used was so broad as the distribution of the observed recovery rates,

which shows that they were not able to explain the large number of the extreme values (0% and 100%). Partly that is why the models aiming explicitly to treat the special distributions did not perform better, than the OLS regression, because most of the predictions were conservative owing to the weak fitting, and ab ovo fell between the theoretical [0;1] limits.
One of the best models was the LAV regression combined with logit transformation, and the other one was the OLS regression without any transformation, but the differences did not prove to be significant in either case. However, the low standard deviations showed that the findings were consistent for all four deal types.

Zhang and Thomas [2009b] compared numerous different modelling procedures from the simple linear regression across the gamma distribution to the decision tree methodologies, and they carried out the testing on the basis of the $\boldsymbol{R^2}$, the Spearman's correlation coefficient, the Mean Absolute Error (MAE) and the Mean Square Error (MSE).

Bastos [2009] used two types of methodologies in the course of calculating the LGD. He classified into the first one the logit transformation and the logistic regression, which provided the delimitation of the LGD into the [0;1] interval, and the other one was the regression methodology based on the decision trees, which aimed at grouping the deals into relatively homogeneous clusters from the viewpoint of LGD.
As regards the testing procedures, he deployed a wider range of devices. For the sake of judging the performance of the models he analysed the accuracy (the mean square error and the mean absolute error) of the estimates by backtesting for numerous time horizons (12, 24, 36, 48 months). In addition to these he also checked whether the estimation error of the values calculated by the models was less than the one of the predicted recovery rates obtained from the historical long-run average.
The author analysed the data of 374 small and medium enterprise loans of the Banco Comercial Portugues defaulted between June 1995 and December 2000, using the "mortality based" approach of *Altman [1989]*. He discounted the recoveries following the default event with the loan-specific contractual lending rates, and put aside the costs arising during the collection process.
He experienced that the cumulative recovery rates followed bimodal distribution on every time horizons which were investigated (12, 24, 36, 48 months). He also made two further remarks on that:

- on the one hand the standard deviation of the recovery rates decreased more and more on the longer time horizons,
- on the other hand the median almost duplicated from the 12^{th} month to the 24^{th} month, while this symptom did not occur in case of the average.

In case of the decision tree based approach the bifurcations were variant on the different time horizons, but we can generally say that the loan amount, the length of the client's relationship with the bank, the contractual lending rate and the rating grade[69] proved to be the most important classifying criteria.

Comparing the performance of the models by backtesting, the author experienced that on the 12- and 24-month time horizons the decision tree based methodology gave better results, while on the longer horizons (36, 48 months) the log-log transformation did. Additionally investigating the predictability of the models he came to the observations that whereas the decision tree methodology performed prominently, the efficiency of the log-log transformation procedure was even worse then the one of the model based on historical averages.

Chalupka and Kopecsni [2009] applied three types of transformation functions in their study about LGD modelling based on the data of the Czech small and medium enterprise (SME) and corporate loan deals: the logit, the log-log ($G(\alpha+\beta' x)=e^{-e^{-\alpha+\beta' x}}$) and the complementary log-log link ($G(\alpha+\beta' x)=1-e^{-e^{-\alpha+\beta' x}}$), and they estimated the parameters of the models with ML procedure.

They cut the LGD values at 0% and 100% away. In addition to the general and the beta distribution based models they also prepared models by ordinal regression for six discrete LGD categories (grades)[70] as alternative procedure, because they experienced that significant differences appeared between the fairly homogeneous groups, which were carried out in this way, concerning the coefficients of the influencing factors and the probability. They found that while the distribution of the recovery rates and the LGD was bimodal for the whole range of the observations, it was more normally distributed within the certain groups.

Roughly speaking, according to their experiences the fitting of the models based on beta distribution slightly fell behind the other models at all points.

[69] The rating grade was not available in case of approximately the half of the deals, thus the author imputed it with the average. Because of the considerable manipulation it can not be eliminated that the results are distorted.

[70] Corresponding with the Moody's limit values the boundaries of the LGD grades are: 0%-10%, 10%-30%, 30%-50%, 50%-70%, 70%-90%, 90%-100%

Altman and Kalotay [2010], by mixing normal distributions, developed a procedure, which is flexible enough to be appropriate for modelling the recovery rate of non-performing loans and bonds. In the very last time other researchers also prepared numerous studies, in the framework of which they also showed the use of mixtures of distributions.

Table 19: Summary of the literary empirical results VI

AUTHORS	METHODOLOGY AND RESULTS
Polívka [2008]	Examining the facilities of the parametric and non-parametric modelling. The distribution of the recovery rates is bimodal or unimodal with fat tails in many cases, with recovery rates as well which are less than 0% or larger than 100%.
Peter [2006]	Using scenarios, logistic regression and Markov-chains.
Thomas et al. [2007a; 2007b]	Investigating and comparing the recoveries deriving from the collection processes done by the bank's internal collection department or by the external agent. Model building by using the classical linear regression, beta and lognormal transformation, Box-Cox procedure and WOE approach. The WOE procedure resulted in the highest R^2 figure.
Caselli et al. [2008]	The linear regression model by using the Ordinal Least Squares (OLS) method proved to be the best predictable.
Bellotti – Crook [2008]	Comparing the performance of the Ordinary Least Squares (OLS) regression, the Least Absolute Value (LAV) method, the Tobit model and the Decision Tree algorithm. Four types of transformation procedures: beta, logit, log and probit. Testing: comparing the predictions to the data observed on the independent test sample. None of the transformation procedures provided consequently the gain in the performance of the models.
Zhang – Thomas [2008]	Numerous different modelling procedures from the simple linear regression across the gamma distribution to the decision tree methodologies. Testing: R^2, Spearman's correlation coefficient, Mean Absolute Error (MAE), Mean Square Error (MSE).
Bastos [2009]	The "mortality based" approach of *Altman [1989]*. Discount rate: loan-specific contractual lending rates. Putting aside the costs which arise during the collection process. Use of logit transformation, logistic regression and regression methodology based on the decision trees. Testing: backtesting for 12, 24, 36 and 48 months time horizons, mean square error, mean absolute error, estimation error. The recovery rates followed bimodal distribution. The most important classifying criteria: loan amount, length of the client's relationship with the bank, contractual lending rate, rating grade.
Chalupka – Kopecsni [2009]	Three types of transformations: logit, log-log and complementary log-log. Cutting the LGD values at 0% and 100% away. Alternative procedure: preparing models by ordinal regression for six discrete LGD categories. The distribution of the recovery rates and the LGD was bimodal, but it was more normally distributed within the certain groups.
Altman – Kalotay [2010]	Mixtures of distributions, mixing normal distributions.

(Self-made table)

Generally speaking the use of more complex methods did not result in better estimations in many cases than the Ordinal Least Squares (OLS). However, in Chapter 6 in the course of my own research's exposition I show as well that the LGD distribution of the portfolio which I examined is bimodal, and the LGD distribution of the categories based on the closing type of the deals merely differs from each other, so their partitioned treatment is reasonable, the assumption of the "classical" normal distribution can not cope.

5.3. Workout LGD methodology concerning the mortgage loan deals

The number of the studies about using the workout LGD methodology for the mortgage loan deals is considerably low, however, some factors can be pointed out, whose effect on the recovery rates the authors of these publications investigated in details.

Numerous studies were created about analysing the relation between the Loan-to-Value (LTV) ratio and the recovery rates, so for example *Clauretie and Herzog [1990]*, Lekkas, Quigley and Van Order *(Lekkas et al. [1993])*, as well as *Calem and LaCour-Little [2004]* observed strict relationship.

The length of the period from the loan's origination to the default, the size of the loan amount and the differences relating to the collection, liquidation processes also proved to be significant influencing factors.

The deficiency that the researchers paid rather little attention to the impacts of the changes in the real estate market's circumstances and the economic recession on the recovery rates or the LGD, partly can be explained with the unavailability of the required data to that. The data scarcity still means a considerable problem in Hungary, so the investigation about it is rather difficult even now.

Important and generally consistent achievements arose in the international literature respecting the investigation of numerous factors' influencing role.

For example *Moral and Garcia-Baena [2002]*, as well as *Moral and Oroz [2002]*, who analysed the recovery data of 1532 non-performing Spanish mortgage loans and 3887 non-performing Spanish retail mortgage loans deriving from specialised banks, basically examined what kind of differences can be experienced between the loss rates of the deals which are involved in legal proceedings and that of the others.

They considered the deals as defaulted which had arrears for more than 90 days. This definition is consistent with the prescription according to the Basel II, but entirely departed from the more general determination applied by *Felsovalyi and Hurt [1998]*, as well as *Eales and Bosworth [1998]*.

The authors used several simplifying premises: for example they calculated with a fix 5% as discount rate, and considered as 0% the negative LGD values. Among all the cost types they paid attention only to the ones relating to the usucaption and the legal costs.

They experienced that the loss rate of the mortgage loans which were involved in legal proceedings were generally higher than the ones of the other non-performing deals: the average LGD of the Spanish mortgage loans arrears for more than 90 days was 12.65%, while the one of those which were involved in legal proceedings was 28.2%. Conversely they found the distribution in both cases asymmetric, left-skewed (the median fell behind the average with 1.1 percentage points in the first case and with 2.45 percentage points in the second case).

At the conference titled "*Workshop and conference on Basel II & Credit Risk Modelling in Consumer Lending"* organised in September 2006 in Southampton *Allan Lucas* suggested the collection process based modelling, the decision tree based approach for calculating the LGD for the mortgage loan deals in his presentation (mentions: *Thomas et al. [2007a]*). In its framework the probability of the repossession and the recovery which may be realised from the sale have to be quantified. The researcher dealt with the adaptability of the decision trees combined with regression procedure and scorecard building for LGD modelling of secured loan products.

Thomas, Mues and Matuszyk *(Thomas et al. [2007a; 2007b])* made known a similar methodology in their study presented earlier, but they focused on the unsecured deals.

Qi and Yang [2007; 2009] investigated the loss rate of high Loan-to-Value (LTV) loans secured by real estate property collateral on the basis of historical loan-level default and recovery data from mortgage insurance companies. They disposed of the data of 241,293 deals referring to the period between 1990 and 2003.

They experienced that the different factors of the loan, the underlying real estate and the collection, liquidation process mainly explained the run of the LGD. The CLTV (Current Loan-to-Value) proved to be the most important influencing factor, which the researchers defined as the quotient of the exposure at default and the property value at the same time.

As a result of examining the impact of the economic cyclicality on the loss rate, they came to the observation that the loss appeared significantly higher on the mortgage loans in unfavourable economic situation then in case of normal market conditions.

In the course of calculating the LGD they considered the accrued interest as well as the legal and property maintenance expenses, sales costs and repairs besides the exposure at default and the following recoveries. On experimental and professional grounds they defined the legal costs as 5% of the exposure at default event, and the property maintenance expenses as 3% of a specially calculated[71] property value.

They took strong simplifications with regard to the discounting: they uniformly used the 1-year LIBOR, which distorted the results of their LGD estimations downwards, since they absolutely disregarded the uncertainty of the cash flows, the risks relating to the non-performing deals.

As indicator of the housing market conditions they used the house price index (HPI – repeat-sales house price index) reported by the OFHEO (Office of Federal Housing Enterprise and Oversight), and they quantified from it the HPR (house price ratio) as the quotient of the current HPI and the HPI 18 months earlier.

The period between 1990 and 1994 was featured by a HPR less then 100%, so the authors considered this period as a downturn from the point of view of the housing market conditions. They experienced that the loss rate changed ceteris paribus in common with the housing market situation and the CLTV in consonance with the former theoretical and empirical achievements.

There was a very strong positive relationship (dominant for the other factors) between the CLTV and the LGD. The HPR correlated moderately negatively with both the LGD and the CLTV, and the same relation appeared between the loan amount at origination and the loss rate as well as the LTV. A weak positive correlation showed up between the LGD and the duration of the procedure, while the loan amount at origination presented a weak negative relation to the length of the period between the deal's origination and the default. The type of the property serving as collateral also proved to be a significant influencing factor.

In case of those properties, which were sold in a shorter time, the LGD showed up to be lower, partially because of the smaller sales costs and repairs, and this result was also consistent to the previous experiences. According to the authors' observations the

[71] The calculation of the property value: the smaller one of (1) the sesquialter of the property value at the deal's origination, and (2) the salvage value net of sales costs and repairs.

dissimilarities between the collection, liquidation, legal etc. processes also influenced the LGD.
The duration from the origination of the loan deal to the default related positively to the loss rate. Investigating this question, different results can be read in the literature: for example *Calem and LaCour-Little [2004]* also found positive correlation, while Lekkas, Quigley and Van Order *(Lekkas et al. [1993])* experienced just the opposite relationship.

Qi and Yang [2007; 2009] prepared a general linear regression model on the basis of the characteristics of the loan and the property, as well as the housing market conditions. When they involved the CLTV as explanatory variable as well, the adjusted coefficient of determination was 0.662, which proved to be considerably high in comparison with the studies of similar topic. This regression equation contained both time-varying (for example CLTV, duration from the origination of the deal, indicator characterising the real estate market situation) and non-time-varying factors (for example LTV, loan amount at origination, type of the real estate).
The authors also investigated that if they got the LTV into the model instead of the CLTV, what kind of consequences would arise to the explanatory power of the model. They experienced that mainly the same variables proved to be significant in this regression, but the value of the adjusted coefficient of determination decreased drastically (from 0.662 to 0.07).
However, significant relation did not appear between the CLTV and the LTV, which derived from the fact that the high LTV loans were overrepresented among the observations, so the LTV scattered in a quite narrow range, while the CLTV was notably influenced also by the further changes of the house prices, and made it more differentiated during the period from the origination of the deal to the default.
The achievements published in the study indicate that the use of the regularly actualised CLTV instead of the LTV is more adequate at all points in the course of calculating the LGD. Considering that the frequent revaluation of the properties is rather expensive the properly actualised data are not available in many cases, at the same time it is indisputable that the appropriate accuracy of the LGD estimate is particularly important for the credit institutions, thus it is worth expanding money upon that.
On the grounds of the experiences of *Qi and Yang [2007; 2009]* I also lay large emphasis on investigating the LGD influencing role of this factor in the course of my empirical research.

6. The empirical research and its results

In the framework of this dissertation I study the specialities of the LGD parameter of the retail mortgage loans, and I take steps to prepare a model with which more exact and more accurate LGD calculation will be possible. In the course of the expounding I do not focus on the calculated LGD values, but on introducing the influencing factors and the models, and valuating their performance.

In the first part of this chapter I demonstrate the database which serves as a basis for my empirical researches, then I make known the terms and assumptions which I used, as well as some methodological decisions, finally I present the concrete analyses and their results.

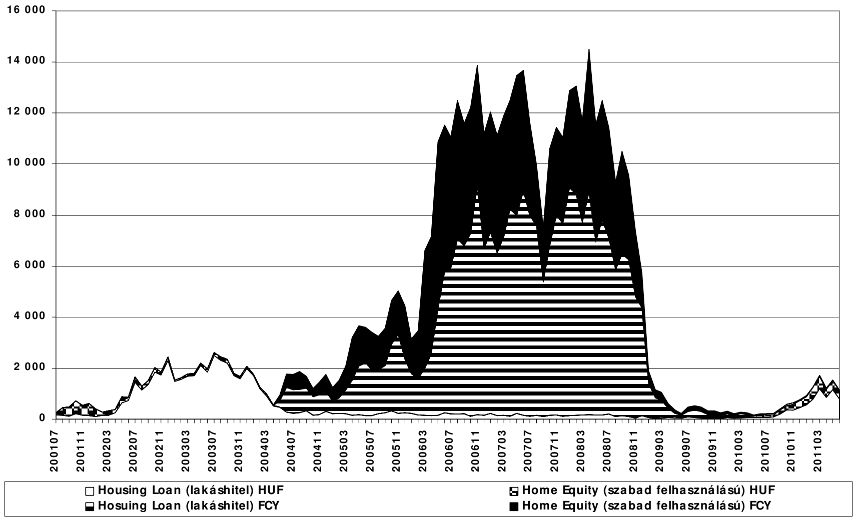
6.1. Database

In the course of my research I applied the data of an anonymous commercial bank's database of closed and non-closed retail mortgage loans. I did not analyse the whole portfolio, but I extracted the deals which were concerned with restructuring or which were secured by life insurance, because these deal types showed very different characteristics from the others.

6.1.1. Portfolio characteristics

The following Figures (Figure 9 and Figure 10) illustrate the run of the new loan originations during the last few years regarding the subportfolio which I analysed. In Figure 9 the total amount of the new retail mortgage deals originated in certain months can be seen divided according to whether the loan is denominated in Hungarian Forint (HUF) or in foreign currency (FCY).

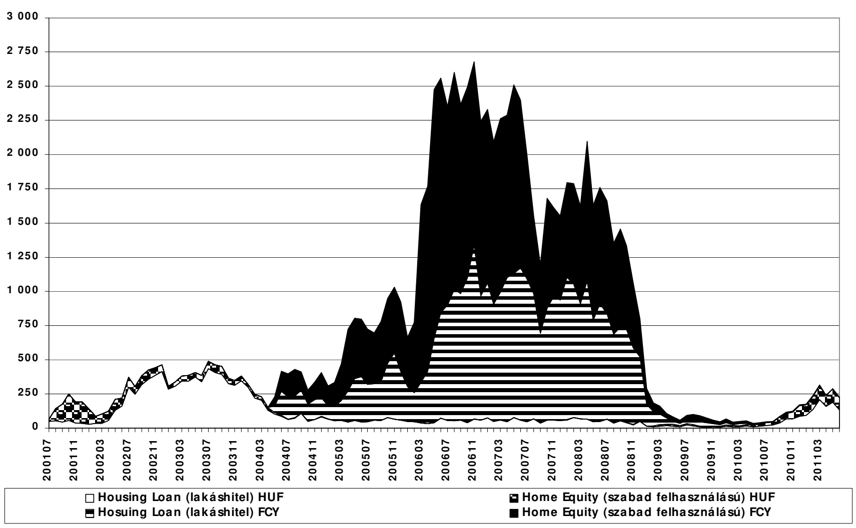
Figure 9: Newly originated deals (million HUF)



(Self-made figure: own calculation results)

Figure 10 shows the quantity of the newly originated deals in each month.

Figure 10: Newly originated deals (counts)



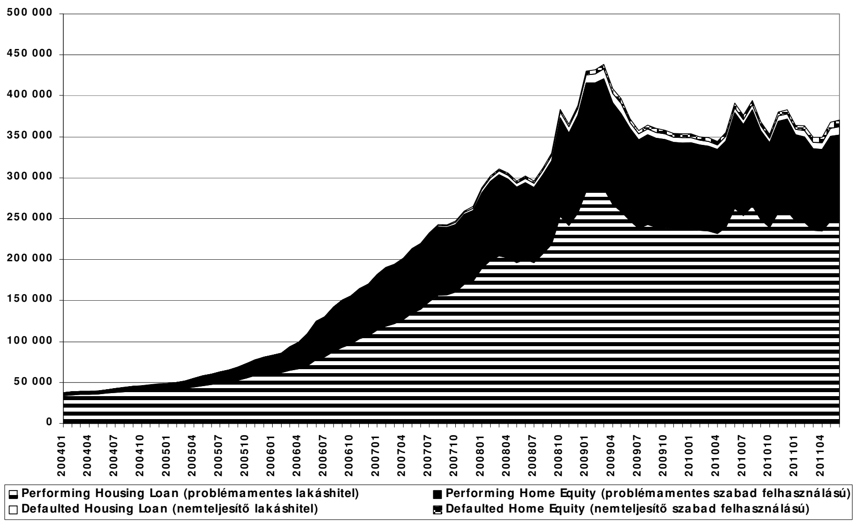
(Self-made figure: own calculation results)

The dynamic growth until 2008, which can be seen in the following figures, is mainly the result of the general boom on the Hungarian lending market during the past few years. The government subsidized mortgage loan program, which started in 2001, intensively increased the credit taking appetite of the people, then in December 2003 when the Government continued to enhance this policy, several credit institutions decided to launch foreign currency credit lending to take the advantages of the low level of the interest rates. Subsequently the foreign currency denominated loans incrementally took the place of HUF loans, almost displaced them.

The turning befell in the autumn of 2008, when the credit institutions executed serious lending restrictions on account of the financial crises. Due to the drastic HUF depreciation CHF credit lending has practically been stopped, and as a consequence of the crisis and the restrictions only minimal new volume had been disbursed during 2009 and 2010.

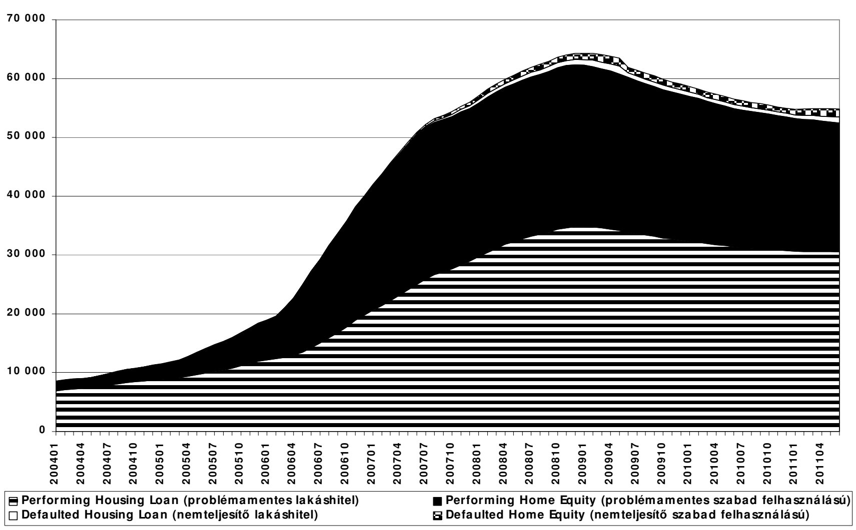
According to these changes in lending policy, the total exposure of the retail mortgage loans did not grow on in the last two years. The following Figures illustrate the run of the total exposure in HUF millions (Figure 11) and the quantity of them (Figure 12).

Figure 11: Total exposure (million HUF)



(Self-made figure: own calculation results)

Figure 12: Total exposure (counts)



(Self-made figure: own calculation results)

The CRD prescribes the use of the so-called downturn LGD in order to calculate the risk weighted assets, in the course of which also the changes arising from the cyclicality of the economic conditions have to be taken into account. Considering that due to the crisis, which started in September 2008, a considerable proportion of the portfolio is derived from the economic downturn period, so in the course of calculating the LGD further adjustment is not necessary to reflect the impact of the economic recession.

6.1.2. Data used

In this subsection I present the data, sorting on the basis of the data sources, which I used in the course of calculating the LGD, then I make known the steps of applying the data and the process of the calculation in Chapter 6.2.

(a) Application data

The first block is composed by the application data, whose majority respects to the clients, who apply for the loan, and the minor part comes from the characteristics of the deals at the origination. On the basis of the greatly expansive dataset which was available, I produced the structured data table containing the following elements:

Table 20: Basic data at the application (known at the date of origination of the deal)

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
deal_id	Deal identification number.
start_term	The duration term of the deal according to the contract (number of months).
loan_purpose	Purpose of the loan: ▪ Real estate purchase, ▪ Real estate construction, ▪ Real estate renovation, ▪ Other.
loan_amount_lcy	Loan amount which was applied for and paid out (in HUF).
coapplicant_flag	Dummy variable which indicates whether there is a co-applicant.
first_instalment	The original monthly repayment amount (in HUF).
full_name	The full name of the client.
gender	The gender of the client: ▪ Male, ▪ Female.
citizenship	The citizenship of the client: ▪ Hungarian, ▪ Other.
birth_settlement	The birth place of the client.
start_age_months	The age of the client at the origination of the deal (number of months).

Table 20 (continuation): Basic data at the application (known at the date of origination of the deal)

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
marital_status	The marital status of the client: ▪ Single, ▪ Married, ▪ Divorced, ▪ Companion, ▪ Widow.
education_level	The education level of the client: ▪ Elementary, ▪ High-school graduate, ▪ Other secondary education, ▪ BSc, ▪ MSc, ▪ Other higher education.
home_settlement	The name of the settlement of the client's living place.
landline_phone_flag	Dummy variable which indicates whether the client has a landline phone.
mobile_phone_flag	Dummy variable which indicates whether the client has mobile phone.
start_address_months	The duration of living at the given permanent address at the origination of the deal (number of months).
empl_industry	The industry of the client's employer: ▪ Agriculture, ▪ Commerce / Entertainment, ▪ Construction, ▪ Education / Medical services / Government, ▪ Finance / Legal services, ▪ Processing industry, ▪ Other.
empl_type	The type of the client's employment: ▪ Employee, ▪ Entrepreneur, ▪ Retired, ▪ Other.
empl_position	The working position of the client: ▪ Blue collar, ▪ Middle leader, ▪ Upper leader, ▪ Other intellectual.
empl_term	The type of the client's labour contract: ▪ Definite - Full time, ▪ Indefinite - Part time, ▪ Indefinite - Full time.
start_work_months	The duration of working for the given employer at the origination of the deal (number of months).
applicant_net_income	The monthly net income of the client.
total_household_income	The total monthly income of the household of the client.
earners_number	The number of earners in the household of the client.
dependents_number	The number of dependents in the household of the client.
existing_ca_flag	Dummy variable which indicates whether the client has a current account.
existing_card_flag	Dummy variable which indicates whether the client has a credit card.
existing_ovd_flag	Dummy variable which indicates whether the client has an overdraft.
existing_loan_flag	Dummy variable which indicates whether the client has another credit.
interest	The original lending rate of the deal.
apr	The Annual Percentage Rate of the deal at the origination.

(Self-made table)

Considering that the Annual Percentage Rate (APR) was not available in case of all deals, imputation became necessary. In the framework of that I quantified the average APR values for each month, currency and deal type according to the purpose of the loan then I refilled the missing values with them.

(b) Behavioural data

While the application data give a static image about the characteristics of the certain deals and clients, the behavioural data show the run of some treats of the deals concerning the whole duration of the loans from time to time.

The bank's database, which was disposable for me, held the behavioural data of the deals relating to the last workday of each month on deal level. Considering that I focused on the retail mortgage loans in the course of my research, I filtered the data according to the type of the client and the product group. In order that to make it possible to investigate the impact of changing the materiality threshold (5th Hypothesis), I defined dummy variables to indicate whether the given deal was voted non-performing in the actual month in case of applying the different materiality thresholds. Beyond that I also constructed indicator codes for the sake of indicating the reason of the default concerning each materiality threshold which I examined.

On the basis of all that I made up the data table with the under-mentioned content (Table 21):

Table 21: The behavioural basic data of the deals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
deal_id	Deal identification number.
basic_number	Client identification number.
product	Type of the product: ▪ Housing Loan: Mortgage loan with home purpose, ▪ Home Equity: Mortgage equity withdrawal.
product_description	Sub-type of the product: ▪ Normal: Normal mortgage loan, ▪ With life insurance: Mortgage loan secured by life insurance, ▪ Restructuring: Restructuring mortgage loan.
application_type	Category according to the type of the application: ▪ Asset-based: Purely collateral-based loan, without income verification, ▪ Income-based: Loan based on income verification.
exposure_lcy	The actual exposure at the end of the month (in HUF).
exposure_ccy	The actual exposure at the end of the month (in the original currency of the deal).
principal_lcy	The actual principal amount at the end of the month (in HUF).
principal_ccy	The actual principal amount at the end of the month (in the original currency of the deal).

Table 21 (continuation): The behavioural basic data of the deals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
start_principal_lcy	The disbursed loan amount (in HUF).
start_principal_ccy	The disbursed loan amount (in the original currency of the deal).
dpd	The number of days past due at the end of the given month.
past_due_amount_lcy	The delayed amount at the end of the given month (in HUF).
past_due_amount_ccy	The delayed amount at the end of the given month (in the original currency of the deal).
defaulted_minwage	Dummy variable which indicates whether the deal has default status in the given month according to the materiality threshold, which is defined by the lowest monthly minimum wage.
default_reason_minwage	The indicator variable which indicates the reason of the default according to the materiality threshold which is defined on the basis of the lowest monthly minimum wage.
defaulted_huf50000	Dummy variable which indicates whether the deal has default status in the given month according to the materiality threshold of HUF 50000.
default_reason_huf50000	The indicator variable which indicates the reason of the default according to the materiality threshold of HUF 50000.
defaulted_huf20000	Dummy variable which indicates whether the deal has default status in the given month according to the materiality threshold of HUF 20000.
default_reason_huf20000	The indicator variable which indicates the reason of the default according to the materiality threshold of HUF 20000.
defaulted_huf2000	Dummy variable which indicates whether the deal has default status in the given month according to the materiality threshold of HUF 2000.
default_reason_huf2000	The indicator variable which indicates the reason of the default according to the materiality threshold of HUF 2000.
defaulted_huf0	Dummy variable which indicates whether the deal has default status in the given month according to the materiality threshold of HUF 0.
default_reason_huf0	The indicator variable which indicates the reason of the default according to the materiality threshold of HUF 0.
write_off_lcy	The loss which has been written off in the given month.
ccy	The original currency of the deal.
start_date	The date of the origination of the deal.
maturity_date	The contractual maturity date of the deal.

(Self-made table)

In the course of working up the subtypes of the deals *(product_description)* I attempted to establish quite homogeneous groups, because I assumed that significant differences can be experienced among their LGD values. The circumscription served the purpose to enable me to filter out the deals from the analysis which were concerned by restructuring or secured by life insurance. I considered as concerned by restructuring not only the deals which the clients claimed for restructuring their already existing loans (successor deals), but the ones as well, which served as ancestors deals. This was necessary, because in the case of these loans the same default definition could not have been applied, thus the testing of the impacts of changing the default definition (1^{st} Hypothesis) would have become impossible. The disposability of the client identification number *(basic_number)* technically enabled me the joining of the concerned deals to each other.

The circumscription of the categories according to the type of the application *(application_type)* was justified by the fact that the maximum LTV-ratio is considerably higher in the case of the loans which are based on income verification, than in the case of the purely collateral-based financings, so I also presupposed significant differences concerning the risk level. I investigated the impact of this feature on the LGD values in the course of the 3rd Hypothesis.

(c) Data referring to the collaterals

Also monthly level data were obtainable for me concerning each collateral underlying the deals. In order to make the recoveries of the loan deals, which were examined by me, comparable with the recoveries of the Hungarian Interbank LGD Database (Chapter 3.5), I tried to construct a data table which possesses equivalent content to the Hungarian Interbank LGD Database (Table 22), according to the pieces of information about the collaterals. In the case of some data fields (for example the floor-space, the number of rooms, the year of the building and the renovation) the lack of data was so considerable that it could not have been handled by imputation reliably, thus finally I left out these variables from the analysis.

Table 22: The basic data referring to the collaterals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
collateral_id	Collateral identification number.
deal_id	Deal Identification number.
appraisaldate	The date of the original appraisal (prior to the disbursement of the loan).
revaluedate	The date of the latest revaluation which is effective in the given month.
priorcharge_amount	The sum of the prior charges on the collateral (in HUF).
start_collvalue	The realization value of the collateral at the origination of the deal.
loancoll_value	The realization value of the collateral at the end of the given month.
start_marketvalue	The market value of the collateral at the origination of the deal.
marketvalue	The market value of the collateral at the end of the given month.
zipcode	The zip code of the real estate which serves as collateral.
settlement	The name of the settlement of the real estate which serves as collateral.
realestate_type	The type of the real estate which serves as collateral: ▪ Detached house, ▪ Owner-occupied block, ▪ Other residential property.
material	The building type of the real estate which serves as collateral: ▪ Brick or stone, ▪ Prefabricated, ▪ Light construction or wood, ▪ Other.

(Self-made table)

All through the categorization according to the type of the real estate *(realestate_type)* and the building type *(material)* I kept the requirement in view that the same grouping should come up as the one which exists in the Hungarian Interbank LGD Database for the sake of making feasible the comparison of the recoveries.

(d) Recoveries and direct costs

I constructed a data table from the recovery amounts and the indirect costs as well. In addition to the deal identification number, the currency and the amounts given in the original currency of the deal I also disposed the date of the paying-up of the recovery and the occurring of the cost, and considering that the whole process of LGD estimation grounds on HUF-amounts, I exchanged the recoveries and the costs from the original currency of the deal to HUF on the exchange rate effective at their emergence date. The table below shows the content of the data table, which was constructed in this manner (Table 23).

Table 23: Recoveries and direct costs

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
deal_id	Deal identification number.
ccy	The original currency of the deal.
repayment_date	The value date of accounting the recovery or the indirect cost.
principal_lcy	The principal recovery amount (in HUF).
interest_lcy	The interest recovery amount (in HUF).
charge_lcy	The charge recovery amount and the accruing direct cost (in HUF).
principal_ccy	The principal recovery amount (in the original currency of the deal).
interest_ccy	The interest recovery amount (in the original currency of the deal).
charge_ccy	The charge recovery amount and the accruing direct cost (in the original currency of the deal).

(Self-made table)

(e) Macroeconomic data

For the sake of investigating the effects of the general macroeconomic situation on the LGD I collected some indicators which I considered as potential LGD influencing factors in the course of my empirical research. The Hungarian Central Statistical Office's (HCSO) STADAT Database served as a source of the majority of the data, while the probabilities of defaults are results from the internal estimations of the bank.

Table 24: Macroeconomic basic data

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
month	The month which the macroeconomic indicators refer to.
unempl_rate	Quarterly average unemployment rate (STADAT 3.10.).
min_wage	The official lowest monthly minimum wage (STADAT 2.1.40.).
avg_netincome	Average monthly net income: until December 2007 the 12-month moving averages calculated from the yearly averages (STADAT 2.1.34.1., STADAT 2.1.34.2.), from January 2008 the monthly figures according to the HCSO (STADAT 2.1.37.).
CPI	Yearly consumer price index: until December 2006 the 12-month moving averages calculated from the yearly averages (STADAT 3.6.1., 2.1.41.), from January 2007 the monthly figures according to the HCSO (STADAT 3.6.1.).
cum_CPI	Fixed-base consumer price index according to the STADAT 3.6.1. (base: January 2001).
realwage_index	Yearly real wage index: the quotient of the 12-month moving average calculated from the change of the average monthly net income (avg_netincome) and the yearly consumer price index (CPI).
cum_realwage_index	Base ratio of the monthly real wage according to the realwage_index (base: January 2001).
cum_GDP_growth	Base ratio of the GDP-growth: base ratio which is calculated from the increasing of the seasonally adjusted GDP values on a quarterly basis (STADAT 3.1.6.), using geometric average (base: January 2001).
GDP_growth	Yearly GDP-growth index: 12-month moving average of the yearly GDP-growth indices which are calculated from the cum_GDP_growth.
HomeEquity_PD	Average PD of the mortgage equity withdrawals at the given month.
HousingLoan_PD	Average PD of the home loans at the given month.
avg_PD	Average PD of the mortgage loans at the given month.

(Self-made table)

In addition to the data enrolled in Table 24 I also used the central bank base rates in the course of estimating the LGD, but considering that they occasionally changed during the month as well, I linked the values of the central bank base rate of the proper currency effective at the time of default event and the values of them effective on 30th June 2001 directly to the certain deals.

In the course of my analyses I made the estimates and built the regression models using the data made known previously. However, before presenting the concrete analyses and their results I consider it as necessary to make known the terms and assumptions which I used, as well as propounding and justifying some methodological decisions. I summarize these aspects in the following subsection.

6.2. Definitions and assumptions

The data tables, which were presented beforehand, contain the deals which are in normal status (not in default status) as well, therefore in the next step I defined the date of all default events of each deal, and I created a data table (Table 21) from the behavioural data which comprehends only the non-performing deals. I think it is important to note that if a certain deal has "cured" after the default, then later on it became non-performing again, I handled all default events separately, so I considered all default events as particular cases from the viewpoint of estimating the LGD.

To select the non-performing deals, in the first step I had to define the mere default event.

6.2.1. The default event

The CRD and the Hungarian prescriptions *(Hkr. 68-69. §)* served as a basis for defining the term "default event".

The calculation of the number of the days past due (DPD) is fundamental to the definition of default. If a client fails to meet one or more instalments of the certain loan, this deal becomes delinquent. The counting of the DPD starts with the first day when an instalment is overdue, so the DPD measures the number of days since the due date of the earliest and currently unpaid past due obligation. If later on the client pays money on his account, then this covers the oldest arrear at first, namely the oldest past due obligation is satisfied foremost, then the other instalments one after the other. If the arrear is paid in full, the deal becomes to normal status again and the DPD is restored to 0.

The establishment of the term "materiality threshold" was needed for the sake of not considering the deals as non-performing in cases when the amounts in arrears are negligible or when the delays occur because of technical reasons. In the basic model the highest delayed amount which is not defined as delinquent (the overdue amount is considered as immaterial) is the minimum of the under-mentioned values:

- the lowest monthly minimum wage effective at the time of becoming delayed,
- 2% of the obligations of the client, and
- one monthly repayment instalment.

It means that counting the days past due (DPD) starts on the day, when the overdue obligations exceed this calculated amount. The most common reason for going into default status is that the DPD for the deal goes above 90, and at the same time the total past due obligation exceeds the prescribed materiality threshold. If the client executes a payment thereafter, and therefore the DPD decreases below 90, then this results in the "recurring" of the deal. The case is an exception to this rule, when the delay of the deal with a material past due amount reaches 181 days, namely in this case the total exposure becomes due, consequently later on the deal is considered as defaulted irrespectively of its current DPD and past due obligation until its closing.

There are two further efficient causes of qualifying the deals as non-performing: the decease of the client and the fraud. The decease of the client results in the deal is becoming to default status, but if the inheritor takes over the loan, then the deal get to normal status again. Also the fraud (for example manipulating the evaluation of the collateral) generates the qualifying as non-performing, but this default status is irrecoverably, it results in the total exposure is becoming due immediately.

So generally speaking a deal is considered as defaulted in the basic model if either of the below conditions holds:

- The client is in delay for more than 90 days with the instalments of the deal, and the past due obligation is more than the lowest monthly minimum wage effective at the time of becoming delayed or 2% of the obligations of the client or one monthly repayment instalment.
- The client was in delay for more than 180 days with instalments of the deal at any time, and the past due obligation exceeded the lowest monthly minimum wage effective at the time of becoming delayed or 2% of the obligations of the client or one monthly repayment instalment.
- It is inferential that the loan will not be paid back, because the client died or a fraud occurred.

If any of these conditions obtain in connection with a loan of a client, then all the other loans of the given client is also considered as non-performing (cross-default), so the term “default status” acts in my empirical analysis as a client-level category.

The 4th Hypothesis was directed towards survey, how the change of the materiality threshold influences the LGD values. For the sake of that I decided to use four different alternative materiality thresholds (HUF 50000, HUF 20000, HUF 2000, HUF 0), but for the comparability I left unchanged the other parameters of the default definition (DPD-counting, cross-default, consideration of the other default reasons), so enabling the separate investigation of the effects derived from modifying the materiality threshold.

Considering that in the course of estimating the LGD the exposure at the date of the default event means the reference point, I quantified both this amount and the reasons behind the non-performing status (Table 25), then I joined them to the behavioural data of the deals.

Table 25: Data about the default

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
default_date	The date of the default event of the deal.
default_month	The period of the default event of the deal (year, month).
months_to_default	The duration from the origination of the deal to the default event (number of months).
defaulted_exposure_lcy	The exposure of the deal at the date of the default event (in HUF).
orig_default_reason_minwage	The indicator variable at the date of the default event which indicates the reason of the default because of arrears according to the materiality threshold which is defined on the basis of the lowest monthly minimum wage.
orig_default_reason_huf50000	The indicator variable at the date of the default event which indicates the reason of the default because of arrears according to the materiality threshold of HUF 50000.
orig_default_reason_huf20000	The indicator variable at the date of the default event which indicates the reason of the default because of arrears according to the materiality threshold of HUF 20000.
orig_default_reason_huf2000	The indicator variable at the date of the default event which indicates the reason of the default because of arrears according to the materiality threshold of HUF 2000.
orig_default_reason_huf0	The indicator variable at the date of the default event which indicates the reason of the default because of arrears according to the materiality threshold of HUF 0.
defaulted_per_start_exposure	The proportion of the exposure at the default and the disbursed amount.
reason_fraud	Dummy variable which indicates whether the deal is considered as defaulted because of fraud.
reason_death	Dummy variable which indicates whether the deal is considered as defaulted because of death.
reason_pastdue_minwage	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold which is defined on the basis of the lowest monthly minimum wage.
reason_pastdue_huf50000	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold of HUF 50000.
reason_pastdue_huf20000	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold of HUF 20000.
reason_pastdue_huf2000	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold of HUF 2000.
reason_pastdue_huf0	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold of HUF 0.
default_age_months	The age of the client at the date of the default event of the deal (number of months).
default_address_months	The duration of living at the given permanent address at the date of the default event (number of months).
default_work_months	The duration of working for the given employer at the date of the default event (number of months).
default_fx_rate	The exchange rate of the deal's currency at the date of the default.
default_unempl_rate	Unemployment rate at the date of the default.
default_min_wage	The lowest monthly minimum wage at the date of the default.
default_avg_netincome	Average monthly net income at the date of the default.
default_realwage_index	Yearly real wage index at the date of the default.
default_CPI	Yearly consumer price index at the date of the default.
default_GDP_growth	Yearly GDP-growth index at the date of the default.

(Self-made table)

6.2.2. Calculating the net recoveries on deal level

The measurement of the recoveries involves all cash recoveries and non-cash items regardless of their source (for example payment from the clients, repossession or selling of the collaterals). Relating to the certain recoveries only the date of the coming-in and the amount were available in the database which I examined, thus the different handling of the distinct types of the recoveries was not feasible, but considering that the Collection Department keeps a separate file about the deal identification numbers of the loans, in case of which the real estate, which served as collateral, has been sold, it became possible for me to compare the recoveries of the Hungarian Interbank LGD Database with the recoveries of the deals which I examined.

I treated the penalty fees and penalty interests as well as internal (for example phone call, reminder letter) and external collection costs as negative cash flows in the course of calculating the LGD. Considering that some costs could not be associated with the individual deals (indirect costs), and therefore the concrete deal-level cost amount is not disposable, I allocated the total collection costs of the given month evenly between the deals which are actually in default status each month. The consideration in the background of this decision is that the portfolio, examined by me, contained only retail mortgage loans, in connection with which the intensity of the collection process was not significantly influenced by either the loan amount, or the exposure at the date of the default event, or other similar factor, on the basis of which the proportioning is practicable and logically reasonable.

In the next step I linked each deal with the obtainable recoveries and direct costs on deal level, as well as the monthly overheads computed from the indirect costs, which I calculated in a way that I divided the total indirect collection costs, which occurred in the certain months, with the quantity of the deals which were in default status in the given month. In order that I will be able to examine the effect of using the different discount rates on the LGD values (4^{th} Hypothesis), I also assigned four types of discount rates to the deals.

Table 26: Data which are needed for calculating the discounted net recoveries

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
recovery	The sum of the recoveries of the deal during the given month (in HUF).
direct_cost	The sum of the direct costs which occurred in connection with collecting the deal during the given month (in HUF).

Table 26 (continuation): Data which are needed for calculating the discounted net recoveries

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
indirect_cost	The indirect collection cost overhead in the given month (in HUF).
Interest	The original lending rate of the deal.
apr	The Annual Percentage Rate of the deal at the origination.
def_rate	The central bank base rate of the original currency of the deal effective at the default of the deal.
curr_rate	The central bank base rate of the original currency of the deal effective on 30th June 2011.

(Self-made table)

After collecting the recoveries and the costs I calculated the net recoveries for each deal on monthly level, then I discounted them back to the date of the default using the following formula:

$$PV_t = \frac{\mathrm{Recovery}_t - \mathit{Direct}\ \mathrm{costs}_t - \mathit{Indirect}\ \mathrm{costs}_t}{(1+r)^{t/12}} \qquad (6.1)$$

where: t: the length of the period from the default (year),

r: discount rate.

In case of the basic model I used the contractual lending rate of each deal as discount rate, because it reflects both the differences between the actual interest levels at the date of the origination of certain deals, and on the other hand it varies according to their currency as well. Nevertheless for the sake of investigating the deviations of the LGDs which derived from using different discount rates I quantified the present values of the net recoveries without discounting and with using the alternative discount rates as well, then I summed up the discounted monthly net recoveries on deal level.

Table 27: The nominal and the discounted net recoveries

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
disc_rec_null_lcy	The sum of the cumulative nominal (not discounted) recoveries of the deal (in HUF).
disc_rec_interest_lcy	The sum of the cumulative recoveries of the deal, discounted by the original lending rate (in HUF).
disc_rec_apr_lcy	The sum of the cumulative recoveries of the deal, discounted by the original Annual Percentage Rate (in HUF).
disc_rec_def_rate_lcy	The sum of the cumulative recoveries of the deal, discounted by the central bank base rate according to the deal's currency at the default (in HUF).
disc_rec_curr_rate_lcy	The sum of the cumulative recoveries of the deal, discounted by the central bank base rate according to the deal's currency on 30th June 2011 (in HUF).

(Self-made table)

In the next step I quantified the cumulative discounted recovery rate relating to each month, dividing the cumulative discounted recoveries by the exposure at the default event:

$$CRM_t = \frac{\sum_{i=1}^{t} PV_i}{EAD} \qquad (6.2)$$

where: CRM_t: cumulative discounted recovery rate t months after the default,

PV_i: discounted net recovery in the i^{th} month after the default,

EAD: the exposure at the time of the default.

As the result of this procedure the monthly series of the cumulative discounted recovery rates for each deal were at my disposal,[72] on the basis of their last items the deal level LGDs have become quantifiable:

$$LGD = \begin{cases} 0, & ha \quad 1 - CRM_{t_{MAX}} \leq 0 \\ 1 - CRM_{t_{MAX}}, & ha \quad 0 < 1 - CRM_{t_{MAX}} < 1 \\ 1, & ha \quad 1 - CRM_{t_{MAX}} \geq 1 \end{cases} \qquad (6.3)$$

where: t_{MAX}: the total length of the recovery period considered.

In the course of my analyses t_{MAX} is the duration from the default of the given deal to its "recurring" or its closing.

It is conspicuous from the above formula that I truncated the deal level LGD values at 0% and 100% in accordance with the procedure which is frequently mentioned by the literature, so I considered that the bank can not lose larger amount than the exposure at the date of the default (the LGD can not exceed 100%), and it can not realize larger cumulative recovery than the exposure at the default (the LGD can not be negative).

6.2.3. Pooling the deals according to the closing type

Generally, the aim of pooling is to split the portfolio into homogenous groups from the point of view of the risk on the basis of the characteristics of the product, the deal, the client and the underlying collateral, which factors are expected to influence significantly the recoveries. My 1st and 2nd Hypotheses have a connection with this fact, in the framework of which I investigated the deviations of the LGD values relating to the subportfolios constructed on the ground of the purpose of the loan *(loan_purpose)* and the type of the application *(application_type)*. In case of all three of them characteristics served as a basis for the grouping, which were already known at the origination of the deal, so the certain deals could be squarely assigned to the proper group.

In this subsection I show another sort of using the categorization: in the course of my empirical research I segmented the deals according to the closing type of the collection

[72] Thereafter I always use the term "recovery rate" for the last member of this series.

process, and for the sake of that I defined the date of closing the deal and some connecting data referring to this date.

Table 28: Data about closing the deals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
write_off_flag	Dummy variable which indicates whether the collection of the deal closed with writing off losses.
woe_month	The period of closing the deal (year, month).
woe_months_since_default	The duration from the origination of the deal to the closing (number of months).
woe_exposure_lcy	The exposure of the deal at the date of the default event (in HUF).
woe_defaulted_minwage	Dummy variable which indicates whether the deal has default status in the period of the closing according to the materiality threshold, which is defined by the lowest monthly minimum wage.
woe_defaulted_huf50000	Dummy variable which indicates whether the deal has default status in the period of the closing according to the materiality threshold of HUF 50000.
woe_defaulted_huf20000	Dummy variable which indicates whether the deal has default status in the period of the closing according to the materiality threshold of HUF 20000.
woe_defaulted_huf2000	Dummy variable which indicates whether the deal has default status in the period of the closing according to the materiality threshold of HUF 2000.
woe_defaulted_huf0	Dummy variable which indicates whether the deal has default status in the period of the closing according to the materiality threshold of HUF 0.
fv_crm_lcy	The cumulative nominal (not discounted) recovery rate of the deal referring to the last period of the recovery process.
real_term	The effective duration of the deal (number of months).

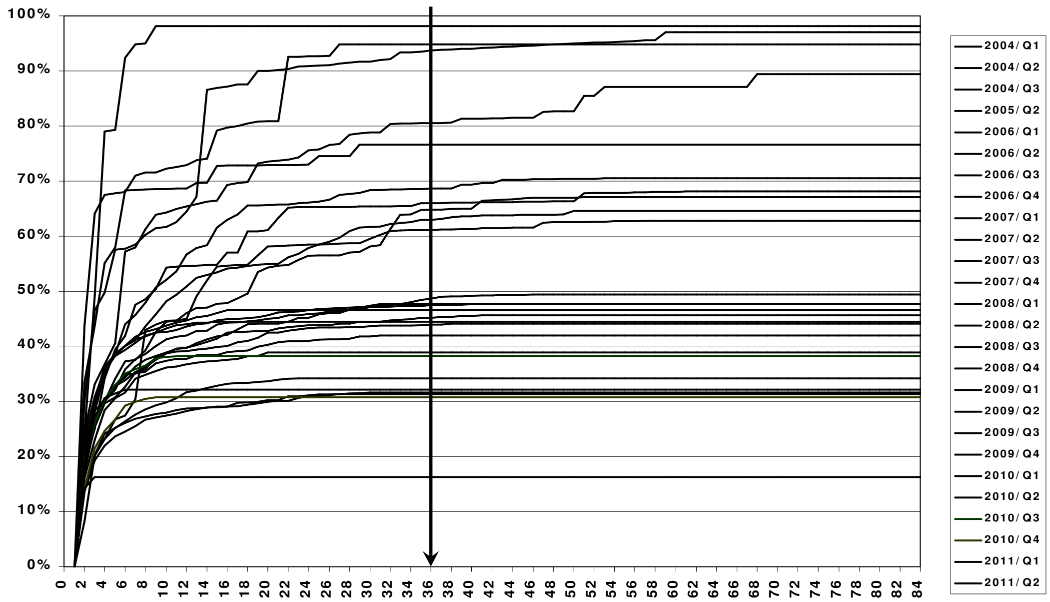
(Self-made table)

Based on all these I worked up the under-mentioned categories *(deal_status)*:

- *"WorkoutEnd"*: The deals which are not in default status any more, because the client has paid back the delayed amount, the exposure has been written off or for example the property which served as underlying collateral has been sold.
- *"NoFurtherRec"*: The deals which are still in default status, since their becoming non-performing longer duration has passed than the effective recovery period, and in case of which at least 90% of the exposure at the date of default has recovered (nominally, without discounting: *fv_crm_lcy*).
- *"NotClosed"*: The deals which can be assigned to neither of the previous categories, in case of which the collection procedure is still in progress.

In the course of my analyses I considered the effective length of the collection period 36 months, because analyzing the data of the database I experienced on the basis of the discounted cumulative recovery rates that regarding the majority of the quarters considerable recoveries did not occur after the first 36 months of the collection process (Figure 13).

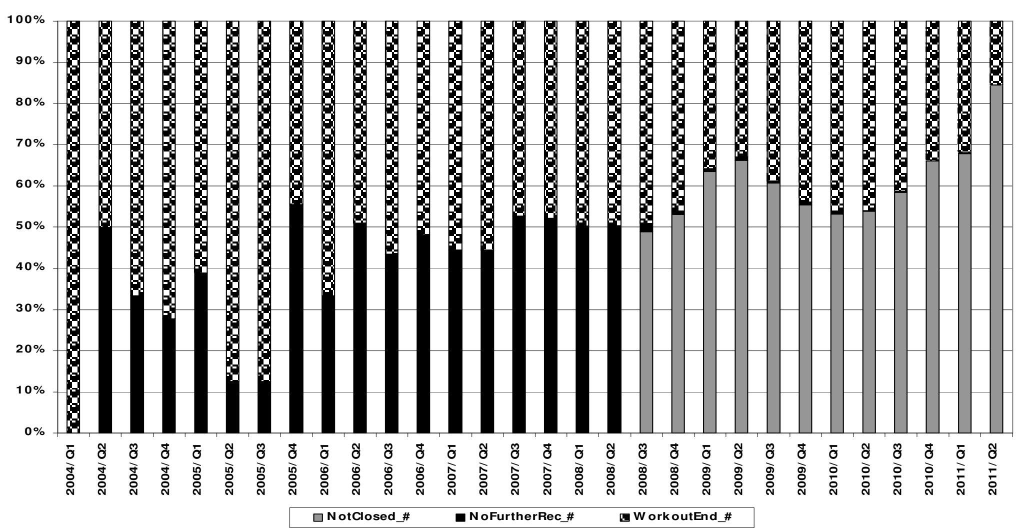
Figure 13: The run of the discounted cumulative recovery rate on a quarterly basis



(Self-made figure: own calculation results)

The unclosed deals in case of which the length of the collection period (the time from the default event) exceeded the 36 months, were classified into the *"NoFurtherRec"* category, increasing the quantity of the deals which can be involved in the analysis, since in case of the basic model I carried out the calculations on the basis of those deals only, which are relating to the first two groups *("WorkoutEnd", "NoFurtherRec")*, I disregarded the deals which were assigned to the third category in the course of quantifying the LGD. The following graph (Figure 14) shows the proportions of the particular deal groups within the portfolio, likewise on a quarterly basis.

Figure 14: The proportions of the particular deal groups within the portfolio on a quarterly basis



(Self-made figure: own calculation results)

It can be considered as natural that the proportion of the latter category rises highly, since the later the given deal has become non-performing, the shorter time was disposable for it to get into another group. According to the historic experiences a large part of these deals gets into the *"WorkoutEnd"* category, because the client will settle his/her arrears, or for example the property which served as underlying collateral will be sold. Otherwise the deal will be assigned to the *"NoFurtherRec"* category maximum 36 months later than becoming non-performing, because the effective recovery period ends at this time, so following that no further considerable recoveries can be expected from it.

To sum it up: sooner or later (in maximum 36 months from the default event) all the deals will be entered into one of the first two categories. Considering that it can be stated according to the results of my analysis that significant difference can be observed between the LGD values of these two groups (1^{st} Hypothesis), I also investigated in my empirical analysis whether any factors can be explored on the basis of which it becomes predictable which category the certain deals will get into (5^{th} Hypothesis), since if we succeed in finding such factors, the deals assigned to the actual *"NotClosed"* category can be involved in the LGD calculation as well.

6.2.4. Calculating the pool level LGD

Considering that in accordance with the Basel regulation a long-term average has to be applied for measuring the LGD on portfolio level, I arranged the deals into so-called cohorts according to the date of non-performing event. I used monthly division, so those deals have been categorized into the same cohort which became non-performing in the same month, then I averaged the deal level LGD values on cohort level for each deal categories (*"WorkoutEnd"*, *"NoFurtherRec"*).

For the sake of carrying out the most possible accurate estimating procedure, I considered the number of the default events as weights in the course of calculating the long run average, because this method takes into consideration the fact that the recovery and cost data of more deals were used for quantifying the LGD values of the cohorts which contain larger quantity of deals, so these are statistically more grounded, thus this methodology will result in larger degree of accuracy of the model.

In the course of the empirical research I quantified the deal category level long run average weighted by number of the non-performing deals according to the under-mentioned formula:

$$LGD_{category} = \frac{\sum_{i=1}^{M} \left[LGD_i * N_i \right]}{\sum_{i=1}^{M} N_i} \qquad (6.4)$$

where: LGD_i: average LGD value of the i^{th} cohort,

M: number of cohorts,

N_i: number of non-performing deals in the i^{th} cohort.

Throughout calculating the LGD I treated the deal categories separately, so it enabled to investigate and compare the LGD values of the certain categories, however, in the final step I averaged the category level LGD values as quantification of the aggregated LGD of the total portfolio which I studied. Being attentive to the requirement of the consistent procedure, the quantity of the deals in the certain categories served as a basis of the weighting in this case as well:

$$LGD = \frac{LGD_{WorkoutEnd} * N_{WorkoutEnd} + LGD_{NoFurther\,\mathrm{Re}c} * N_{NoFurther\,\mathrm{Re}c}}{N_{WorkoutEnd} + N_{NoFurther\,\mathrm{Re}c}} \qquad (6.5)$$

6.2.5. Data used for investigating the influencing factors

In the framework of my empirical research I also probed what kind of characteristics influence significantly the run of the LGD. For the sake of establishing these analyses first I constructed a table from the available data about the underlying collaterals of the deals, which contains the following data for each deal and each default event:

Table 29: The secondary data about the collaterals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
deal_id	Deal identification number.
start_month	The period of the origination of the deal (year, month).
default_month	The period of the default event of the deal (year, month).
start_value_month	The period of defining the collateral value effective at the origination of the deal (year, month).
default_value_month	The period of defining the collateral value effective at the default of the deal (year, month).
priorcharge_rate	The quotient of the sum of the prior charges on the collateral and the realization value at the origination of the deal.
start_collvalue	The realization value of the collateral at the origination of the deal.
default_collvalue	The realization value of the collateral at the default of the deal.
start_marketvalue	The market value of the collateral at the origination of the deal.
default_marketvalue	The market value of the collateral at the default of the deal.
start_LTV	The proportion of the loan amount and the market value of the collateral at the origination.
current_LTV	The proportion of the exposure at the default and the market value of the collateral at the default.

Table 29 (continuation): The secondary data about the collaterals

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
zipcode	The zip code of the real estate which serves as collateral.
settlement	The name of the settlement of the real estate which serves as collateral.
region	The region of the real estate which serves as collateral: ▪ Budapest & environs, ▪ Central-Western, ▪ Eastern, ▪ North-Eastern, ▪ North-Western, ▪ South-Central, ▪ South-Eastern, ▪ South-Western, ▪ Western.
county	The county of the real estate which serves as collateral: ▪ Baranya, ▪ Borsod-Abaúj-Zemplén, ▪ Budapest, ▪ Bács-Kiskun, ▪ Békés, Csongrád, ▪ Fejér, ▪ Győr-Moson-Sopron, ▪ Hajdu-Bihar, ▪ Heves, ▪ Jász-Nagykun-Szolnok, ▪ Komárom-Esztergom, ▪ Nógrád, ▪ Pest, ▪ Somogy, ▪ Szabolcs-Szatmár-Bereg, ▪ Tolna, ▪ Vas, ▪ Veszprém, ▪ Zala.
settlement_type	The type of the settlement of the real estate which serves as collateral: ▪ Budapest & environs, ▪ County town & environs, ▪ Other city & environs, ▪ Village, ▪ Small village.
realestate_type	The type of the real estate which serves as collateral: ▪ Detached house, ▪ Owner-occupied block, ▪ Other residential property.
material	The building type of the real estate which serves as collateral: ▪ Brick or stone, ▪ Prefabricated, ▪ Light construction or wood, ▪ Other.

(Self-made table)

In the course of that I considered the latter from the date of the original appraisal *(appraisaldate)* and the last revaluation date *(revaluedate)* effective at the origination and the default of the deal, as the date of defining the collateral values at the origination and the default of the deal. The cases were exceptions to that, when the values at the

default differed from the values of the origination of the deal, in such cases I considered the date of the default event as the defining date of the values at the default.
In those cases when the date of the original appraisal was not disposable, then I imputed them with the date of the origination of the deal, in case of lack of the values at the default I filled them up with the values at the origination of the deal *(start_collvalue, start_marketvalue)*.
Considering that more than one real estate can lie under certain deals, and the same real estate can serve as collateral for several deals as well, I had to carry out the allocation of the collateral values and the market values of the collaterals. For the sake of that I summed up the values of the collaterals on deal level, and I linked the characteristics of the real estate with the highest value to each deal in all cases. If the same collateral referred to more than one deal, then I linked the values to the single deals allocated according to the proportion of the exposure at the default. Following that there was only one record for each deal in the data table, to which I was already able to join the region, county and type of the settlement according to the location of the real estate, so the data shown in Table 29 occurred at last.

The formerly illustrated data structures enabled me to join the secondary data referring to the collaterals to the data about the clients and the deals, and to develop the data table which grounds the analysis of the influencing factors of the LGD. In the table below (Table 30) I make known the data fields with which in this latest step I supplemented the final data table, which served as a basis for regression building.

Table 30: Macroeconomic secondary data

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
start_fx_rate	The exchange rate of the deal's currency at the origination.
start_unempl_rate	Unemployment rate at the origination of the deal.
start_min_wage	The lowest monthly minimum wage at the origination of the deal.
start_avg_netincome	Average monthly net income at the origination of the deal.
start_realwage_index	Yearly real wage index at the origination of the deal.
start_CPI	Yearly consumer price index at the origination of the deal.
start_GDP_growth	Yearly GDP-growth index at the origination of the deal.
fx_index_ds	The index of the exchange rate of the currency at the default and the origination (ratio).
collvalue_index_ds	The index of the realization value of the collateral at the default and the origination (ratio).
marketvalue _index_ds	The index of the market value of the collateral at the default and the origination (ratio).
unempl_rate_index_ds	The index of the unemployment rate at the default and the origination (ratio).
min_wage_index_ds	The index of the lowest monthly minimum wage at the default and the origination (ratio).

Table 30 (continuation): Macroeconomic secondary data

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
avg_netincome_index_ds	The index of the average monthly net income at the default and the origination (ratio).
cum_realwage_index_ds	The ratio of the real wages at the default and the origination.
cum_CPI_ds	The index of the consumer prices at the default and the origination (the quotient of the cumulative consumer price indices).
GDP_growth_index_ds	The index of the GDP at the default and the origination (the quotient of the cumulative GDP-growth indices).

(Self-made table)

6.3. Hypotheses and results

After the review of the features of the examined portfolio, the data used as well as the definitions and assumptions applied in the basic model I enter upon the presentation of the concrete calculations.

6.3.1. Product types according to the purpose of the loan

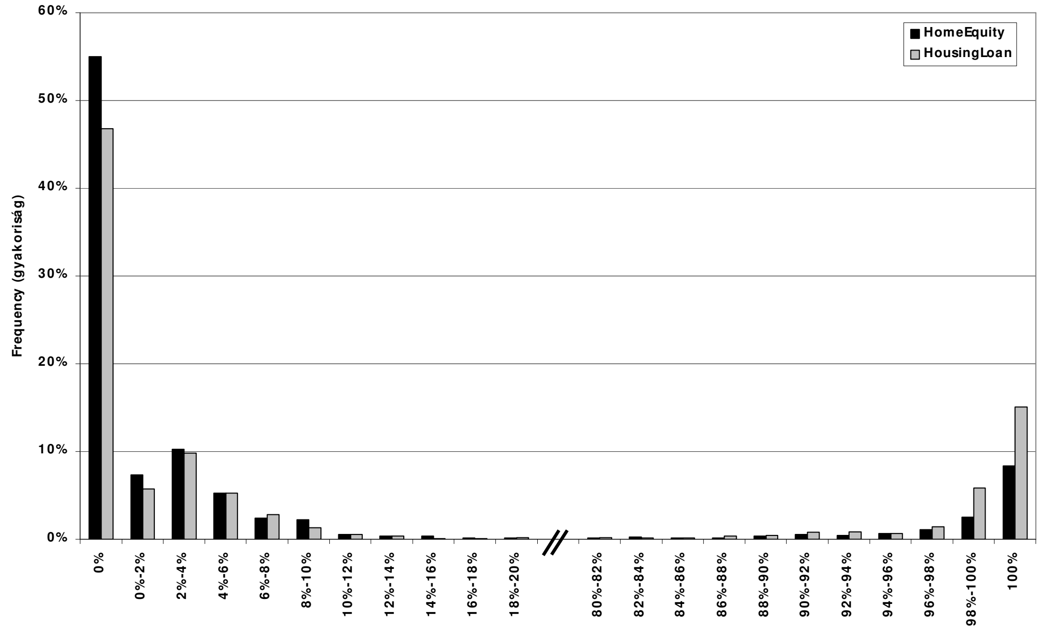
The subject of my 1st Hypothesis was the connection between the purpose of the loan and the LGD. According to my anticipative expectations in the case of the deals, where the purpose of the loan is the construction or purchase of the real estate which serves as collateral, larger recoveries can be expected in comparison with the mortgage equity withdrawals. The assumption lies behind this hypothesis that the clients presume less to take the risk of losing their home in the case, if they had decided to take up the loan exactly for the sake of its obtainment. In connection with that I think it is worth mentioning the study of Grippa, Ianotti and Leandri *(Grippa et al. [2005])* as well, in which the authors reported that in case of the loans examined by them the realized recovery of the loans with home purpose proved to be higher than the one of the loans with other purpose.

> ***1st Hypothesis: The LGD values of the loans with home purpose are lower than the LGD values of the mortgage equity withdrawals.***

For the sake of justifying this hypothesis I sorted the deals, which appeared in the database examined by me, into two groups: the one category is composed of 4278 mortgage equity withdrawals (*product=HomeEquity*), and the other one is of 2558 loans with home purpose (*product=HousingLoan*). As the first step I compared the LGD distributions of the two groups graphically illustrating with bar-chart (Figure 15).

Considering that the LGDs of 98.4% of the deals were below 20% or above 80%, I represented only these ranges on the graph for the sake of the better perspicuity.[73]

Figure 15: The LGD distribution of the loans with different purposes



(Self-made figure: own calculation results)

The bar-chart seems to contradict the statement composed by my hypothesis, since albeit both distributions are bimodal, the proportion of the loans with home purpose compared to the mortgage equity withdrawals is considerably smaller in the lower LGD bands, and is larger in the higher LGD bands. It can be seen in Table 31, that the average LGD of the mortgage equity withdrawals (16.263%) is much lower than the average LGD of the loans with home purpose (27.117%), as well as significant difference can be experienced regarding the sample variance: the average deviation of the LGD values from the average is smaller in the group of the mortgage equity withdrawals. The indices of skewness and kurtosis also reflect the strong difference of the distributions both from each other and from the normal distribution.

[73] The graph which represents the whole range can be found in the Appendix.

Table 31: The descriptive statistics of the deal groups formed on the basis of the purpose of the loans

	HOMEEQUITY (mortgage equity withdrawals)	*HOUSINGLOAN (loans with home purpose)*
Mean	0.162627035	0.27117426
Standard Error	0.005268231	0.008408898
Median	0	0.0087174
Mode	0	0
Standard Deviation	0.344576112	0.425294106
Sample Variance	0.118732697	0.180875077
Kurtosis	1.729229435	-0.828533239
Skewness	1.905276031	1.064302719
Range	1	1
Minimum	0	0
Maximum	1	1
Count	4278	2558

(Self-made table: own calculation results)

In the next step I carried out Homogeneity Analysis referring to the equivalence of the LGD distributions of the deals which are sorted into two different categories (Table 32). So in this case the null hypothesis states that the distributions are identical, while the rejection of the null hypothesis can deny only the equivalence, but it does not give any kind of concrete information concerning the type of the distributions. I created 16 LGD bands (classes) altogether, but I did not define their broadness equally, instead I considered narrow intervals on the segments near 0% and 100%, and broader intervals on the middle section as separate LGD bands, moreover I worked up distinct classes for LGD values of 0% and 100% with respect to the large quantity of the extreme values.

Table 32: Homogeneity Analysis referring to the distributions of the LGD values of the deals with different loan purpose

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	2352	1197	3549	0.5497896213	0.4679437060	0.0000018875
0%-2%	312	146	458	0.0729312763	0.0570758405	0.0000005489
2%-4%	438	251	689	0.1023842917	0.0981235340	0.0000000263
4%-6%	224	134	358	0.0523609163	0.0523846755	0.0000000000
6%-8%	101	72	173	0.0236091632	0.0281469898	0.0000001190
8%-10%	95	33	128	0.0222066386	0.0129007037	0.0000006766
10%-20%	66	29	95	0.0154277700	0.0113369820	0.0000001762
20%-50%	29	21	50	0.0067788686	0.0082095387	0.0000000409
50%-80%	41	18	59	0.0095839177	0.0070367475	0.0000001100
80%-90%	40	29	69	0.0093501636	0.0113369820	0.0000000572
90%-92%	22	20	42	0.0051425900	0.0078186083	0.0000001705
92%-94%	17	21	38	0.0039738195	0.0082095387	0.0000004721
94%-96%	29	16	45	0.0067788686	0.0062548866	0.0000000061
96%-98%	47	37	84	0.0109864423	0.0144644253	0.0000001440
98%-100%	108	149	257	0.0252454418	0.0582486317	0.0000042382
100%	357	385	742	0.0834502104	0.1505082095	0.0000060603
Total	4278	2558	6836	1	1	0.0000147339

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i} + n_{X_i}} \left(\frac{n_{Y_i}}{n_Y} - \frac{n_{X_i}}{n_X} \right)^2 = 4278 * 2558 * 0.0000147339 = 161.2346342$$

The degree of freedom is $\boldsymbol{v}$=16-1=15, so the upper critical value is $\boldsymbol{c_f}$= 25.0197923046055 at significance level of $\boldsymbol{\alpha}$=5%, therefore the null hypothesis has to be rejected. The $\boldsymbol{p}$-value is 0, so the equivalence of the distributions can not be stated at any popular significance levels.

Regarding the considerably large quantity of elements I examined the null hypothesis with asymptotic z-test that the average LGD of the mortgage equity withdrawals equals to the average LGD of the loans of home purpose (Table 33). When I considered as alternative hypothesis the statement that the average LGD of the loans with home purpose exceeds the average LGD value of the mortgage equity withdrawals (one-sided asymptotic z-test), the null hypothesis proved to be ignorable at significance level of 5%, moreover it seemed to be acceptable at none of the popular significance levels, the $\boldsymbol{p}$-value was 0. I examined the null hypothesis which states the equality of the average LGD of the two deal categories also against the alternative hypothesis that the averages differ from each other (two-sided asymptotic z-test), and I got similar result in this case as well: the average LGDs of the two deal groups could be considered as equal at none of the popular significance levels, 0 arose as $\boldsymbol{p}$-value in this case as well.

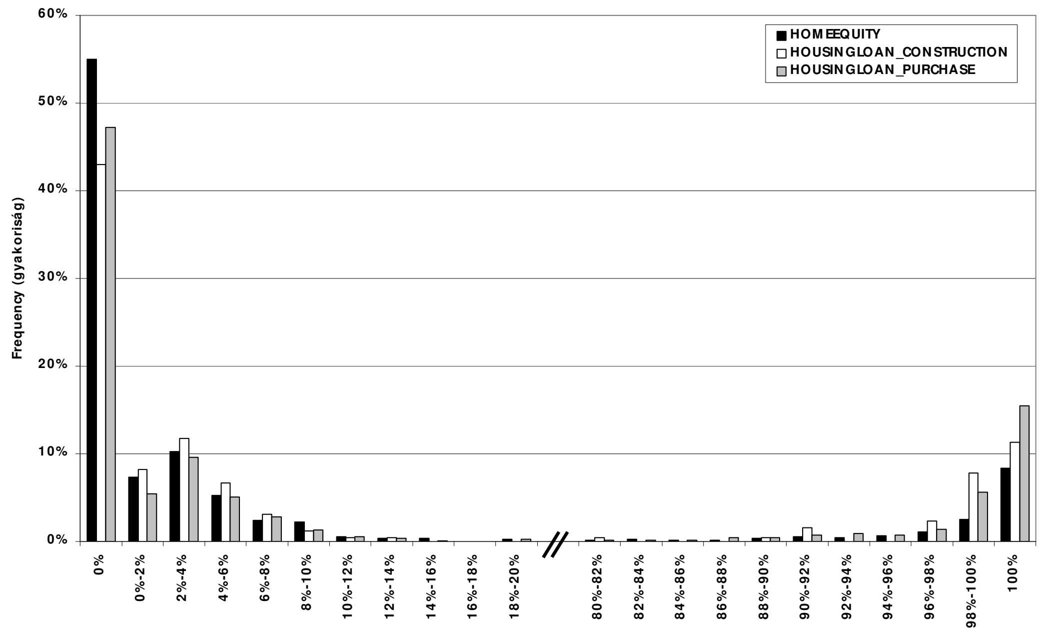
Table 33: Two-sample asymptotic *z*-test referring to the equality of the average LGD values of the loans with different loan purposes (α=0.05)

	HOMEEQUITY (mortgage equity withdrawals)	***HOUSINGLOAN (loans with home purpose)***
Mean	0.162627035	0.27117426
Known Variance	0.118732697	0.180875077
Observations	4278	2558
Hypothesized Mean Difference	0	
Z	10.93906937	
P(Z<=z) one-tail	0.000000000	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.000000000	
z Critical two-tail	1.959963985	

(Self-made table: own calculation results)

I went on dividing the group of the loans with home purpose according to whether the purpose of the loan was home construction or home purchase, then I carried out the same analyses which I also did in the course of comparing the LGD values of the loans with home purpose and the mortgage equity withdrawals.

Figure 16: The LGD distribution of the loans with different purposes (detailed)



(Self-made figure: own calculation results)

Both the bar-chart which graphically illustrates the LGD distributions (Figure 16) and the descriptive statistics (Table 34) show that the differences between the two subgroups of the loans with home purpose (construction and purchase) are more significant than the deviations which are experienced between the loans with home purpose and the mortgage equity withdrawals. This is equally true regarding the mean values, the indices of dispersion, kurtosis and skewness.

Table 34: The descriptive statistics of the deal groups formed on the basis of the purpose of the loans (detailed)

	HOMEEQUITY (mortgage equity withdrawals)	*HOUSINGLOAN (loans with home purpose)*		
		CONSTRUCTION	*PURCHASE*	*∑ HOUSINGLOAN*
Mean	0.162627035	*0.253701927*	*0.273117317*	0.27117426
Standard Error	0.005268231	*0.025928438*	*0.008888925*	0.008408898
Median	0	*0.01536913*	*0.00765416*	0.0087174
Mode	0	*0*	*0*	0
Standard Deviation	0.344576112	*0.414855002*	*0.426483166*	0.425294106
Sample Variance	0.118732697	*0.172104673*	*0.181887891*	0.180875077
Kurtosis	1.729229435	*-0.551732077*	*-0.855203487*	-0.828533239
Skewness	1.905276031	*1.184816173*	*1.052163548*	1.064302719
Range	1	*1*	*1*	1
Minimum	0	*0*	*0*	0
Maximum	1	*1*	*1*	1
Count	4278	*256*	*2302*	2558

(Self-made table: own calculation results)

Herewith I composed the equivalence of the LGD distributions of the mortgage loans with the purpose of home construction and home purchase as the null hypothesis of the Homogeneity Analysis (Table 35). I created 16 LGD bands (classes) also in this case, in the same way as previously presented. Considering that only 256 loans with home construction purpose appeared in the sample which I examined, rather few deals got into certain classes in this way as well, but I did not consider the further merging as necessarily reasonable.

Table 35: Homogeneity Analysis referring to the distributions of the LGD values of the deals with different loan purpose (detailed)

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	110	1087	1197	0.4296875000	0.4721980886	0.0000015097
0%-2%	21	125	146	0.0820312500	0.0543006082	0.0000052670
2%-4%	30	221	251	0.1171875000	0.0960034752	0.0000017879
4%-6%	17	117	134	0.0664062500	0.0508253692	0.0000018117
6%-8%	8	64	72	0.0312500000	0.0278019114	0.0000001651
8%-10%	3	30	33	0.0117187500	0.0130321460	0.0000000523
10%-20%	2	27	29	0.0078125000	0.0117289314	0.0000005289
20%-50%	3	18	21	0.0117187500	0.0078192876	0.0000007241
50%-80%	1	17	18	0.0039062500	0.0073848827	0.0000006723
80%-90%	2	27	29	0.0078125000	0.0117289314	0.0000005289
90%-92%	4	16	20	0.0156250000	0.0069504778	0.0000037624
92%-94%	0	21	21	0.0000000000	0.0091225022	0.0000039629
94%-96%	0	16	16	0.0000000000	0.0069504778	0.0000030193
96%-98%	6	31	37	0.0234375000	0.0134665508	0.0000026870
98%-100%	20	129	149	0.0781250000	0.0560382276	0.0000032740
100%	29	356	385	0.1132812500	0.1546481321	0.0000044447
Total	256	2302	2558	1	1	0.0000341982

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}}\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 256*2302*0.0000341982 = 20.15342246$$

The degree of freedom is $\boldsymbol{v}$=16-1=15 again, the upper critical value is $\boldsymbol{c_f}$= 25.0197923046055 at significance level of $\boldsymbol{\alpha}$=5%, therefore the null hypothesis proved to be acceptable. The $\boldsymbol{p}$-value is 0.16612589, so 16.613% is the lowest significance level where the statement of the equivalence of the distributions could be disproved, which is higher than the generally used significance levels. Accepting the null hypothesis means that the LGD distributions of the groups of the loans with home construction and home purchase purpose can be considered as identical, but exact information does not derive from the Homogeneity Analysis referring to the type and the characteristics of the distribution.

I investigated also the null hypothesis with asymptotic *z*-test whether the average LGD of the mortgage loans with the purpose of home construction and home purchase can be considered as equal (Table 36). At significance level of 5% I was able to accept the null hypothesis against the alternative hypothesis which states the dissimilarity of the averages (two-sided asymptotic *z*-test), the ***p***-value is 0.4787356, so the statement referring to the equality of the averages can be considered as true at all the significance levels which are lower than 47.87%.

Table 36: Two-sample asymptotic *z*-test referring to the equality of the average LGD values of the loans with different loan purposes (*α*=0.05) (detailed)

	CONSTRUCTION	***PURCHASE***
Mean	0.253701927	0.273117317
Known Variance	0.172104673	0.181887891
Observations	256	2302
Hypothesized Mean Difference	0	
Z	-0.708337653	
P(Z<=z) one-tail	0.2393678	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.4787356	
z Critical two-tail	1.959963985	

(Self-made table: own calculation results)

However, it is also necessary to note that because of the notable dissimilarities from the normal distribution and the large differences between the quantities of the deals (the relatively small number of the loans with the purpose of home construction) the results of the Homogeneity Analysis and the asymptotic *z*-test directed towards verifying the equality of the averages have to be accepted under reserves.

To summarize the lessons from the analyses: my 1st Hypothesis did not prove to be true, the LGD values of the loans with home purpose seemed lower than the LGD values of the mortgage equity withdrawals at none of the popular significance levels, the results of the tests show just the opposite of that. The analyses also clarified that the LGD distributions of the two groups defined within the loans with home purpose (home building and home purchase) differ much less from each other than the LGD distributions of the mortgage loans with home purpose and the mortgage equity withdrawals, thus the separate treating has relevance only in the case of the two latter groups in the course of the categorization, the application of more detailed parcelling does not have any notable added value.

6.3.2. Deal groups defined according to the type of the application

In the framework of my 2nd Hypothesis I investigated whether the LGD values of the 4171 purely collateral-based loans without income verification (*application_type=AssetBased*) and the 2665 loans based on income verification (*application_type=IncomeBased*) in the sample differ from each other significantly. According to my presumption only lower recoveries can be expected from the deals which belong to the former group, following the occasional default event, because the income of the clients who have resort to this kind of loan is supposedly lower and less steady in comparison with the ones who are prepared to give free run of their income certificate to the bank at the application.

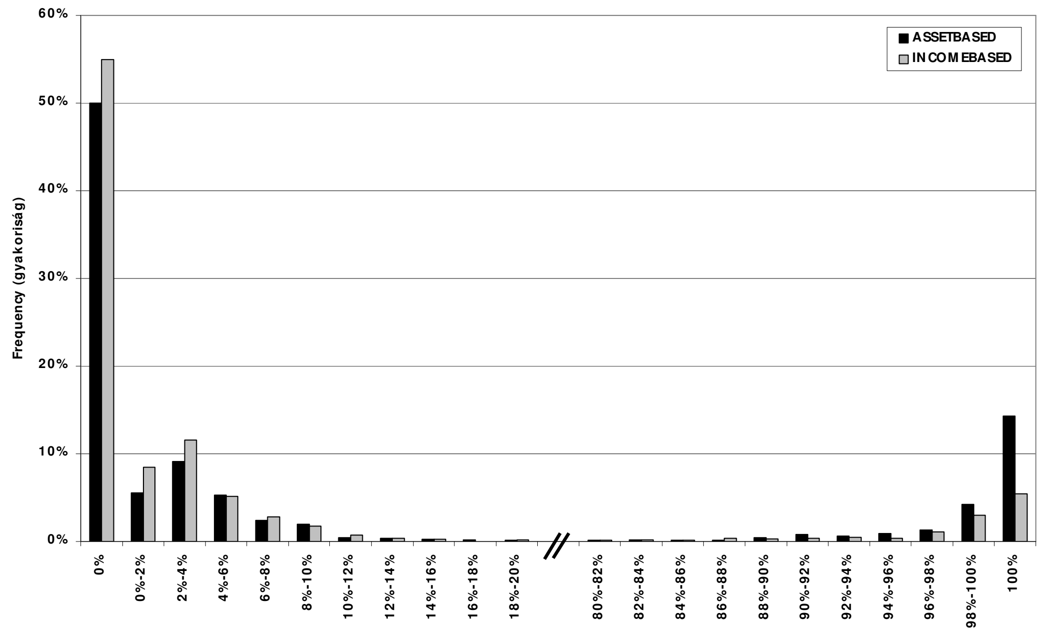
If my expectation proves to be true, it can even account for the experienced differences between the LGD values of the deal groups which are defined on the basis of the loan purposes. I will return to this question after having tested the 2nd Hypothesis, analysing what proportions are represented by the purely collateral-based loans without income verification and the loans based on income verification within the deal categories defined according to the loan purpose.

2nd Hypothesis: The purely collateral-based loans without income verification are characterized by higher LGDs than the loans based on income verification.

During testing the hypothesis I applied the simplifications which I made known previously, so I represented the LGD ranges only below 20% and above 80% in the course of the graphical illustration, moreover despite the notable deviations from the normal distribution (bimodality) I applied the asymptotic z-test and the Homogeneity Analysis, taking the relatively large quantity of elements into consideration in each deal category.

The following bar-chart illustrates the LGD distributions of the purely collateral-based loans without income verification and the loans based on income verification on the basis of the examined sample (Figure 17). The outlined picture seems to confirm my hypothesis, since the proportion of the LGDs of 100% or near 100% proved to be considerably higher in case of the purely collateral-based loans without income verification in comparison with the loans based on income verification.

Figure 17: The LGD distribution of the loans with different types of application



(Self-made figure: own calculation results)

It is indicated by the descriptive statistics (Table 37) as well, since the average LGD of the purely collateral-based loans without income verification was 24.502%, while the one of the loans based on income verification was 13.786%. Again this latter category proved to be better from the viewpoint of homogeneity as well, I calculated 10.022% as its sample variance, while 16.871% in case of the loans without income verification.

Table 37: The descriptive statistics of the subportfolios with different types of application

	ASSETBASED ***(purely collateral-based)***	***INCOMEBASED*** ***(based on income verification)***
Mean	0.245021546	0.137860167
Standard Error	0.006359844	0.006132507
Median	0.00008417	0
Mode	0	0
Standard Deviation	0.410739577	0.316582337
Sample Variance	0.168707	0.100224376
Kurtosis	-0.451306855	2.97670009
Skewness	1.227357148	2.195753619
Range	1	1
Minimum	0	0
Maximum	1	1
Count	4171	2665

(Self-made table: own calculation results)

Relating to the equivalence of the LGD distributions of the two deal categories I carried out Homogeneity Analysis (Table 38) again using the 16 LGD bands which I made known previously. Herewith I composed the statement as null hypothesis that the LGD values of the purely collateral-based loans without income verification and the deals based on income verification follow the same distribution. Approving that, it would mean withal the rejection of my 2nd Hypothesis.

Table 38: Homogeneity Analysis referring to the LGD distributions of the deals with different types of application

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	2085	1464	3549	0.4998801247	0.5493433396	0.0000006894
0%-2%	232	226	458	0.0556221530	0.0848030019	0.0000018592
2%-4%	381	308	689	0.0913450012	0.1155722326	0.0000008519
4%-6%	222	136	358	0.0532246464	0.0510318949	0.0000000134
6%-8%	99	74	173	0.0237353153	0.0277673546	0.0000000940
8%-10%	82	46	128	0.0196595541	0.0172607880	0.0000000450
10%-20%	55	40	95	0.0131862863	0.0150093809	0.0000000350
20%-50%	24	26	50	0.0057540158	0.0097560976	0.0000003203
50%-80%	26	33	59	0.0062335171	0.0123827392	0.0000006409
80%-90%	41	28	69	0.0098297770	0.0105065666	0.0000000066
90%-92%	33	9	42	0.0079117718	0.0033771107	0.0000004896
92%-94%	26	12	38	0.0062335171	0.0045028143	0.0000000788
94%-96%	36	9	45	0.0086310237	0.0033771107	0.0000006134
96%-98%	55	29	84	0.0131862863	0.0108818011	0.0000000632
98%-100%	177	80	257	0.0424358667	0.0300187617	0.0000005999
100%	597	145	742	0.1431311436	0.0544090056	0.0000106086
Total	4171	2665	6836	1	1	0.0000170094

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}} \left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 4171*2665*0.0000170094 = 189.07117095$$

The degree of freedom is ν=16-1=15, so the upper critical value is c_f=25.0197923 at significance level of α=5%, so the null hypothesis has to be rejected, therefore it can be stated at significance level of 5% that the LGD distributions of the two deal categories differ from each other. The p-value is 0, so the null hypothesis, namely the equivalence of the distributions, can be accepted at none of the popular significance levels.

Considering that the Homogeneity Analysis exposes nothing about the type and other characteristics of the distributions, I carried out a two-sample asymptotic z-test to investigate the null hypothesis that the average LGD values of the two examined deal categories are the same (Table 39). The test carried out against both the one-sided and the two-sided alternative hypothesis showed that the null hypothesis can be disproved at all the popular significance levels (the p-value is 0), so it can be stated with large confidence that the average LGD values differ from each other.

Table 39: Two-sample asymptotic *z*-test referring to the equality of the average LGD values of the loans with different types of application (α=0.05)

	ASSETBASED (purely collateral-based)	*INCOMEBASED (based on income verification)*
Mean	0.245021546	0.137860167
Known Variance	0.168707	0.100224376
Observations	4171	2665
Hypothesized Mean Difference	0	
Z	12.12934108	
P(Z<=z) one-tail	0.000000000	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.000000000	
z Critical two-tail	1.959963985	

(Self-made table: own calculation results)

All the tests, which were carried out, indicate that the LGD values of the purely collateral-based loans without income verification and of the deals based on income verification differ from each other significantly, the graphical illustration of the distributions and the descriptive statistics clearly show that the LGD values of the latter category are lower in the examined portfolio. These results uniformly seem to justify my 2nd Hypothesis.

I mentioned earlier in this subsection that this can withal serve as an explanation even for the deviations between the LGD values of the deal groups which are defined on the basis of the loan purposes. Considering that the LGD values of the deals based on income verification proved to be significantly lower than the LGD values of the purely collateral-based loans without income verification, if the deals pertaining to the latter category dominate among the loans with home purpose, then this can partly explain why the statement which is composed in the 1st Hypothesis did not prove to be watertight. For the purpose of its checking I recorded the average LGD values of the deal groups, which are defined on the basis of the loan purpose and the type of the application, in Table 40.

Table 40: The average LGD values of the deal groups which are defined on the basis of the loan purpose and the type of the application

Product \ Application type	*ASSETBASED (purely collateral-based)*	*INCOMEBASED (based on income verification)*
HOMEEQUITY (mortgage equity withdrawals)	0.1878344	0.1138721
HOUSINGLOAN (loans with home purpose)	0.3643907	0.1668367

(Self-made table: own calculation results)

It is conspicuous that the average LGD values of the loans with home purpose (*product=HousingLoan*) are higher in case of both deal categories which are defined on the basis of the type of the application, in comparison with the ones of the mortgage equity withdrawals (*product=HomeEquity*), so it does not give any explanation why the statement composed in the 1st Hypothesis did not pass the test. Moreover the fact that in the examined portfolio the purely collateral-based loans without income verification represent larger proportion within the group of the mortgage equity withdrawals (65.919%), than within the category of the loans with home purpose (52.815%), would reason intuitively exactly the fact that the mortgage equity withdrawals should be featured by higher LGD values. Therefore it does not serve either proper explanation for proving the 1st Hypothesis false.

6.3.3. The function of the applied discount rate

Actually neither the CRD nor the national regulation contains particular prescriptions regarding what kind of method the discount rate should be defined with, moreover a uniform standpoint did not emerge among the researchers either, thus there are numerous variants in the literature, from which I made known the most observable ones in Chapter 3.2.1.

I consider as important to mention again from the "early" works the study of Altman, Haldeman and Narayanan *(Altman et al. [1977])*, in which the authors did not use discounting. *Keisman and Marshella [2009]* chose the same procedure, but as alternative solution they carried out their calculations with discounting by the interest rate which existed before the default event. *Eales and Bosworth [1998]* applied the alternative cost of capital calculated according to CAPM and the borrowing rate, moreover *Asarnow and Edwards [1995]* decided to use this latter variant as well. *Bastos [2009]* also discounted the recoveries following the default event with the loan-specific contractual lending rates. *Moral and Garcia-Baena [2002]*, *Moral and Oroz [2002]*, as well as *Grunert and Weber [2005; 2009]* calculated with a flat discount rate of 5%, while *Qi and Yang [2007; 2009]* used the 1-year LIBOR as discount rate.

From the review of the more sophisticated procedures I also emphasize the work of *Carty and Liebermann [1996]* who quantified the discount rate in a way that they added a spread, which was calculated on the basis of the deal's risk, to the LIBOR; as well as the study of *Chalupka and Kopecsni [2009]* in which the authors defined the discount rate as the sum of the risk-free rate and the risk premium which is different for each asset class, and they used a flat risk premium of 0%, 1%, …, 8% and 9% as alternative solution.

Hereinafter in the framework of investigating my 3rd Hypothesis I present the notability of the deviations deriving from the use of the different discount rates.

3rd Hypothesis: The type of the applied discount rate influences the calculated LGD value considerably.

In case of the basic model I used the contractual lending rate *("interest")* of each deal as discount rate, because it reflects both the differences between the actual interest levels at the date of the origination of certain deals, and on the other hand it varies according to their currency as well. In the course of investigating the present hypothesis I compared the LGD values, which were calculated with the alternative discount rates, to the LGD values of the basic model in all cases, namely I carried out research into what extent the given alternatives divert the LGD values from the ones of the basic model.

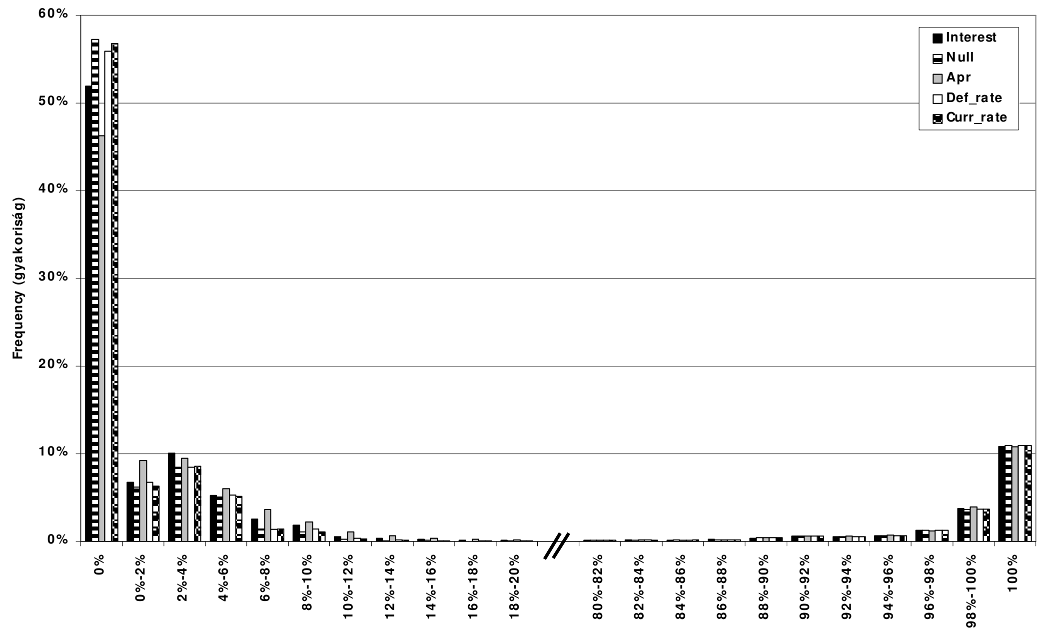
I examined the use of the following alternative discount rates:

- *"null"*: discount rate of 0% (disregarding the time value and the risks of the recoveries),
- *"apr"*: the contractual Annual Percentage Rate of the given deals,
- *"def_rate"*: the central bank base rate of the currency of the deal (CHF, EUR, HUF) effective at the default,
- *"curr_rate"*: the central bank base rate of the currency of the deal (CHF, EUR, HUF) effective on 30th June 2011.

During testing the hypothesis I applied the simplifications which I made known previously, so I represented the LGD ranges only below 20% and above 80% in the course of the graphical illustration, moreover despite the notable deviations from the normal distribution (bimodality) I applied the asymptotic z-test and the Homogeneity Analysis, taking the relatively large quantity of elements into consideration in each deal category.

The following bar-chart illustrates the distributions of the LGD values, which are calculated with using the five different discount rates, in the ranges below 20% and above 80% on the basis of the investigated sample (Figure 18). It is visible that, though in the high LGD range large differences did not appear between the proportions of the LGD values, which are calculated with the given alternative discount rates, considerable deviations can be experienced by 0% and in the LGD bands which are near that.

Figure 18: The distribution of the LGD values calculated with using the different discount rates



(Self-made figure: own calculation results)

Both the graphical illustration (Figure 18) and the descriptive statistics (Table 41) showed that the use of the 0% discount rate and the contractual Annual Percentage Rate diverted the LGD values the most considerably from the ones of the basic model. While using the 0% discount rate resulted in a 0.5 percentage point lower average LGD, applying the contractual Annual Percentage Rate modified the average almost to the same extent in the opposite direction.

Table 41: The descriptive statistics of the LGD values quantified with using the different discount rates

	INTEREST (contractual lending rate)	*NULL*	*APR (contractual Annual Percentage Rate)*	*DEF_RATE (central bank base rate at the default)*	*CURR_RATE (central bank base rate on 30th June 2011)*
Mean	0.203244911	0.198844196	0.207594251	0.200291612	0.199231004
Standard Error	0.004601158	0.004609843	0.004592045	0.004609268	0.004608961
Median	0	0	0.00340063	0	0
Mode	0	0	0	0	0
Standard Deviation	0.380424211	0.38114232	0.379670788	0.381094755	0.381069366
Sample Variance	0.14472258	0.145269468	0.144149907	0.145233212	0.145213862
Kurtosis	0.403478081	0.442640378	0.366144612	0.425334024	0.439336471
Skewness	1.529258321	1.545757602	1.512888225	1.538929405	1.544303567
Range	1	1	1	1	1
Minimum	0	0	0	0	0
Maximum	1	1	1	1	1
Count	6836	6836	6836	6836	6836

(Self-made table: own calculation results)

Using again the 16 LGD bands created on the basis of the logic laid previously, I carried out Homogeneity Analyses pair-wise referring to the equivalence of the LGD distributions (Table 42-Table 45). As null hypothesis I always composed the statement that the distribution of the LGD values calculated with some alternative discount rates equals to the one of the LGD values derived from using the contractual lending rate, namely I carried out the comparison with the basic model in all cases. The alternative hypothesis meant the simple negation of the equivalence of the distributions, so if the null hypothesis has to be rejected, this confirms the considerable LGD influencing role of the used discount rate.

First I did the Homogeneity Analysis referring to the discount rates (the 0% discount rate and the contractual Annual Percentage Rate) which diverted the LGD values from the ones in the basic model in the most notable measure according to the graphical illustration (Figure 18) and the descriptive statistics (Table 41).

Table 42: Homogeneity Analysis referring to the distribution of the LGD values which are calculated with using the contractual lending rate ("interest") and the 0% discount rate ("null")

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	3549	3916	7465	0.519163253	0.57284962	0.0000003861
0%-2%	458	424	882	0.066998245	0.062024576	0.0000000280
2%-4%	689	577	1266	0.100789936	0.084406085	0.0000002120
4%-6%	358	345	703	0.052369807	0.05046811	0.0000000051
6%-8%	173	96	269	0.025307197	0.0140433	0.0000004717
8%-10%	128	72	200	0.0187244	0.010532475	0.0000003355
10%-20%	95	35	130	0.0138970158	0.0051199532	0.0000005926
20%-50%	50	45	95	0.0073142188	0.0065827970	0.0000000056
50%-80%	59	51	110	0.0086307782	0.0074605032	0.0000000125
80%-90%	69	71	140	0.0100936220	0.0103861908	0.0000000006
90%-92%	42	42	84	0.006143944	0.006143944	0.0000000000
92%-94%	38	35	73	0.005558806	0.005119953	0.0000000026
94%-96%	45	45	90	0.006582797	0.006582797	0.0000000000
96%-98%	84	85	169	0.012287888	0.012434172	0.0000000001
98%-100%	257	249	506	0.037595085	0.03642481	0.0000000027
100%	742	748	1490	0.108543008	0.109420714	0.0000000005
Total	6836	6836	13672	1	1	0.0000020558

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}}\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 6838*6838*0.0000020558 = 96.06875698$$

Table 43: Homogeneity Analysis referring to the distribution of the LGD values which are calculated with using the contractual lending rate ("interest") and the contractual Annual Percentage Rate ("apr")

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	3549	3160	6709	0.519163253	0.462258631	0.0000004827
0%-2%	458	632	1090	0.066998245	0.092451726	0.0000005944
2%-4%	689	648	1337	0.100789936	0.094792276	0.0000000269
4%-6%	358	410	768	0.052369807	0.059976594	0.0000000753
6%-8%	173	250	423	0.025307197	0.036571094	0.0000002999
8%-10%	128	150	278	0.0187244	0.021942657	0.0000000373
10%-20%	95	169	264	0.0138970158	0.0247220597	0.0000004439
20%-50%	50	73	123	0.0073142188	0.0106787595	0.0000000920
50%-80%	59	65	124	0.0086307782	0.0095084845	0.0000000062
80%-90%	69	68	137	0.0100936220	0.0099473376	0.0000000002
90%-92%	42	40	82	0.006143944	0.005851375	0.0000000010
92%-94%	38	39	77	0.005558806	0.005705091	0.0000000003
94%-96%	45	47	92	0.006582797	0.006875366	0.0000000009
96%-98%	84	83	167	0.012287888	0.012141603	0.0000000001
98%-100%	257	267	524	0.037595085	0.039057929	0.0000000041
100%	742	735	1477	0.108543008	0.107519017	0.0000000007
Total	6836	6836	13672	1	1	0.0000020659

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}}\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 6838 * 6838 * 0.0000020659 = 96.54286238$$

The degree of freedom is $\boldsymbol{v}$=16-1=15, so the upper critical value is $\boldsymbol{c_f}$=25.0197923 at significance level of $\boldsymbol{\alpha}$=5%, therefore the null hypothesis has to be rejected at significance level of 5%, namely the LGD distributions can not be considered as identical. The ***p***-value is 0 in both cases, namely the results were not astonishing, since they showed that the presumption of the equivalence of the distributions could be confirmed at none of the popular significance levels.

In the following step I also carried out the Homogeneity Analysis referring to the two other alternative discount rates, in case of which the LGD influencing role appeared to be smaller beforehand.

Table 44: Homogeneity Analysis referring to the distribution of the LGD values which are calculated with using the contractual lending rate ("interest") and the central bank base rate of the currency of the deals effective at the default ("def_rate")

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	3549	3819	7368	0.519163253	0.558660035	0.0000002117
0%-2%	458	458	916	0.066998245	0.066998245	0.0000000000
2%-4%	689	579	1268	0.100789936	0.084698654	0.0000002042
4%-6%	358	361	719	0.052369807	0.05280866	0.0000000003
6%-8%	173	92	265	0.025307197	0.013458163	0.0000005298
8%-10%	128	97	225	0.0187244	0.014189585	0.0000000914
10%-20%	95	48	143	0.0138970158	0.0070216501	0.0000003306
20%-50%	50	53	103	0.0073142188	0.0077530720	0.0000000019
50%-80%	59	53	112	0.0086307782	0.0077530720	0.0000000069
80%-90%	69	69	138	0.0100936220	0.0100936220	0.0000000000
90%-92%	42	41	83	0.006143944	0.005997659	0.0000000003
92%-94%	38	38	76	0.005558806	0.005558806	0.0000000000
94%-96%	45	46	91	0.006582797	0.006729081	0.0000000002
96%-98%	84	84	168	0.012287888	0.012287888	0.0000000000
98%-100%	257	251	508	0.037595085	0.036717379	0.0000000015
100%	742	747	1489	0.108543008	0.109274429	0.0000000004
Total	6836	6836	13672	1	1	0.0000013791

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}}\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 6838 * 6838 * 0.0000013791 = 64.44589542$$

Table 45: Homogeneity Analysis referring to the distribution of the LGD values which are calculated with using the contractual lending rate ("interest") and the central bank base rate of the currency of the deals effective on 30th June 2011 ("curr_rate")

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	3549	3881	7430	0.519163253	0.567729666	0.0000003175
0%-2%	458	432	890	0.066998245	0.063194851	0.0000000163
2%-4%	689	587	1276	0.100789936	0.085868929	0.0000001745
4%-6%	358	350	708	0.052369807	0.051199532	0.0000000019
6%-8%	173	97	270	0.025307197	0.014189585	0.0000004578
8%-10%	128	75	203	0.0187244	0.010971328	0.0000002961
10%-20%	95	40	135	0.0138970158	0.0058513751	0.0000004795
20%-50%	50	47	97	0.0073142188	0.0068753657	0.0000000020
50%-80%	59	52	111	0.0086307782	0.0076067876	0.0000000094
80%-90%	69	71	140	0.0100936220	0.0103861908	0.0000000006
90%-92%	42	42	84	0.006143944	0.006143944	0.0000000000
92%-94%	38	35	73	0.005558806	0.005119953	0.0000000026
94%-96%	45	45	90	0.006582797	0.006582797	0.0000000000
96%-98%	84	84	168	0.012287888	0.012287888	0.0000000000
98%-100%	257	251	508	0.037595085	0.036717379	0.0000000015
100%	742	747	1489	0.108543008	0.109274429	0.0000000004
Total	6836	6836	13672	1	1	0.0000017601

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}}\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 6838 * 6838 * 0.0000017601 = 82.24972273$$

In this case the results also spoke of the considerable differences of the distributions at all the popular significance levels, namely they confirmed the statement which I composed in my 3rd Hypothesis.

Although the distributions differ notably from the normal distribution, I thought that because of the large quantity of the elements the equality of the LGD values calculated with the different discount rates can be examined with paired two-sample ***t***-tests. In all cases the average LGD quantified with applying the contractual lending rate served as a basis for the comparison, I compared the average LGD values calculated with the alternative discount rates to it.

Table 46: Paired two-sample *t*-test referring to the equality of the average LGD values calculated with using the different discount rates (*α*=0.05)

	INTEREST (contractual lending rate)	***NULL***	***APR (contractual Annual Percentage Rate)***	***DEF_RATE (central bank base rate at the default)***	***CURR_RATE (central bank base rate on 30th June 2011)***
Mean	0.203244911	0.198844196	0.207594251	0.2002916	0.199231004
Variance	0.14472258	0.145269468	0.144149907	0.1452332	0.145213862
Observations	6836	6836	6836	6836	6836
Pearson Correlation		0.998835294	0.999469259	0.99956728	0.999281453
Hypothesized Mean Difference		0	0	0	0
Df		6835	6835	6835	6835
t Stat		19.78300145	-28.98863658	21.76018117	22.96975192
P(T<=t) two-tail		0.000000000	0.000000000	0.000000000	0.000000000
t Critical two-tail		1.960311067	1.960311067	1.960311067	1.960311067

(Self-made table: own calculation results)

Concerning all the alternative discount rates the results indicate that the null hypothesis, which states the equality of the averages, can be accepted at none of the popular significance levels, the ***p***-value is 0 in all cases, namely not only the distributions but also the average LGD values can not be considered as equal.

On the basis of the presented results my 3rd Hypothesis can be voted as justified, since none of the tests which were carried out confirmed the equality of the LGD values.

6.3.4. The importance of choosing the materiality threshold

I made known the prescriptions relating to defining the materiality threshold in connection with the circumscription of the default event in Chapter 2.3.1, also touching upon that the credit institutions are also allowed to use criteria which are different from the prescriptions *(Hkr. 68. § (5)-(7) Paragraph)*, if they are able to justify its necessity, reasonability.

Regarding the definition of the default event the difference between the empirical studies appears mostly from that point, in case of how long delay they consider the deals as non-performing: the study of *Dermine and Neto de Carvalho [2003; 2005]* can be mentioned as an example in which the authors considered already the 1-day delay as default event as well, while for example Thomas, Mues, Matuszyk and Moore *(Thomas et al. [2007a; 2007b])* handled only the deals with more than 180-day delay as non-performing.

The materiality threshold was not generally emphasized in the literature, however, I think that it is an important question, thus my 4th Hypothesis is in connection with it, according to which the lowering of the materiality threshold, applied in the basic model, does not affect the result of the LGD calculation considerably.

According to my anticipative expectations the low-amount arrears are quite rare in case of the mortgage loans, since in many cases even one monthly repayment instalment exceeds the lowest monthly minimum wage, so it has a relatively small probability that the clients delay with an amount which is smaller than it.[74]

> ***4th Hypothesis: The lowering of the materiality threshold used in the basic model does not affect the result of the LGD calculation considerably in case of the retail mortgage loans.***

For the sake of testing my hypothesis I investigated the effect of using four alternative thresholds, in addition to the materiality threshold in the basic model, on the results of the LGD calculation.

In the basic model the highest delayed amount which is not defined as delinquent (the overdue amount is considered as immaterial) is the minimum of the under-mentioned values:

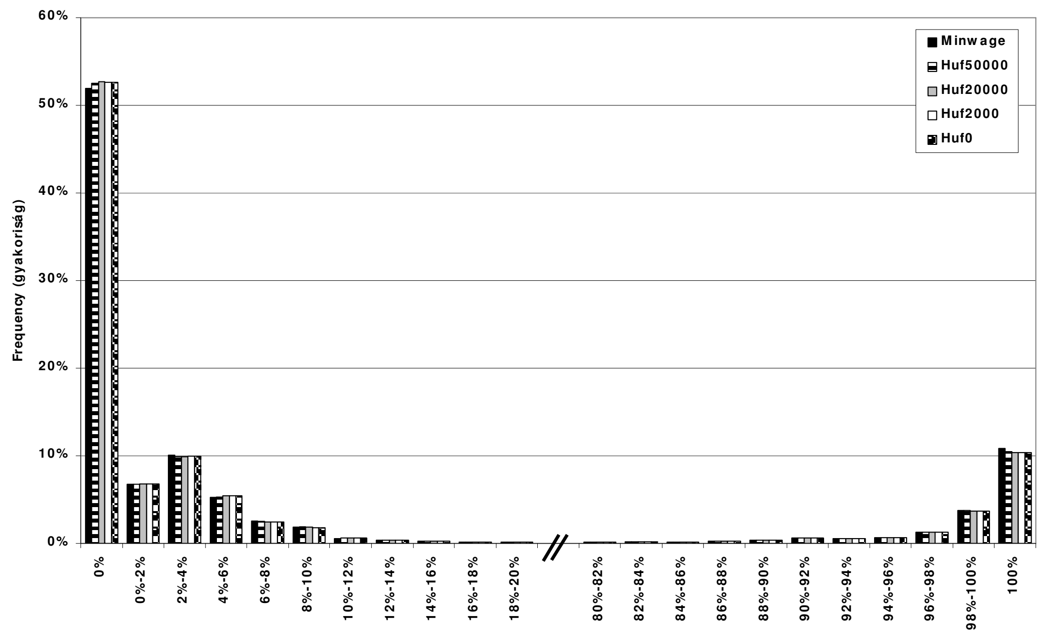
- the lowest monthly minimum wage effective at the time of becoming delayed,
- 2% of the obligations of the client, and
- one monthly repayment instalment.

I applied 50000 HUF, 20000 HUF, 2000 HUF and 0 HUF as alternative threshold, and I investigated whether a notable role can be put down to the use of the lower materiality thresholds from the viewpoint of the result of the LGD calculation. At the first glance

[74] In case of other loan types choosing the materiality threshold supposedly takes a larger LGD influencing role.

no significant difference can be seen among the distributions which are illustrated with the bar-chart hereunder (Figure 19).

Figure 19: The distribution of the LGD values in case of using the different materiality thresholds



(Self-made figure: own calculation results)

The descriptive statistics (Table 47) indicates that the lower materiality thresholds resulted in slightly lower average LGD, and in case of them the deviation of the LGD values also proved to be somewhat lower, but the differences do not seem considerable.

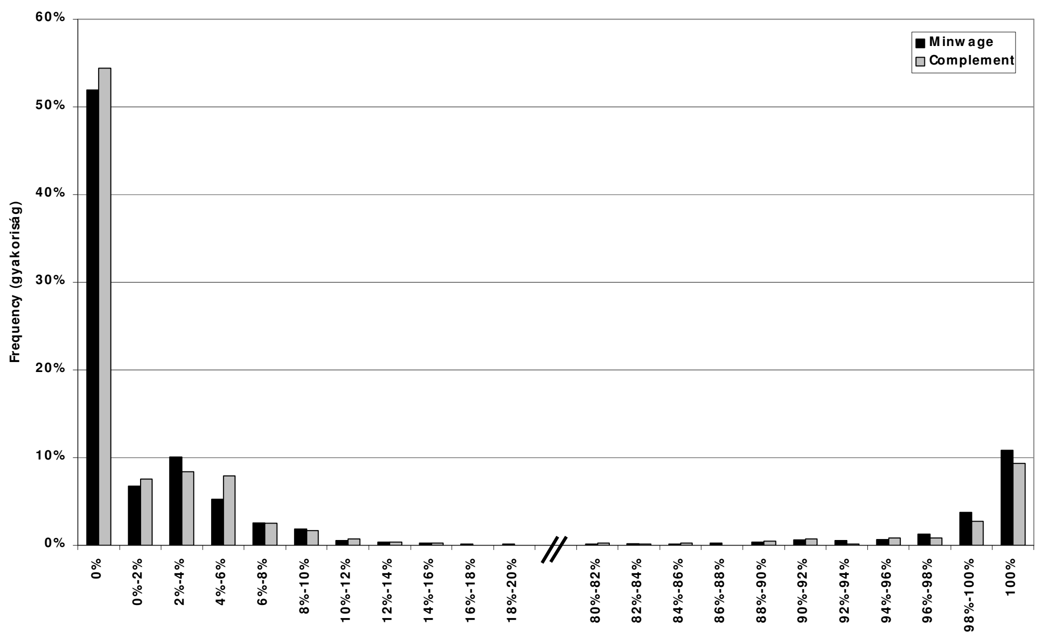
Table 47: The descriptive statistics of the LGD values quantified with using the different materiality thresholds

	MINWAGE (basic model)	***HUF50000***	***HUF20000***	***HUF2000***	***HUF0***
Mean	0.203244911	0.198818676	0.196525283	0.196592032	0.196592032
Standard Error	0.004601158	0.004478672	0.004433066	0.004431299	0.004431299
Median	0	0	0	0	0
Mode	0	0	0	0	0
Standard Deviation	0.380424211	0.377113715	0.375478313	0.375485556	0.375485556
Sample Variance	0.14472258	0.142214754	0.140983963	0.140989403	0.140989403
Kurtosis	0.403478081	0.515390649	0.573990891	0.572648524	0.572648524
Skewness	1.529258321	1.565071542	1.583713466	1.583178996	1.583178996
Range	1	1	1	1	1
Minimum	0	0	0	0	0
Maximum	1	1	1	1	1
Count	6836	7090	7174	7180	7180

(Self-made table: own calculation results)

These results disprove the statement composed in the 4th Hypothesis not by a long, but they require further researches, thus in the following step I separated the "technical defaults", namely those which are not considered as non-performing according to the definition of the basic model, but they did according to the materiality threshold of 0 HUF. In the course of the further examinations I compared the LGD values of this subportfolio (*"complement"*) with the LGD values in the basic model (*"minwage"*).

Figure 20: The distribution of the LGD values in the basic model and the LGD values of the "technical defaults"



(Self-made figure: own calculation results)

This comparison indicates far more considerable differences than the former ones, considering both the graphical illustration (Figure 20) and the descriptive statistics (Table 48). The distribution proved to be bimodal in case of both the subportfolios, but the proportion of the "technical defaults" was notably larger in the lower LGD bands and smaller in the higher LGD bands in comparison with the ones in the basic model. A difference of approximately 3.5 percentage points appeared between the average LGDs of the two categories: while the average LGD in the basic model was 20.324%, the average LGD of the "technical defaults" was only 16.877%.

Table 48: The descriptive statistics of the LGD values in the basic model and the LGD values of the "technical defaults"

	MINWAGE (basic model)	*COMPLEMENT ("technical defaults")*
Mean	0.203244911	0.16877126
Standard Error	0.004601158	0.012237682
Median	0	0
Mode	0	0
Standard Deviation	0.380424211	0.353836236
Sample Variance	0.14472258	0.125200082
Kurtosis	0.403478081	1.472245433
Skewness	1.529258321	1.847190108
Range	1	1
Minimum	0	0
Maximum	1	1
Count	6836	836

(Self-made table: own calculation results)

Analogously with the former ones I tested the equivalence of the distributions with Homogeneity Analysis. As null hypothesis I composed that the distribution of the LGD values of the "technical defaults" is the same as the distribution of the LGD values calculated with using the materiality threshold of the basic model. The rejection of the null hypothesis would mean that the distributions are different, that is the lowering of the materiality threshold influences the result of the LGD calculation considerably.

Table 49: Homogeneity Analysis referring to the distribution of the LGD values in the basic model and the LGD values of the "technical defaults"

	n_{Yi}	n_{Xi}	$n_{Yi}+n_{Xi}$	g_{Yi}	g_{Xi}	$\frac{1}{n_{Y_i}+n_{X_i}}*\left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2$
0%	3549	455	4004	0.5191632534	0.5442583732	0.0000001573
0%-2%	458	63	521	0.0669982446	0.0753588517	0.0000001342
2%-4%	689	70	759	0.1007899356	0.0837320574	0.0000003834
4%-6%	358	66	424	0.0523698069	0.0789473684	0.0000016660
6%-8%	173	21	194	0.0253071972	0.0251196172	0.0000000002
8%-10%	128	14	142	0.0187244002	0.0167464115	0.0000000276
10%-20%	95	11	106	0.0138970158	0.0131578947	0.0000000052
20%-50%	50	2	52	0.0073142188	0.0023923445	0.0000004659
50%-80%	59	3	62	0.0086307782	0.0035885167	0.0000004101
80%-90%	69	9	78	0.0100936220	0.0107655502	0.0000000058
90%-92%	42	6	48	0.0061439438	0.0071770335	0.0000000222
92%-94%	38	1	39	0.0055588063	0.0011961722	0.0000004880
94%-96%	45	7	52	0.0065827970	0.0083732057	0.0000000616
96%-98%	84	7	91	0.0122878877	0.0083732057	0.0000001684
98%-100%	257	23	280	0.0375950848	0.0275119617	0.0000003631
100%	742	78	820	0.1085430076	0.0933014354	0.0000002833
Total	6836	836	7672	1	1	0.0000046421

(Self-made table: own calculation results)

$$\chi^2 = n_Y n_X \sum_{i=1}^{k} \frac{1}{n_{Y_i}+n_{X_i}} \left(\frac{n_{Y_i}}{n_Y}-\frac{n_{X_i}}{n_X}\right)^2 = 6836*836*0.0000046421 = 26.52901346$$

The degree of freedom is $\boldsymbol{v}$=16-1=15, so the upper critical value is $\boldsymbol{c_f}$=25.0197923 at significance level of $\boldsymbol{\alpha}$=5%, so the null hypothesis had to be rejected at this confidence level, namely it can not be stated that the distribution of the LGD values of the "technical defaults" does not differ considerably from the distribution of the LGD values calculated with using the materiality threshold of the basic model. The $\boldsymbol{p}$-value is 0.0328165, so the statement referring to the equivalence of the distributions can be considered as true only at the significance levels which are lower than 3.28%.

Despite the fact that the distributions are quite special, they differ notably from the normal distribution, I thought that considering the relatively large quantity of the elements the adequacy of the null hypothesis which states the equality of the average LGD values can be examined against the alternative hypothesis, which composes the dissimilarity of the averages, by two-sample asymptotic z-test in this case as well (two-sided asymptotic z-test, Table 50).

Table 50: Two-sample asymptotic z-test referring to the equality of the average LGD values in the basic model and the LGD values of the "technical defaults" (α=0.05)

	MINWAGE (basic model)	***COMPLEMENT ("technical defaults")***
Mean	0.203244911	0.16877126
Known Variance	0.14472258	0.125200082
Observations	6836	836
Hypothesized Mean Difference	0	
Z	2.636794038	
P(Z<=z) two-tail	0.008369363	
z Critical two-tail	1.959963985	

(Self-made table: own calculation results)

As the result of the test the null hypothesis had to be rejected at significance level of 5%, and the considerably low $\boldsymbol{p}$-value (0.008369363) indicated that the statement about the equality of the average LGD of the "technical defaults" and the average LGD in the basic model can be considered as true only at the significance levels which are lower then 0.837%.

Considering the speciality of the distributions and the low $\boldsymbol{p}$-values of the Homogeneity Analysis and the two-sample asymptotic z-test, the results are worthy to be handled with reservations. However, it can be laid down that according to the examinations, which were carried out, the statement composed in my 4th Hypothesis, according to which using the different materiality thresholds does not cause considerable affect on the result of the LGD calculation, can be accepted only at quite low significance levels.

6.3.5. Categories according to the closing type of the deals

In Subsection 6.2.3 I presented the segmentation of the deals according to the closing type *(deal_status)*, in the framework of which I classified the default events of the examined portfolio into three groups *("WorkoutEnd", "NoFurtherRec", "NotClosed")*. In connection with this is my 5th Hypothesis whose subject is the searching for the features of the categories defined according to the closing type of the deals, applying the logistic regression methodology which is reviewed in Chapter 4.4.2. According to my anticipative expectations the characteristics of the cases which compose the categories of the different closing types are insomuch diverse, that they are properly classifiable with using statistical methods. Considering that only the cases which are classified into the *"WorkoutEnd"* and the *"NoFurtherRec"* categories are involved in the LGD calculation, I examined my present hypothesis referring to them as well.

> ***5th Hypothesis: The LGD values of the categories according to the closing type of the deals differ strongly from each other, and the elements of the two groups which have closed recovery process ("NoFurtherRec", "WorkoutEnd") can be properly separated with using logistic regression.***

In the course of the segmentation I classified the ones into the *"WorkoutEnd"* category which are not in default status any more, because the client paid back the delayed amount, the exposure was written off as loss or the property which serves as underlying collateral was sold. Conversely the *"NoFurtherRec"* category contains the deals still in default status, in case of which more than 36 months passed from becoming non-performing or at least 90% of the exposure at the default event recovered (nominally, not being discounted).

Before applying the regression methodology I ascertained that the LGD values of this group really differ strongly from each other.

Both the bar-chart which represents the distributions (Figure 21) and the descriptive statistics (Table 51) confirm the statement which is composed in the 5th Hypothesis, according to which the LGD values of the categories according to the closing type of the deals strongly differ from each other. While the average LGD of the *"NoFurtherRec"* category was 88.563%, only 1.647% arose as average in the *"WorkoutEnd"* category. Beyond that both the dispersion and the kurtosis and the skewness indices speak about the considerably large diversities of the distributions as well.

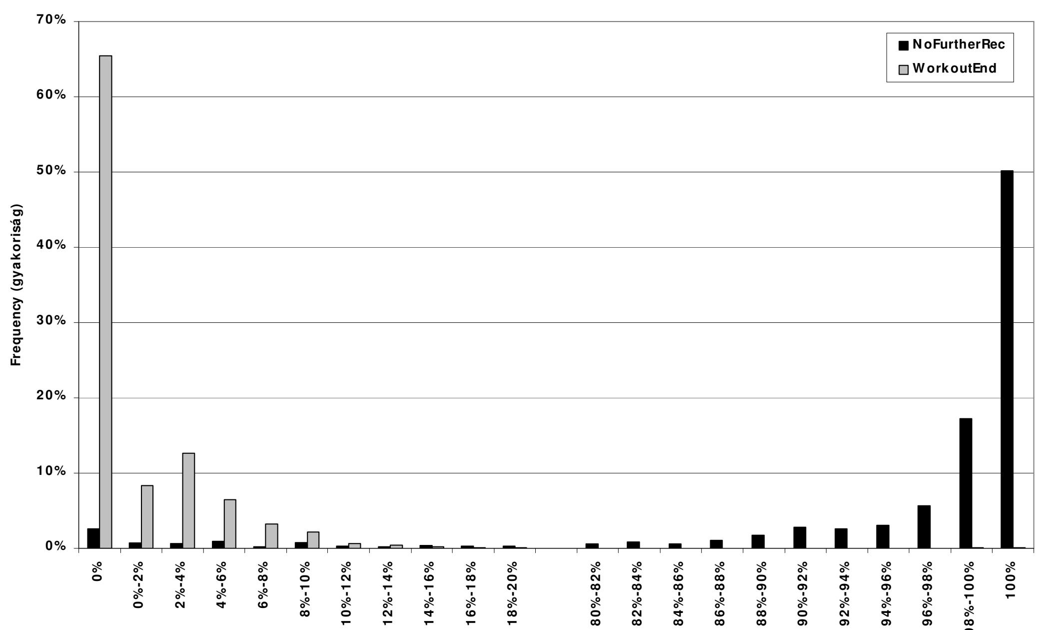
Table 51: The descriptive statistics of the LGD values of the deals with different closing types

	NOFURTHERREC	*WORKOUTEND*
Mean	0.885630279	0.016469412
Standard Error	0.006875633	0.000744533
Median	1	0
Mode	1	0
Standard Deviation	0.263526086	0.054544343
Sample Variance	0.069445998	0.002975085
Kurtosis	4.954960454	215.6652553
Skewness	-2.51700625	12.94456878
Range	1	1
Minimum	0	0
Maximum	1	1
Count	1469	5367

(Self-made table: own calculation results)

Analyzing the bar-chart it is also conspicuous that the elements of the *"WorkoutEnd"* group prevail clearly in the low LGD bands, whereas the dominance of exactly the other category features the high LGD range.

Figure 21: The distribution of the LGD values of the deals with different closing types



(Self-made figure: own calculation results)

On the basis of the above-mentioned results I drew the conclusion that the two distributions differ really strongly from each other, and the successfulness of the regression methodology in the filed of carrying out the classification can be assumed.

I built the logistic regression with SAS Enterprise Miner^{TM} 5.2 applying stepwise model selecting procedure. I applied only the variables[75] as inputs which were already known

[75] The variables are enlisted in the Appendix.

at the date of the default event, since I purposed to configure a model which is able to find out with large probability how the recovery period of the given deal will end.

Having tested numerous model types and transformation procedures in the end I came to the result that the simple logistic regressions built on standardised variables gave the best performance: the model variants which apply logit or probit link without any transformation proved to be the most promising ones. The standardisation of the input variables did not influence the performance of the logistic models, but I judged it necessary for the sake of valuing the role of the variables and promoting the final model selecting decisions.

Table 52 contains the fit statistics of the regression models which provide the best performance, displaying with light grey colour the indices which differ from each other in case of the two models.

Table 52: The fit statistics of the selected regression models

	LOGIT LINK	*PROBIT LINK*
Kolmogorov-Smirnov Statistic	0.69125	0.68515
Akaike's Information Criterion	4274.34	4325.86
Average Squared Error	0.10	0.10
Roc Index	0.90207	0.90208
Average Error Function	0.31	0.31
Capture Response	6.29402	6.29402
Degrees of Freedom for Error	6824.00	6809.00
Model Degrees of Freedom	12.00	27.00
Total Degrees of Freedom	6836.00	6836.00
Divisor for ASE	13672.00	13672.00
Error Function	4250.34	4271.86
Final Prediction Error	0.10	0.10
Gain	25.88	25.88
Gini Coefficient	0.80	0.80
Bin-Based Two-Way Kolmogorov-Smirnov Statistic	0.68225	0.67492
Lift	1.25880	1.25880
Maximum Absolute Error	1.00	1.00
Misclassification Rate	0.15579	0.15448
Mean Square Error	0.10019	0.10129
Sum of Frequencies	6836.00	6836.00
Number of Estimate Weights	12.00	27.00
Root Average Sum of Squares	0.32	0.32
Percent Response	98.8297	98.8297
Root Final Prediction Error	0.32	0.32
Root Mean Squared Error	0.32	0.32
Schwarz's Bayesian Criterion	4356.30	4510.27
Sum of Squared Errors	1369.77	1384.81
Sum of Case Weights Times Freq	13672.00	13672.00

(Self-made table: own calculation results)

It can be stated on the basis of the facts mentioned above that only a marginal difference can be observed between these two model variants from the viewpoint of the performance, nevertheless in the end I decided upon the model which applies logit link, because it proved to be at least as good as the other one considering all the fit statistics with the exception of the ROC-index and the Misclassification Rate, and its drawback in regard to these two indices is not notable either.

In Table 53 I recorded the result of the Maximum Likelihood estimation: the explanatory variables of the model which uses the logit link, the $\boldsymbol{\beta}$ coefficients of some variables, the standard errors of the estimations, the value of Wald's $\boldsymbol{F}$-test and the connecting $\boldsymbol{p}$-values as well as the $\boldsymbol{e^{\beta}}$s.

Adequately to Subsection 4.4.2 the $\boldsymbol{e^{\beta j}}$ factor shows how much partial multiplicative affect the unity absolute increment of the $\boldsymbol{x_j}$ explanatory variable has to the odds ceteris paribus, namely to the quotient of the probability of the supervention of the given event and the complementary event. Consequently in case of the variables, whose $\boldsymbol{\beta}$ coefficient is positive, the high value indicates that the deal will get into the *"NoFurtherRec"* category with large probability, while in case of the variables with negative $\boldsymbol{\beta}$ even contrarily: the low value of the given variable is accompanied by the larger probability of getting into the *"NoFurtherRec"* category.

Table 53: The variables of the logistic regression with logit link

Parameter	*Estimate*	*Standard Error*	*Wald* χ^2	*Pr >* χ^2 *(p-value)*	*Exp (Estimate)*
Intercept	2.4906	0.1940	164.76	<.0001	12.069
REASON_DEATH = 0	1.0071	0.1595	39.88	<.0001	2.738
REASON_PASTDUE = 0	0.9365	0.2344	15.96	<.0001	2.551
SETTLEMENT_TYPE = Budapest & environs	0.3630	0.1383	6.89	0.0087	1.438
SETTLEMENT_TYPE = County town & environs	0.4630	0.1228	14.22	0.0002	1.589
SETTLEMENT_TYPE = Other city & environs	-0.0689	0.1024	0.45	0.5010	0.933
SETTLEMENT_TYPE = Small village	-0.4707	0.1395	11.38	0.0007	0.625
STD_CUM_REALWAGE_INDEX_DS_Y	0.4041	0.0531	57.82	<.0001	1.498
STD_DEFAULT_REALWAGE_INDEX	-0.2433	0.0616	15.58	<.0001	0.784
STD_GDP_GROWTH_INDEX_DS_Y	-1.3934	0.0983	200.95	<.0001	0.248
STD_MONTHS_TO_DEFAULT	1.1343	0.1064	113.74	<.0001	3.109
STD_START_LTV	-0.3588	0.0671	28.55	<.0001	0.699

(Self-made table: own calculation results)

So, concerning the categorisation of the default event altogether 8 variables proved to be significant, and I make them known classified according to their kind.

The reasons of the default

The *reason_death* dummy variable indicates whether the deal is considered as non-performing because of death, while the *reason_pastdue* dummy serves as indicator whether the delay is the reason of the default status. The $\boldsymbol{e^{\beta}}$ values in the last column of Table 53 can be interpreted in a manner that ceteris paribus the death (reason_death=1) heightens the quotient of the probability of falling into the *"WorkoutEnd"* category and the probability of getting into the *"NoFurtherRec"* category (odds) to its 2.738-fold, compared to the case when the default occurred because of other reason than death. Similarly, the default status deriving from delay heightens the odds to its 2.551-fold ceteris paribus.

The settlement type of the real estate

The settlement type of the real estate which serves as collateral was also qualified as significant influencing factor. As compared with the rural properties (*settlement_type=Village*) in case of the real estates which are in Budapest and its environs (*settlement_type=Budapest & environs*), or in a county town and its environs (*settlement_type=County town & environs*) the quotient of the probability of falling into the *"WorkoutEnd"* category and the probability of getting into the *"NoFurtherRec"* category is approximately half as much again (1.438, 1.589). In case of the other settlement types the relation is just the opposite, that is the probability of the *"WorkoutEnd"* closing type of the deal is even less than that of the rural properties. The consequence can be drawn from this that compared to the others, the deals secured by real estates which are in large cities and their environs close more often in a way that the client will repay its debt, or there is a larger probability that the real estate will be sold or the obligation of the client will be written off in accounting meaning.

Macroeconomic factors

The yearly measure which is calculated form the index of the GDP at the default and at the origination (*GDP_growth_index_ds_y*: the yearly average growth quantified from the quotient of the cumulative GDP-growth indices) was also quantified as significant variable in the model. The e^{β} =0.248 shows that the faster the GDP grew in the period from the disbursement of the loan to the becoming non-performing of the deal, the larger the probability of falling into the *"NoFurtherRec"* category is, namely with smaller chance did the client repay large amount or was the real estate sold (or in special cases the loan amount was written off) later on. This result partly contradicts my anticipative expectations, since the GDP-growth generally pertains with the growth of the value of the real estates, so in the one hand the selling price can be higher, and on the other hand the client is also more strongly interested in repaying his/her debt, since he/she does not want to hazard the outvaluing real estate. It is possible that other factors lie in the background, for example the fact that the measure of the severity of the credit approval requirements differed in the distinctive phases of the economic development, or the circumstance that the growth of the wages was not always able to keep up with the inflation in the phase of intensive economic boom. So in the background of qualifying this variable as significant can be the strong correlation with other influencing factors.

In case of the yearly average real wage index (*cum_realwage_index_ds_y*), which is calculated from the ratio of the real wages at the default and the origination, the e^{β} =1.498 indicates that the faster the real wage grew in the period from the disbursement of the loan to the becoming non-performing of the deal, the larger the probability of falling into the *"WorkoutEnd"* category is. The background of this result can be that if the real wages grew fast in the period from the borrowing to the default, relatively large recoveries could be excepted following the default event, namely in many cases the clients were able to repay the majority of their existing debt (for example from their savings or the unforced sale of the real estate), or the bank succeeded in selling the credit deal or the underlying collateral.

The *default_realwage_index* shows the yearly real wage index at the date of the default, namely it is the quotient of the 12-month moving average calculated from the yearly change of the average monthly net income and the yearly consumer price index. The value of e^{β} (0.784), which is less than 1, indicates that in case of the loans whose default event happened in a period when the real wage decreased intensively, the proportion of the *"NoFurtherRec"* closing type was lower: the client repaid his/her debt with larger probability, or the real estate got sold with larger probability either by the bank or by the client. One of its possible explanations is that in these economic circumstances many clients recognised that they can solve their situation only with moving into a smaller flat, thus they voluntarily sold the real estate and from it they repaid their debt against the bank.

Loan-to-value ratio: start_LTV

The e^{β} =0.699, which belongs to the ratio of the loan amount and the market value of the collateral at the origination (*start_LTV*), shows that in case of the deals with high LTV-ratio the proportion of getting into the *"NoFurtherRec"* category is ceteris paribus larger, namely it has a smaller probability that the client will repay his/her debt later on. In such cases the loan amount is considerable in proportion to the value of the real estate, thus the client has less impetus to endeavour himself/herself to arrange the situation.

Paying history: months_to_default

The results indicate that the length of the period from the origination of the deal to the default event (*months_to_default*) influenced the closing type of the deal quite strongly (e^{β} =3.109). In cases when the client properly met his/her paying obligations for a long time before the default event, finally the deal got into the *"WorkoutEnd"* category with larger probability. The explanation for it can be that generally not the bad discipline to pay lay in the background, [76] but some kind of unexpected event in the consequence of which the client got into a tight situation, but considering also that he/she has already repaid notable amounts as instalment since the borrowing, he/she did the best for the sake of repaying his/her debts later on as well.

As comparison I present the results of the Maximum Likelihood estimation concerning the variables of the model with probit link in the following table.

Table 54: The variables of the logistic regression with probit link

Parameter	*Estimate*	*Standard Error*	*Wald χ^2*	*Pr > χ^2 (p-value)*
Intercept	1.2822	0.1098	136.31	<.0001
EMPL_INDUSTRY = Agriculture	0.0819	0.1456	0.32	0.5736
EMPL_INDUSTRY = Commerce, Entertainment	-0.1589	0.0726	4.79	0.0286
EMPL_INDUSTRY = Construction	-0.1932	0.0968	3.98	0.0461
EMPL_INDUSTRY = Education, Medical services, Gov	0.4436	0.1071	17.16	<.0001
EMPL_INDUSTRY = Finance, Legal services	-0.1503	0.1685	0.80	0.3725
EMPL_INDUSTRY = Other	-0.0713	0.0891	0.64	0.4233
PRODUCT = Home Equity	0.1363	0.0359	14.45	0.0001
REASON_DEATH = 0	0.5971	0.0867	47.47	<.0001
REASON_PASTDUE = 0	0.5073	0.1262	16.16	<.0001
REGION = Budapest & environs	0.2233	0.1446	2.39	0.1224
REGION = Central-Western	0.0124	0.0997	0.02	0.9012
REGION = Eastern	-0.4134	0.1049	15.52	<.0001
REGION = North-Eastern	-0.2266	0.0772	8.63	0.0033
REGION = North-Western	0.0202	0.1079	0.04	0.8516
REGION = South-Central	0.1153	0.1394	0.68	0.4082
REGION = South-Eastern	0.1327	0.1213	1.20	0.2740
REGION = South-Western	0.1097	0.1382	0.63	0.4272
SETTLEMENT_TYPE = Budapest & environs	0.0364	0.1388	0.07	0.7934
SETTLEMENT_TYPE = County town & environs	0.3116	0.0816	14.58	0.0001
SETTLEMENT_TYPE = Other city & environs	-0.0203	0.0683	0.09	0.7667
SETTLEMENT_TYPE = Small village	-0.1990	0.0871	5.22	0.0224
STD_CUM_CPI_DS_Y	-0.1572	0.0349	20.26	<.0001
STD_DEFAULT_CPI	-0.1328	0.0444	8.95	0.0028
STD_FIRST_INSTALMENT	-0.1215	0.0349	12.11	0.0005
STD_GDP_GROWTH_INDEX_DS_Y	-0.5156	0.0404	163.04	<.0001
STD_MONTHS_TO_DEFAULT	0.6218	0.0535	135.32	<.0001

(Self-made table: own calculation results)

The variables which are in connection with the rate of growth of the real wage (*cum_realwage_index_ds_y*, *default_realwage_index*) do not appear in this model, but

[76] In case of bad discipline to pay the default generally occurs earlier.

as a quasi compensation the indices which measure the changing of the consumer prices (*cum_CPI_ds_y*, *default_CPI*) proved to be significant. Similarly, the model with probit link does not contain the *start_LTV*, which shows the ratio of the loan amount and the market value of the collateral at the origination, but the product type (*product*), which is in tight connection with this variable, was qualified as significant. The industry of the client's employer, the region of the property and the amount of the first instalment occurred as further variables.
It is an important lesson that the reasons of the default (*reason_death*, *reason_pastdue*), the settlement type of the real estate which serves as collateral (*settlement_type*), the yearly average rate of the GDP-growth in the period from the disbursement of the loan to the default (*GDP_growth_index_ds_y*), and the length of the period to the default (*months_to_default*) proved to be significant in case of this model as well, and regarding these variables the direction of the connections are the same as the ones in the model with the logit link. Roughly speaking it can be stated that the two models show considerable cognateness concerning both the scale of the influencing factors and the direction of the connections.

On the basis of all these results my 5th Hypothesis can be considered as justified, since the tests carried out confirm the considerable dissimilarity of the LGD values of the categories which are defined according to the closing type of the deals, moreover with logistic regression it managed to establish a model in case of which the Misclassification Rate is only 15.5%, the Gini-coefficient is 0.902 and the ROC-index is 0.80, which can also be considered as advantageous (Table 52).

6.3.6. The Hungarian Interbank LGD Database

According to the prescriptions of the CRD and the Hkr.[77] the estimation of the LGD has to be founded basically on the internal data, but external or even common data can be used as well, if it is provable that there is not any significant difference between the internal and external data regarding for example the composition of the data, or if the differences can be adjusted properly for the sake of completing representativity. In connection with this circumstance my 6th Hypothesis was directed towards the examination of the Hungarian Interbank LGD Database which is presented in Chapter

[77] I made known the details in Subsection 2.3.3.

3.5. My goal was to make a survey of the factors which are able to predict statistically confidently the length of the period which is needed for the recoveries from selling the collateral or the debt, and to predict the recovery rate itself.

Similarly to the procedure presented in Chapter 6.2.5 about the data used for investigating the influencing factors for the purpose of allocating properly the collateral values and the market values of the collaterals I carried out deal level summing-up here as well, and I linked the characteristics of the real estate with the highest value to each deal in all cases. If the same collateral referred to more than one deal, then I linked the values to the single deals allocated according to the proportion of the exposure at the default.

> ***6th Hypothesis: With the linear regression models on the basis of the Hungarian Interbank LGD Database, the deals of the "NotClosed" category can also be involved in the calculation, and a more exact and more accurate deal level LGD estimation becomes possible.***

In the framework of the present Hypothesis I took steps with using the data of the Hungarian Interbank LGD Database to develop a complex model, with which the deals of the *"NotClosed"* category can also be involved in the calculations. However, it is its very important condition to prepare regressions, with which a precise prediction can be made referring to the expected length of the recovery period of the deals of the *"NotClosed"* category and the recovery rate deriving from the selling, on the basis of the data which are available at the default. If I manage to find the factors which explain the large proportion of the variance of these two target variables (expected length of the recovery period, recovery rate deriving from the selling), then my 6th Hypothesis can be considered as justified, since on the basis of these regressions a more accurate prediction can be made for the deal level LGDs, in comparison with the case when these pieces of information are left out of consideration.

The length of the recovery period

For the purpose of justifying my hypothesis in the first step I built a linear regression referring to the expected length of the recovery period. The target variable was the number of months from the default event to the selling of the deal or the underlying property (*months_to_cut_off*), whereas the following ones performed as input variables:

- the data fields of the Hungarian Interbank LGD Database, whose filling-in is compulsory, which were already known at the date of the default event and which are also available in the database of the bank,[78]
- makroeconomic data which I was able to connect to the single deals on the basis of the date of the origination or the default event.

Deriving from the fact that the filling-in of a considerable part of the data fields in the database is only optional, and/or it is not available in the used database of the bank, the number of the data fields which can be involved in the modelling proved to be rather small, since the large data deficiency made unviable the use of numerous factors which seemed to be promising otherwise. Appendix 10 contains the elements of the set of the variables which served as a basis for modelling: I applied these and their transformed variables as potential factors in the course of model building.

I built the linear regression with stepwise procedure using SAS Enterprise Miner™ 5.2, then because the model which was established in this manner had considerably weak explanatory power, I made modifications on expert base, but I was not able to improve the performance of the model notably with any correcting steps. In the following I present the results of the Wald test (Table 55) and the ***F***-tests which examine the significance of the single variables (Table 56), referring to the model which proved to be the best according to the adjusted coefficient of determination.

Table 55: The Wald test of the model developed for the length of the recovery period

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F (p-value)*
Model	28	12561	448.604493	5.82	.0001
Error	1517	116992	77.120925		
Corrected Total	1545	129553			

(Self-made table: own calculation results)

Table 56: The significance tests of the variables of the model developed for the length of the recovery period

Effect	*DF*	*Sum of Squares*	*F Value*	*Pr > F (p-value)*
COUNTY	19	3777.0358	2.58	0.0002
LOAN_PURPOSE	3	1137.6503	4.92	0.0021
CUM_CPI_DS_Y	1	1635.1704	21.20	<.0001
UNEMPL_RATE_INDEX_DS_Y	1	312.0183	4.05	0.0445
DEFAULT_CPI	1	523.9663	6.79	0.0092
DEFAULT_MIN_WAGE	1	1896.4627	24.59	.0001
DEFAULTED_PER_START_EXPOSURE	1	1178.1484	15.28	<.0001
CURRENT_LTV	1	2524.8006	32.74	<.0001

(Self-made table: own calculation results)

[78] This condition is necessary so that the model can be applied for the own database of the bank as well.

Though low ***p***-values derived from both the Wald test and the significance test of the single explanatory variables, altogether the model is still rather weak, since the adjusted coefficient of determination is only 0.0803 (the raw coefficient of determination is 0.0970), namely the factors can explain only a very small proportion of the variance of the recovery period's length.

After all I think it is worth reviewing which factors proved to be considerable in the course of the modelling the length of the recovery period on the basis of the Maximum Likelihood estimation, thus in Table 57 I indicated the explanatory variables, the estimated parameters, their standardised values, their standard errors as well as the results of the ***t***-test and the ***p***-values relating to them. A part of the variables of the prepared model is in connection with the deal itself or with the underlying collateral, whereas the other part of them consists of the macroeconomic changes in the period from the origination of the deal and the characteristics of the macroeconomic situation at the default.

Table 57: The variables of the regression developed for the length of the recovery period

Parameter	*Estimate (non-standar-dised)*	*Estimate (standar-dised)*	*Standard Error (standar-dised)*	*t Value*	*Pr > |t| (p-value)*
Intercept	-100.7	14.9128	0.4085	36.51	<.0001
COUNTY=Baranya	1.1004	1.1004	1.3274	0.83	0.4072
COUNTY=Borsod-Abaúj-Zemplén	-2.4692	-2.4692	0.9468	-2.61	0.0092
COUNTY=Budapest	1.8953	1.8953	0.7837	2.42	0.0157
COUNTY=Bács-Kiskun	1.0652	1.0652	1.2701	0.84	0.4018
COUNTY=Békés	-3.7576	-3.7576	1.2714	-2.96	0.0032
COUNTY=Csongrád	3.2227	3.2227	1.1520	2.80	0.0052
COUNTY=Fejér	0.1592	0.1592	1.3159	0.12	0.9037
COUNTY=Győr-Moson-Sopron	0.4562	0.4562	1.5834	0.29	0.7733
COUNTY=Hajdu-Bihar	-0.3116	-0.3116	1.1118	-0.28	0.7793
COUNTY=Heves	-1.5188	-1.5188	0.9731	-1.56	0.1188
COUNTY=Jász-Nagykun-Szolnok	-0.2926	-0.2926	0.9262	-0.32	0.7521
COUNTY=Komárom-Esztergom	-0.4112	-0.4112	1.1715	-0.35	0.7256
COUNTY=Nógrád	0.7071	0.7071	0.9255	0.76	0.4450
COUNTY=Pest	-0.1437	-0.1437	0.6565	-0.22	0.9268
COUNTY=Somogy	-0.3566	-0.3566	0.9087	-0.39	0.6948
COUNTY=Szabolcs-Szatmár-Bereg	-1.6533	-1.6533	0.6420	-2.58	0.0101
COUNTY=Tolna	5.6115	5.6115	2.2564	2.49	0.0130
COUNTY=Vas	-2.7035	-2.7035	2.6723	-1.01	0.3118
COUNTY=Veszprém	0.7974	0.7974	1.5420	0.52	0.6052
LOAN_PURPOSE=Other	0.3749	0.3749	0.4136	0.91	0.3649
LOAN_PURPOSE=Real estate construction	1.5622	1.5622	0.6627	2.36	0.0185
LOAN_PURPOSE=Real estate purchase	-1.2811	-1.2811	0.4434	-2.89	0.0039
CUM_CPI_DS_Y	88.6327	1.2608	0.2738	4.60	<.0001
UNEMPL_RATE_INDEX_DS_Y	4.4630	0.5840	0.2930	2.01	0.0445
DEFAULT_CPI	42.6944	0.7877	0.3022	2.61	0.0092
DEFAULT_MIN_WAGE	-0.00033	-1.4144	0.2852	-4.96	<.0001
DEFAULTED_PER_START_EXPOSURE	-10.1692	-1.2241	0.3132	-3.91	<.0001
CURRENT_LTV	8.7642	1.4951	0.2613	5.72	<.0001

(Self-made table: own calculation results)

Considering the single counties (*county*) quite important differences appeared: in case of uniformity of all the other factors even a 9-month difference can occur between the lengths of the recovery periods in case of two real estates which lie in other counties.[79]

Regarding the purpose of the loan (*loan_purpose*) the real estate renovation constituted the base of comparison, compared to it the recovery period is ceteris paribus shorter in case of the loans with the purpose of real estate purchase (*loan_purpose=Real estate purchase*), while it is longer in case of the other purposes of the loans. Its explanation can be that because of its individual character there is a smaller demand for the property which is built or renovated according to the own taste of the client, while more liquid "common real estates" lie as collateral behind an important part of the loans with the purpose of real estate purchase.

Among the macroeconomic factors the yearly average growth of the consumer prices and the unemployment rate from the origination of the deal to the default event (*cum_CPI_ds_y*, *unempl_rate_index_ds_y*) as well as the consumer price index and the minimum wage at the default (*default_CPI*, *default_min_wage*) occurred in the model. Generally speaking while according to the model the inflation pressure and the heavy unemployment increase the length of the recovery period to a certain extent, the high level of minimum wage affects the length of the recovery period into the direction of decrease.

Both the proportion of the exposure at the default and the disbursed amount (*defaulted_per_start_exposure*), and the ratio of the exposure at the default and the value of the collateral at the same time (*current_LTV*) proved to be significant as well. These factors can take effect on the length of the recovery period mainly through influencing the intensity of the collection procedure. Since if the bank realizes that the debt of the client has not decreased considerably since the origination of the deal (or it could even increase for example in case of the notable change of the currency exchange rate), then it handles the deal in the "normal" collection process for a shorter time and it settles upon the selling earlier. On the contrary to it, if the loan amount is very considerable in comparison with the value of the collateral, then the selling can take a longer time, because on the one hand the bank announces a competition for the buyers for the sake of realising the appropriate recovery, and on the other hand the deferring of the selling is a rational decision having confidence in the increasing of the collateral value as well.

[79] One of the extremes is Tolna, the other is Békés: the difference of the coefficients is 5.6115-(-3.7576)=9.3691 months.

It can be generally said that all the variables can be interpreted logically easily, after all the explanatory power of the model is insofar low that it does not justify the statement composed in the 6th Hypothesis, since using the Hungarian Interbank LGD Database I did not manage to built a linear regression model which has good explanatory power.

The recovery rate deriving from the selling

In the next step I constructed a linear regression for the recovery rate also using stepwise procedure. In this case the target variable was the proportion of the recovery deriving from the selling discounted to the date of the default and the exposure at the default (*disc_nr_interest*). I applied the contractual borrowing rate as discount rate, which I defined in a way that on the basis of the database of the bank I quantified for each currency the average contractual borrowing rate of the deals which originated in the same month, and I connected them to the deals of the Hungarian Interbank LGD Database. The scale of the input variables corresponded with the one which was used in the framework of modelling the length of the recovery period (Appendix 10).
I performed the model building using SAS Enterprise Miner™ 5.2 in this case as well. The established model showed better performance (the adjusted coefficient of determination is 0.2272) in comparison with the model created for the length of the recovery period, but it even fell short of my anticipative expectations. The following tables contain the results of the Wald test (Table 58) and the values of the ***F***-tests which examine the significance of the single variables (Table 59).

Table 58: The Wald test of the model developed for the recovery rate deriving from the selling

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F (p-value)*
Model	30	23.131504	0.771050	16.14	<.0001
Error	1515	72.378447	0.047775		
Corrected Total	1545	95.509951			

(Self-made table: own calculation results)

Table 59: The significance tests of the variables of the model developed for the recovery rate deriving from the selling

Effect	*DF*	*Sum of Squares*	*F Value*	*Pr > F (p-value)*
COUNTY	19	1.8659	2.06	0.0047
SETTLEMENT_TYPE	4	0.4738	2.48	0.0423
CUM_CPI_DS_Y	1	1.7868	37.40	<.0001
UNEMPL_RATE_INDEX_DS_Y	1	0.3569	7.47	0.0063
CURRENT_LTV	1	9.9364	207.98	<.0001
PRIORCHARGE_RATE	1	2.4968	52.26	<.0001
LOAN_PURPOSE	3	0.6484	4.52	0.0036

(Self-made table: own calculation results)

It is conspicuous that numerous ones among the explanatory variables appear also in the model created for the length of the recovery period, namely there is a large overlapping between the factors of the two models: as a matter of fact very similar factors influence the length of the recovery period and the recovery rate deriving from the selling. In Table 60 I summarised the results of the Maximum Likelihood estimation of the regression developed for the recovery rate deriving from the selling: the explanatory variables, the estimated parameters, their standardised values, their standard error as well as the results of the ***t***-tests and the ***p***-values relating to them.

Table 60: The variables of the regression developed for the recovery rate deriving from the selling

Parameter	*Estimate (non-standar-dised)*	*Estimate (standar-dised)*	*Standard Error (standar-dised)*	*t Value*	*Pr > |t| (p-value)*
Intercept	3.7596	0.5866	0.0112	52.37	<.0001
COUNTY=Baranya	-0.0772	-0.0772	0.0334	-2.31	0.0208
COUNTY=Borsod-Abaúj-Zemplén	0.0192	0.0192	0.0237	0.81	0.4175
COUNTY=Budapest	0.1468	0.1468	0.0363	4.04	<.0001
COUNTY=Bács-Kiskun	-0.0192	-0.0192	0.0317	-0.60	0.5462
COUNTY=Békés	-0.0134	-0.0134	0.0318	-0.42	0.6737
COUNTY=Csongrád	-0.0444	-0.0444	0.0291	-1.52	0.1279
COUNTY=Fejér	0.0194	0.0194	0.0329	0.59	0.5547
COUNTY=Győr-Moson-Sopron	0.000177	0.000177	0.0395	0.00	0.9964
COUNTY=Hajdu-Bihar	0.0314	0.0314	0.0278	1.13	0.2591
COUNTY=Heves	-0.0360	-0.0360	0.0244	-1.48	0.1399
COUNTY=Jász-Nagykun-Szolnok	-0.0246	-0.0246	0.0323	-1.06	0.2885
COUNTY=Komárom-Esztergom	-0.00372	-0.00372	0.0293	-0.13	0.8988
COUNTY=Nógrád	-0.0253	-0.0253	0.0234	-1.08	0.2810
COUNTY=Pest	0.0125	0.0125	0.0190	0.66	0.5120
COUNTY=Somogy	-0.00909	-0.00909	0.0228	-0.40	0.6902
COUNTY=Szabolcs-Szatmár-Bereg	-0.0377	-0.0377	0.0160	-2.35	0.0187
COUNTY=Tolna	0.0299	0.0299	0.0563	0.53	0.5960
COUNTY=Vas	-0.0377	-0.0377	0.0667	-0.57	0.5719
COUNTY=Veszprém	0.0735	0.0735	0.0385	1.91	0.0565
SETTLEMENT_TYPE=Budapest & environs	0.00621	0.00621	0.0266	0.23	0.8155
SETTLEMENT_TYPE=County town & environs	0.0353	0.0353	0.0144	2.45	0.0145
SETTLEMENT_TYPE=Other city & environs	-0.0117	-0.0117	0.0113	-1.03	0.3021
SETTLEMENT_TYPE=Small village	-0.0180	-0.0180	0.0136	-1.32	0.1858
CUM_CPI_DS_Y	-2.5920	-0.0369	0.00603	-6.12	<.0001
UNEMPL_RATE_INDEX_DS_Y	-0.1342	-0.0176	0.00642	-2.73	0.0063
CURRENT_LTV	-0.5953	-0.1016	0.00704	-14.42	<.0001
PRIORCHARGE_RATE	-0.3176	-0.0498	0.00688	-7.23	<.0001
LOAN_PURPOSE=Other	0.0185	0.0185	0.00998	1.85	0.0643
LOAN_PURPOSE=Real estate construction	-0.0566	-0.0566	0.0166	-3.41	0.0007
LOAN_PURPOSE=Real estate purchase	0.0257	0.0257	0.0112	2.30	0.0216

(Self-made table: own calculation results)

The recovery rate deriving from the selling is considerably different in terms of the single counties (*county*). According to the model the largest deviation is between Budapest and Baranya County, in case of uniformity of all the other factors the recovery rate deriving from the selling is approximately 22.4 percentage points higher in

Budapest than in Baranya County.[80] In addition to Baranya, Csongrád, Szabolcs-Szatmár-Bereg and Heves are also classed among the counties which can be characterised with considerably low recovery rate, whereas the recoveries are ceteris paribus quite high for example in Veszprém.

The type of the settlement (*settlement_type*) also proved to be an important factor. The base of comparison was formed in case of this variable by the deals in whose background rural property lies as collateral. Compared to them, the recovery rate deriving from the selling is ceteris paribus 3.53 percentage points higher in case of the deals which are secured by property situated on a county town or its environs, whereas for example in case of the deals secured by real estate situated on small village 1.8 percentage points lower recovery can be expected according to the model.

Among the macroeconomic factors the yearly average growth of the consumer prices and the unemployment rate from the origination of the deal to the default event (*cum_CPI_ds_y*, *unempl_rate_index_ds_y*) proved to be important. Both of the mentioned factors are in inverse relation with the recovery rate deriving from the selling, namely both the inflation pressure and the heavy unemployment ceteris paribus decrease the recovery.

The proportion of the exposure at the default and the collateral value at the same time (*current_LTV*) also proved to be a fairly important factor. However, *Qi and Yang [2007; 2009]* investigated the relationship between the CLTV and the LGD in their studies[81] which are made known in Chapter 5.3, a parallel can be drawn between the results of the referenced authors and the model developed by me: the higher the LTV at the default event is, ceteris paribus the lower the recovery rate deriving from the selling is, namely the higher LGD can be expected, taking some simplifications.

The quotient of the prior charges on the collateral and the realization value at the origination of the deal (*priorcharge_rate*) appears as a further factor among the explanatory variables which proved to be significant. Similarly to the LTV at the default (*current_LTV*), this is also in negative relationship with the recovery rate deriving from the selling.

Neither the qualifying of the purpose of the loan (*loan_purpose*) as significant is surprising. It can serve as a logical explanation for the lower recovery rate of the deals with the purpose of real estate construction (*loan_purpose=Real estate construction*) that if the paying difficulties already existed during the building as well, then the

---

[80] The difference of the coefficients is 0.1468-(-0.0772)=0.2240, namely 22.4 percentage points.

[81] In contrast to the studies of *Qi and Yang [2007; 2009]* the present hypothesis does not focus on investigating the LGD, but on examining the recovery rate deriving from the selling.

standards of the execution probably fell short of the requirements, moreover if the construction is not finished yet, then this circumstance continues to reduce the expected recovery. In contrast to this the recovery rate deriving from the selling is approximately 8.23 percentage points higher is case of the loans with the purpose of real estate purchase ceteris paribus.[82]

Similarly to the regression developed for the length of the recovery period, it can be stated here as well that the model gives a good account of itself, but its explanatory power proved to be low. I think that I did not manage to justify the statement composed in the 6th Hypothesis, according to which the deals of the *"NotClosed"* category can also be involved in the calculations with using the data of the Hungarian Interbank LGD Database, and through this a more exact and more precise LGD estimation becomes possible: the explanatory power of the linear regression models which were built on the data of the Hungarian Interbank LGD Database proved to be considerably low, thus their applicability for the purpose of prediction is questionable.

6.3.7. The influencing factors of the LGD value of the different categories

Further respects to my researches were provided by the studies of Thomas, Mues, Matuszyk and Moore *(Thomas et al. [2007a; 2007b])*[83], in which the authors explored the determinative factors of the recoveries from collection processes done by the bank's internal collection department and by the external agent. Both studies indicated that the scopes of the influencing factors differed considerably from each other in case of using the two different collection channel.

The studies of the referenced authors inspired my present hypothesis, in the framework of which I investigated whether the influencing factors of the LGD values of the deals with different closing types differ considerably from each other.

> ***7th Hypothesis: Different factors influence the LGD values of the deals with different closing types ("WorkoutEnd", "NoFurtherRec"), thus it is inappropriate to handle these categories together in the course of modelling the deal level LGD.***

[82] The difference of the coefficients is 0.0257-(-0.0566)=0.0823, namely 8.23 percentage points.
[83] I dealt with these studies in Chapter 5.2 which presents the analyses referring to loans on the basis of workout LGD methodology.

For the purpose of justifying my hypothesis I built linear regression models with stepwise procedure using SAS Enterprise Miner™ 5.2 separately for the categories defined according to the closing type of the deals, and on the basis of them I investigated the factors which proved to be significant. In the models the target variable was the deal level LGD calculated discounting with the contractual borrowing rate of the deal (*account_lgd_interest*), whereas all the data fields which were presented in Chapter 6.1.2 were included as input factors, which were already available at the date of the default event of the single deals. In the first step I involved the currency (*ccy*) in the modelling as well, but the difference between the estimated parameters relating to certain currencies was insofar large, that for example in case of the *"NoFurtherRec"* deal category it indicated ceteris paribus an approximately 40 percentage points LGD difference between the CHF-loans and the EUR-loans. This is presumably a consequence of underlying factors, thus considering also that this variable became dominant in case of both deal categories (it explained 80-90% of the variance), finally I left it out from the input variables.[84] Appendix 11 contains the elements of the set of variables which served as a basis for the modelling.

The influencing factors of the LGD of the "*WorkoutEnd" deal category*

The *"WorkoutEnd"* category contains the deals which are not in default status any more, because the client has paid back the delayed amount, the exposure has been written off or the property which served as underlying collateral has been sold. So this deal class is considerably heterogeneous, and it is not surprising that here I also met similar problem than in the course of modelling the length of the recovery period in the Hungarian Interbank LGD Database: the regression, built with stepwise procedure using SAS Enterprise Miner™ 5.2, had a rather small explanatory power. The adjusted coefficient of determination is 0.0751 altogether (the raw coefficient of determination is 0.0836), namely the explanatory variables can explain only a small proportion of the variance of the recovery period's length, after all the results of the Wald test (Table 61) and the ***F***-tests which examine the significance of the single variables (Table 62) show appropriately low ***p***-values. So the model is significant, even though its applicability for predicting purposes is questionable.

[84] Because of the same reason I left out the index of the currency exchange rate at the default and at the origination (*fx_index_ds*) from the input variables as well.

Table 61: The Wald test of the model developed for the LGD of the "WorkoutEnd" deal category

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F (p-value)*
Model	27	0.754840	0.027957	9.81	0.0001
Error	2905	8.275638	0.002849		
Corrected Total	2932	9.030478			

(Self-made table: own calculation results)

Table 62: The significance tests of the variables of the model developed for the LGD of the "WorkoutEnd" deal category

Effect	*DF*	*Sum of Squares*	*F Value*	*Pr > F (p-value)*
COUNTY	19	0.1339	2.47	0.0004
DEFAULTED_EXPOSURE_LCY	1	0.0211	7.42	0.0065
DEFAULTED_PER_START_EXPOSURE	1	0.0304	10.67	0.0011
PRIORCHARGE_RATE	1	0.0188	6.59	0.0103
AVG_PD	1	0.0704	24.71	<.0001
DEFAULT_AVG_NETINCOME	1	0.0113	3.97	0.0464
DEFAULT_CPI	1	0.1010	35.45	<.0001
DEFAULT_REALWAGE_INDEX	1	0.0838	29.42	<.0001
DEFAULT_UNEMPL_RATE	1	0.0195	6.84	0.0090

(Self-made table: own calculation results)

It is conspicuous from the Table 62 that none of the client characteristics proved to be significant influencing factor in spite of the fact that considerable such data fields existed among the input variables. The factors which describe the macroeconomic situation at the default as well as some deal and collateral characteristics played the most dominant role.

In Table 63 I summarised the results of the Maximum Likelihood estimation of the regression developed for the LGD of the *"WorkoutEnd"* deal category: the explanatory variables, the estimated parameters, their standardised values, their standard error as well as the results of the ***t***-tests and the ***p***-values relating to them.

The county of the property which serves as collateral (*county*) appeared also in this model as important influencing factor. The estimated parameters indicate that the divergence between the LGD of the deals which are secured by real estate situated in different counties can be even 2.492 percentage points. One of the extremes is Tolna, whereas the other one is Csongrád County: according to the model the LGD of the deal secured by a real estate situated in Tolna County is ceteris paribus 2.492 percentage points higher than that of the deal which is secured by a real estate situated in Csongrád County.

If we also take into consideration that the average LGD value of the *"WorkoutEnd"* category is only 1.647%, this divergence is not negligible at all.

It is an interesting result that while according to the data of the Hungarian Interbank LGD Database Csongrád County belonged to the counties which were featured by low recovery rate, in present case just its opposite arose as result. I consider it as important to emphasize that while the Hungarian Interbank LGD Database contains only the data in case of which selling occurred, the examination of the present hypothesis lies on the database of the bank in which a large proportion is represented by the deals which are not in default status any more because for example the client has already paid back the delayed amount.

Table 63: The variables of the regression developed for the LGD of the "WorkoutEnd" category

Parameter	*Estimate (non-standar-dised)*	*Estimate (standar-dised)*	*Standard Error (standar-dised)*	*t Value*	*Pr > |t| (p-value)*
Intercept	0.7496	0.0186	0.00123	15.10	<.0001
COUNTY=Baranya	-0.00725	-0.00725	0.00551	-1.32	0.1885
COUNTY=Borsod-Abaúj-Zemplén	-0.00111	-0.00111	0.00394	-0.28	0.7776
COUNTY=Budapest	-0.00671	-0.00671	0.00283	-2.37	0.0179
COUNTY=Bács-Kiskun	-0.00444	-0.00444	0.00417	-1.06	0.2870
COUNTY=Békés	-0.00560	-0.00560	0.00452	-1.24	0.2151
COUNTY=Csongrád	-0.00802	-0.00802	0.00494	-1.62	0.1048
COUNTY=Fejér	-0.00648	-0.00648	0.00419	-1.55	0.1223
COUNTY=Győr-Moson-Sopron	-0.00099	-0.00099	0.00439	-0.22	0.8220
COUNTY=Hajdu-Bihar	-0.00417	-0.00417	0.00407	-1.02	0.3055
COUNTY=Heves	0.00716	0.00716	0.00567	1.26	0.2069
COUNTY=Jász-Nagykun-Szolnok	-0.00251	-0.00251	0.00618	-0.41	0.6854
COUNTY=Komárom-Esztergom	-0.00220	-0.00220	0.00409	-0.54	0.5914
COUNTY=Nógrád	0.00510	0.00510	0.00584	-0.87	0.3824
COUNTY=Pest	-0.00101	-0.00101	0.00290	-0.35	0.7277
COUNTY=Somogy	-0.00658	-0.00658	0.00683	-0.96	0.3354
COUNTY=Szabolcs-Szatmár-Bereg	-0.00509	-0.00509	0.00384	-1.33	0.1847
COUNTY=Tolna	0.0169	0.0169	0.00831	2.03	0.0421
COUNTY=Vas	-0.00375	-0.00375	0.00708	-0.53	0.5965
COUNTY=Veszprém	-0.00213	-0.00213	0.00400	-0.53	0.5947
DEFAULTED_EXPOSURE_LCY	5.28E-10	0.00275	0.00101	2.72	0.0065
DEFAULTED_PER_START_EXPOSURE	0.0216	0.00399	0.00122	3.27	0.0011
PRIORCHARGE_RATE	0.0173	0.00250	0.000973	2.57	0.0103
AVG_PD	-0.6131	-0.0153	0.00307	-4.97	<.0001
DEFAULT_AVG_NETINCOME	-4.08E-7	-0.00349	0.00175	-1.99	0.0464
DEFAULT_CPI	-0.5301	-0.00774	0.00130	-5.95	<.0001
DEFAULT_REALWAGE_INDEX	-0.1616	-0.00587	0.00108	-5.42	<.0001
DEFAULT_UNEMPL_RATE	0.5051	0.00805	0.00308	2.61	0.0090

(Self-made table: own calculation results)

Numerous researchers reported strong positive relationship between the loan amount at the origination and the LGD (for example *Felsovalyi and Hurt [1998], Dermine and Neto de Carvalho [2003; 2005]*, as well as *Thomas et al. [2007a; 2007b]*), whereas we can find precedent also for its opposite: for example *Eales and Bosworth [1998]*, as well as *Qi and Yang [2007; 2009]* mentioned the negative correlation between the loan amount at the origination and the LGD. In the model developed by me instead of the loan amount at the origination the exposure at the date of the default (*defaulted_exposure_lcy*) proved to be an important factor, and I experienced a positive relation in line with the study of *Bellotti and Crook [2008]*.

Among the data fields related to the deal or the underlying property, the proportion of the exposure at the default and the disbursed amount (*defaulted_per_start_exposure*), as well as the quotient of the prior charges on the collateral and the realization value at the origination of the deal (*priorcharge_rate*) were qualified as further important influencing factors. We can say on the basis of the estimated parameter values that the high values of both factors take effect towards increasing the deal level LGD, namely ceteris paribus the less the existing loan amount decreased since the origination of the deal and the larger the charging on the property which serves as collateral is, the higher the LGD is (the lower the recovery rate is) according to the model.

All the further significant explanatory variables feature the macroeconomic situation at the default: the average default rate of the mortgage loans (*avg_PD*), the average net income (*default_avg_netincome*), the consumer price index (*default_CPI*), the yearly growth index of the real wages (*default_realwage_index*) as well as the unemployment rate (*default_unempl_rate*). I consider it necessary to emphasize the negative sign of the estimated parameter of the default rate (*avg_PD*), since we can usually read in the literature about the positive correlation between the LGD and the default rate (for example *Grunert and Weber [2005; 2009]*, *Brady et al. [2007]*, *Bellotti and Crook [2008]*), or in some cases about independency respectively (*Carey – Gordy [2003]*). However, in case of the other factors the results were not surprising: while ceteris paribus the high wages, the fast growth of the real wage and the high consumer price index decrease the LGD, the high unemployment rate takes effect towards increasing the LGD.

The influencing factors of the LGD of the "*NoFurtherRec*" *deal category*

The *"NoFurtherRec"* category consists of the deals which are still in default status, since their becoming non-performing longer than 36 months duration has passed, and in case of which at least 90% of the exposure at the date of default has recovered (nominally, without discounting). According to my anticipative expectations, this group is much more homogeneous in comparison with the *"WorkoutEnd"* category, and the influencing factors of the deal level LGD can be better defined.

The following tables contain the results of the Wald test (Table 64) and the values of the ***F***-tests which examine the significance of the single variables (Table 65) of the regression which is built with stepwise procedure using SAS Enterprise Miner™ 5.2.

Table 64: The Wald test of the model developed for the LGD of the "NoFurtherRec" deal category

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F (p-value)*
Model	20	25.154705	1.257735	28.55	<.0001
Error	503	22.158698	0.044053		
Corrected Total	523	47.313403			

(Self-made table: own calculation results)

Table 65: The significance tests of the variables of the model developed for the LGD of the "NoFurtherRec" deal category

Effect	*DF*	*Sum of Squares*	*F Value*	*Pr > F (p-value)*
DEFAULT_AGE_MONTHS	1	0.2438	5.54	0.0190
LANDLINE_PHONE_FLAG	1	0.1830	4.15	0.0421
REASON_PASTDUE	1	1.4432	32.76	<.0001
MONTHS_TO_DEFAULT	1	1.6908	38.38	<.0001
REGION	8	0.7113	2.02	0.0426
SETTLEMENT_TYPE	4	0.5189	2.94	0.0200
CUM_REALWAGE_INDEX_DS	1	0.9748	22.13	<.0001
CUM_CPI_DS	1	0.2262	5.13	0.0239
GDP_GROWTH_INDEX_DS	1	6.2741	142.42	<.0001
AVG_PD	1	0.4073	9.25	0.0025

(Self-made table: own calculation results)

In line with my anticipative expectations this model is really stronger, the adjusted coefficient of determination is 0.5130 (the raw coefficient of determination is 0.5317), which is considerably high even in comparison with the results published in the literature.

Table 66: The variables of the regression developed for the LGD of the "NoFurtherRec" deal category

Parameter	*Estimate (non-standar-dised)*	*Estimate (standar-dised)*	*Standard Error (standar-dised)*	*t Value*	*Pr > |t| (p-value)*
Intercept	-7.9544	0.7094	0.0323	21.96	<.0001
DEFAULT_AGE_MONTHS	-0.00018	-0.0251	0.0106	-2.35	0.0190
LANDLINE_PHONE_FLAG=0	0.0209	0.0209	0.0103	2.04	0.0421
REASON_PASTDUE=0	-0.1702	-0.1702	0.0297	-5.72	<.0001
MONTHS_TO_DEFAULT	-0.0304	-0.2216	0.0358	-6.20	<.0001
REGION=Budapest & environs	-0.0347	-0.0347	0.0388	-0.89	0.3714
REGION=Central-Western	-0.0308	-0.0308	0.0283	-1.09	0.2766
REGION=Eastern	0.0284	0.0284	0.0263	1.08	0.2811
REGION=North-Eastern	0.0488	0.0488	0.0198	2.46	0.0142
REGION=North-Western	-0.0580	-0.0580	0.0305	-1.90	0.0577
REGION=South-Central	0.0361	0.0361	0.0407	0.89	0.3751
REGION=South-Eastern	-0.0347	-0.0347	0.0326	-1.06	0.2878
REGION=South-Western	0.000328	0.000328	0.0422	0.01	0.9938
SETTLEMENT_TYPE=Budapest & environs	0.0253	0.0253	0.0383	0.66	0.5091
SETTLEMENT_TYPE=County town & environs	-0.0721	-0.0721	0.0227	-3.17	0.0016
SETTLEMENT_TYPE=Other city & environs	0.00688	0.00688	0.0180	0.38	0.7034
SETTLEMENT_TYPE=Small village	0.0298	0.0298	0.0234	1.27	0.2030
CUM_REALWAGE_INDEX_DS	-2.7167	-0.0573	0.0122	-4.70	<.0001
CUM_CPI_DS	2.3466	0.0884	0.0390	2.27	0.0239
GDP_GROWTH_INDEX_DS	9.0760	0.1762	0.0148	11.93	<.0001
AVG_PD	3.6371	0.0417	0.0137	3.04	0.0025

(Self-made table: own calculation results)

Table 66 makes known the results of the Maximum Likelihood estimation of the regression developed for the LGD of the *"NoFurtherRec"* deal category: the explanatory variables, the estimated parameters, their standardised values, their standard error as well as the results of the ***t***-tests and the ***p***-values relating to them.

Viewing the above table it is conspicuous that in contrast with the regression developed for the *"WorkoutRec"* deal category the client characteristics also took an important role in this model. Ceteris paribus the age of the client at the default (*default_age_months*) influences the LGD slightly negatively, namely the youngest the client is at the default in case of uniformity of all the other factors, the highest the LGD is. Relatively few researchers investigated this factor in the literature, but for example *Thomas et al. [2007a; 2007b]*, as well as *Bellotti and Crook [2008]* experienced the opposite of my results. However, as regards the LGD influencing role of the dummy which indicates the landline phone (*landline_phone_flag*), I came to the same conclusion as the studies of *Thomas et al. [2007a; 2007b]*: the LGD is ceteris paribus higher for the deals in case of which the client does not possess landline phone. According to my model, in case of uniformity of all the other factors the LGD showed approximately 2.09 percentage points lower in case of the clients who possess landline phone.

The circumstance whether the deal became non-performing because of delay (*reason_pastdue* dummy), as well as the length of the period from the origination of the deal to the default event (*months_to_default*) also proved to be significant factors. Similarly to for example the studies of *Thomas et al. [2007a; 2007b]*, as well as *Bellotti and Crook [2008]* the longest the period has passed since the origination of the deal, the LGD is ceteris paribus the lowest according to my model as well. It can lie in its background that if the client properly met his/her paying obligations for a long time before the default event, he/she will probably make more strenuous efforts for the sake of arranging his/her situation later on as well, than the clients whose discipline to pay is worse in advance.

The region (*region*) and the type of the settlement (*settlement_type)* of the property which serves as collateral proved to be further important factors. It is worth to take notice of the fact that these factors did not appear in any models presented earlier, but the county (*county*) was qualified as a significant factor in all of them. It can be its explanation that in case of this deal category the recovery does not derive from the underlying properties, it is much rather influenced by the client's paying ability (and the paying willingness), thus it has a greater importance for example what kind of opportunities of work the region has in which the client lives. The fact can not be left out of consideration that in many cases the permanent address and the workplace is not in the same settlement, but in the same region, thus the regional characteristics have larger influencing power.

According to the model, ceteris paribus the LGDs of the deals which are secured by properties situated in the North-Western region are the lowest, and the ones from the North-Eastern region are the highest. On the basis of the estimated parameters the divergence between the deal level LGDs is approximately 10.68 percentage points[85] in case of uniformity of all the other factors, namely it is quite considerable. In case of the types of the settlements the difference is in the similar order of magnitude[86]: the base of comparison is formed by the deals secured by village-property, in comparison to this the LGD of the deals secured by properties which are in county town and its environs is ceteris paribus 7.21 percentage points lower, whereas the ones situated in small villages proved to be 2.98 percentage points higher.

The latest group of the explanatory variables is formed by the macroeconomic characteristics. A part of them indicates the changes occurred in the period from the origination of the deal to the default: the growth of the real wage (*cum_realwage_index_ds*), the changing of the consumer prices (*cum_CPI_ds*) and the growth of the GDP (*GDP_growth_index_ds*). According to the model while the growing real wages take effect towards the decreasing the LGD in case of uniformity of all the other factors, the two other factors influence the LGD towards exactly the opposite direction, namely the inflation pressure and the faster economic growth ceteris paribus result in higher LGD. Among the macroeconomic characteristics can be classed the average default rate of the mortgage loans (*avg_PD*) as well, which lies in positive relationship with the LGD in case of this model, similarly to the results published in the literature.

On the basis of reviewing the linear regression models developed for the *"WorkoutEnd"* and the *"NoFurtherRec"* deal categories, it can be said summing up that the results support the statement composed in my 7th Hypothesis according to which different factors influence the LGD values of the deals with different closing types, thus it is inappropriate to handle them together in the course of modelling the deal level LGD.

---

[85] The difference of the coefficients is 0.0488-(-0.0580)=0.1068, namely 10.68 percentage points.
[86] The difference of the coefficients is 0.0298-(-0.0721)=0.1019, namely 10.19 percentage points.

6.4. Summary: the applicability of the results in practice

In the past period the questions in connection with the capital adequacy received high priority for the credit institutions. All the ingravescent economic problems, the increasing risks and the aggravation of the capital adequacy prescriptions have the affect that the capital available for the institutions is tighter and tighter. Under such conditions the proper capital management and portfolio management are essential, thus the exact quantification of the credit risk parameters also has an increasing importance.

Taking into consideration this aspect as well, the notability of prudential definition of the pooling criteria is undoubted, since the divergence of the risk parameters results in different capital requirements withal, thus the credit risk parameters serve as important input factors for the decisions in connection with the portfolio management as well. In case of the portfolio examined by me I experienced significant differences between the LGD values of the certain subportfolios in the course of the categorization according to both the purpose of the loan and the type of the application. Naturally the appropriate pooling criteria can differ from each other institute by institute, moreover they can change in time as well, thus the dynamic approach and the systematic revisions are essential in the course of their use. For that matter the CRD prescribes as well that concerning the statistical models comprehensive supervision has to be made at least annually, which has to include the monitoring of the predictive power, the freeness of distortion and the stability, the review of specifications, the comparison of predicted and real realized results (Back Testing). For the objectivity and exploration of the model's deficiencies the requirement of a review by professional evaluation is a further prescription *(EC [2011c] Article 170; Hkr. 63. §)*.

The result that the applied discount rate has significant LGD-influencing role is important because actually neither the CRD nor the national regulation contains particular prescriptions regarding what kind of method the rate should be defined with. In my opinion the contractual lending rate of the deals can be considered as the most appropriate one, since it reflects both the differences between the actual interest levels at the date of the origination of certain deals, and on the other hand it varies according to their currency as well. However, the empirical results indicate that the definition significantly influences the calculated LGD values.

The appropriate choice of the materiality threshold is generally important, because it promotes the elimination of numerous technical default events, since therefore the delay of "insignificant amounts" does not result automatically in getting into non-performing status. On the other hand according to my empirical researches the decision about the materiality threshold takes notable effect on the calculated value of the risk parameters in case of the mortgage loans as well, so the opportunity composed in *Hkr. 68. § (5)-(7) Paragraph* according to which the credit institutions are allowed to use criteria which are different from the prescriptions has great importance.

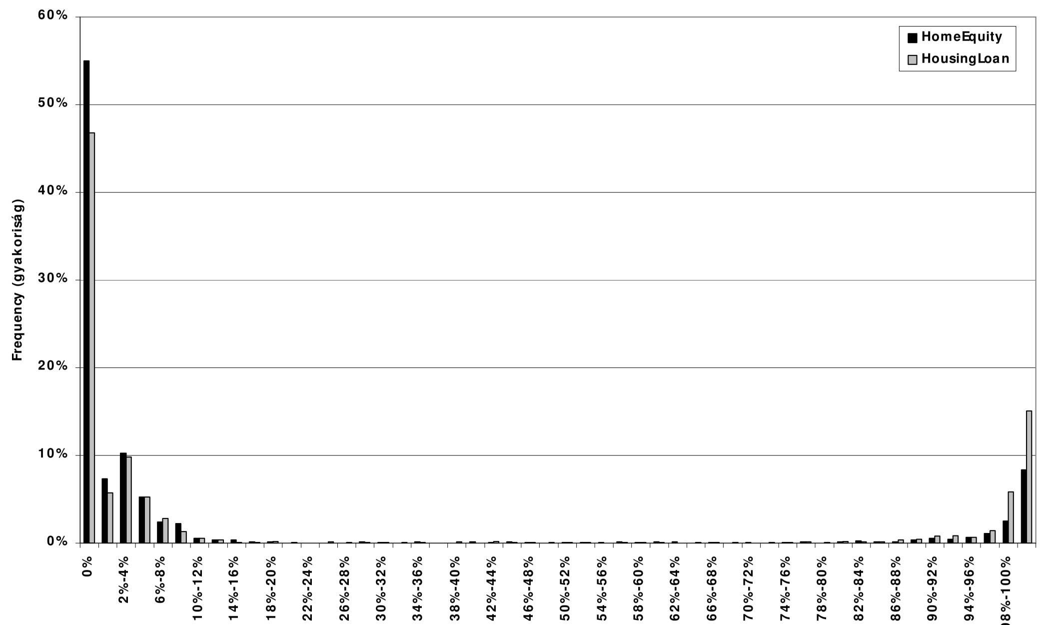
In the course of my researches I did not manage to build regression models having so large explanatory power on the basis of the data of the Hungarian Interbank LGD Database that this way the deals can also be involved in the calculations whose collection process has not been closed yet, but according to my expectations this will also be possible later when the quantity of the deals in the database increases. Considering that the use of the data deriving from the common database can provide advantages for all the credit institutions, it would be expediential that more institutions join it and create a relatively large and variegated database by historic uploading their data, which also enables the consideration of the individual characteristics of their portfolios applying the proper filters.

Keeping it in view I focused in the framework of my dissertation on the analysis of the categories according to the closing type of the deals: I examined the differences between the LGD values of the individual groups, the opportunities of the classification as well as the influencing factors of the LGD values of the certain categories, and all of my results confirmed my anticipative expectations according to which the separate handling of these groups is reasonable. I managed to build a logistic regression with which the futural closing type of the deals with not yet closed collection process can also be predicted with quite great reliability, thus I count great potential to the data of the Hungarian Interbank LGD Database concerning its futural use in the predictions regarding the recovery rates from collateral selling.

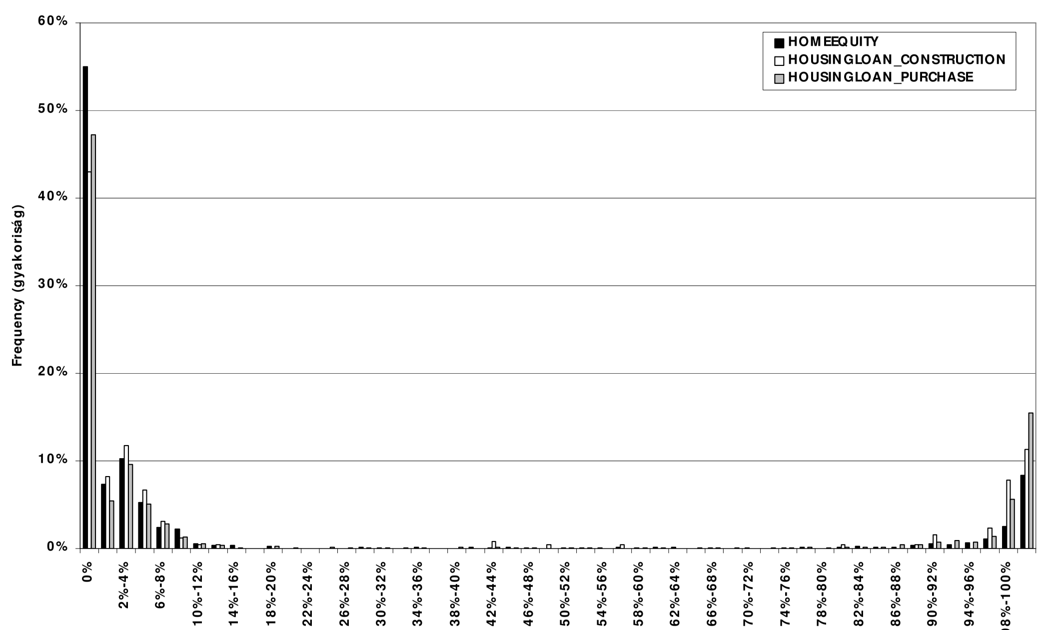
7. Appendices

7.1. The graphs of the distributions

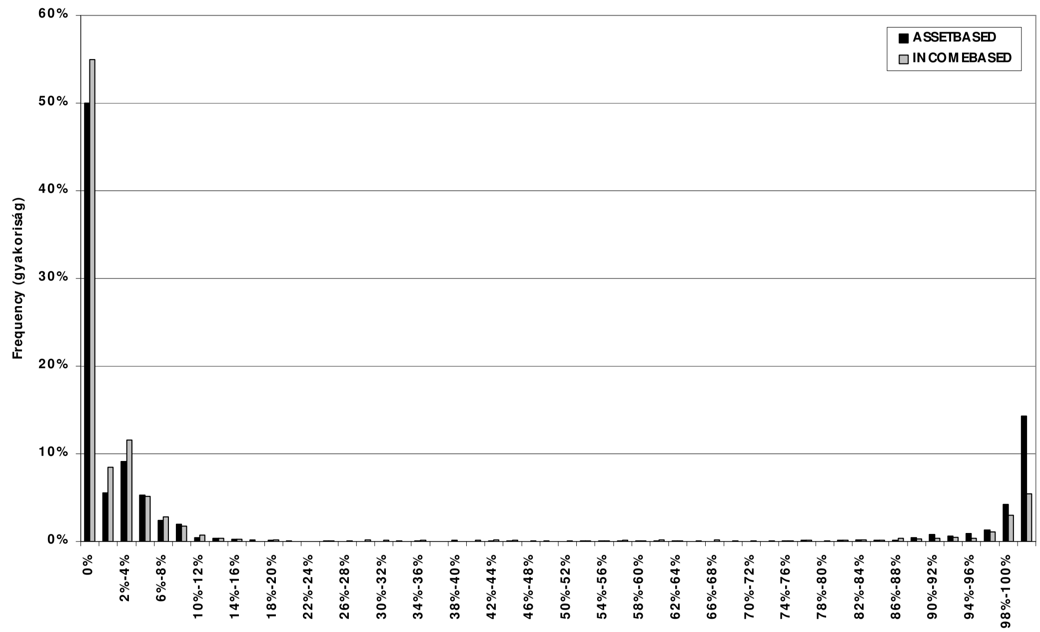
Appendix 1: The LGD distribution of the loans with different purposes



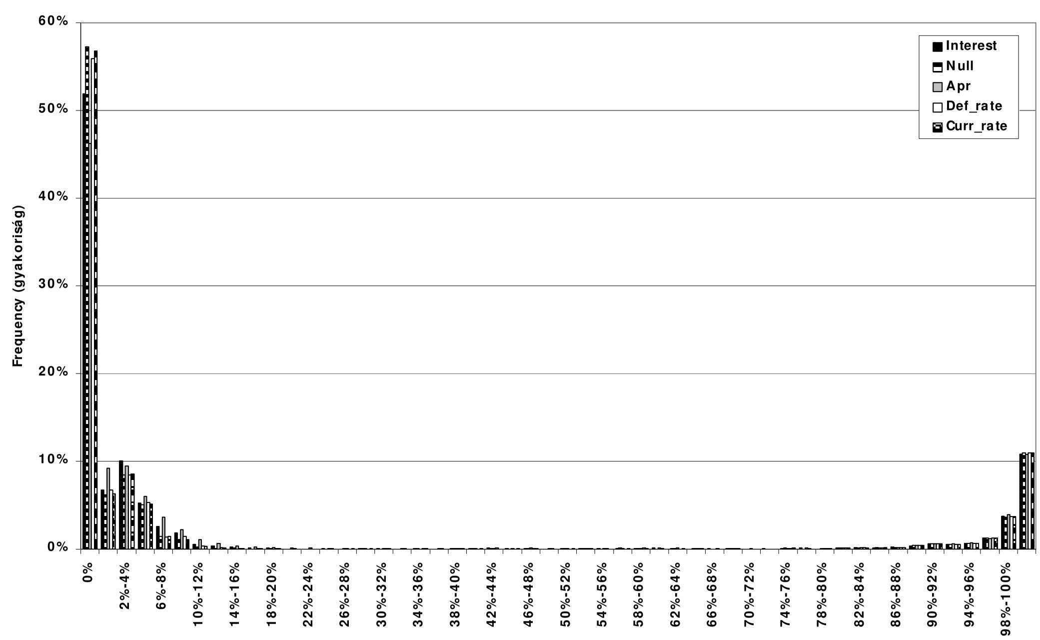
Appendix 2: The LGD distribution of the loans with different purposes (detailed)



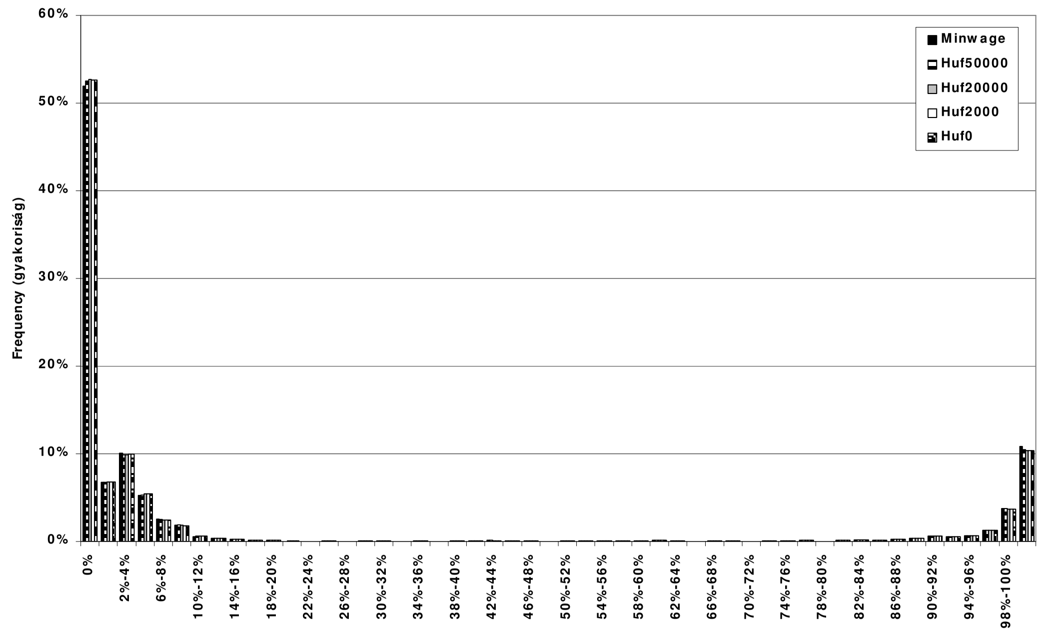
Appendix 3: The LGD distribution of the loans with different types of application



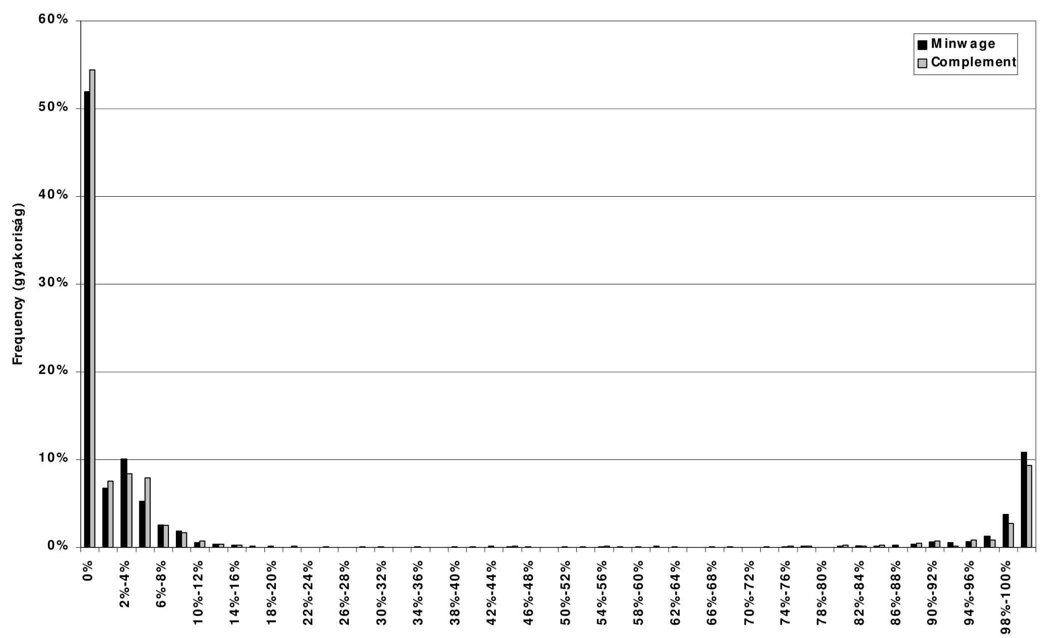
Appendix 4: The distribution of the LGD values calculated with using the different discount rates



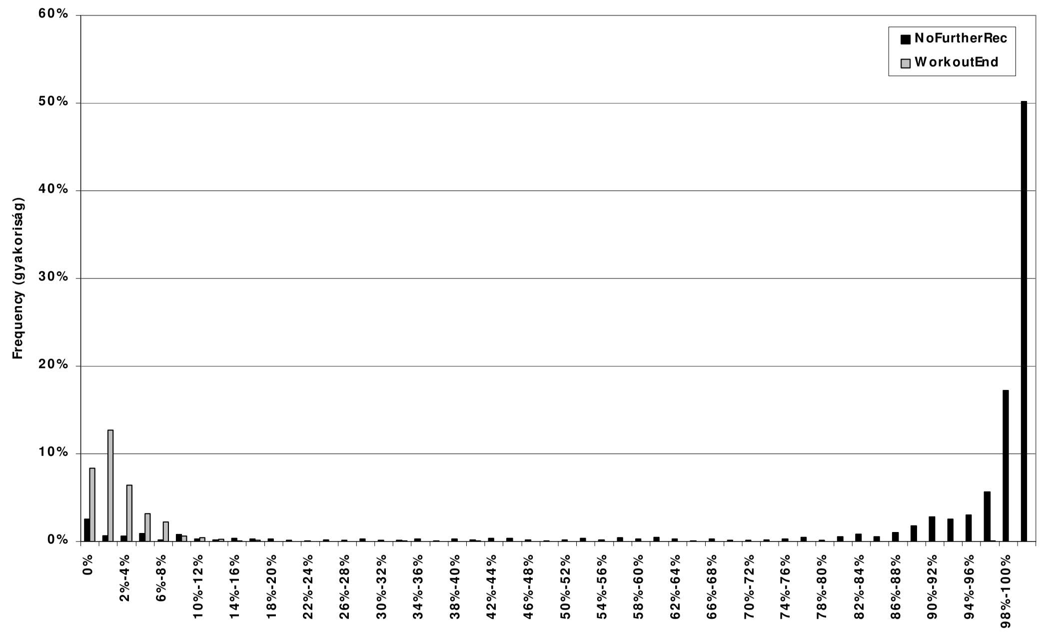
Appendix 5: The distribution of the LGD values in case of using the different materiality thresholds



Appendix 6: The distribution of the LGD values in the basic model and the LGD values of the "technical defaults"



Appendix 7: The distribution of the LGD values of the deals with different closing types



7.2. The variables which serve as a basis for the modelling

Appendix 8: The whole set of variable

NAME OF THE DATA FIELD	CONTENT OF THE DATA FIELD
deal_id	Deal identification number.
product	Type of the product: ▪ Housing Loan: Mortgage loan with home purpose, ▪ Home Equity: Mortgage equity withdrawal.
application_type	Category according to the type of the application: ▪ Asset-based: Purely collateral-based loan, without income verification, ▪ Income-based: Loan based on income verification.
ccy	The original currency of the deal.
deal_status	The closing type of the recovery period of the deal.
start_date	The date of the origination of the deal.
default_date	The date of the default event of the deal.
start_principal_lcy	The disbursed loan amount (in HUF).
defaulted_exposure_lcy	The exposure of the deal at the date of the default event (in HUF).
start_month	The period of the origination of the deal (year, month).
default_month	The period of the default event of the deal (year, month).
woe_month	The period of closing the deal (year, month).
start_term	The duration term of the deal according to the contract (number of months).
loan_purpose	Purpose of the loan: ▪ Real estate purchase, ▪ Real estate construction, ▪ Real estate renovation, ▪ Other.
coapplicant_flag	Dummy variable which indicates whether there is a co-applicant.
first_instalment	The original monthly repayment amount (in HUF).
basic_number	Client identification number.
full_name	The full name of the client.
gender	The gender of the client: ▪ Male, ▪ Female.
citizenship	The citizenship of the client: ▪ Hungarian, ▪ Other.
birth_settlement	The birth place of the client.
start_age_months	The age of the client at the origination of the deal (number of months).
default_age_months	The age of the client at the date of the default event of the deal (number of months).
marital_status	The marital status of the client: ▪ Single, ▪ Married, ▪ Divorced, ▪ Companion, ▪ Widow.
education_level	The education level of the client: ▪ Elementary, ▪ High-school graduate, ▪ Other secondary education, ▪ BSc, ▪ MSc, ▪ Other higher education.
home_settlement	The name of the settlement of the client's living place.

landline_phone_flag	Dummy variable which indicates whether the client has a landline phone.
start_address_months	The duration of living at the given permanent address at the origination of the deal (number of months).
default_address_months	The duration of living at the given permanent address at the date of the default event (number of months).
mobile_phone_flag	Dummy variable which indicates whether the client has mobile phone.
empl_industry	The industry of the client's employer: ▪ Agriculture, ▪ Commerce / Entertainment, ▪ Construction, ▪ Education / Medical services / Government, ▪ Finance / Legal services, ▪ Processing industry, ▪ Other.
empl_type	The type of the client's employment: ▪ Employee, ▪ Entrepreneur, ▪ Retired, ▪ Other.
empl_position	The working position of the client: ▪ Blue collar, ▪ Middle leader, ▪ Upper leader, ▪ Other intellectual.
empl_term	The type of the client's labour contract: ▪ Definite - Full time, ▪ Indefinite - Part time, ▪ Indefinite - Full time.
start_work_months	The duration of working for the given employer at the origination of the deal (number of months).
default_work_months	The duration of working for the given employer at the date of the default event (number of months).
applicant_net_income	The monthly net income of the client.
total_household_income	The total monthly income of the household of the client.
earners_number	The number of earners in the household of the client.
dependents_number	The number of dependents in the household of the client.
existing_ca_flag	Dummy variable which indicates whether the client has a current account.
existing_card_flag	Dummy variable which indicates whether the client has a credit card.
existing_ovd_flag	Dummy variable which indicates whether the client has an overdraft.
existing_loan_flag	Dummy variable which indicates whether the client has another credit.
real_term	The effective duration of the deal (number of months).
defaulted_per_start_exposure	The proportion of the exposure at the default and the disbursed amount.
reason_fraud	Dummy variable which indicates whether the deal is considered as defaulted because of fraud.
reason_death	Dummy variable which indicates whether the deal is considered as defaulted because of death.
reason_pastdue	Dummy variable which indicates whether the deal is considered as defaulted according to the materiality threshold.
months_to_default	The duration from the origination of the deal to the default event (number of months).
woe_months_since_default	The duration from the origination of the deal to the closing (number of months).
start_fx_rate	The exchange rate of the deal's currency at the origination.
default_fx_rate	The exchange rate of the deal's currency at the date of the default.
fx_index_ds	The index of the exchange rate of the currency at the default and the origination (ratio).

start_value_month	The period of defining the collateral value effective at the origination of the deal (year, month).
default_value_month	The period of defining the collateral value effective at the default of the deal (year, month).
priorcharge_amount	The sum of the prior charges on the collateral (in HUF).
priorcharge_rate	The quotient of the sum of the prior charges on the collateral and the realization value at the origination of the deal.
start_collvalue	The realization value of the collateral at the origination of the deal.
default_collvalue	The realization value of the collateral at the default of the deal.
start_marketvalue	The market value of the collateral at the origination of the deal.
default_marketvalue	The market value of the collateral at the default of the deal.
zipcode	The zip code of the real estate which serves as collateral.
settlement	The name of the settlement of the real estate which serves as collateral.
region	The region of the real estate which serves as collateral: ▪ Budapest & environs, ▪ Central-Western, ▪ Eastern, ▪ North-Eastern, ▪ North-Western, ▪ South-Central, ▪ South-Eastern, ▪ South-Western, ▪ Western.
county	The county of the real estate which serves as collateral: ▪ Baranya, ▪ Borsod-Abaúj-Zemplén, ▪ Budapest, ▪ Bács-Kiskun, ▪ Békés, Csongrád, ▪ Fejér, ▪ Győr-Moson-Sopron, ▪ Hajdu-Bihar, ▪ Heves, ▪ Jász-Nagykun-Szolnok, ▪ Komárom-Esztergom, ▪ Nógrád, ▪ Pest, ▪ Somogy, ▪ Szabolcs-Szatmár-Bereg, ▪ Tolna, ▪ Vas, ▪ Veszprém, ▪ Zala.
settlement_type	The type of the settlement of the real estate which serves as collateral: ▪ Budapest & environs, ▪ County town & environs, ▪ Other city & environs, ▪ Village, ▪ Small village.
realestate_type	The type of the real estate which serves as collateral: ▪ Detached house, ▪ Owner-occupied block, ▪ Other residential property.
material	The building type of the real estate which serves as collateral: ▪ Brick or stone, ▪ Prefabricated, ▪ Light construction or wood, ▪ Other.
start_LTV	The proportion of the loan amount and the market value of the collateral at the origination.
current_LTV	The proportion of the exposure at the default and the market value of

	the collateral at the default.
collvalue_index_ds	The index of the realization value of the collateral at the default and the origination (ratio).
marketvalue _index_ds	The index of the market value of the collateral at the default and the origination (ratio).
HomeEquity_PD	Average PD of the mortgage equity withdrawals at the given month.
HousingLoan_PD	Average PD of the home loans at the given month.
avg_PD	Average PD of the mortgage loans at the given month.
start_unempl_rate	Unemployment rate at the origination of the deal.
default_unempl_rate	Unemployment rate at the date of the default.
start_min_wage	The lowest monthly minimum wage at the origination of the deal.
default_min_wage	The lowest monthly minimum wage at the date of the default.
start_avg_netincome	Average monthly net income at the origination of the deal.
default_avg_netincome	Average monthly net income at the date of the default.
start_realwage_index	Yearly real wage index at the origination of the deal.
default_realwage_index	Yearly real wage index at the date of the default.
start_CPI	Yearly consumer price index at the origination of the deal.
default_CPI	Yearly consumer price index at the date of the default.
start_GDP_growth	Yearly GDP-growth index at the origination of the deal.
default_GDP_growth	Yearly GDP-growth index at the date of the default.
unempl_rate_index_ds	The index of the unemployment rate at the default and the origination (ratio).
min_wage_index_ds	The index of the lowest monthly minimum wage at the default and the origination (ratio).
avg_netincome_index_ds	The index of the average monthly net income at the default and the origination (ratio).
cum_realwage_index_ds	The ratio of the real wages at the default and the origination.
cum_CPI_ds	The index of the consumer prices at the default and the origination (the quotient of the cumulative consumer price indices).
GDP_growth_index_ds	The index of the GDP at the default and the origination (the quotient of the cumulative GDP-growth indices).
unempl_rate_index_ds_y	The yearly average growth of the unemployment rate in the period from the origination of the deal to the default.
min_wage_index_ds_y	The yearly average growth of the lowest monthly minimum wage in the period from the origination of the deal to the default.
avg_netincome_index_ds_y	The yearly average growth of the average monthly net income in the period from the origination of the deal to the default.
cum_realwage_index_ds_y	The yearly average growth of the real wages in the period from the origination of the deal to the default.
cum_CPI_ds_y	The yearly average growth of the customer prices in the period from the origination of the deal to the default.
GDP_growth_index_ds_y	The yearly average rate of the GDP-growth in the period from the origination of the deal to the default.
apr	The Annual Percentage Rate of the deal at the origination.
interest	The original lending rate of the deal.
def_rate	The central bank base rate of the original currency of the deal effective at the default of the deal.
curr_rate	The central bank base rate of the original currency of the deal effective on 30th June 2011.
account_lgd_null	The deal level LGD calculated without discounting.
account_lgd_interest	The deal level LGD calculated with discounting with the original lending rate.
account_lgd_apr	The deal level LGD calculated with discounting with the original Annual Percentage Rate.
account_lgd_def_rate	The deal level LGD calculated with discounting with the central bank base rate according to the deal's currency at the default.
account_lgd_curr_rate	The deal level LGD calculated with discounting with the central bank base rate according to the deal's currency on 30th June 2011.

Appendix 9: The set of variables of the 5th Hypothesis

NAME OF THE DATA FIELD	ROLE	SCALE OF MEASUREMENT
APPLICANT_NET_INCOME	input	interval
APR	input	interval
AVG_NETINCOME_INDEX_DS	input	interval
AVG_NETINCOME_INDEX_DS_Y	input	interval
AVG_PD	input	interval
CCY	input	nominal
CITIZENSHIP	input	binary
COAPPLICANT_FLAG	input	binary
COLLVALUE_INDEX_DS	input	interval
COUNTY	input	nominal
CUM_CPI_DS	input	interval
CUM_CPI_DS_Y	input	interval
CUM_REALWAGE_INDEX_DS	input	interval
CUM_REALWAGE_INDEX_DS_Y	input	interval
CURRENT_LTV	input	interval
CURR_RATE	input	interval
DEAL_STATUS	target	binary
DEFAULTED_EXPOSURE_LCY	input	interval
DEFAULTED_PER_START_EXPOSURE	input	interval
DEFAULT_ADDRESS_MONTHS	input	interval
DEFAULT_AGE_MONTHS	input	interval
DEFAULT_AVG_NETINCOME	input	interval
DEFAULT_COLLVALUE	input	interval
DEFAULT_CPI	input	interval
DEFAULT_GDP_GROWTH	input	interval
DEFAULT_MARKETVALUE	input	interval
DEFAULT_MIN_WAGE	input	interval
DEFAULT_REALWAGE_INDEX	input	interval
DEFAULT_UNEMPL_RATE	input	interval
DEFAULT_WORK_MONTHS	input	interval
DEF_RATE	input	interval
DEPENDENTS_NUMBER	input	nominal
EARNERS_NUMBER	input	nominal
EDUCATION_LEVEL	input	nominal
EMPL_INDUSTRY	input	nominal
EMPL_POSITION	input	nominal
EMPL_TERM	input	nominal
EMPL_TYPE	input	nominal
EXISTING_CARD_FLAG	input	binary
EXISTING_CA_FLAG	input	binary
EXISTING_LOAN_FLAG	input	binary
EXISTING_OVD_FLAG	input	binary
FIRST_INSTALMENT	input	interval
FX_INDEX_DS	input	interval
GDP_GROWTH_INDEX_DS	input	interval
GDP_GROWTH_INDEX_DS_Y	input	interval
GENDER	input	binary
INTEREST	input	interval
LANDLINE_PHONE_FLAG	input	binary
APPLICATION_TYPE	input	binary
LOAN_PURPOSE	input	nominal
MARITAL_STATUS	input	nominal
MARKETVALUE_INDEX_DS	input	interval
MATERIAL	input	nominal
MIN_WAGE_INDEX_DS	input	interval

MIN_WAGE_INDEX_DS_Y	input	interval
MOBILE_PHONE_FLAG	input	binary
MONTHS_TO_DEFAULT	input	interval
PRIORCHARGE_AMOUNT	input	interval
PRIORCHARGE_RATE	input	interval
PRODUCT	input	binary
REALESTATE_TYPE	input	nominal
REASON_DEATH	input	binary
REASON_FRAUD	input	binary
REASON_PASTDUE	input	binary
REGION	input	nominal
SETTLEMENT_TYPE	input	nominal
START_ADDRESS_MONTHS	input	interval
START_AGE_MONTHS	input	interval
START_AVG_NETINCOME	input	interval
START_COLLVALUE	input	interval
START_CPI	input	interval
START_GDP_GROWTH	input	interval
START_LTV	input	interval
START_MARKETVALUE	input	interval
START_MIN_WAGE	input	interval
START_PRINCIPAL_LCY	input	interval
START_REALWAGE_INDEX	input	interval
START_TERM	input	interval
START_UNEMPL_RATE	input	interval
START_WORK_MONTHS	input	interval
TOTAL_HOUSEHOLD_INCOME	input	interval
UNEMPL_RATE_INDEX_DS	input	interval
UNEMPL_RATE_INDEX_DS_Y	input	interval

Appendix 10: The set of variables of the 6th Hypothesis

NAME OF THE DATA FIELD	ROLE IN THE REGRESSION MODELLING THE LENGTH OF THE RECOVERY PERIOD	ROLE IN THE REGRESSION MODELLING THE RECOVERY RATE DERIVING FROM THE SELLING	SCALE OF MEASUREMENT
AVG_NETINCOME_INDEX_DS	input	input	interval
AVG_NETINCOME_INDEX_DS_Y	input	input	interval
AVG_PD	input	input	interval
CCY	input	input	nominal
COLLVALUE_INDEX_DS	input	input	interval
COUNTY	input	input	nominal
CUM_CPI_DS	input	input	interval
CUM_CPI_DS_Y	input	input	interval
CUM_REALWAGE_INDEX_DS	input	input	interval
CUM_REALWAGE_INDEX_DS_Y	input	input	interval
CURRENT_LTV	input	input	interval
DEFAULTED_EXPOSURE_LCY	input	input	interval
DEFAULTED_PER_START_EXPOSURE	input	input	interval
DEFAULT_AVG_NETINCOME	input	input	interval
DEFAULT_COLLVALUE	input	input	interval
DEFAULT_CPI	input	input	interval
DEFAULT_GDP_GROWTH	input	input	interval
DEFAULT_MARKETVALUE	input	input	interval
DEFAULT_MIN_WAGE	input	input	interval
DEFAULT_REALWAGE_INDEX	input	input	interval
DEFAULT_UNEMPL_RATE	input	input	interval
DISC_NR_INTEREST	-	target	interval
GDP_GROWTH_INDEX_DS	input	input	interval
GDP_GROWTH_INDEX_DS_Y	input	input	interval
LOAN_PURPOSE	input	input	nominal
MARKETVALUE_INDEX_DS	input	input	interval
MATERIAL	input	input	nominal
MIN_WAGE_INDEX_DS	input	input	interval
MIN_WAGE_INDEX_DS_Y	input	input	interval
MONTHS_TO_CUT_OFF	target	-	interval
MONTHS_TO_DEFAULT	input	input	interval
PRIORCHARGE_AMOUNT	input	input	interval
PRIORCHARGE_RATE	input	input	interval
REALESTATE_TYPE	input	input	nominal
REGION	input	input	nominal
SETTLEMENT_TYPE	input	input	nominal
START_AVG_NETINCOME	input	input	interval
START_COLLVALUE	input	input	interval
START_CPI	input	input	interval
START_GDP_GROWTH	input	input	interval
START_MARKETVALUE	input	input	interval
START_MIN_WAGE	input	input	interval
START_PRINCIPAL_LCY	input	input	interval
START_REALWAGE_INDEX	input	input	interval
START_TERM	input	input	interval
START_UNEMPL_RATE	input	input	interval
UNEMPL_RATE_INDEX_DS	input	input	interval
UNEMPL_RATE_INDEX_DS_Y	input	input	interval

Appendix 11: The set of variables of the 7th Hypothesis

NAME OF THE DATA FIELD	ROLE	SCALE OF MEASUREMENT
ACCOUNT_LGD_INTEREST	target	interval
APPLICANT_NET_INCOME	input	interval
APR	input	interval
AVG_NETINCOME_INDEX_DS	input	interval
AVG_NETINCOME_INDEX_DS_Y	input	interval
AVG_PD	input	interval
CCY	-	nominal
CITIZENSHIP	input	binary
COAPPLICANT_FLAG	input	binary
COLLVALUE_INDEX_DS	input	interval
COUNTY	input	nominal
CUM_CPI_DS	input	interval
CUM_CPI_DS_Y	input	interval
CUM_REALWAGE_INDEX_DS	input	interval
CUM_REALWAGE_INDEX_DS_Y	input	interval
CURRENT_LTV	input	interval
DEFAULTED_EXPOSURE_LCY	input	interval
DEFAULTED_PER_START_EXPOSURE	input	interval
DEFAULT_ADDRESS_MONTHS	input	interval
DEFAULT_AGE_MONTHS	input	interval
DEFAULT_AVG_NETINCOME	input	interval
DEFAULT_COLLVALUE	input	interval
DEFAULT_CPI	input	interval
DEFAULT_GDP_GROWTH	input	interval
DEFAULT_MARKETVALUE	input	interval
DEFAULT_MIN_WAGE	input	interval
DEFAULT_REALWAGE_INDEX	input	interval
DEFAULT_UNEMPL_RATE	input	interval
DEFAULT_WORK_MONTHS	input	interval
DEPENDENTS_NUMBER	input	nominal
EARNERS_NUMBER	input	nominal
EDUCATION_LEVEL	input	nominal
EMPL_INDUSTRY	input	nominal
EMPL_POSITION	input	nominal
EMPL_TERM	input	nominal
EMPL_TYPE	input	nominal
EXISTING_CARD_FLAG	input	binary
EXISTING_CA_FLAG	input	binary
EXISTING_LOAN_FLAG	input	binary
EXISTING_OVD_FLAG	input	binary
FIRST_INSTALMENT	input	interval
FX_INDEX_DS	-	interval
GDP_GROWTH_INDEX_DS	input	interval
GDP_GROWTH_INDEX_DS_Y	input	interval
GENDER	input	binary
INTEREST	input	interval
LANDLINE_PHONE_FLAG	input	binary
APPLICATION_TYPE	input	binary
LOAN_PURPOSE	input	nominal
MARITAL_STATUS	input	nominal
MARKETVALUE_INDEX_DS	input	interval
MATERIAL	input	nominal
MIN_WAGE_INDEX_DS	input	interval
MIN_WAGE_INDEX_DS_Y	input	interval
MOBILE_PHONE_FLAG	input	binary

MONTHS_TO_DEFAULT	input	interval
PRIORCHARGE_AMOUNT	input	interval
PRIORCHARGE_RATE	input	interval
PRODUCT	input	binary
REALESTATE_TYPE	input	nominal
REASON_DEATH	input	binary
REASON_FRAUD	input	binary
REASON_PASTDUE	input	binary
REGION	input	nominal
SETTLEMENT_TYPE	input	nominal
START_ADDRESS_MONTHS	input	interval
START_AGE_MONTHS	input	interval
START_AVG_NETINCOME	input	interval
START_COLLVALUE	input	interval
START_CPI	input	interval
START_GDP_GROWTH	input	interval
START_LTV	input	interval
START_MARKETVALUE	input	interval
START_MIN_WAGE	input	interval
START_PRINCIPAL_LCY	input	interval
START_REALWAGE_INDEX	input	interval
START_TERM	input	interval
START_UNEMPL_RATE	input	interval
START_WORK_MONTHS	input	interval
TOTAL_HOUSEHOLD_INCOME	input	interval
UNEMPL_RATE_INDEX_DS	input	interval
UNEMPL_RATE_INDEX_DS_Y	input	interval

7.3. The logistic regression with logit link

Appendix 12: Logit Link: Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood (Intercept Only)	-2 Log Likelihood (Intercept & Covariates)	Likelihood Ratio χ^2	DF	Pr > χ^2
2941.451	1783.345	1158.1052	11	<.0001

Appendix 13: Logit Link: Analysis of Effects

Effect	DF	Wald χ^2	Pr > χ^2
CUM_REALWAGE_INDEX_DS_Y	1	57.8225	<.0001
DEFAULT_REALWAGE_INDEX	1	15.5777	<.0001
GDP_GROWTH_INDEX_DS_Y	1	200.9459	<.0001
MONTHS_TO_DEFAULT	1	113.7396	<.0001
REASON_DEATH	1	39.8771	<.0001
REASON_PASTDUE	1	15.9568	<.0001
SETTLEMENT_TYPE	4	28.2323	<.0001
START_LTV	1	28.5511	<.0001

Appendix 14: Logit Link: Analysis of ML-estimates (non-standardised)

Parameter	Estimate	Standard Error	Wald χ^2	Pr > χ^2	Exp (Estimate)
Intercept	42.0667	4.6443	82.04	<.0001	999.000
CUM_REALWAGE_INDEX_DS_Y	31.6436	4.1614	57.82	<.0001	999.000
DEFAULT_REALWAGE_INDEX	-6.7027	1.6982	15.58	<.0001	0.001
GDP_GROWTH_INDEX_DS_Y	-65.4953	4.6203	200.95	<.0001	0.000
MONTHS_TO_DEFAULT	0.0750	0.00703	113.74	<.0001	1.078
REASON_DEATH = 0	1.0071	0.1595	39.88	<.0001	2.738
REASON_PASTDUE = 0	0.9365	0.2344	15.96	<.0001	2.551
SETTLEMENT_TYPE = Budapest & environs	0.3630	0.1383	6.89	0.0087	1.438
SETTLEMENT_TYPE = County town & environs	0.4630	0.1228	14.22	0.0002	1.589
SETTLEMENT_TYPE = Other city & environs	-0.0689	0.1024	0.45	0.5010	0.933
SETTLEMENT_TYPE = Small village	-0.4707	0.1395	11.38	0.0007	0.625
START_LTV	-1.9218	0.3597	28.55	<.0001	0.146

Appendix 15: Logit Link: Analysis of ML-estimates (standardised)

Parameter	Estimate	Standard Error	Wald χ^2	Pr > χ^2	Exp (Estimate)
Intercept	2.4906	0.1940	164.76	<.0001	12.069
REASON_DEATH = 0	1.0071	0.1595	39.88	<.0001	2.738
REASON_PASTDUE = 0	0.9365	0.2344	15.96	<.0001	2.551
SETTLEMENT_TYPE = Budapest & environs	0.3630	0.1383	6.89	0.0087	1.438
SETTLEMENT_TYPE = County town & environs	0.4630	0.1228	14.22	0.0002	1.589
SETTLEMENT_TYPE = Other city & environs	-0.0689	0.1024	0.45	0.5010	0.933
SETTLEMENT_TYPE = Small village	-0.4707	0.1395	11.38	0.0007	0.625
STD_CUM_REALWAGE_INDEX_DS_Y	0.4041	0.0531	57.82	<.0001	1.498
STD_DEFAULT_REALWAGE_INDEX	-0.2433	0.0616	15.58	<.0001	0.784
STD_GDP_GROWTH_INDEX_DS_Y	-1.3934	0.0983	200.95	<.0001	0.248
STD_MONTHS_TO_DEFAULT	1.1343	0.1064	113.74	<.0001	3.109
STD_START_LTV	-0.3588	0.0671	28.55	<.0001	0.699

Appendix 16: Logit Link: Odds Ratio Estimates

Effect	*Point Estimate (non-standardised)*	*Point Estimate (standardised)*
CUM_REALWAGE_INDEX_DS_Y	999.000	1.498
DEFAULT_REALWAGE_INDEX	0.001	0.784
GDP_GROWTH_INDEX_DS_Y	<0.001	0.248
MONTHS_TO_DEFAULT	1.078	3.109
REASON_DEATH (0 vs 1)	7.495	7.495
REASON_PASTDUE (0 vs 1)	6.508	6.508
SETTLEMENT_TYPE (Budapest & environs vs Village)	1.914	1.914
SETTLEMENT_TYPE (County town & environs vs Village)	2.116	2.116
SETTLEMENT_TYPE (Other city & environs vs Village)	1.243	1.243
SETTLEMENT_TYPE (Small village vs Village)	0.832	0.832
START_LTV	0.146	0.699

Appendix 17: Logit Link: Fit Statistics

Fit statistics	*Statistics Label*	*Train*
AIC	Akaike's Information Criterion	4274.34
ASE	Average Squared Error	0.10
AVERR	Average Error Function	0.31
DFE	Degrees of Freedom for Error	6824.00
DFM	Model Degrees of Freedom	12.00
DFT	Total Degrees of Freedom	6836.00
DIV	Divisor for ASE	13762.00
ERR	Error Function	4250.34
FPE	Final Prediction Error	0.10
MAX	Maximum Absolute Error	1.00
MSE	Mean Square Error	0.10
NOBS	Sum of Frequencies	6836.00
_NW_	Number of Estimate Weights	12.00
RASE	Root Average Sum of Squares	0.32
RFPE	Root Final Prediction Error	0.32
RMSE	Root Mean Squared Error	0.32
SBC	Schwarz's Bayesian Criterion	4356.30
SSE	Sum of Squared Errors	1369.77
SUMW	Sum of Case Weights Times Freq	13692.00
MISC	Misclassification Rate	0.16

Appendix 18: Logit Link: Classification Table

Target	*Outcome*	*Target Percentage*	*Outcome Percentage*	*Count*	*Total Percentage*
NoFurtherRec	NoFurtherRec	68.1004	51.7359	760	11.1176
WorkoutEnd	NoFurtherRec	31.8996	6.6331	256	5.2077
NoFurtherRec	WorkoutEnd	12.3951	48.2641	709	10.3716
WorkoutEnd	WorkoutEnd	87.6049	93.3669	5011	73.3031

Appendix 19: Logit Link: Event Classification Table

Target	*False Negative*	*True Negative*	*False Positive*	*True Positive*
DEAL_STATUS	356	760	709	5011

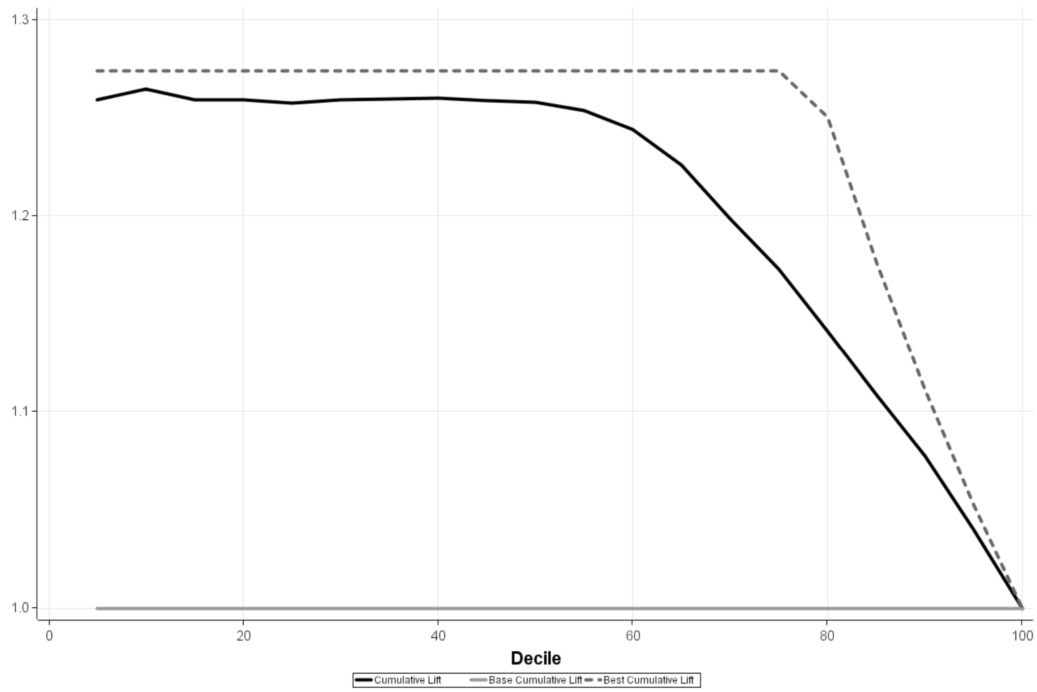
Appendix 20: Logit Link: Assessment Score Rankings

Decile	*Gain*	*Lift*	*Cumulative Lift*	*% Response*	*Cumulative % Response*	*Observation Number*	*Posterior Probability Mean*
0	.	.	.	.	.	.	.
5	25.8804	1.25880	1.25880	98.8297	98.8297	341.8	0.99668
10	26.4394	1.26998	1.26439	99.7074	99.2686	341.8	0.99451
15	25.8804	1.24762	1.25880	97.9520	98.8297	341.8	0.99282
20	25.8804	1.25880	1.25880	98.8297	98.8297	341.8	0.99112
25	25.7313	1.25135	1.25731	98.2446	98.7127	341.8	0.98930
30	25.8804	1.26626	1.25880	99.4149	98.8297	341.8	0.98732
35	25.9336	1.26253	1.25934	99.1223	98.8715	341.8	0.98517
40	25.9735	1.26253	1.25974	99.1223	98.9029	341.8	0.98236
45	25.8390	1.24762	1.25839	97.9520	98.7972	341.8	0.97811
50	25.7686	1.25135	1.25769	98.2446	98.7420	341.8	0.96946
55	25.3383	1.21036	1.25338	95.0263	98.4042	341.8	0.94304
60	24.3587	1.13583	1.24359	89.1750	97.6351	341.8	0.87662
65	22.5552	1.00913	1.22555	79.2276	96.2191	341.8	0.79416
70	19.8382	0.84516	1.19838	66.3546	64.0859	341.8	0.71438
75	17.2598	0.81163	1.17260	63.7215	92.0616	341.8	0.64857
80	14.0768	0.66331	1.14077	52.0772	89.5626	341.8	0.58344
85	10.8398	0.58878	1.10830	46.2259	87.0134	341.8	0.51625
90	7.7242	0.54928	1.07724	43.1246	84.5751	341.8	0.44431
95	3.9883	0.36743	1.03988	28.8473	81.6421	341.8	0.36608
100	0.0000	0.24222	1.00000	19.0170	78.5108	341.8	0.23192

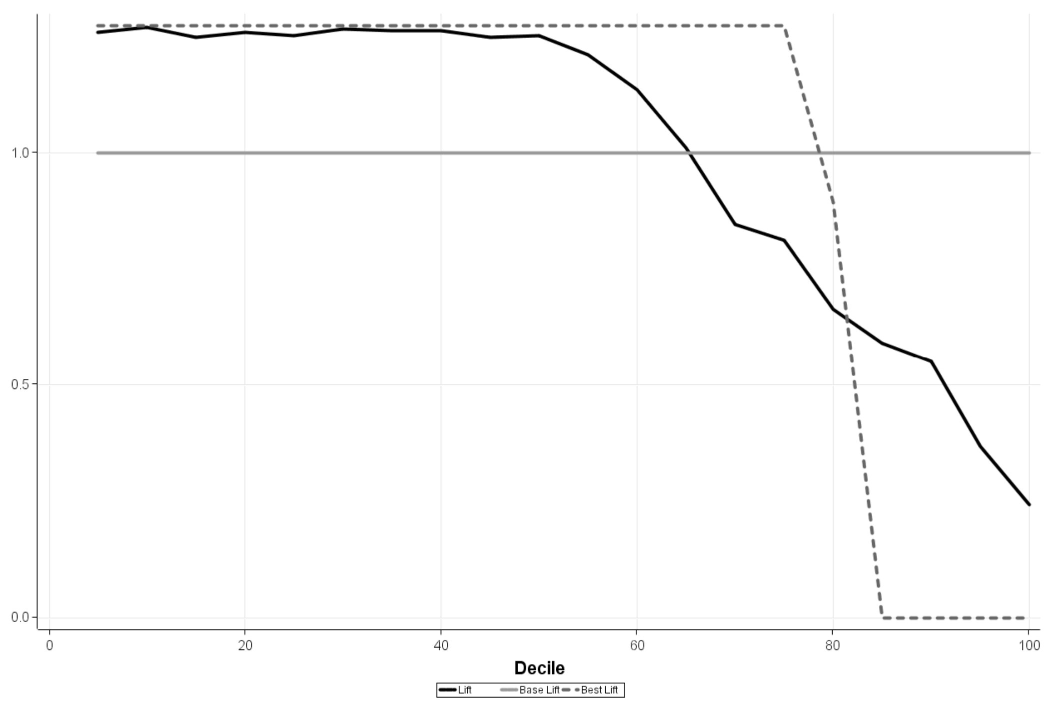
Appendix 21: Logit Link: Assessment Score Distribution

Posterior Probability Range	*Number of Events*	*Number of Nonevents*	*Posterior Probability Mean*	*Percentage*
0.95 – 1.00	3506	45	0.98554	51.9456
0.90 – 0.95	260	20	0.92765	4.0960
0.85 – 0.90	193	22	0.87524	3.1451
0.80 – 0.85	175	32	0.82375	3.0281
0.75 – 0.80	145	51	0.77555	2.8672
0.70 – 0.75	163	73	0.72440	3.4523
0.65 – 0.70	165	98	0.67384	3.8473
0.60 – 0.65	156	105	0.62592	3.8180
0.55 – 0.60	133	120	0.57624	0.7010
0.50 – 0.55	115	143	0.52621	3.7741
0.45 – 0.50	102	126	0.47552	3.3353
0.40 – 0.45	103	137	0.42589	3.5108
0.35 – 0.40	66	135	0.37674	2.9403
0.30 – 0.35	31	122	0.32578	2.2382
0.25 – 0.30	24	101	0.27548	1.8286
0.20 – 0.25	12	61	0.23130	1.0679
0.15 – 0.20	10	40	0.17610	0.7314
0.10 – 0.15	7	14	0.12892	0.3072
0.05 – 0.10	0	18	0.07530	0.2633
0.00 – 0.05	1	6	0.04272	0.1024

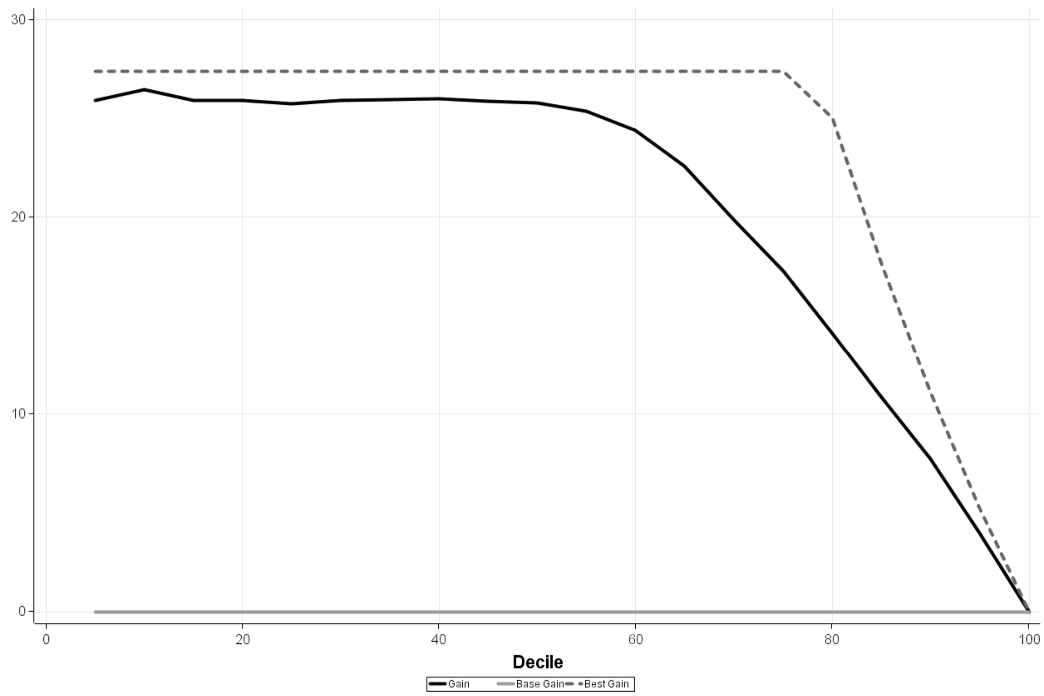
Appendix 22: Logit Link: Cumulative Lift



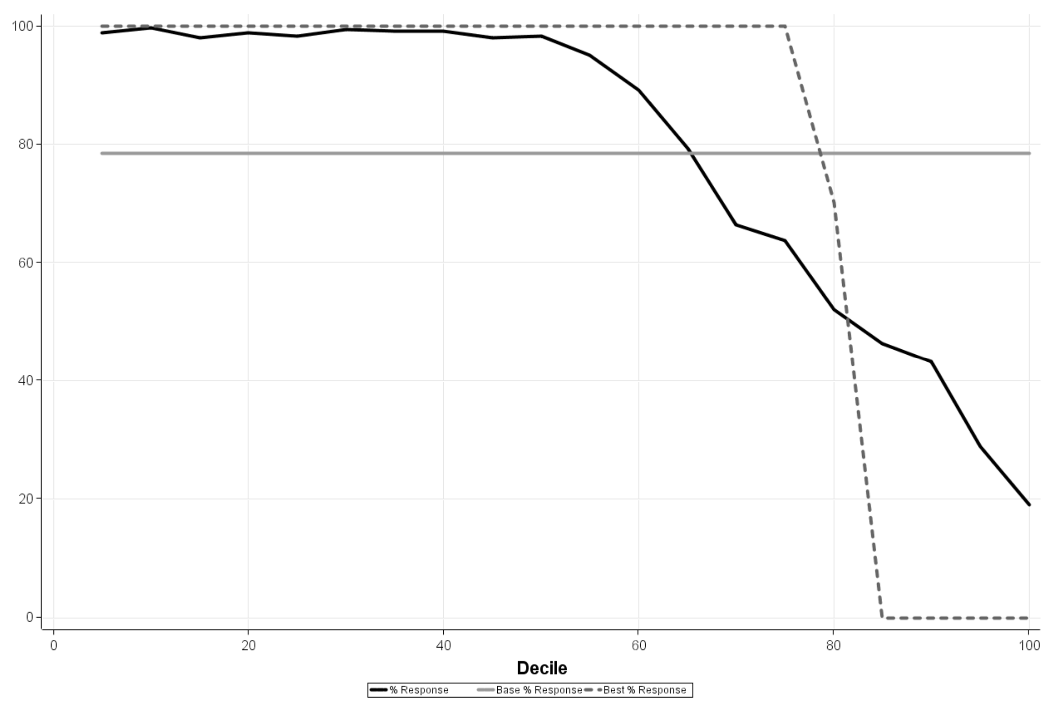
Appendix 23: Logit Link: Lift



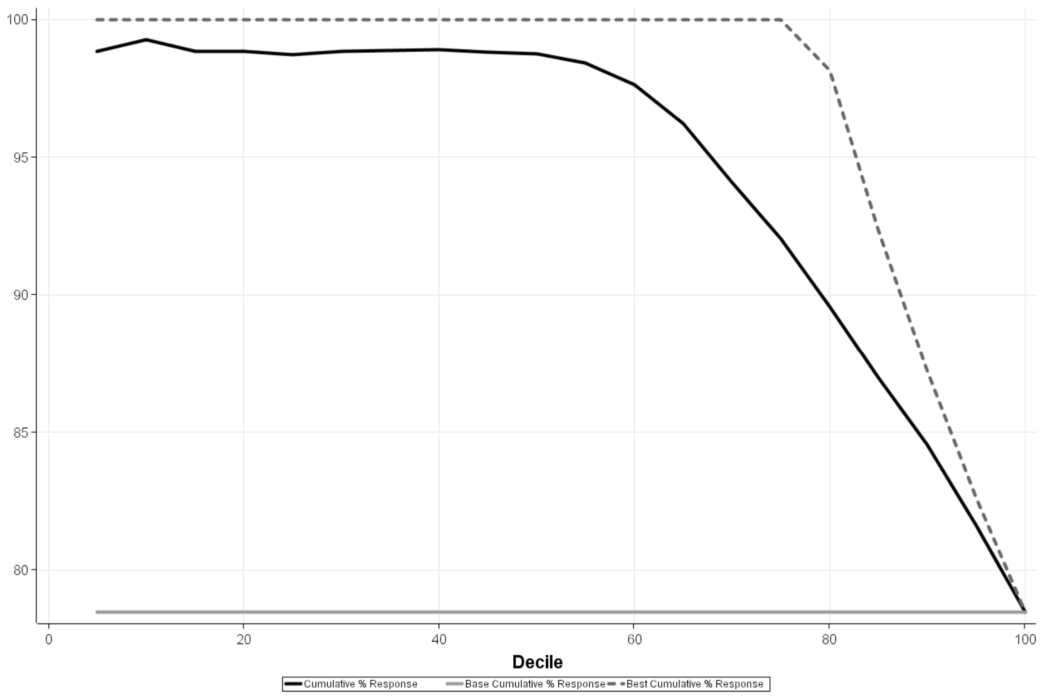
Appendix 24: Logit Link: Gain



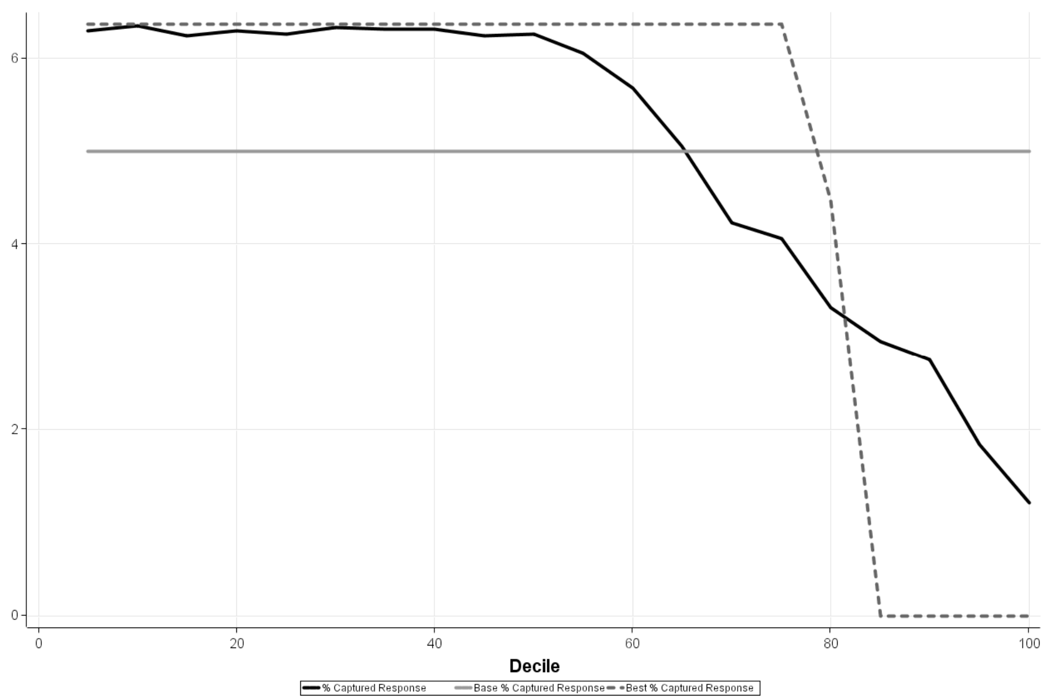
Appendix 25: Logit Link: % Response



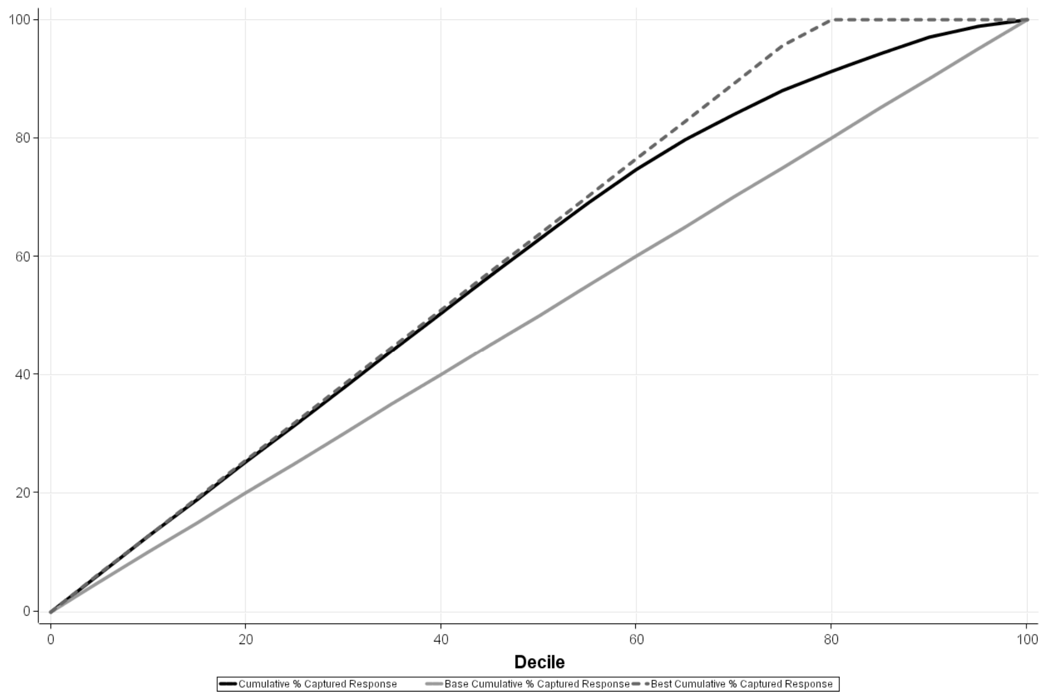
Appendix 26: Logit Link: Cumulative % Response



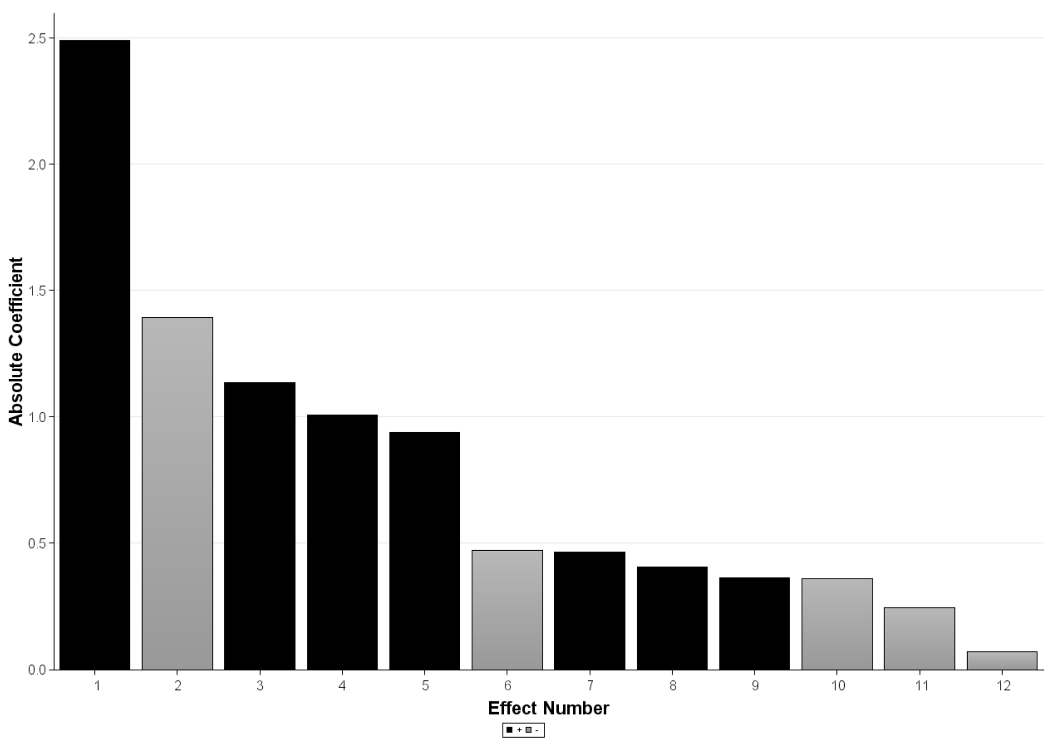
Appendix 27: Logit Link: % Captured Response



Appendix 28: Logit Link: Cumulative % Captured Response



Appendix 29: Logit Link: Effects Plot (standardised)



7.4. The logistic regression with probit link

Appendix 30: Probit Link: Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood (Intercept Only)	*-2 Log Likelihood (Intercept & Covariates)*	*Likelihood Ratio χ^2*	*DF*	*Pr > χ^2*
2941.451	1745.971	1195.4792	26	<.0001

Appendix 31: Probit Link: Analysis of Effects

Effect	*DF*	*Wald χ^2*	*Pr > χ^2*
CUM_CPI_DS_Y	1	20.2583	<.0001
DEFAULT_CPI	1	8.9536	0.0028
EMPL_INDUSTRY	6	25.5083	0.0003
FIRST_INSTALMENT	1	12.1104	0.0005
GDP_GROWTH_INDEX_DS_Y	1	163.0441	<.0001
MONTHS_TO_DEFAULT	1	135.3249	<.0001
PRODUCT	1	14.4498	0.0001
REASON_DEATH	1	47.4718	<.0001
REASON_PASTDUE	1	16.1598	<.0001
REGION	8	26.6244	0.0008
SETTLEMENT_TYPE	4	21.8685	0.0002

Appendix 32: Probit Link: Analysis of ML-estimates (non-standardised)

Parameter	*Estimate*	*Standard Error*	*Wald χ^2*	*Pr > χ^2*
Intercept	48.5854	3.8321	160.74	<.0001
CUM_CPI_DS_Y	-14.1127	3.1355	20.26	<.0001
DEFAULT_CPI	-8.6062	2.8762	8.95	0.0028
EMPL_INDUSTRY=Agriculture	0.0819	0.1456	0.32	0.5736
EMPL_INDUSTRY=Commerce, Entertainment	-0.1589	0.0726	4.79	0.0295
EMPL_INDUSTRY=Construction	-0.1932	0.0968	3.98	0.0461
EMPL_INDUSTRY=Education, Medical services, Gov	0.4436	0.1071	17.16	<.0001
EMPL_INDUSTRY=Finance, Legal services	-0.1503	0.1685	0.80	0.3725
EMPL_INDUSTRY=Other	-0.0713	0.0891	0.64	0.4233
FIRST_INSTALMENT	-4.31E-6	1.24E-6	12.11	0.0005
GDP_GROWTH_INDEX_DS_Y	-24.2347	1.8980	163.04	<.0001
MONTHS_TO_DEFAULT	0.0411	0.00353	135.32	<.0001
PRODUCT=Home Equity	0.1363	0.0359	14.45	0.0001
REASON_DEATH=0	0.5971	0.0867	47.47	<.0001
REASON_PASTDUE=0	0.5073	0.1262	16.16	<.0001
REGION=Budapest & environs	0.2233	0.1446	2.39	0.1224
REGION=Central-Western	0.0124	0.0997	0.02	0.9012
REGION=Eastern	-0.4134	0.1049	15.52	<.0001
REGION=North-Eastern	-0.2266	0.0772	8.63	0.0033
REGION=North-Western	0.0202	0.1079	0.04	0.8516
REGION=South-Central	0.1153	0.1394	0.68	0.4082
REGION=South-Eastern	0.1327	0.1213	1.20	0.2740
REGION=South-Western	0.1097	0.1382	0.63	0.4272
SETTLEMENT_TYPE=Budapest & environs	0.0364	0.1388	0.07	0.7934
SETTLEMENT_TYPE=County town & environs	0.3116	0.0816	14.58	0.0001
SETTLEMENT_TYPE=Other city & environs	-0.0203	0.0683	0.09	0.7667
SETTLEMENT_TYPE=Small village	-0.1990	0.0871	5.22	0.0224

Appendix 33: Probit Link: Analysis of ML-estimates (standardised)

Parameter	*Estimate*	*Standard Error*	*Wald χ^2*	*Pr > χ^2*
Intercept	1.2822	0.1098	136.31	<.0001
EMPL_INDUSTRY=Agriculture	0.0819	0.1456	0.32	0.5736
EMPL_INDUSTRY=Commerce, Entertainment	-0.1589	0.0726	4.79	0.0295
EMPL_INDUSTRY=Construction	-0.1932	0.0968	3.98	0.0461
EMPL_INDUSTRY=Education, Medical services, Gov	0.4436	0.1071	17.16	<.0001
EMPL_INDUSTRY=Finance, Legal services	-0.1503	0.1685	0.80	0.3725
EMPL_INDUSTRY=Other	-0.0713	0.0891	0.64	0.4233
PRODUCT=Home Equity	0.1363	0.0359	14.45	0.0001
REASON_DEATH=0	0.5971	0.0867	47.47	<.0001
REASON_PASTDUE=0	0.5073	0.1262	16.16	<.0001
REGION=Budapest & environs	0.2233	0.1446	2.39	0.1224
REGION=Central-Western	0.0124	0.0997	0.02	0.9012
REGION=Eastern	-0.4134	0.1049	15.52	<.0001
REGION=North-Eastern	-0.2266	0.0772	8.63	0.0033
REGION=North-Western	0.0202	0.1079	0.04	0.8516
REGION=South-Central	0.1153	0.1394	0.68	0.4082
REGION=South-Eastern	0.1327	0.1213	1.20	0.2740
REGION=South-Western	0.1097	0.1382	0.63	0.4272
SETTLEMENT_TYPE=Budapest & environs	0.0364	0.1388	0.07	0.7934
SETTLEMENT_TYPE=County town & environs	0.3116	0.0816	14.58	0.0001
SETTLEMENT_TYPE=Other city & environs	-0.0203	0.0683	0.09	0.7667
SETTLEMENT_TYPE=Small village	-0.1990	0.0871	5.22	0.0224
STD_CUM_CPI_DS_Y	-0.1572	0.0349	20.26	<.0001
STD_DEFAULT_CPI	-0.1328	0.0444	8.95	0.0028
STD_FIRST_INSTALMENT	-0.1215	0.0349	12.11	0.0005
STD_GDP_GROWTH_INDEX_DS_Y	-0.5156	0.0404	163.04	<.0001
STD_MONTHS_TO_DEFAULT	0.6218	0.0535	135.32	<.0001

Appendix 34: Probit Link: Fit Statistics

Fit statistics	*Statistics Label*	*Train*
AIC	Akaike's Information Criterion	4325.86
ASE	Average Squared Error	0.10
AVERR	Average Error Function	0.31
DFE	Degrees of Freedom for Error	6809.00
DFM	Model Degrees of Freedom	27.00
DFT	Total Degrees of Freedom	6836.00
DIV	Divisor for ASE	13672.00
ERR	Error Function	4271.86
FPE	Final Prediction Error	0.10
MAX	Maximum Absolute Error	1.00
MSE	Mean Square Error	0.10
NOBS	Sum of Frequencies	6836.00
_NW_	Number of Estimate Weights	27.00
RASE	Root Average Sum of Squares	0.32
RFPE	Root Final Prediction Error	0.32
RMSE	Root Mean Squared Error	0.32
SBC	Schwarz's Bayesian Criterion	4510.27
SSE	Sum of Squared Errors	1384.81
SUMW	Sum of Case Weights Times Freq	13672.00
MISC	Misclassification Rate	0.15

Appendix 35: Probit Link: Classification Table

Target	*Outcome*	*Target Percentage*	*Outcome Percentage*	*Count*	*Total Percentage*
NoFurtherRec	NoFurtherRec	68.4870	52.0762	765	11.1908
WorkoutEnd	NoFurtherRec	31.5130	6.5586	352	5.1492
NoFurtherRec	WorkoutEnd	12.3098	47.9238	704	10.2984
WorkoutEnd	WorkoutEnd	87.6902	93.4414	5015	73.3616

Appendix 36: Probit Link: Event Classification Table

Target	False Negative	True Negative	False Positive	True Positive
DEAL_STATUS	352	465	704	5015

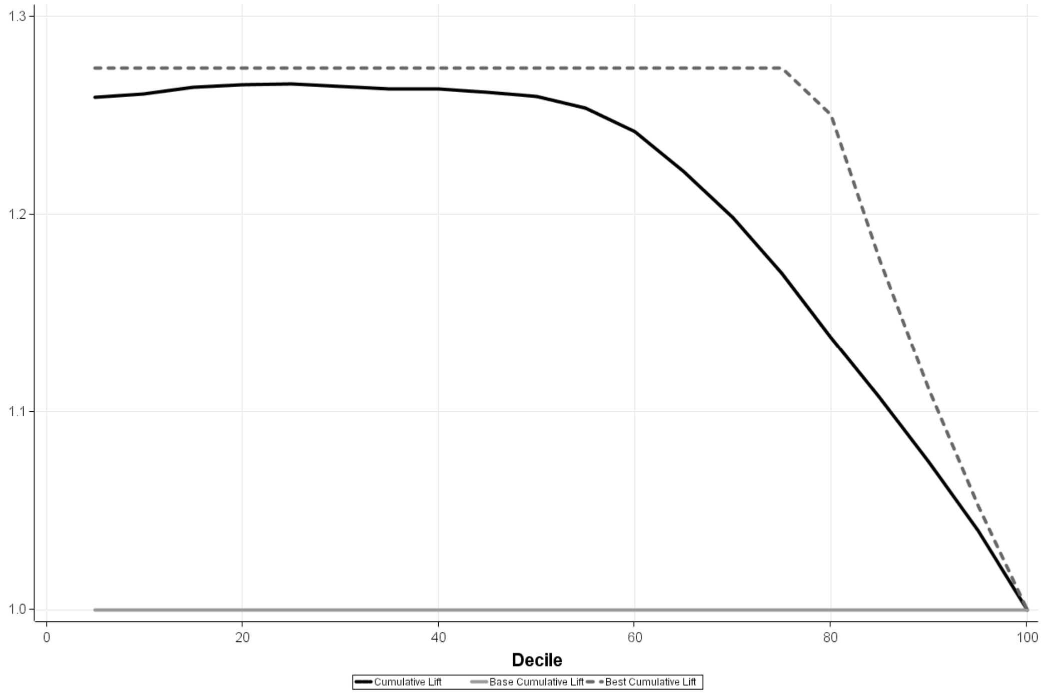
Appendix 37: Probit Link: Assessment Score Rankings

Decile	Gain	Lift	Cumulative Lift	% Response	Cumulative % Response	Observation Number	Posterior Probability Mean
0	.	.	.	.	.	.	.
5	25.8804	1.25880	1.25880	98.8297	98.8297	341.8	0.99928
10	26.0667	1.26253	1.26067	99.1223	98.9760	341.8	0.99793
15	26.3772	1.26998	1.26377	99.7074	99.2198	341.8	0.99658
20	26.5325	1.26998	1.26533	99.7074	99.3417	341.8	0.99506
25	26.5511	1.26626	1.26551	99.4149	99.3563	341.8	0.99322
30	26.4394	1.25880	1.26439	98.8297	99.2686	341.8	0.99071
35	26.3063	1.25508	1.26306	98.5372	99.1641	341.8	0.98767
40	26.2996	1.26253	1.26300	99.1223	99.1589	341.8	0.98396
45	26.1288	1.24762	1.26129	97.9520	99.0248	341.8	0.97815
50	25.9549	1.24390	1.25955	97.6594	98.8882	341.8	0.96646
55	25.3383	1.19173	1.25338	93.5635	98.4042	341.8	0.93301
60	24.1724	1.11347	1.24172	87.4195	97.4888	341.8	0.86006
65	22.1424	0.97783	1.22142	76.7700	95.8950	341.8	0.77967
70	19.8115	0.89510	1.19812	70.2750	94.0650	341.8	0.71236
75	17.0114	0.77809	1.17011	61.0884	91.8666	341.8	0.64792
80	13.7460	0.64766	1.13746	50.8484	89.3030	341.8	0.58190
85	10.6895	0.61785	1.10690	48.5079	86.9032	341.8	0.51567
90	7.4467	0.52320	1.07447	41.0767	84.3573	341.8	0.44268
95	3.9883	0.41737	1.03988	32.7677	81.6421	341.8	0.35975
100	0.0000	0.24222	1.00000	19.0170	78.5108	341.8	0.21592

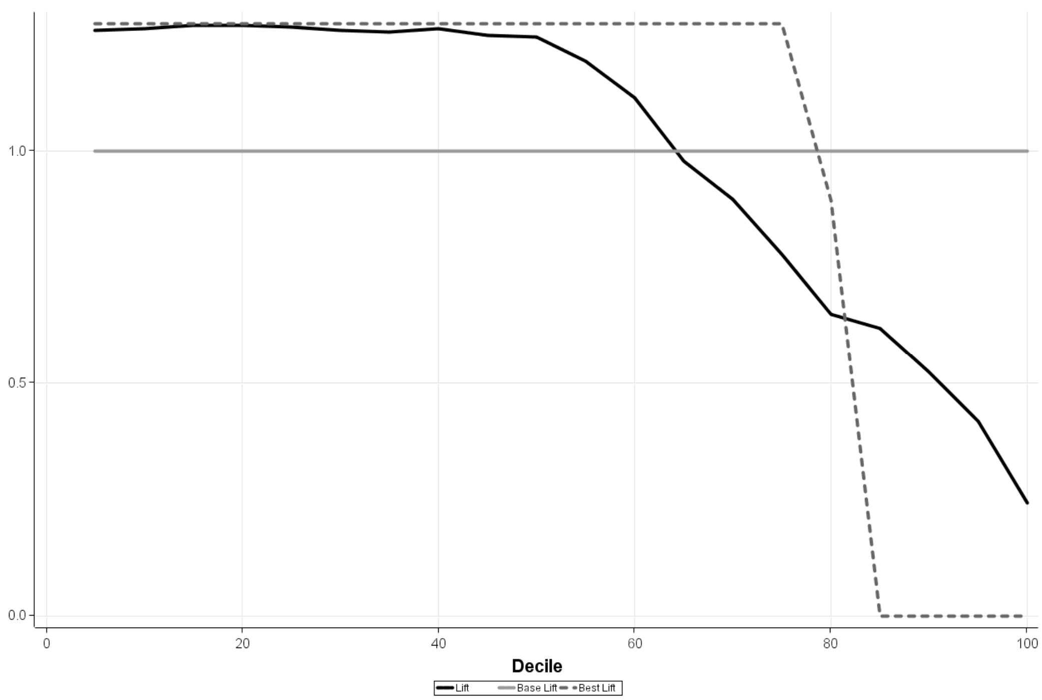
Appendix 38: Probit Link: Assessment Score Distribution

Posterior Probability Range	Number of Events	Number of Nonevents	Posterior Probability Mean	Percentage
0.95 – 1.00	3441	42	0.98823	50.9508
0.90 – 0.95	266	18	0.92762	4.1545
0.85 – 0.90	183	27	0.87442	3.0720
0.80 – 0.85	174	25	0.82408	2.9111
0.75 – 0.80	183	66	0.77341	3.6425
0.70 – 0.75	179	72	0.72539	3.6717
0.65 – 0.70	185	91	0.67328	4.0374
0.60 – 0.65	145	109	0.62528	3.7156
0.55 – 0.60	122	130	0.57524	3.6864
0.50 – 0.55	137	124	0.52569	3.8180
0.45 – 0.50	89	138	0.47548	3.3207
0.40 – 0.45	89	136	0.42586	3.2914
0.35 – 0.40	76	123	0.37564	2.9111
0.30 – 0.35	35	102	0.32568	2.0041
0.25 – 0.30	31	93	0.27422	1.8139
0.20 – 0.25	13	79	0.22564	1.3458
0.15 – 0.20	10	49	0.17573	0.8631
0.10 – 0.15	7	18	0.12370	0.3657
0.05 – 0.10	1	16	0.07503	0.2487
0.00 – 0.05	1	11	0.03549	0.1755

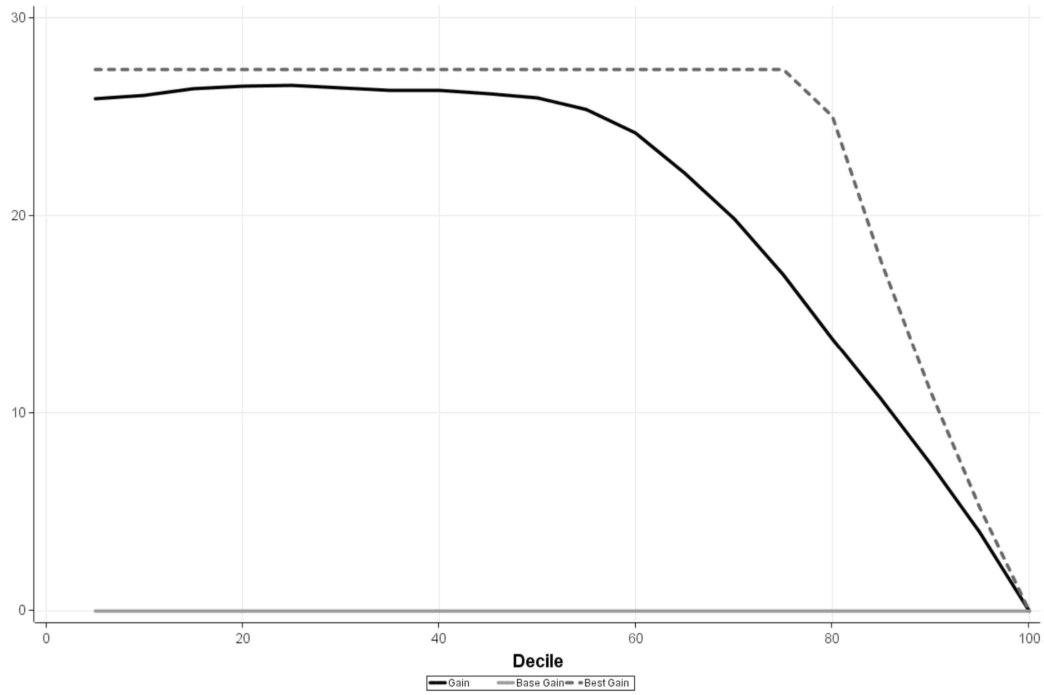
Appendix 39: Probit Link: Cumulative Lift



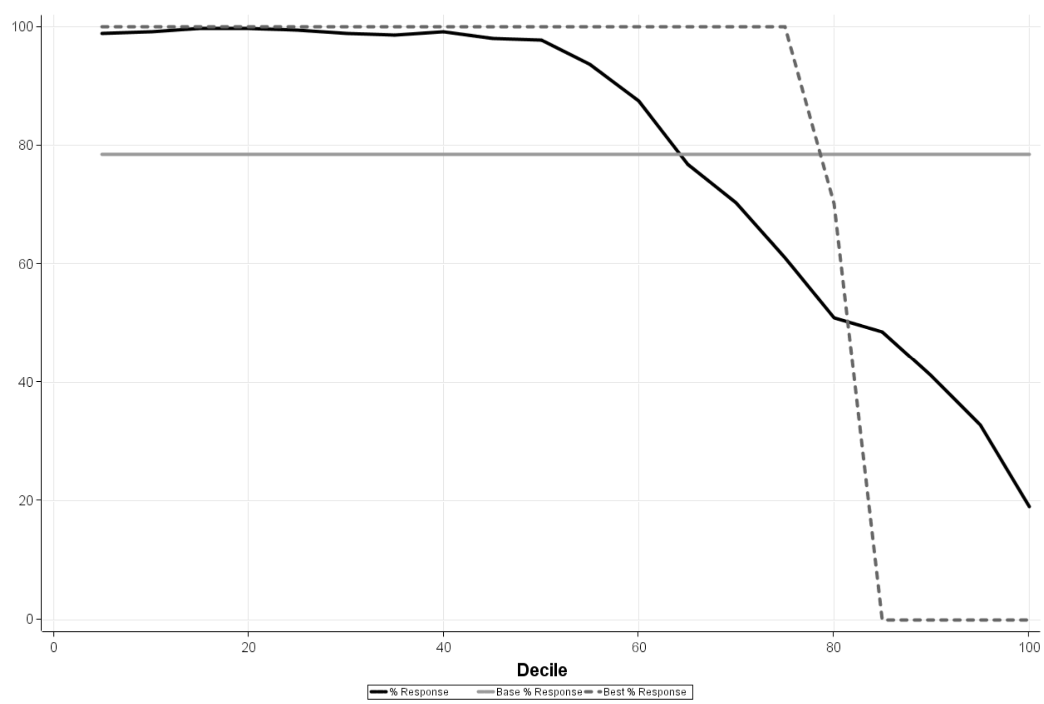
Appendix 40: Probit Link: Lift



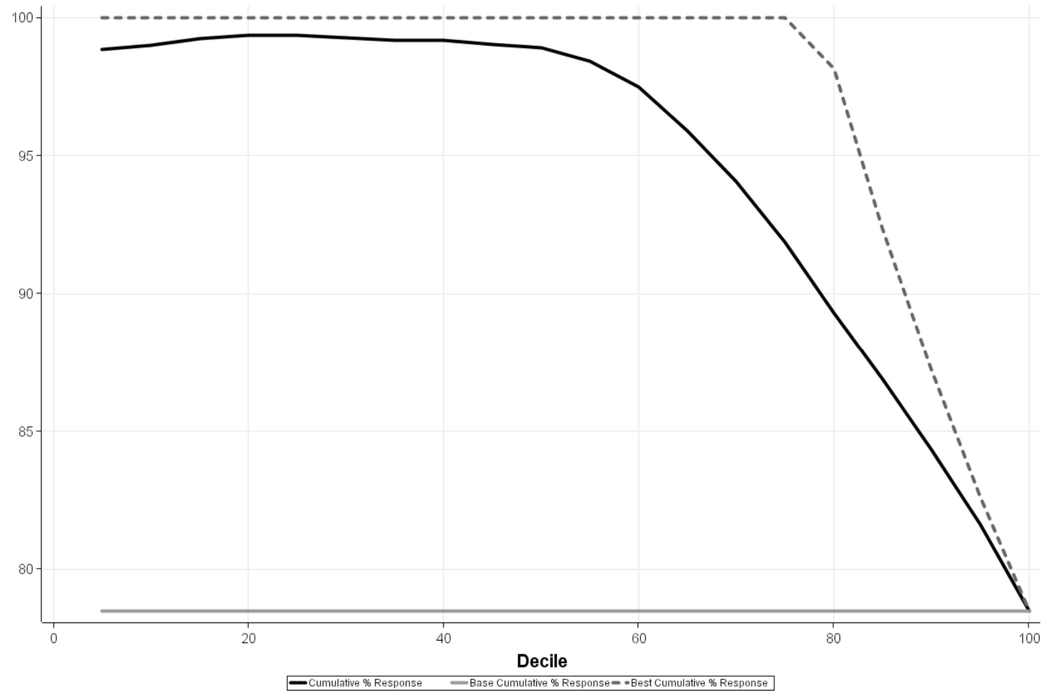
Appendix 41: Probit Link: Gain



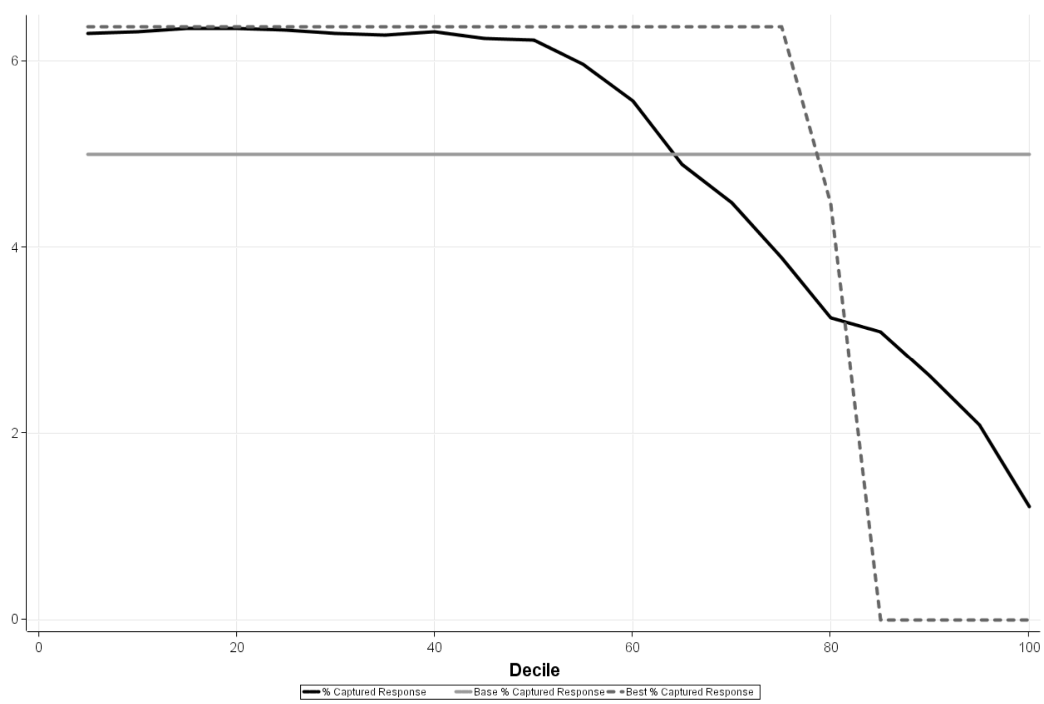
Appendix 42: Probit Link: % Response



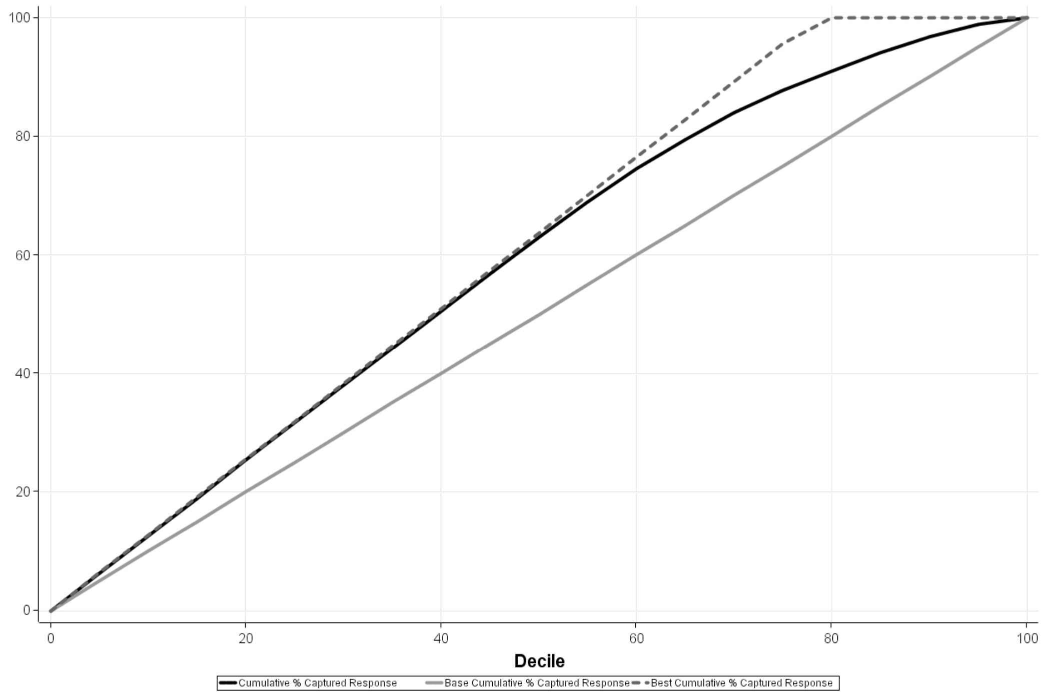
Appendix 43: Probit Link: Cumulative % Response



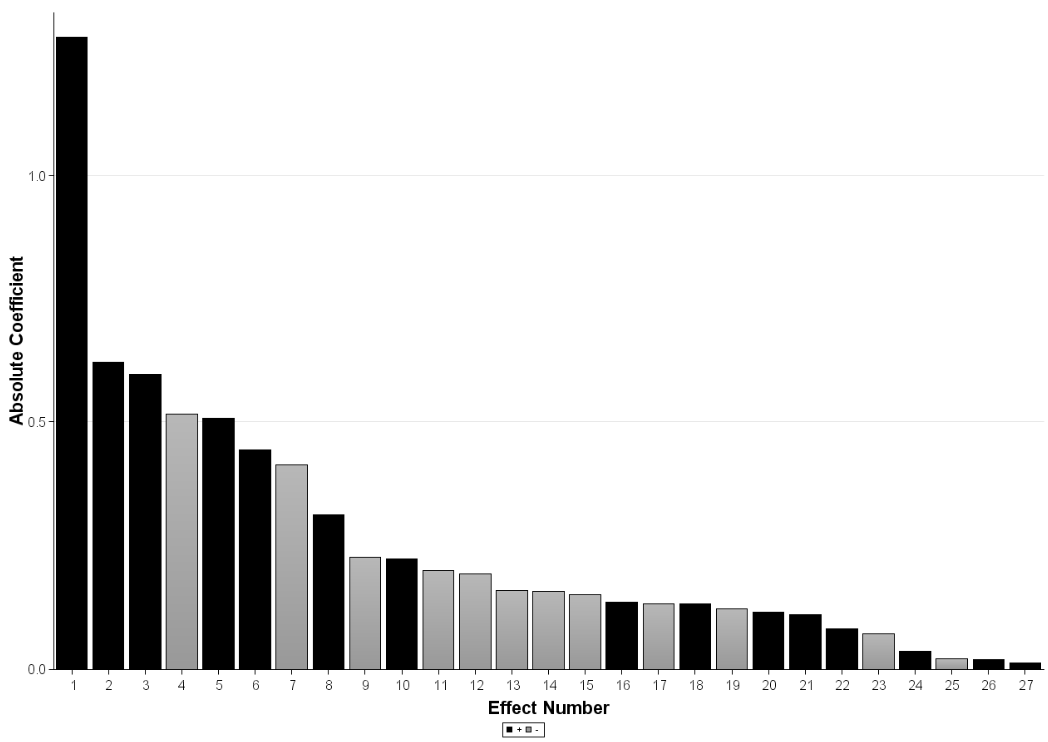
Appendix 44: Probit Link: % Captured Response



Appendix 45: Probit Link: Cumulative % Captured Response



Appendix 46: Probit Link: Effects Plot (standardised)



7.5. The regression modelling the length of the recovery period

Appendix 47: Recovery Period: Global Wald test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	28	12561	448.604493	5.82	.0001
Error	1517	116992	77.120925		
Corrected Total	1545	129553			

Appendix 48: Recovery Period: Model Fit Statistics

R-Square	0.0970		Adj R-Sq	0.0803
AIC	6746.6738		BIC	749.3490
SBC	6901.6331		C(p)	40.2011

Appendix 49: Recovery Period: Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
COUNTY	19	3777.0358	2.58	0.0002
CUM_CPI_DS_Y	1	1635.1704	21.20	<.0001
CURRENT_LTV	1	2524.8006	32.74	<.0001
DEFAULTED_PER_START_EXPOSURE	1	1178.1484	15.28	<.0001
DEFAULT_CPI	1	523.9663	6.79	0.0092
DEFAULT_MIN_WAGE	1	1896.4627	24.59	.0001
UNEMPL_RATE_INDEX_DS_Y	1	312.0183	4.05	0.0445
LOAN_PURPOSE	3	1137.6503	4.92	0.0021

Appendix 50: Recovery Period: Analysis of ML-estimates (non-standardised)

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-100.7	21.3343	36.51	<.0001
COUNTY=Baranya	1.1004	1.3274	0.83	0.4072
COUNTY=Borsod-Abaúj-Zemplén	-2.4692	0.9468	-2.61	0.0092
COUNTY=Budapest	1.8953	0.7837	2.42	0.0157
COUNTY=Bács-Kiskun	1.0652	1.2701	0.84	0.4018
COUNTY=Békés	-3.7576	1.2714	-2.96	0.0032
COUNTY=Csongrád	3.2227	1.1520	2.80	0.0052
COUNTY=Fejér	0.1592	1.3159	0.12	0.9037
COUNTY=Győr-Moson-Sopron	0.4562	1.5834	0.29	0.7733
COUNTY=Hajdu-Bihar	-0.3116	1.1118	-0.28	0.7793
COUNTY=Heves	-1.5188	0.9731	-1.56	0.1188
COUNTY=Jász-Nagykun-Szolnok	-0.2926	0.9262	-0.32	0.7521
COUNTY=Komárom-Esztergom	-0.4112	1.1715	-0.35	0.7256
COUNTY=Nógrád	0.7071	0.9255	0.76	0.4450
COUNTY=Pest	-0.1437	0.6565	-0.22	0.9268
COUNTY=Somogy	-0.3566	0.9087	-0.39	0.6948
COUNTY=Szabolcs-Szatmár-Bereg	-1.6533	0.6420	-2.58	0.0101
COUNTY=Tolna	5.6115	2.2564	2.49	0.0130
COUNTY=Vas	-2.7035	2.6723	-1.01	0.3118
COUNTY=Veszprém	0.7974	1.5420	0.52	0.6052
CUM_CPI_DS_Y	88.6327	19.2486	4.60	<.0001
CURRENT_LTV	8.7642	1.5317	5.72	<.0001
DEFAULTED_PER_START_EXPOSURE	-10.1692	2.6018	-3.91	<.0001
DEFAULT_CPI	42.6944	16.3797	2.61	0.0092
DEFAULT_MIN_WAGE	-0.00033	0.000067	-4.96	<.0001
UNEMPL_RATE_INDEX_DS_Y	4.4630	2.2188	2.01	0.0445
LOAN_PURPOSE=Other	0.3749	0.4136	0.91	0.3649
LOAN_PURPOSE=Real estate construction	1.5622	0.6627	2.36	0.0185
LOAN_PURPOSE=Real estate purchase	-1.2811	0.4434	-2.89	0.0039

Appendix 51: Recovery Period: Analysis of ML-estimates (standardised)

Parameter	*Estimate*	*Standard Error*	*t Value*	*Pr > |t|*
Intercept	14.9128	0.4085	36.51	<.0001
COUNTY=Baranya	1.1004	1.3274	0.83	0.4072
COUNTY=Borsod-Abaúj-Zemplén	-2.4692	0.9468	-2.61	0.0092
COUNTY=Budapest	1.8953	0.7837	2.42	0.0157
COUNTY=Bács-Kiskun	1.0652	1.2701	0.84	0.4018
COUNTY=Békés	-3.7576	1.2714	-2.96	0.0032
COUNTY=Csongrád	3.2227	1.1520	2.80	0.0052
COUNTY=Fejér	0.1592	1.3159	0.12	0.9037
COUNTY=Győr-Moson-Sopron	0.4562	1.5834	0.29	0.7733
COUNTY=Hajdu-Bihar	-0.3116	1.1118	-0.28	0.7793
COUNTY=Heves	-1.5188	0.9731	-1.56	0.1188
COUNTY=Jász-Nagykun-Szolnok	-0.2926	0.9262	-0.32	0.7521
COUNTY=Komárom-Esztergom	-0.4112	1.1715	-0.35	0.7256
COUNTY=Nógrád	0.7071	0.9255	0.76	0.4450
COUNTY=Pest	-0.1437	0.6565	-0.22	0.9268
COUNTY=Somogy	-0.3566	0.9087	-0.39	0.6948
COUNTY=Szabolcs-Szatmár-Bereg	-1.6533	0.6420	-2.58	0.0101
COUNTY=Tolna	5.6115	2.2564	2.49	0.0130
COUNTY=Vas	-2.7035	2.6723	-1.01	0.3118
COUNTY=Veszprém	0.7974	1.5420	0.52	0.6052
STD_CUM_CPI_DS_Y	1.2608	0.2738	4.60	<.0001
STD_CURRENT_LTV	1.4951	0.2613	5.72	<.0001
STD_DEFAULTED_PER_START_EXPOSURE	-1.2241	0.3132	-3.91	<.0001
STD_DEFAULT_CPI	0.7877	0.3022	2.61	0.0092
STD_DEFAULT_MIN_WAGE	-1.4144	0.2852	-4.96	<.0001
STD_UNEMPL_RATE_INDEX_DS_Y	0.5840	0.2930	2.01	0.0445
LOAN_PURPOSE=Other	0.3749	0.4136	0.91	0.3649
LOAN_PURPOSE=Real estate construction	1.5622	0.6627	2.36	0.0185
LOAN_PURPOSE=Real estate purchase	-1.2811	0.4434	-2.89	0.0039

Appendix 52: Recovery Period: Fit Statistics

Fit statistics	*Statistics Label*	*Train*
AIC	Akaike's Information Criterion	6775.04
ASE	Average Squared Error	75.79
AVERR	Average Error Function	75.79
DFE	Degrees of Freedom for Error	1523.00
DFM	Model Degrees of Freedom	29.00
DFT	Total Degrees of Freedom	1552.00
DIV	Divisor for ASE	1552.00
ERR	Error Function	117629.20
FPE	Final Prediction Error	78.68
MAX	Maximum Absolute Error	50.07
MSE	Mean Square Error	77.24
NOBS	Sum of Frequencies	1552.00
_NW_	Number of Estimate Weights	29.00
RASE	Root Average Sum of Squares	8.71
RFPE	Root Final Prediction Error	8.87
RMSE	Root Mean Squared Error	8.79
SBC	Schwarz's Bayesian Criterion	6930.12
SSE	Sum of Squared Errors	117629.20
SUMW	Sum of Case Weights Times Freq	1552.00

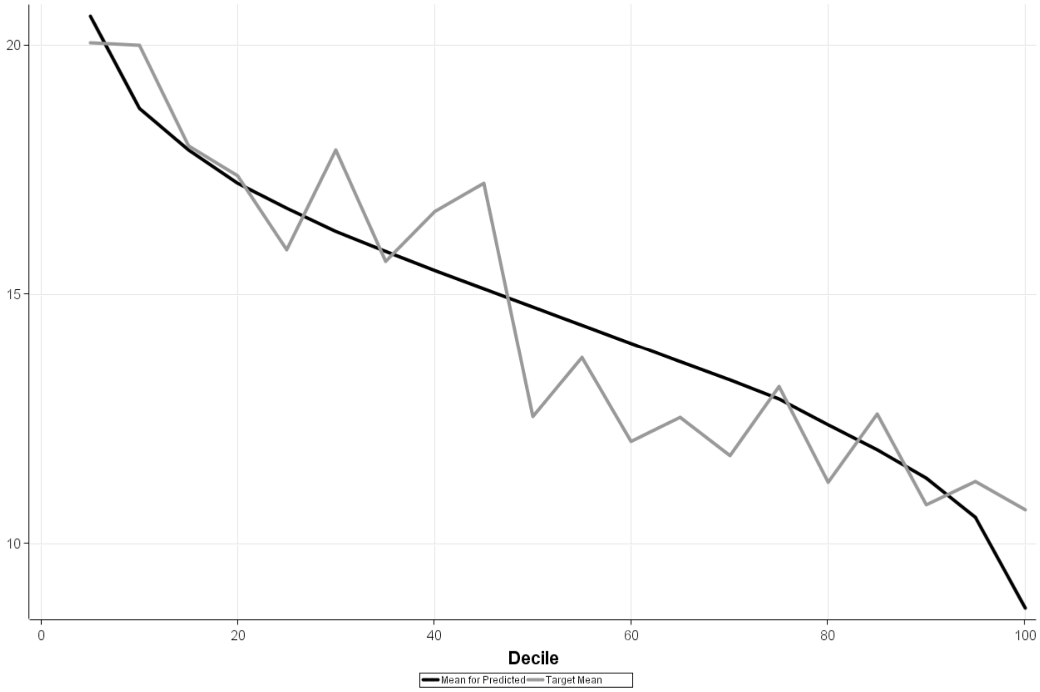
Appendix 53: Recovery Period: Assessment Score Rankings

Decile	*Observation Number*	*Target Mean*	*Mean for Predicted*
0	.	.	.
5	77.6	20.4012	20.5847
10	77.6	19.9974	18.7322
15	77.6	17.9794	17.9003
20	77.6	17.3789	17.2282
25	77.6	15.9046	16.7373
30	77.6	17.8918	16.2676
35	77.6	15.6701	15.8587
40	77.6	16.6701	15.4862
45	77.6	17.2242	15.1170
50	77.6	12.5567	14.7570
55	77.6	13.7345	14.3907
60	77.6	12.0593	14.0098
65	77.6	12.5387	13.6420
70	77.6	11.7706	13.2773
75	77.6	13.1443	12.8969
80	77.6	11.2294	12.3833
85	77.6	12.5928	11.8863
90	77.6	10.7887	11.3196
95	77.6	11.2552	10.5359
100	77.6	10.6804	8.7298

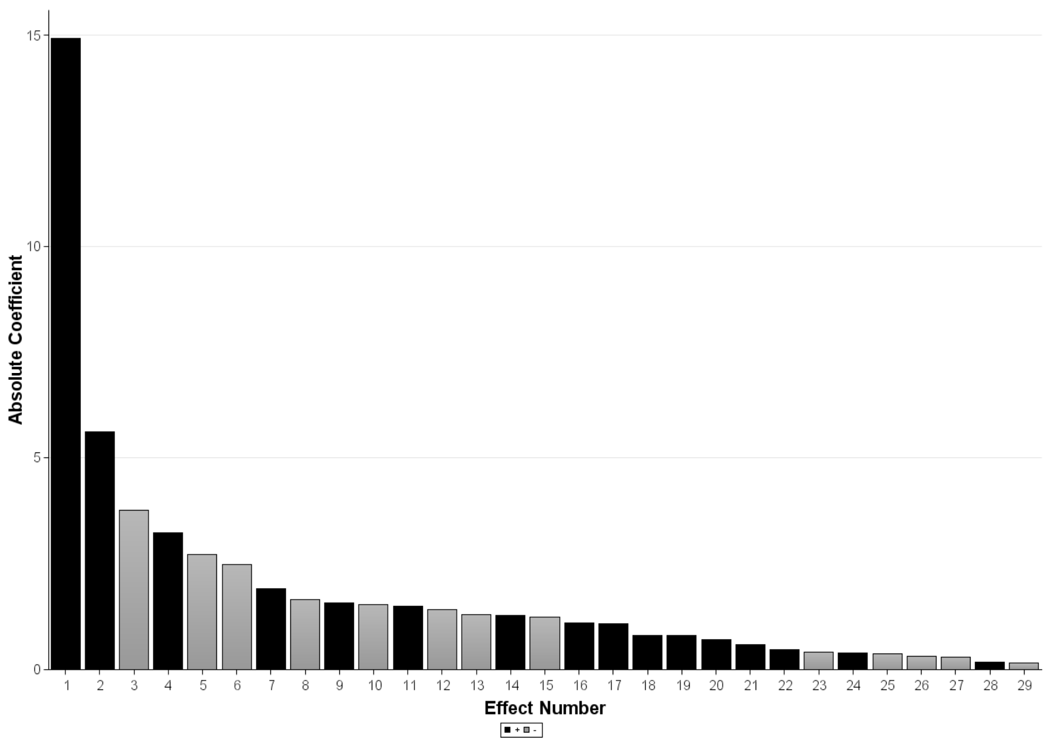
Appendix 54: Recovery Period: Assessment Score Distribution

Range for Predicted	*Target Mean*	*Mean for Predicted*	*Number of Observations*	*Model Score*
24.03 – 25.04	27.5000	24.7497	2	24.5378
23.02 – 24.03	12.0000	23.9645	1	23.5285
22.01 – 23.02	20.7778	11.6437	9	22.5191
21.00 – 22.01	18.4000	21.4217	10	21.5097
20.00 – 21.00	22.2174	20.3675	23	20.5003
18.99 – 20.00	19.4898	19.4069	49	19.4909
17.98 – 18.99	19.7753	18.4702	89	18.4815
16.97 – 17.98	16.9545	17.4273	132	17.4721
15.96 – 16.97	17.0947	16.4332	169	16.4628
14.95 – 15.96	16.4952	15.4601	208	15.4534
13.94 – 14.95	12.6402	14.4609	214	14.4440
12.93 – 13.94	12.5802	13.4447	212	13.4346
11.92 – 12.93	12.2138	12.4339	159	12.4252
10.91 – 11.92	11.0880	11.4490	125	11.4158
9.90 – 10.91	11.5696	10.4643	79	10.4064
8.89 – 9.90	10.8158	9.3983	38	9.3971
7.88 – 8.89	10.4737	8.3719	19	8.3877
6.87 – 7.88	9.7500	7.2970	8	7.3783
5.86 – 6.87	14.0000	6.5622	4	6.3689
4.85 – 5.86	5.0000	5.2765	2	5.3595

Appendix 55: Recovery Period: Score Ranking Overlay



Appendix 56: Recovery Period: Effects Plot (standardised)



7.6. The regression modelling the recovery rate deriving from the selling

Appendix 57: Recovery Rate: Global Wald test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	30	23.131504	0.771050	16.14	<.0001
Error	1515	72.378447	0.047775		
Corrected Total	1545	95.509951			

Appendix 58: Recovery Rate: Model Fit Statistics

R-Square	0.2422		Adj R-Sq	0.2272
AIC	-4671.1063		BIC	-4668.4698
SBC	-4505.4601		C(p)	46.2833

Appendix 59: Recovery Rate: Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
COUNTY	19	1.8659	2.06	0.0047
CUM_CPI_DS_Y	1	1.7868	37.40	<.0001
CURRENT_LTV	1	9.9364	207.98	<.0001
SETTLEMENT_TYPE	4	0.4738	2.48	0.0423
UNEMPL_RATE_INDEX_DS_Y	1	0.3569	7.47	0.0063
LOAN_PURPOSE	3	0.6484	4.52	0.0036
PRIORCHARGE_RATE	1	2.4968	52.26	<.0001

Appendix 60: Recovery Rate: Analysis of ML-estimates (non-standardised)

Parameter	Estimate	Standard Error	t Value	Pr > $\lvert t\rvert$
Intercept	3.7596	0.4603	52.37	<.0001
COUNTY=Baranya	-0.0772	0.0334	-2.31	0.0208
COUNTY=Borsod-Abaúj-Zemplén	0.0192	0.0237	0.81	0.4175
COUNTY=Budapest	0.1468	0.0363	4.04	<.0001
COUNTY=Bács-Kiskun	-0.0192	0.0317	-0.60	0.5462
COUNTY=Békés	-0.0134	0.0318	-0.42	0.6737
COUNTY=Csongrád	-0.0444	0.0291	-1.52	0.1279
COUNTY=Fejér	0.0194	0.0329	0.59	0.5547
COUNTY=Győr-Moson-Sopron	0.000177	0.0395	0.00	0.9964
COUNTY=Hajdu-Bihar	0.0314	0.0278	1.13	0.2591
COUNTY=Heves	-0.0360	0.0244	-1.48	0.1399
COUNTY=Jász-Nagykun-Szolnok	-0.0246	0.0323	-1.06	0.2885
COUNTY=Komárom-Esztergom	-0.00372	0.0293	-0.13	0.8988
COUNTY=Nógrád	-0.0253	0.0234	-1.08	0.2810
COUNTY=Pest	0.0125	0.0190	0.66	0.5120
COUNTY=Somogy	-0.00909	0.0228	-0.40	0.6902
COUNTY=Szabolcs-Szatmár-Bereg	-0.0377	0.0160	-2.35	0.0187
COUNTY=Tolna	0.0299	0.0563	0.53	0.5960
COUNTY=Vas	-0.0377	0.0667	-0.57	0.5719
COUNTY=Veszprém	0.0735	0.0385	1.91	0.0565
CUM_CPI_DS_Y	-2.5920	0.4238	-6.12	<.0001
CURRENT_LTV	-0.5953	0.0413	-14.42	<.0001
SETTLEMENT_TYPE=Budapest & environs	0.00621	0.0266	0.23	0.8155
SETTLEMENT_TYPE=County town & environs	0.0353	0.0144	2.45	0.0145
SETTLEMENT_TYPE=Other city & environs	-0.0117	0.0113	-1.03	0.3021
SETTLEMENT_TYPE=Small village	-0.0180	0.0136	-1.32	0.1858
UNEMPL_RATE_INDEX_DS_Y	-0.1342	0.0491	-2.73	0.0063
PRIORCHARGE_RATE	-0.3176	0.0439	-7.23	<.0001
LOAN_PURPOSE=Other	0.0185	0.00998	1.85	0.0643
LOAN_PURPOSE=Real estate construction	-0.0566	0.0166	-3.41	0.0007
LOAN_PURPOSE=Real estate purchase	0.0257	0.0112	2.30	0.0216

Appendix 61: Recovery Rate: Analysis of ML-estimates (standardised)

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.5866	0.0112	52.37	<.0001
COUNTY=Baranya	-0.0772	0.0334	-2.31	0.0208
COUNTY=Borsod-Abaúj-Zemplén	0.0192	0.0237	0.81	0.4175
COUNTY=Budapest	0.1468	0.0363	4.04	<.0001
COUNTY=Bács-Kiskun	-0.0192	0.0317	-0.60	0.5462
COUNTY=Békés	-0.0134	0.0318	-0.42	0.6737
COUNTY=Csongrád	-0.0444	0.0291	-1.52	0.1279
COUNTY=Fejér	0.0194	0.0329	0.59	0.5547
COUNTY=Győr-Moson-Sopron	0.000177	0.0395	0.00	0.9964
COUNTY=Hajdu-Bihar	0.0314	0.0278	1.13	0.2591
COUNTY=Heves	-0.0360	0.0244	-1.48	0.1399
COUNTY=Jász-Nagykun-Szolnok	-0.0246	0.0323	-1.06	0.2885
COUNTY=Komárom-Esztergom	-0.00372	0.0293	-0.13	0.8988
COUNTY=Nógrád	-0.0253	0.0234	-1.08	0.2810
COUNTY=Pest	0.0125	0.0190	0.66	0.5120
COUNTY=Somogy	-0.00909	0.0228	-0.40	0.6902
COUNTY=Szabolcs-Szatmár-Bereg	-0.0377	0.0160	-2.35	0.0187
COUNTY=Tolna	0.0299	0.0563	0.53	0.5960
COUNTY=Vas	-0.0377	0.0667	-0.57	0.5719
COUNTY=Veszprém	0.0735	0.0385	1.91	0.0565
SETTLEMENT_TYPE=Budapest & environs	0.00621	0.0266	0.23	0.8155
SETTLEMENT_TYPE=County town & environs	0.0353	0.0144	2.45	0.0145
SETTLEMENT_TYPE=Other city & environs	-0.0117	0.0113	-1.03	0.3021
SETTLEMENT_TYPE=Small village	-0.0180	0.0136	-1.32	0.1858
STD_CUM_CPI_DS_Y	-0.0369	0.00603	-6.12	<.0001
STD_CURRENT_LTV	-0.1016	0.00704	-14.42	<.0001
STD_UNEMPL_RATE_INDEX_DS_Y	-0.0176	0.00642	-2.73	0.0063
STD_PRIORCHARGE_RATE	-0.0498	0.00688	-7.23	<.0001
LOAN_PURPOSE=Other	0.0185	0.00998	1.85	0.0643
LOAN_PURPOSE=Real estate construction	-0.0566	0.0166	-3.41	0.0007
LOAN_PURPOSE=Real estate purchase	0.0257	0.0112	2.30	0.0216

Appendix 62: Recovery Rate: Fit Statistics

Fit statistics	Statistics Label	Train
AIC	Akaike's Information Criterion	-4688.97
ASE	Average Squared Error	0.05
AVERR	Average Error Function	0.05
DFE	Degrees of Freedom for Error	1521.00
DFM	Model Degrees of Freedom	31.00
DFT	Total Degrees of Freedom	1552.00
DIV	Divisor for ASE	1552.00
ERR	Error Function	72.68
FPE	Final Prediction Error	0.05
MAX	Maximum Absolute Error	1.69
MSE	Mean Square Error	0.05
NOBS	Sum of Frequencies	1552.00
_NW_	Number of Estimate Weights	31.00
RASE	Root Average Sum of Squares	0.22
RFPE	Root Final Prediction Error	0.22
RMSE	Root Mean Squared Error	0.22
SBC	Schwarz's Bayesian Criterion	-4523.21
SSE	Sum of Squared Errors	72.68
SUMW	Sum of Case Weights Times Freq	1552.00

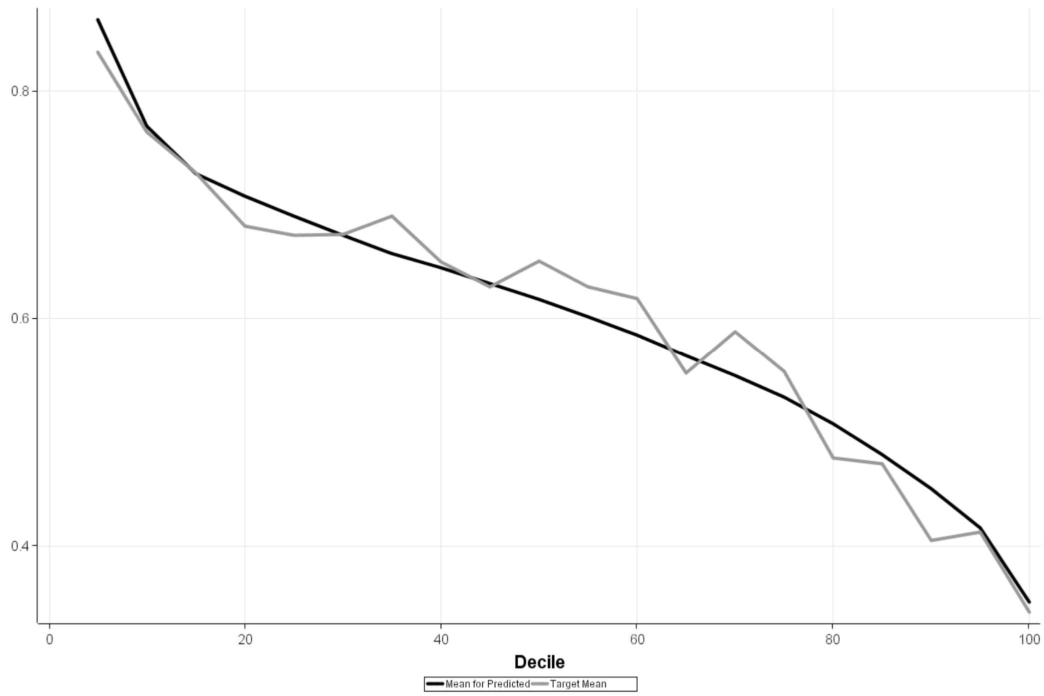
Appendix 63: Recovery Rate: Assessment Score Rankings

Decile	Observation Number	Target Mean	Mean for Predicted
0	.	.	.
5	77.6	0.83422	0.86296
10	77.6	0.76374	0.76888
15	77.6	0.72815	0.72771
20	77.6	0.68111	0.70769
25	77.6	0.67338	0.69004
30	77.6	0.67369	0.67322
35	77.6	0.68985	0.65724
40	77.6	0.64983	0.64449
45	77.6	0.62827	0.63071
50	77.6	0.65047	0.61666
55	77.6	0.62779	0.60131
60	77.6	0.61805	0.58530
65	77.6	0.55163	0.56705
70	77.6	0.58833	0.54943
75	77.6	0.55319	0.53047
80	77.6	0.47747	0.50696
85	77.6	0.47243	0.48015
90	77.6	0.40520	0.45007
95	77.6	0.41246	0.41604
100	77.6	0.34248	0.35119

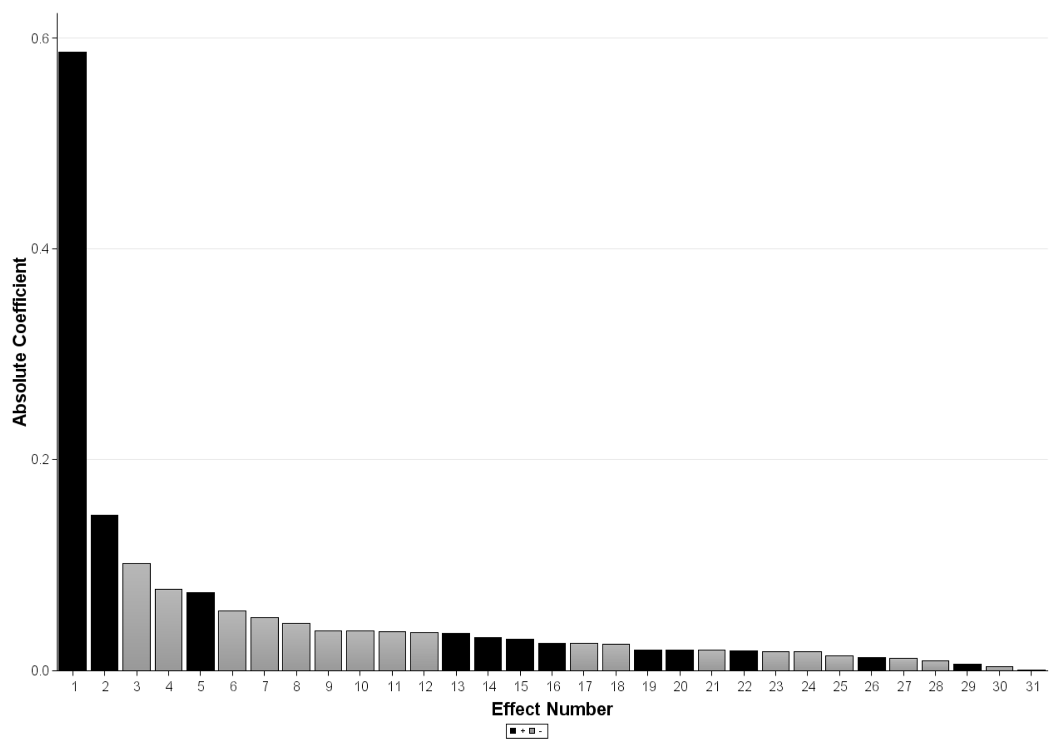
Appendix 64: Recovery Rate: Assessment Score Distribution

Range for Predicted	Target Mean	Mean for Predicted	Number of Observations	Model Score
0.97 – 1.01	1.01950	0.99688	3	0.99023
0.93 – 0.97	1.05269	0.64864	3	0.94988
0.89 – 0.93	0.98393	0.90820	11	0.90954
0.85 – 0.89	0.79922	0.86957	27	0.86920
0.81 – 0.85	0.78791	0.82735	27	0.82886
0.77 – 0.81	0.76431	0.78718	48	0.78852
0.73 – 0.77	0.74078	0.74268	73	0.74818
0.69 – 0.73	0.69195	0.70666	169	0.70784
0.65 – 0.69	0.67296	0.66619	203	0.66749
0.61 – 0.65	0.64318	0.62780	223	0.62715
0.57 – 0.61	0.60573	0.58811	181	0.58681
0.53 – 0.57	0.57559	0.54754	173	0.54647
0.49 – 0.53	0.48440	0.50617	127	0.50613
0.45 – 0.49	0.44672	0.46576	100	0.46579
0.41 – 0.45	0.41957	0.42765	87	0.42545
0.36 – 0.41	0.34977	0.38973	56	0.38511
0.32 – 0.36	0.32933	0.34572	25	0.24476
0.28 – 0.32	0.37856	0.30578	10	0.60442
0.24 – 0.28	0.33938	0.26295	4	0.26408
0.20 – 0.24	0.19039	0.22226	2	0.22374

Appendix 65: Recovery Rate: Score Ranking Overlay



Appendix 66: Recovery Rate: Effects Plot (standardised)



7.7. The regression developed for the LGD of the "WorkoutEnd" deal category

Appendix 67: WorkoutEnd: Global Wald test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	0.754840	0.027957	9.81	0.0001
Error	2905	8.275638	0.002849		
Corrected Total	2932	9.030478			

Appendix 68: WorkoutEnd: Model Fit Statistics

R-Square	0.0836		Adj R-Sq	0.0751
AIC	-17162.0739		BIC	-17159.7831
SBC	-16994.5281		C(p)	40.8459

Appendix 69: WorkoutEnd: Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
AVG_PD	1	0.0704	24.71	<.0001
COUNTY	19	0.1339	2.47	0.0004
DEFAULTED_EXPOSURE_LCY	1	0.0211	7.42	0.0065
DEFAULTED_PER_START_EXPOSURE	1	0.0304	10.67	0.0011
DEFAULT_AVG_NETINCOME	1	0.0113	3.97	0.0464
DEFAULT_CPI	1	0.1010	35.45	<.0001
DEFAULT_REALWAGE_INDEX	1	0.0838	29.42	<.0001
DEFAULT_UNEMPL_RATE	1	0.0195	6.84	0.0090
PRIORCHARGE_RATE	1	0.0188	6.59	0.0103

Appendix 70: WorkoutEnd: Analysis of ML-estimates (non-standardised)

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.7496	0.1119	15.10	<.0001
AVG_PD	-0.6131	0.1233	-4.97	<.0001
COUNTY=Baranya	-0.00725	0.00551	-1.32	0.1885
COUNTY=Borsod-Abaúj-Zemplén	-0.00111	0.00394	-0.28	0.7776
COUNTY=Budapest	-0.00671	0.00283	-2.37	0.0179
COUNTY=Bács-Kiskun	-0.00444	0.00417	-1.06	0.2870
COUNTY=Békés	-0.00560	0.00452	-1.24	0.2151
COUNTY=Csongrád	-0.00802	0.00494	-1.62	0.1048
COUNTY=Fejér	-0.00648	0.00419	-1.55	0.1223
COUNTY=Győr-Moson-Sopron	-0.00099	0.00439	-0.22	0.8220
COUNTY=Hajdu-Bihar	-0.00417	0.00407	-1.02	0.3055
COUNTY=Heves	0.00716	0.00567	1.26	0.2069
COUNTY=Jász-Nagykun-Szolnok	-0.00251	0.00618	-0.41	0.6854
COUNTY=Komárom-Esztergom	-0.00220	0.00409	-0.54	0.5914
COUNTY=Nógrád	0.00510	0.00584	-0.87	0.3824
COUNTY=Pest	-0.00101	0.00290	-0.35	0.7277
COUNTY=Somogy	-0.00658	0.00683	-0.96	0.3354
COUNTY=Szabolcs-Szatmár-Bereg	-0.00509	0.00384	-1.33	0.1847
COUNTY=Tolna	0.0169	0.00831	2.03	0.0421
COUNTY=Vas	-0.00375	0.00708	-0.53	0.5965
COUNTY=Veszprém	-0.00213	0.00400	-0.53	0.5947
DEFAULTED_EXPOSURE_LCY	5.28E-10	1.94E-10	2.72	0.0065
DEFAULTED_PER_START_EXPOSURE	0.0216	0.00662	3.27	0.0011
DEFAULT_AVG_NETINCOME	-4.08E-7	2.048E-7	-1.99	0.0464
DEFAULT_CPI	-0.5301	0.0890	-5.95	<.0001
DEFAULT_REALWAGE_INDEX	-0.1616	0.0298	-5.42	<.0001
DEFAULT_UNEMPL_RATE	0.5051	0.1932	2.61	0.0090
PRIORCHARGE_RATE	0.0173	0.00675	2.57	0.0103

Appendix 71: WorkoutEnd: Analysis of ML-estimates (standardised)

Parameter	*Estimate*	*Standard Error*	*t Value*	*Pr > |t|*
Intercept	0.0186	0.00123	15.10	<.0001
COUNTY=Baranya	-0.00725	0.00551	-1.32	0.1885
COUNTY=Borsod-Abaúj-Zemplén	-0.00111	0.00394	-0.28	0.7776
COUNTY=Budapest	-0.00671	0.00283	-2.37	0.0179
COUNTY=Bács-Kiskun	-0.00444	0.00417	-1.06	0.2870
COUNTY=Békés	-0.00560	0.00452	-1.24	0.2151
COUNTY=Csongrád	-0.00802	0.00494	-1.62	0.1048
COUNTY=Fejér	-0.00648	0.00419	-1.55	0.1223
COUNTY=Győr-Moson-Sopron	-0.00099	0.00439	-0.22	0.8220
COUNTY=Hajdu-Bihar	-0.00417	0.00407	-1.02	0.3055
COUNTY=Heves	0.00716	0.00567	1.26	0.2069
COUNTY=Jász-Nagykun-Szolnok	-0.00251	0.00618	-0.41	0.6854
COUNTY=Komárom-Esztergom	-0.00220	0.00409	-0.54	0.5914
COUNTY=Nógrád	0.00510	0.00584	-0.87	0.3824
COUNTY=Pest	-0.00101	0.00290	-0.35	0.7277
COUNTY=Somogy	-0.00658	0.00683	-0.96	0.3354
COUNTY=Szabolcs-Szatmár-Bereg	-0.00509	0.00384	-1.33	0.1847
COUNTY=Tolna	0.0169	0.00831	2.03	0.0421
COUNTY=Vas	-0.00375	0.00708	-0.53	0.5965
COUNTY=Veszprém	-0.00213	0.00400	-0.53	0.5947
STD_AVG_PD	-0.0153	0.00307	-4.97	<.0001
STD_DEFAULTED_EXPOSURE_LCY	0.00275	0.00101	2.72	0.0065
STD_DEFAULTED_PER_START_EXPOSURE	0.00399	0.00122	3.27	0.0011
STD_DEFAULT_AVG_NETINCOME	-0.00349	0.00175	-1.99	0.0464
STD_DEFAULT_CPI	-0.00774	0.00130	-5.95	<.0001
STD_DEFAULT_REALWAGE_INDEX	-0.00587	0.00108	-5.42	<.0001
STD_DEFAULT_UNEMPL_RATE	0.00805	0.00308	2.61	0.0090
STD_PRIORCHARGE_RATE	0.00250	0.000973	2.57	0.0103

Appendix 72: WorkoutEnd: Fit Statistics

Fit statistics	*Statistics Label*	*Train*
AIC	Akaike's Information Criterion	-31522.67
ASE	Average Squared Error	0.00
AVERR	Average Error Function	0.00
DFE	Degrees of Freedom for Error	5339.00
DFM	Model Degrees of Freedom	28.00
DFT	Total Degrees of Freedom	5367.00
DIV	Divisor for ASE	5367.00
ERR	Error Function	14.94
FPE	Final Prediction Error	0.00
MAX	Maximum Absolute Error	0.98
MSE	Mean Square Error	0.00
NOBS	Sum of Frequencies	5367.00
_NW_	Number of Estimate Weights	28.00
RASE	Root Average Sum of Squares	0.05
RFPE	Root Final Prediction Error	0.05
RMSE	Root Mean Squared Error	0.05
SBC	Schwarz's Bayesian Criterion	-31338.20
SSE	Sum of Squared Errors	14.94
SUMW	Sum of Case Weights Times Freq	5367.00

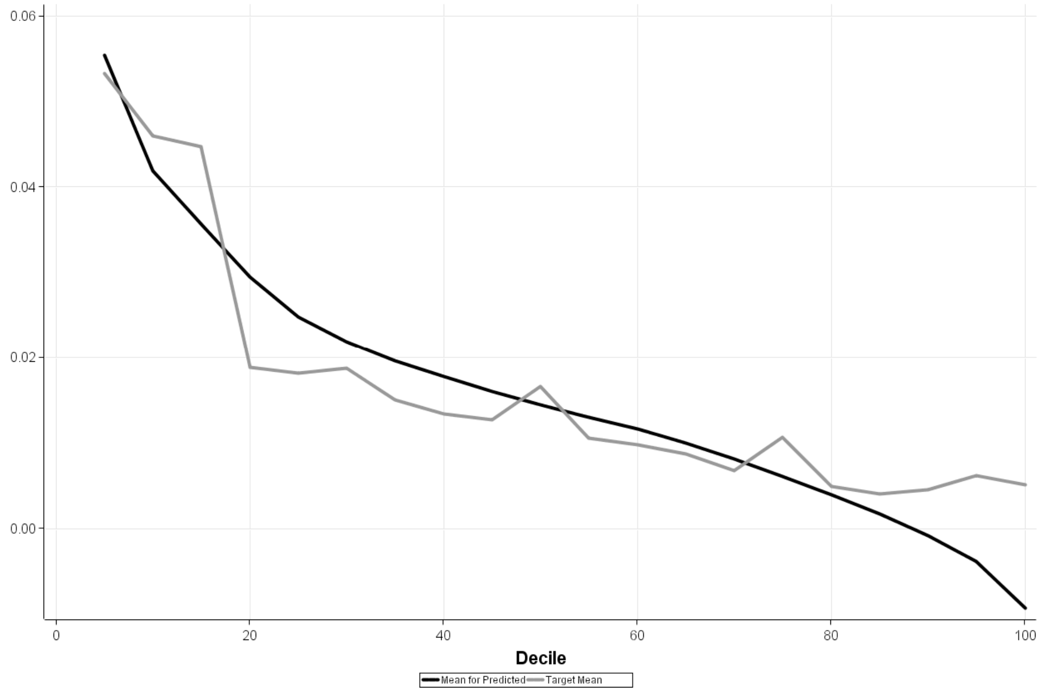
Appendix 73: WorkoutEnd: Assessment Score Rankings

Decile	Observation Number	Target Mean	Mean for Predicted
0	.	.	.
5	268.35	0.053219	0.055356
10	268.35	0.045962	0.041871
15	268.35	0.044686	0.035652
20	268.35	0.018838	0.029448
25	268.35	0.018169	0.024823
30	268.35	0.018796	0.021861
35	268.35	0.015111	0.019641
40	268.35	0.013384	0.017747
45	268.35	0.012735	0.016029
50	268.35	0.016640	0.014529
55	268.35	0.010643	0.013073
60	268.35	0.009788	0.011643
65	268.35	0.008797	0.009996
70	268.35	0.006851	0.008194
75	268.35	0.010694	0.006131
80	268.35	0.004933	0.003989
85	268.35	0.004066	0.001728
90	268.35	0.004608	-0.000733
95	268.35	0.006276	-0.003777
100	268.35	0.005192	-0.009168

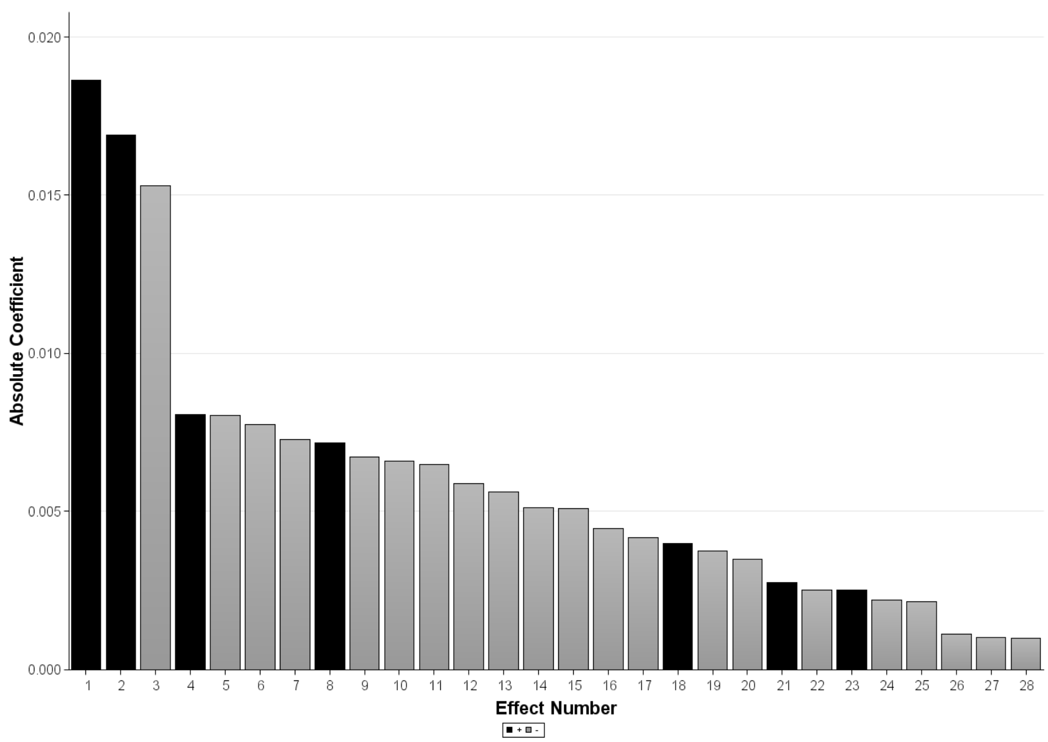
Appendix 74: WorkoutEnd: Assessment Score Distribution

Range for Predicted	Target Mean	Mean for Predicted	Number of Observations	Model Score
0.09 – 0.10	0.19227	0.095915	6	0.096226
0.09 – 0.09	0.21967	0.089844	5	0.089712
0.08 – 0.09	0.50903	0.082158	2	0.083198
0.07 – 0.08	0.01677	0.075706	13	0.076685
0.07 – 0.07	0.02901	0.070009	10	0.070171
0.06 – 0.07	0.04732	0.062626	20	0.063657
0.05 – 0.06	0.04353	0.056981	45	0.057144
0.05 – 0.05	0.04735	0.050025	110	0.050630
0.04 – 0.05	0.04363	0.043829	232	0.044116
0.03 – 0.04	0.05051	0.037700	280	0.037602
0.03 – 0.03	0.02275	0.030953	295	0.031089
0.02 – 0.03	0.01866	0.024155	509	0.024575
0.01 – 0.02	0.01478	0.017854	966	0.018061
0.01 – 0.01	0.00996	0.011704	1126	0.011548
0.00 – 0.01	0.00706	0.005090	800	0.005034
-0.00 – 0.00	0.00508	-0.001165	620	-0.001480
-0.01 – -0.00	0.00536	-0.007296	284	-0.007993
-0.02 – -0.01	0.00559	-0.013802	33	-0.014507
-0.02 – -0.02	0.00132	-0.019939	8	-0.021021
-0.03 – -0.02	0.00510	-0.028864	3	-0.027534

Appendix 75: WorkoutEnd: Score Ranking Overlay



Appendix 76: WorkoutEnd: Effects Plot (standardised)



7.8. The regression developed for the LGD of the "NoFurtherRec" deal category

Appendix 77: NoFurtherRec: Global Wald test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	25.154705	1.257735	28.55	<.0001
Error	503	22.158698	0.044053		
Corrected Total	523	47.313403			

Appendix 78: NoFurtherRec: Model Fit Statistics

R-Square	0.5317		Adj R-Sq	0.5130
AIC	-1615.5491		BIC	-1610.3592
SBC	-1526.0578		C(p)	4.8866

Appendix 79: NoFurtherRec: Analysis of Effects

Effect	DF	Sum of Squares	F Value	Pr > F
AVG_PD	1	0.4073	9.25	0.0025
CUM_CPI_DS	1	0.2262	5.13	0.0239
CUM_REALWAGE_INDEX_DS	1	0.9748	22.13	<.0001
DEFAULT_AGE_MONTHS	1	0.2438	5.54	0.0190
GDP_GROWTH_INDEX_DS	1	6.2741	142.42	<.0001
LANDLINE_PHONE_FLAG	1	0.1830	4.15	0.0421
MONTHS_TO_DEFAULT	1	1.6908	38.38	<.0001
REASON_PASTDUE	1	1.4432	32.76	<.0001
REGION	8	0.7113	2.02	0.0426
SETTLEMENT_TYPE	4	0.5189	2.94	0.0200

Appendix 80: NoFurtherRec: Analysis of ML-estimates (non-standardised)

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-7.9544	1.2036	21.96	<.0001
AVG_PD	3.6371	1.1961	3.04	0.0025
CUM_CPI_DS	2.3466	1.0357	2.27	0.0239
CUM_REALWAGE_INDEX_DS	-2.7167	0.5775	-4.70	<.0001
DEFAULT_AGE_MONTHS	-0.00018	0.000078	-2.35	0.0190
GDP_GROWTH_INDEX_DS	9.0760	0.7605	11.93	<.0001
LANDLINE_PHONE_FLAG=0	0.0209	0.0103	2.04	0.0421
MONTHS_TO_DEFAULT	-0.0304	0.00490	-6.20	<.0001
REASON_PASTDUE=0	-0.1702	0.0297	-5.72	<.0001
REGION=Budapest & environs	-0.0347	0.0388	-0.89	0.3714
REGION=Central-Western	-0.0308	0.0283	-1.09	0.2766
REGION=Eastern	0.0284	0.0263	1.08	0.2811
REGION=North-Eastern	0.0488	0.0198	2.46	0.0142
REGION=North-Western	-0.0580	0.0305	-1.90	0.0577
REGION=South-Central	0.0361	0.0407	0.89	0.3751
REGION=South-Eastern	-0.0347	0.0326	-1.06	0.2878
REGION=South-Western	0.000328	0.0422	0.01	0.9938
SETTLEMENT_TYPE=Budapest & environs	0.0253	0.0383	0.66	0.5091
SETTLEMENT_TYPE=County town & environs	-0.0721	0.0227	-3.17	0.0016
SETTLEMENT_TYPE=Other city & environs	0.00688	0.0180	0.38	0.7034
SETTLEMENT_TYPE=Small village	0.0298	0.0234	1.27	0.2030

Appendix 81: NoFurtherRec: Analysis of ML-estimates (standardised)

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.7094	0.0323	21.96	<.0001
LANDLINE_PHONE_FLAG=0	0.0209	0.0103	2.04	0.0421
REASON_PASTDUE=0	-0.1702	0.0297	-5.72	<.0001
REGION=Budapest & environs	-0.0347	0.0388	-0.89	0.3714
REGION=Central-Western	-0.0308	0.0283	-1.09	0.2766
REGION=Eastern	0.0284	0.0263	1.08	0.2811
REGION=North-Eastern	0.0488	0.0198	2.46	0.0142
REGION=North-Western	-0.0580	0.0305	-1.90	0.0577
REGION=South-Central	0.0361	0.0407	0.89	0.3751
REGION=South-Eastern	-0.0347	0.0326	-1.06	0.2878
REGION=South-Western	0.000328	0.0422	0.01	0.9938
SETTLEMENT_TYPE=Budapest & environs	0.0253	0.0383	0.66	0.5091
SETTLEMENT_TYPE=County town & environs	-0.0721	0.0227	-3.17	0.0016
SETTLEMENT_TYPE=Other city & environs	0.00688	0.0180	0.38	0.7034
SETTLEMENT_TYPE=Small village	0.0298	0.0234	1.27	0.2030
STD_AVG_PD	0.0417	0.0137	3.04	0.0025
STD_CUM_CPI_DS	0.0884	0.0390	2.27	0.0239
STD_CUM_REALWAGE_INDEX_DS	-0.0573	0.0122	-4.70	<.0001
STD_DEFAULT_AGE_MONTHS	-0.0251	0.0106	-2.35	0.0190
STD_GDP_GROWTH_INDEX_DS	0.1762	0.0148	11.93	<.0001
STD_MONTHS_TO_DEFAULT	-0.2216	0.0358	-6.20	<.0001

Appendix 82: NoFurtherRec: Fit Statistics

Fit statistics	Statistics Label	Train
AIC	Akaike's Information Criterion	-4777.62
ASE	Average Squared Error	0.04
AVERR	Average Error Function	0.04
DFE	Degrees of Freedom for Error	1448.00
DFM	Model Degrees of Freedom	21.00
DFT	Total Degrees of Freedom	1469.00
DIV	Divisor for ASE	1469.00
ERR	Error Function	55.23
FPE	Final Prediction Error	0.04
MAX	Maximum Absolute Error	1.03
MSE	Mean Square Error	0.04
NOBS	Sum of Frequencies	1469.00
_NW_	Number of Estimate Weights	21.00
RASE	Root Average Sum of Squares	0.19
RFPE	Root Final Prediction Error	0.20
RMSE	Root Mean Squared Error	0.20
SBC	Schwarz's Bayesian Criterion	-4666.48
SSE	Sum of Squared Errors	55.23
SUMW	Sum of Case Weights Times Freq	1469.00

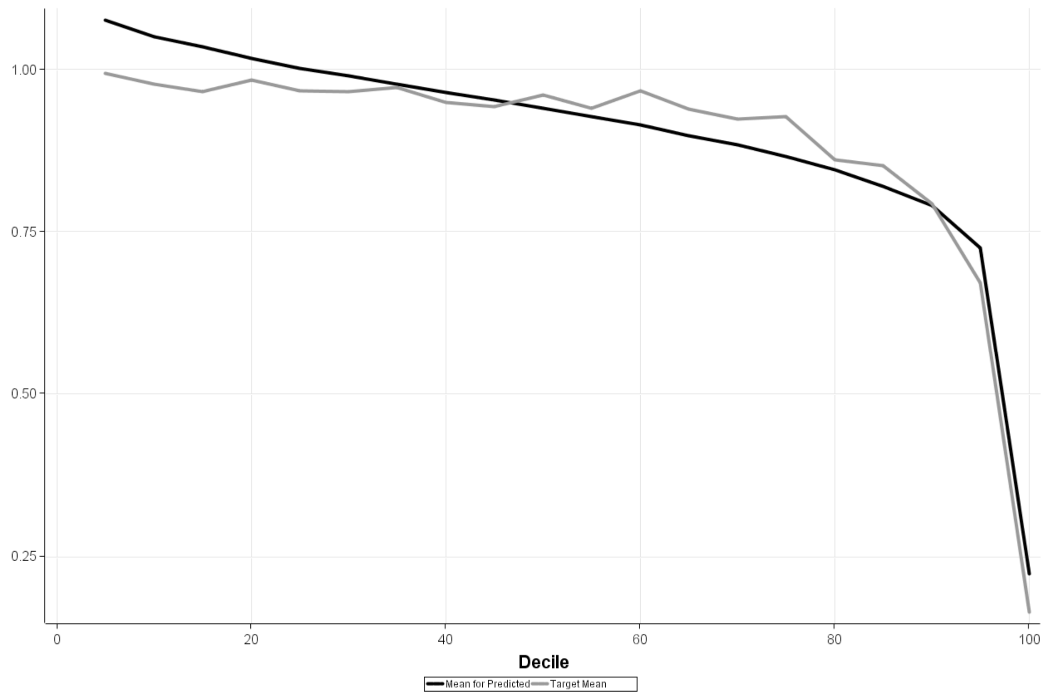
Appendix 83: NoFurtherRec: Assessment Score Rankings

Decile	Observation Number	Target Mean	Mean for Predicted
0	.	.	.
5	73.45	0.99288	1.07581
10	73.45	0.97736	1.04989
15	73.45	0.96484	1.03409
20	73.45	0.98309	1.01595
25	73.45	0.96701	1.00124
30	73.45	0.96518	0.98958
35	73.45	0.97115	0.97658
40	73.45	0.94901	0.96443
45	73.45	0.94288	0.95220
50	73.45	0.96047	0.94006
55	73.45	0.93984	0.92685
60	73.45	0.96712	0.91356
65	73.45	0.93889	0.89788
70	73.45	0.92361	0.88346
75	73.45	0.92690	0.86568
80	73.45	0.86066	0.84568
85	73.45	0.85155	0.81977
90	73.45	0.79253	0.79005
95	73.45	0.67185	0.72490
100	73.45	0.16580	0.22351

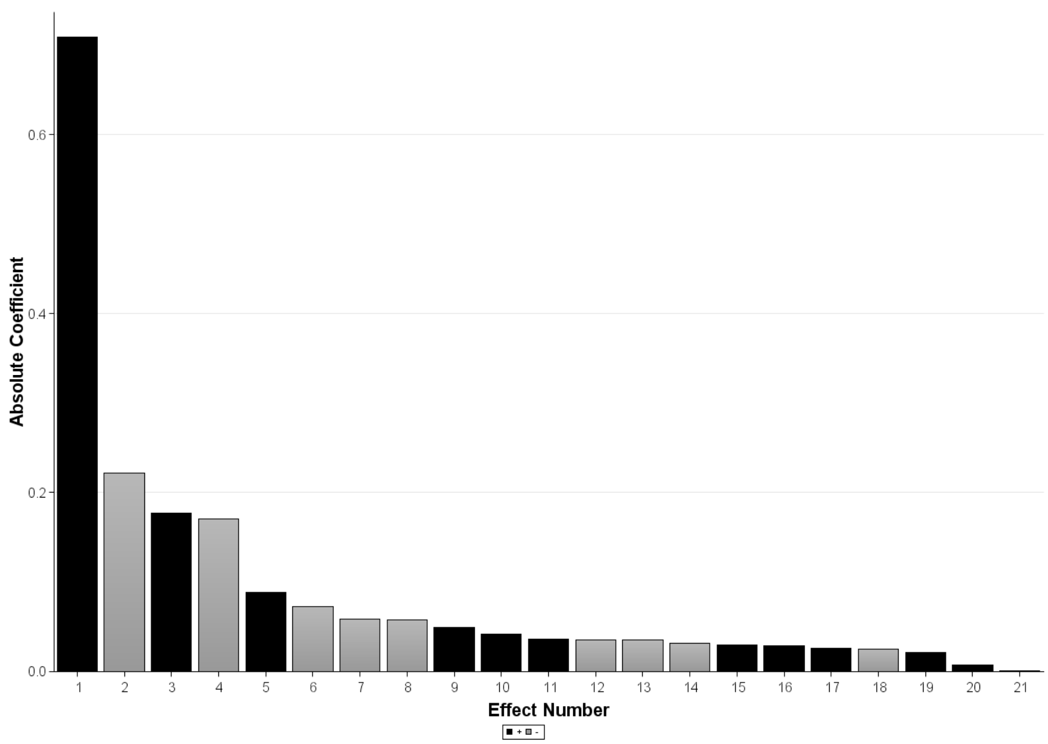
Appendix 84: NoFurtherRec: Assessment Score Distribution

Range for Predicted	Target Mean	Mean for Predicted	Number of Observations	Model Score
1.04 – 1.11	0.98459	1.05995	168	1.07609
0.96 – 1.04	0.96858	0.99637	402	0.99987
0.89 – 0.96	0.94790	0.92477	415	0.92365
0.81 – 0.89	0.88349	0.85077	257	0.84744
0.73 – 0.81	0.77405	0.77898	121	0.77122
0.66 – 0.73	0.69909	0.70778	25	0.69500
0.58 – 0.66	0.28740	0.61686	9	0.61879
0.50 – 0.58	0.41501	0.54149	10	0.54257
0.43 – 0.50	0.41334	0.47369	10	0.46635
0.35 – 0.43	0.19902	0.39261	8	0.39013
0.28 – 0.35	0.14902	0.31482	4	0.31392
0.20 – 0.28	0.02823	0.24203	2	0.23770
0.12 – 0.20	0.06576	0.15429	6	0.16148
0.05 – 0.12	0.05551	0.08693	10	0.08526
-0.03 – 0.05	0.02005	0.01383	12	0.00905
-0.11 – -0.03	0.02798	-0.04340	5	-0.06717
-0.18 – -0.11	0.05340	-0.11119	2	-0.14339
-0.26 – -0.18	0.00236	-0.25013	1	-0.21960
-0.33 – -0.26	0.05864	-0.33379	1	-0.29582
-0.41 – -0.33	0.02265	-0.41015	1	-0.37204

Appendix 85: NoFurtherRec: Score Ranking Overlay



Appendix 86: NoFurtherRec: Effects Plot (standardised)



Literature

ACHARYA, V. V. – BHARATH, S. T. – SRINIVASAN, A. [2003]: Understanding the Recovery Rates of Defaulted Securities. Centre for Economic Policy Research Discussion Papers, Working Paper, London Business School, London. Mimeo, pp. 1-35.

ACHARYA, V. V. – BHARATH, S. T. – SRINIVASAN, A. [2007]: Does Industry-wide Distress Affect Defaulted Firms? – Evidence from Creditor Recoveries. *Journal of Financial Economics*, Vol. 85, No. 3, pp. 787-821.

ACOCK, A. C. [2008]: A Gentle Introduction to Stata. 2nd Edition, Stata Press, Texas

AGRESTI, A. [2002]: Categorical Data Analysis. 2nd Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, New Jersey

AKAIKE, H. [1970]: Statistical Predictor Identification. *Annals of Institute of Statistical Mathematics*, Vol. 22, No. 1, pp. 203-217.

ALLEN, L. – DELONG, G. – SAUNDERS, A. [2004]: Issues in the Credit Risk Modeling of Retail Markets. *Journal of Banking and Finance*, Vol. 28, Issue 4, pp. 727-752.

ALLISON, P. D. [1998]: Survival Analysis Using SAS: A Practical Guide, SAS Publishing

ALTMAN, E. I. [1989]: Measuring Corporate Bond Mortality and Performance. *Journal of Finance*, Vol. 44, No. 4, pp. 909-922.

ALTMAN, E. I. [2001]: Altman High Yield Bond and Default Study. US Fixed Income High Yield Report, Salomon Smith Barney, July

ALTMAN, E. I. [2007]: Global Debt Markets in 2007: New Paradigm or Great Credit Bubble. *Journal of Applied Corporate Finance*, Summer, pp. 17-31.

ALTMAN, E. I. [2009]: Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence. Working Paper, New York University, Stern School of Business

ALTMAN, E. I. – BRADY, B. – RESTI, A. – SIRONI, A. [2005a]: The Link Between Default and Recovery Rates: Theory, Empirical Evidence and Implications. New York University, Salomon Center Working Paper Series #S-03-4, *Journal of Business*, Vol. 78, No. 6, November, pp. 2203-2227.

ALTMAN, E. – EBERHART, A. [1994]: Do Seniority Provisions Protect Bondholders' Investments? *Journal of Portfolio Management*, Summer, pp. 67-75.

ALTMAN, E. – FANJUL, G. [2004]: Defaults and Returns in the High Yield Bond Market: Analysis Through 2003. NYU Salomon Center Working Paper, January

ALTMAN, E. – HALDEMAN, R. – NARAYANAN, P. [1977]: ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance*, Vol. 1, No. 1, July, pp. 29-54.

ALTMAN, E. – KALOTAY, E. [2010]: A Flexible Approach to Modeling Ultimate Recoveries on Defaulted Loans and Bonds. 10th May 2010. URL: http://pages.stern.nyu.edu/~ealtman/FlexibleRecovery_v1.1.pdf (downloaded: 11.09.2010.)

ALTMAN, E. I. – KARLIN, B. [2009]: Defaults and Returns in the High-Yield Bond Market and Market Outlook: 2009 First-Half Report. NYU Salomon Center, Stern School of Business, August

ALTMAN, E. I. – KISHORE, V. M. [1996]: Almost Everything You Wanted to Know About Recoveries on Defaulted Bonds. *Financial Analyst Journal*, Vol. 52, No. 6, November/December, pp. 57-64.

ALTMAN, E. I. – RESTI, A. – SIRONI, A. [2001]: Analyzing and Explaining Default Recovery Rates. ISDA Research Report, London, December

ALTMAN, E. I. – RESTI, A. – SIRONI, A. [2005b]: Loss Given Default: A Review of the Literature. In: ALTMAN, E. I. – RESTI, A. – SIRONI, A. (eds.): Recovery Risk. The Next Challenge in Credit Risk Management. Risk Books, London. NYU Salomon Center and NYU Stern School of Business; Bocconi University, pp. 41-59.

ALTMAN, E. I.– RESTI, A. – SIRONI, A. [2006]: Default Recovery Rates: A Review of the Literature and Recent Empirical Evidence. *Journal of Finance Literature*, Vol. 2, Winter, pp. 21-45.

AMEMIYA, T. [1981]: Qualitative Response Models: A Survey. *Journal of Economic Literature*, Vol. 19, December, pp. 1483-1536.

AMEMIYA, T. [1985]: Advanced Econometrics. Harvard University Press. Large Sample Theory, pp. 80-104.

ANDERSON, T. W. [1984]: An introduction to multivariate statistical analysis. 2nd Edition, Wiley, New York

ARATEN, M. – JACOBS, JR., – VARSHEY, P. [2004]: Measuring LGD on Commercial Loans: An 18-Year Internal Study. *The RMA Journal*, Vol. 86, No. 8, May, pp. 28-35.

ASARNOW, E. – EDWARDS, D. [1995]: Measuring Loss on Defaulted Bank Loans: A 24 Year Study. *Journal of Commercial Lending*, Vol. 77, No. 7, March, pp. 11-23.

AZEN, S. – VAN GUILDER, M. [1981]: Conclusions regarding algorithms for handling incomplete data. Proceedings Statistical Computing Section, American Statistical Association, pp. 53-56.

BAKSHI, G. – MADAN, D. – ZHANG, F. [2001]: Recovery in Default Risk Modeling: Theoretical Foundations and Empirical Applications. Working Paper, University of Maryland
BAKSHI, G. – MADAN, D. – ZHANG, F. [2006a]: Understanding the Role of Recovery in Default Risk Models: Empirical Comparisons and Implied Recovery Rates. FDIC Center for Financial Research Working Paper, University of Maryland, 6th September 2006.
BAKSHI, G. – MADAN, D. – ZHANG, F. [2006b]: Investigating the role of systematic and firm-specific factors in default risk: Lessons from empirically evaluating credit risk models. *Journal of Business* Vol. 79, pp. 1955-1988.
BARANYI, A. – SZÉLES, ZS. [2010]: Egy hitelintézet kockázatvállalása és a bázeli szabályozás korlátai. (Risk Undertaking of a Credit Institution and the Barriers of the Basel Regulation) *Pénzügyi Szemle*, Vol. 55, No. 1, pp. 168-180.
BARCO, M. [2007]: Going Downturn. *Risk Magazine*, Vol. 20, No. 8, pp. 70-75.
BARNARD, J. – RUBIN, D. B. [1999]: Small-sample degrees of freedom with multiple imputation. *Biometrika*, Vol. 86, No. 4, pp. 949-955.
BARTLETT, M. S. [1996]: Multivariate Analysis. *Journal of the Royal Statistical Society*, Series B, Vol. 9, pp. 176-197.
BASTOS, JOÃO, A. [2009]: Forecasting bank loans loss-given-default. Working Paper, September. URL: http://cemapre.iseg.utl.pt/archive/preprints/380.pdf (downloaded: 10.08.2010.)
BCBS [1988]: International convergence of capital measurement and capital standards. Basel Committee on Banking Supervision, July 1988. URL: http://www.bis.org/publ/bcbs04a.pdf (downloaded: 28.07.2010.)
BCBS [2004]: International Convergence of Capital Measurement and Capital Standards. Revised Framework, Basel Committee on Banking Supervision, Bank of International Settlement, June 2004. URL: http://www.bis.org/publ/bcbs128.pdf (downloaded: 28.07.2010.)
BCBS [2009a]: History of the Basel Committee and its Membership. Basel Committee on Banking Supervision, Bank of International Settlement, August 2009. URL: http://www.bis.org/bcbs/history.pdf (downloaded: 28.07.2010.)
BCBS [2009b]: Enhancements to the Basel II framework. Basel Committee on Banking Supervision, July 2009. URL: http://www.bis.org/publ/bcbs157.pdf (downloaded: 28.07.2010.)
BCBS [2009c]: Revisions to the Basel II market risk framework. Basel Committee on Banking Supervision, July 2009. URL: http://www.bis.org/publ/bcbs158.pdf (downloaded: 28.07.2010.)
BCBS [2009d]: Guidelines for computing capital for incremental risk in the trading book. Basel Committee on Banking Supervision, July 2009. URL: http://www.bis.org/publ/bcbs159.pdf (downloaded: 28.07.2010.)
BCBS [2009e]: Strengthening the resilience of the banking sector. Basel Committee on Banking Supervision, Consultative document, December 2009. URL: http://www.bis.org/publ/bcbs164.pdf (downloaded: 28.07.2010.)
BCBS [2009f]: International framework for liquidity risk measurement, standards and monitoring. Basel Committee on Banking Supervision, Consultative document, December 2009. URL: http://www.bis.org/publ/bcbs165.pdf (downloaded: 28.07.2010.)
BCBS [2010]: Basel III: International framework for liquidity risk measurement, standards and monitoring. Basel Committee on Banking Supervision, December 2010. URL: http://www.bis.org/publ/bcbs188.pdf (downloaded: 25.07.2011.)
BCBS [2011]: Basel III: A global regulatory framework for more resilient banks and banking systems. Basel Committee on Banking Supervision, December 2010, Rev. June 2011. URL: http://www.bis.org/publ/bcbs189.pdf (downloaded: 31.07.2011.)
BELLOTTI, T. – CROOK, J. [2008]: Modelling and estimating Loss Given Default for credit cards. Credit Research Centre, University of Edinburgh Business School, 10th November 2008
BELLOTTI, T. – CROOK, J. [2009]: Calculating LGD for credit cards. GFRMC Conference on Risk Management in the Personal Financial Services Sector. London
BERGER, A. N. – UDELL, G. F. [1990]: Collateral, Loan Quality, and Bank Risk. *Journal of Monetary Economics*, Vol. 25, pp. 21-42.
BHATIA, M. [2006]: Credit Risk Management & Basel II. – An Implication Guide. Risk Books, Navarra
BODZÁSI, B. DR. [2010]: Az európai jelzálogpiacok integrálása és az ehhez kapcsolódó nemzetközi magánjogi kérdések. (The Integration of the European Mortgage Markets and the Relating International Common-Law Issues) *Külgazdaság*, Vol. 7, No. 3-4, Jogi melléklet, pp. 23-46.
BONFIM, D. [2009]: Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. *Journal of Banking & Finance*, Vol. 33, pp. 281-299.
BOS, R. J. – KELHOFFER, K. – KEISMAN, D. [2002]: Recovery Research. Ultimate Recovery in an Era of Record Defaults. Standard & Poor's, July
BOX, G. E. – COX, D. R. [1964]: An analysis of transformation. *Journal of Royal Statistical Society*, Series B, Vol. 26, pp. 211-246.

BRADY, B. – CHANG, P. – MIU, P. – OZDEMIR, B. – SCHWARTZ, D. C. [2007]: Discount Rate for Workout Recoveries: An Empirical Study. Social Science Research Network, Working Paper Series, August

BUSE, A. [1982]: The Likelihood Ratio, Wald, and Lagrange-Multiplier Test: An Expository Note. *The American Statistician*, August, pp. 153-157.

CALEM, P. S. – LACOUR-LITTLE, M. [2004]: Risk-Based Capital Requirements for Mortgage Loans. *Journal of Banking & Finance*, Vol. 28, pp. 647-672.

CAMPBELL, J. Y. – COCCO, J. F. [2003]: Household Risk Management and Optimal Mortgage Choice. *Quarterly Journal of Economics*, Vol. 188, November, pp. 1449-1494.

CAREY, M. – GORDY, M. [2003]: Systematic Risk in Recoveries on Defaulted Debt. Mimeo, Federal Reserve Board, Washington, DC.

CARTY, L. V. – GATES, D. – GUPTON, G. M. [2000]: Bank Loan Loss Given Default. Moody's Investor Service, November

CARTY, L. V. – LIEBERMANN, D. [1996]: Defaulted Bank Loan Recoveries. Moody's Investors Service, Moody's Special Comment, November

CASELLI, S. – GATTI, S. – QUERCI, F. [2008]: The Sensitivity of the Loss Given Default Rate to Systemic Risk: New Empirical Evidence on Bank Loans. *Journal of Financial Services Research*, Vol. 34, No. 1, pp. 1-34.

CEBS [2008]: CEBS's technical advice to the European commission on options and national discretions. Committee of European Banking Supervisors, June 2008. URL: http://www.eba.europa.eu/getdoc/354c6e4c-f22a-46a0-9025-0c55f460a5a6/2008-17-10-Final-Advice-on-options-and-national-di.aspx (downloaded: 25.07.2011.)

CEBS [2009]: CEBS's second advice on options and national discretions. Committee of European Banking Supervisors, June 2009. URL: http://www.eba.europa.eu/getdoc/8d794a1e-6ecb-4db5-97bd-55c1760095ef/CEBS-s-follow-up-advice-on-national-discretions.aspx (downloaded: 25.07.2011.)

CEBS [2010]: Results of the comprehensive quantitative impact study. Committee of European Banking Supervisors, December 2010. URL: http://www.eba.europa.eu/cebs/media/Publications/Other%20Publications/QIS/EU-QIS-report-2.pdf (downloaded: 25.07.2011.)

CHABAANE, A. – LAURENT, J. P. – SALOMON, J. [2004]: Double Impact: Credit Risk Assessment and Collateral Value. Working Paper

CHALUPKA, R. – KOPECSNI, J. [2009]: Modelling Bank Loan LGD of Corporate and SME Segment: A Case Study. Charles University in Prague, Faculty of Social Sciences, *Czech Journal of Economics and Finance*, Vol. 59, No. 4. (IES Working Paper, No. 27, 2008)

CHATFIELD, C. – COLLINS, A. J. [1980]: Introduction to Multivariate Analysis. London, Chapman & Hall

CHAVA, S. – STEFANESCU, C. – TURNBULL, S. M. [2008]: Modeling the loss distribution. Mays School of Business at Texas A&M University, London Business School , Bauer College of Business at University of Houston. Working Paper, 21st April

CLAURETIE, T. M. [1990]: A Note on Mortgage Risk: Default vs. Loss Rate. *EREZEA Journal*, Vol. 18, No. 2, pp. 202-206.

CLAURETIE, T. M. – HERZOG, T. [1990]: The Effect of State Foreclosure Laws on Loan Losses: Evidence form Mortgage Insurance Industry. *Journal of Money, Credit, and Banking*, Vol. 22, No. 2, pp. 221-233.

COLLIN-DUFRESNE, P. – GOLDSTEIN, R. – HUGONNIER, J. [2004]: A general formula for valuing defaultable securities. *Econometrica*, Vol. 72, Issue 5, pp. 1377-1407.

COX, D. R. [1970]: Analysis of Binary Data. London, Methuen

COX, D. R. [1984]: Analysis of Survival Data. Chapman & Hall, London

CRAWFORD, G. W. – ROSENBLATT, E. [1995]: Efficient Mortgage Default Option Exercise: Evidence form Loss Severity. *The Journal of Real Estate Research*, Vol. 10, No. 5, pp. 543-555.

CROOK, J. – BELLOTTI, T. – ANDREEWA, G. – ANSELL, J. [2007]: New Methods of Estimating LGD for Consumer Loans. Symposium on Risk Management in the Retail Financial Services Sector, Imperial College London

CROUHY, M – GALAI, D. – MARK, R. [2000]: A Comparative Analysis of Current Credit Risk Models. *Journal of Banking and Finance*, Vol. 24, pp. 59-117.

CROUHY, M – GALAI, D. – MARK, R. [2001]: Risk Management. McGraw-Hill, New York

CSERNYÁK, L. DR. (ed.) [1998]: Valószínűségszámítás. Matematika közgazdászoknak. (Calculus of Probability. Mathematics for Economists.) Nemzeti Tankönyvkiadó, Budapest

DAS, S. R. – HANOUNA, P. [2009]: Implied Recovery. Working Paper, Santa Clara University, 2nd May 2009

DAVIDSON, R. – MACKINNON, J. G. [1981]: Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica*, Vol. 49, pp. 781-793.

DE LAROSIÈRE, J. [2009]: The High-level group on financial supervision in the EU: Report. Brussels, 25th February 2009. URL: http://ec.europa.eu/internal_market/finances/docs/de_larosiere_report_en.pdf (downloaded: 25.07.2011.)

DE SERVIGNY, A. – OLIVER, R. [2004]: Measuring and Managing Credit Risk. McGraw Hill, Boston
DEMPSTER, A. P. – LAIRD, N. M. – RUBIN, D. B. [1977]: Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, Series B. Methodological, Vol. 39, pp. 1-38.
DENG, Y. – QUIGLEY, J. – VAN ORDER, R. [2000]: Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options. *Econometrica*, Vol. 68, March, pp. 275-307.
DERKSEN, S. – KESELMAN, J. H. [1992]: Backward, Forward, and Stepwise Automated Subset Selection Algorithms. Frequency of Obtaining Authentic and Noise Variables. *British Journal of Mathematical and Statistical Psychology*, Vol. 45, pp. 265-282.
DERMINE, J. – NETO DE CARVALHO, C. [2003]: Bank Loan Losses-given-default – Empirical Evidence. First draft, 20th October 2003, Mimeo
DERMINE, J. – NETO DE CARVALHO, C. [2005]: How to Measure Recoveries and Provisions on Bank Lending: Methodology and Empirical Evidence. In: ALTMAN, E. I. – RESTI, A. – SIRONI, A. (eds.): Recovery Risk. The Next Challenge in Credit Risk Management. Risk Books, London. NYU Salomon Center and NYU Stern School of Business; Bocconi University, pp. 101-119.
DERMINE, J. – NETO DE CARVALHO, C. [2006]: Bank loan losses-given-default: A Case Study. *Journal of Banking and Finance*, Vol. 30, Issue 4, pp. 1219-1243.
DILLON, W. R. – GOLDSTEIN, M. [1987]: Multivariate Analysis. Methods and Applications. John Wiley & Sons, Inc., New York
DONCHEV, T. [2009]: Modeling Defaults in Residential Mortgage Backed Securities: An Intensity Based Approach. NICB, The Merchant Bank of Choice, August
DRAPER, N. – SMITH, H. [1981]: Applied Regression Analysis. John Wiley & Sons, Inc., New York
DUFFEE, G. R. [1999]: Estimating the Price of Default Risk. *Review of Financial Studies*, Vol. 12, No. 1, Spring, pp. 197-225.
DUFFIE, D. – SINGLETON, K. J. [1999]: Modeling the Term Structures of Defaultable Bonds. *Review of Financial Studies*, Vol. 12, pp. 687-720.
DUFFIE, G. R. – SINGLETON, K. [2003]: Credit Risk: Pricing, Measurement, and Management. Princeton University Press
DÜLLMANN, K. – TRAPP, M. [2004]: Systematic Risk in Recovery Rates. An Empirical Analysis of US corporate credit exposures. Working Paper, University of Mannheim
DURRETT, R. [2005]: Probability: Theory and Examples. Brooks / Cole, Thomson Learning, Inc., Belmont, C.A.
EALES, R. – BOSWORTH, E. [1998]: Severity of Loss in the Event of Default in Small Business and Large Consumer Loans. *The Journal of Lending and Credit Risk Management*, May, pp. 58-65.
EC [2001]: Commission Recommendation 2001/193/EC of 1st March 2001 on pre-contractual information to be given to customers by lenders offering home loans. European Commission, Journal of the European Union L69, 10th March 2001
EC [2005]: Green Paper – Mortgage Credit in the EU. Commission of the European Communities. URL: http://eur-lex.europa.eu/LexUriServ/site/en/com/2005/com2005_0327en01.pdf (downloaded: 19.07.2011.)
EC [2007]: White Paper on the Integration of EU Mortgage Credit Markets. Commission of the European Communities. URL: http://www.mfcr.cz/cps/rde/xbcr/mfcr/White_Paper_MC.pdf (downloaded: 19.07.2011.)
EC [2008a]: Public consultation on possible changes to the Capital Requirements Directive (CRD, consisting of Directives 2006/48/EC and 2006/49/EC). European Commission. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/consultation_en.pdf (downloaded: 25.07.2011.)
EC [2008b]: Second public consultation on possible changes to the Capital Requirements Directive (CRD, consisting of Directives 2006/48/EC and 2006/49/EC). CRD potential changes on securitisation. European Commission. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/consultation2_en.pdf (downloaded: 25.07.2011.)
EC [2008c]: Proposed changes to Trading Book Capital Requirements. European Commission, Commission Services Staff Working Document. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/feedback_en.pdf (downloaded: 25.07.2011.)
EC [2009a]: Commission Directive 2009/27/EC of 7th April 2009 amending certain Annexes to Directive 2006/49/EC of the European Parliament and of the Council as regards technical provisions concerning risk management. European Commission, Journal of the European Union L94, 8th April 2009.
EC [2009b]: Commission Directive 2009/83/EC of 27th July 2009 amending certain Annexes to Directive 2006/48/EC of the European Parliament and of the Council as regards technical provisions concerning risk management. European Commission, Journal of the European Union L196, 28th July 2009.
EC [2009c]: Public Consultation regarding further possible changes to the Capital Requirements Directive (“CRD”). Possible further changes to the Capital Requirements Directive. European

Commission, Commission Services Staff Working Document, 24th July 2009. URL: http://ec.europa.eu/internal_market/consultations/docs/2009/capital_requirements_directive/CRD_consultation_document_en.pdf (downloaded: 25.07.2011.)
EC [2009d]: Proposal for a Directive of the European Parliament and of the Council amending Directives 2006/48/EC and 2006/49/EC as regards capital requirements for the trading book and for re-securitisations, and the supervisory review of remuneration policies SEC(2009) 974 final SEC(2009) 975 final. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/com2009/Leg_Proposal_Adopted_1307.pdf (downloaded: 25.07.2011.)
EC [2010a]: Possible further changes to the Capital Requirements Directive. European Commission, Commission Services Staff Working Document, 26th February 2010. URL: http://ec.europa.eu/internal_market/consultations/2010/crd4_en.htm (downloaded: 31.07.2011.)
EC [2010b]: Countercyclical Capital Buffer. European Commission, Consultation Document, 22nd October 2010. URL: http://ec.europa.eu/internal_market/consultations/docs/2010/capitalbuffer/consultation_paper_en.pdf (downloaded: 31.07.2011.)
EC [2010c]: Public consultation on possible measures to strengthen bank capital requirements for counterparty credit risk. European Commission, Consultation Document, 9th February 2011. URL: http://ec.europa.eu/internal_market/consultations/docs/2011/credit_risk/consultation_paper_en.pdf (downloaded: 31.07.2011.)
EC [2011a]: Proposal for a Directive of the European Parliament and of the Council on credit agreements relating to residential property. European Commission, Brussels, 31st March 2011. URL: http://www.europolitics.info/pdf/gratuit_en/291224-en.pdf (downloaded: 10.08.2011.)
EC [2011b]: Impact Assessment. Accompanying the document Regulation on the European Parliament and of the Council on credit agreements relating to residential property. Commission Staff Working Paper, 31st March 2011. URL: http://ec.europa.eu/internal_market/finservices-retail/docs/credit/mortgage/sec_2011_356-ia_en.pdf (downloaded: 10.08.2011.)
EC [2011c]: Proposal for a Regulation of the European Parliament and of the Council on prudential requirements for credit institutions and investment firms. European Commission, Brussels, 20th July 2011. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/CRD4_reform/20110720_regulation_proposal_part1_en.pdf; http://ec.europa.eu/internal_market/bank/docs/regcapital/CRD4_reform/20110720_regulation_proposal_part2_en.pdf; http://ec.europa.eu/internal_market/bank/docs/regcapital/CRD4_reform/20110720_regulation_proposal_part3_en.pdf (downloaded: 31.07.2011.)
EC [2011d]: Proposal for a Directive of the European Parliament and of the Council on the access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms and amending Directive 2002/87/EC of the European Parliament and of the Council on the supplementary supervision of credit institutions, insurance undertakings and investment firms in a financial conglomerate. European Commission, Brussels, 20th July 2011. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0453:FIN:EN:PDF (downloaded: 31.07.2011.)
EC [2011e]: Impact Assessment. Accompanying the document Regulation on the European Parliament and of the Council on prudential requirements for credit institutions and investment firms. Commission Staff Working Paper, 20th July 2011. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/CRD4_reform/IA_regulation_en.pdf (downloaded: 31.07.2011.)
EC [2011f]: Impact Assessment. Accompanying the document Proposal for a Directive of the European Parliament and of the Council on the access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms and amending Directive 2002/87/EC of the European Parliament and of the Council on the supplementary supervision of credit institutions, insurance undertakings and investment firms in a financial conglomerate. Commission Staff Working Paper, 20th July 2011. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/CRD4_reform/IA_directive_en.pdf (downloaded: 31.07.2011.)
EEC [1977]: Council Directive 77/780/EEC of 12th December 1977 on the coordination of laws, regulations and administrative provisions relating to the taking up and pursuit of the business of credit institutions. The Council of the European Communities. URL: http://www.gbld.org/index.asp?mode=21&country=43 (downloaded: 25.07.2011.)
EEC [1977]: Second Council Directive 89/646/EEC of 15th December 1989 on the coordination of laws, regulations and administrative provisions relating to the taking up and pursuit of the business of credit

institutions and amending Directive 77/780/EEC. The Council of the European Communities. URL: http://www.gbld.org/index.asp?mode=21&country=43 (downloaded: 25.07.2011.)
EEC [1993]: Council Directive 93/6/EEC of 15th March 1993 on the capital adequacy on investments firms and credit institutions. The Council of the European Communities, Official Journal of the European Union L141, 11th June 1993
EFRON, B. – TIBSHIRANI, R. J. [1993]: An Introduction to the Bootstrap. Chapman & Hall, New York
ELUL, R. [2006]: Residential Mortgage Default. *Business Review*, Q3, pp. 21-30.
EMERY, K. – CANTOR, R. – KEISMAN, D. – OU, S. [2007]: Moody's Ultimate Recovery Database. Moody's Investors Service, Global Credit Research, Special Comment, April
ENGELMANN, B. – RAUHMEIER, R. (eds.) [2006]: The Basel II Risk Parameters. Estimation, Validation, and Stress Testing. Springer Verlag, Heidelberg/Berlin
ENGLE, R. F. [1982]: A General Approach to Lagrangian Multiplier Diagnostics. *Annals of Econometrics*, Vol. 20, pp. 83-104.
ENGLE, R. F. [1984]: Wald, Likelihood-Ratio and Lagrangian Multiplier Tests in Econometrics. In: GRILICHES, Z. – INTRILIGATOR, M. D. (eds.): Handbook of Econometrics. Elsevier, New York. Chapter 13, pp. 775-826.
ENGLE, R. F. – BROWN, S. [1985]: Model Selection for Forecasting. J. Computation in Statistics
EPC [2000]: Directive 2000/12/EC of the European Parliament and of the Council of 20th March 2000 relating to the taking up and pursuit of the business of credit institutions. European Parliament and of the Council, Official Journal of the European Union L126, 26th May 2000
EPC [2006a]: Directive 2006/48/EC of the European Parliament and of the Council of 14th June 2006 relating to the taking up and pursuit of the business of credit institutions (recast). European Parliament and of the Council, Official Journal of the European Union L177, 30th June 2006
EPC [2006b]: Directive 2006/49/EC of the European Parliament and of the Council of 14th June 2006 on the capital adequacy of investment firms and credit institutions (recast). European Parliament and of the Council, Official Journal of the European Union L177, 30th June 2006
EPC [2008]: Directive 2008/48/EC of the European Parliament and of the Council of 23rd April 2008 on credit agreements for consumers and repealing Council Directive 87/102/EEC. European Parliament and of the Council, Official Journal of the European Union L133, 22nd May 2008
EPC [2009]: Directive 2009/111/EC of the European Parliament and of the Council of 16th September 2009 amending directives 2006/48/EC, 2006/49/EC and 2007/64/EC as regards banks affiliated to central credit institutions, certain own funds items, large exposures, supervisory arrangements, and crisis management. European Parliament and of the Council, Official Journal of the European Union L302, 17th November 2009
EPC [2010a]: Directive 2010/76/EU of the European Parliament and of the Council of 24th November 2010 amending Directives 2006/48/EC and 2006/49/EC as regards capital requirements for the trading book and for re-securitisations, and the supervisory review of remuneration policies. European Parliament and of the Council, Official Journal of the European Union L329, 14th December 2010
EPC [2010b]: Regulation (EU) No. 1093/2010 of the European Parliament and of the Council of 24th November 2010 establishing a European Supervisory Authority (European Banking Authority), amending Decision No. 716/2009/EC and repealing Commission Decision 2009/78/EC Official Journal of the European Union L331, 15th December 2010
EPC [2010c]: Regulation (EU) No. 1094/2010 of the European Parliament and of the Council of 24th November 2010 establishing a European Supervisory Authority (European Insurance and Occupational Pensions Authority), amending Decision No. 716/2009/EC and repealing Commission Decision 2009/79/EC, Official Journal of the European Union L331, 15th December 2010
EPC [2010d]: Regulation (EU) No. 1095/2010 of the European Parliament and of the Council of 24th November 2010 establishing a European Supervisory Authority (European Securities and Markets Authority), amending Decision No. 716/2009/EC and repealing Commission Decision 2009/77/EC, Official Journal of the European Union L331, 15th December 2010
EPC [2010e]: Regulation (EU) No. 1092/2010 of the European Parliament and of the Council of 24th November 2010 on European Union macro-prudential oversight of the financial system and establishing a European Systemic Risk Board Official Journal of the European Union L331, 15th December 2010
EPC [2010f]: Council Regulation (EU) No. 1096/2010 of 17th November 2010 conferring specific tasks upon the European Central Bank concerning the functioning of the European Systemic Risk Board, Official Journal of the European Union L331, 15th December 2010
EPC [2010g]: Directive 2010/78/EU of the European Parliament and of the Council of 24th November 2010 amending Directives 98/26/EC, 2002/87/EC, 2003/6/EC, 2003/41/EC, 2003/71/EC, 2004/39/EC, 2004/109/EC, 2005/60/EC, 2006/48/EC, 2006/49/EC and 2009/65/EC in respect of the powers of the European Supervisory Authority (European Banking Authority), the European Supervisory Authority (European Insurance and Occupational Pensions Authority) and the European Supervisory Authority

(European Securities and Markets Authority), Official Journal of the European Union L331, 15th December 2010
EVERITT, B. S. [2002]: The Cambridge Dictionary of Statistics. 2nd Edition, Cambridge University Press
FELSOVALYI, A. – HURT, L. [1998]: Measuring Loss on Latin American Defaulted Bank Loans: A 27-Year Study of 27 Countries. *Journal of Lending and Credit Risk Management*, Vol. 81, No. 2, pp. 41-46.
FINGER, C. [1999]: Conditional Approaches for CreditMetrics® Portfolio Distributions. CreditMetrics® Monitor, April
FISHER, R. A. [1936]: The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, pp. 179-188.
FRIDSON, M. – GARMAN, M. C. – OKASHIMA, K. [2000]: Recovery Rates: The Search for Meaning. Merrill Lynch, March
FRIEDMAN, C. – SANDOW, S. [2003]: Ultimate Recoveries. *Risk Magazine*, Vol. 16, August, pp. 69-73.
FRYE, J. [2000a]: Collateral Damage. A Source of Systematic Credit Risk. *Risk Magazine*, Vol. 13, No. 4, April, pp. 91-94.
FRYE, J. [2000b]: Collateral Damage Detected. Federal Reserve Bank of Chicago Working Paper, Emerging Issues Series, October, pp. 1-14.
FRYE, J. [2000c]: Depressing Recoveries. *Risk Magazine*, Vol. 13, No. 11, November, pp. 108-111.
FRYE, J. [2003]: A False Sense of Security. LGD in High Default Years. *Risk Magazine*, Vol. 16, No. 8, August, pp. 63-67.
FRYE, J. [2004]: Recovery Risk and Economic Capital. Federal Reserve Bank of Chicago. In: DEV, A. (ed.) [2004]: Economic Capital. A Practitioner Guide. Risk Books, pp. 49-67.
FÜSTÖS, L. – KOVÁCS, E. – MESZÉNA, GY. – SIMONNÉ, M. N. [2004]: Alakfelismerés. Sokváltozós statisztikai módszerek. (Pattern Recognition. Multivariate Statistical Methods) Új Mandátum Kiadó, Budapest
GLÖBNER, P. – STEINBAUER, A. – IANOVA, V. [2006]: International LGD Estimation in Practice. *WILMOTT Magazine*, Vol. 1, pp. 86-91.
GORDY, M. [2000]: A Comparative Anatomy of Credit Risk Models. *Journal of Banking and Finance*, Vol. 24, No. 1-2, January, pp. 119-149.
GREENE, W. H. [2003]: Econometric Analysis. 5th Edition (International student), Prentice Hall, New Jersey
GRIPPA, P. S. – IANNOTTI, F. – LEANDRI, F. [2005]: Recovery rates in the banking industry: stylised facts emerging from the Italian experience. In: ALTMAN, E. I. – RESTI, A. – SIRONA, A. (eds.): Recovery Risk. Risk Books, London, pp. 121-141.
GRUENSTEIN, J. [1995]: Predicting Residential Mortgage Defaults. Paper presented at the American Real Estate and Urban Economics Meeting, January
GRUNERT, J. – WEBER, M. [2005]: Recovery Rates of Bank Loans: Empirical Evidence for Germany. Department of Banking and Finance, University of Mannheim, Working Paper, March
GRUNERT, J. – WEBER, M. [2009]: Recovery Rates of Commercial Lending: Empirical Evidence for German Companies. *Journal of Banking & Finance*, Vol. 33, pp. 505-513.
GUPTON, G. – FINGER, C. – BHATIA, M. [1997]: CreditMetrics™, JP Morgan & Co Technical Document
GUPTON, G. M. – GATES, D. – CARTY, L. V. [2000]: Bank Loan Loss Given Default. Global Credit Research, Moody's Investor Service, November
GUPTON, G. M. – STEIN, R. M. [2002]: Losscalc™: Model for Predicting Loss Given Default (LGD). Global Credit Research, Moody's KMV, Moody's Investor Service, New York, February
GUPTON, G. M. – STEIN, R. M. [2005]: Losscalc V2: Dynamic Prediction of LGD, Modelling Methodology. Moody's KMV Company, Moody's Investor Service, January, pp. 1-44.
HAITOVSKY, Y. [1969]: A Note on the Maximization of $\overline{R}^2$. *American Statistician*, Vol. 23, February, pp. 20-21.
HAJDU, O. [2003]: Többváltozós statisztikai számítások. Statisztikai módszerek a társadalmi és gazdasági elemzésekben. (Multivariate Statistical Calculations. Statistical Methods in the Social and Economic Analyses) Központi Statisztikai Hivatal (Hungarian Central Statistical Office), Budapest
HAJDU, O. (ed.) [2004]: Statisztika III. (Statistics III) Egyetemi jegyzet, Budapest
HAMILTON, D. T. – GUPTON, G. M. – BERTHAULT, A. [2001]: Default and Recovery Rates of Corporate Bond Issuers: 2000. Moody's Investors Service, February
HAMILTON, D. T. – VARMA, P. – OU, S. – CANTOR, R. [2003]: Loss. Characteristics of Commercial Mortgage Foreclosure
HARPER, W. L. – HOOKER, C. A. (eds.) [1976]: Foundations of Probability Theory, Statistical Inference, and Statistical Theory of Science. Vol. 2, Boston, D. Riedel
HARREL, F. E. [2001]: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer Series in Statistics. Springer-Verlag, New York
HARTWIG, F. – DEARING, B. E. [1979]: Exploratory Data Analysis. Sage Publications, Beverly Hills, London

HECKMAN, J. [1976]: The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Sample Estimator for such Models. *Annals of Economic and Social Measurement*, Vol. 5, No. 4, pp. 475-492.

HFSA [2005]: A hitelintézetek és befektetési vállalkozások új tőkekövetelmény szabályaira (CRD) vonatkozó szakmai anyagok. (Professional Documents Referring to the New Capital Requirement Rules of the Credit Institutions and the Investments Firms) 2. átdolgozott változat (2nd revised Edition), Pénzügyi Szervezete Állami Felügyelete (Hungarian Financial Supervisory Authority), July. URL: http://www.pszaf.hu/data/cms179662/bazel2_konzcrd_v2.pdf (downloaded: 25.07.2011.)

HFSA [2008a]: A tőkemegfelelés belső értékelési folyamata (ICAAP). Útmutató a felügyelt intézmények részére. (Internal Capital Adequacy Assessment Process (ICAAP) – Guidelines for Supervised Institutions.) Pénzügyi Szervezetek Állami Felügyelete (Hungarian Financial Supervisory Authority), January

HFSA [2008b]: A Szolvencia II keretdirektíva-javaslat és a CRD elemeinek összehasonlítása. (Comparison of the Elements of the Directive Proposal Solvency II and of the CRD) Pénzügyi Szervezetek Állami Felügyelete (Hungarian Financial Supervisory Authority), January

HFSA [2008c]: Validációs Kézikönyv. A belső minősítésen alapuló módszerek és a működési kockázat fejlett mérési módszereinek (AMA) bevezetéséről, értékeléséről, jóváhagyásáról. I. rész: A belső minősítésen alapuló módszer. (Validation Guidelines. On the implementation, assessment and approval of Internal Ratings Based (IRB) Approaches and Advanced Measurement Approaches (AMA). Part I: International Ratings Based Approach.) Pénzügyi Szervezetek Állami Felügyelete (Hungarian Financial Supervisory Authority), June

HFSA [2009]: Összefoglaló a tőkekövetelmény direktíva módosítására vonatkozó nyilvános konzultációra bocsátott bizottsági munkaanyagról. (Summary of the Committee Working Paper Taken under Public Consultation about Modifying the Capital Requirement Directive) Pénzügyi Szervezete Állami Felügyelete (Hungarian Financial Supervisory Authority). URL: http://www.pszaf.hu/data/cms2039719/CRD_mod_konz.pdf (downloaded: 25.07.2010.)

HFSA [2010]: A hitelintézetek és befektetési vállalkozások tőkekövetelmény szabályozásának (CRD) a közelmúltban elfogadott és jelenleg folyamatban lévő uniós módosításai. (The Union's Recently Accepted or Actually Ongoing Modifications of the Capital Requirement Regulation of Credit Institutions and Investment Firms (CRD)) Pénzügyi Szervezete Állami Felügyelete (Hungarian Financial Supervisory Authority). URL: http://www.pszaf.hu/data/cms2109746/CRD_I_IV_aktualizalt_internetre.pdf (downloaded: 25.07.2010.)

HLAWATSCH, S. – OSTROWSKI, S. [2010]: Simulation and Estimation of Loss Given Default. Otto-von-Guericke-University Magdeburg, FEMM Working Paper, No. 10, Faculty of Economics and Management Magdeburg, Working Paper Series, March

HMA [2008]: LGD-adatbázis rendszerterv. (LGD Database System plan) Magyar Jelzálogbank Egyesület (Hungarian Mortgage Association), 7th March 2008

HU, Y.-T. – PERRAUDIN, W. [2002]: The Dependence of Recovery Rates and Defaults. Working Paper, BirkBeck College, Mimeo, February

HUANG, X. – OOSTERLEE, C. W. [2008]: Generalized Beta Regression Models for Random Loss-Given-Default. Delft University of Technology Report 08-10.

HULL, J. – WHITE, A. [1995]: The Impact of Default Risk on the Prices of Options and Other Derivative Securities. *Journal of Banking and Finance*, Vol. 19, pp. 299-322.

HUNYADI, L. – VITA, L. [2004]: Statisztika közgazdászoknak. Statisztikai módszerek a társadalmi és gazdasági elemzésekben. (Statistics for Economists. Statistical Methods in the Social and Economic Analyses) 3rd Edition, Központi Statisztikai Hivatal (Hungarian Central Statistical Office), Budapest

INFO-DATAX [2006]: Összbanki LGD adatbázis adatmodellje. (Data Model of the Interbank LGD Database). Tanulmány a Pénzügyi Szervezetek Állami Felügyelete „A lakossági és vállalati kockázatok minősítésének fejlesztési koncepciója, figyelemmel Bázel II. követelményrendszerére" c. pályázata keretében, Info-Datax Kft., July 2006

ISDA [2003]: European Loss Given Study, Summary Information Package. International Swaps and Derivatives Association. URL: http://www.isda.org/c_and_a/pdf/european-loss-given-default.pdf (downloaded: 02.08.2010.)

JARROW, R. A. – LANDO, D. – TURNBULL, S. M. [1997]: A Markov Model for the Term Structure of Credit Risk Spreads. *Review of Financial Studies*, Vol. 10, pp. 481-523.

JARROW, R. – PROTTER, P. [2004]: Structural vs Reduced Form Models: A New Information Based Perspective. *Journal of Investment Management*, Vol. 2, No. 2, pp. 34-43.

JARROW, R. A. – TURNBULL, S. M. [1995]: Pricing Derivatives on Financial Securities Subject to Credit Risk. *Journal of Finance*, Vol. 50, pp. 53-86.

JOHNSTON, J. [1984]: Econometric Methods. McGraw-Hill, New York

JOKIVUOLLE, E. – PEURA, S. [2003]: A Model for Estimating Recovery Rates and Collateral Haircuts for Bank Loans. *European Financial Management*, Vol. 9, No. 3, pp. 299–314.

KARDOSNÉ, V. ZS. [2010]: Várható változások az európai tőkeszabályozásban. (Expected Changes in the European Capital Regulation) *Hitelintézeti Szemle*, Vol. 9, No. 3, pp. 236-248.
KEISMAN, D. – VAN DE CASTLE, K. [1999]: Recovering Your Money: Insights into Losses from Defaults. Standard & Poor's Creditweek, June
KEISMAN, D. [2004]: Ultimate Recovery Rates on Bank Loan and Bond Defaults. Standard & Poor's, Loss Stats, New York
KEISMAN, D. – MARSHELLA, T. [2009]: Recoveries on Defaulted Debt in an Era of Black Swans. Moody's Global Corporate Finance, Moody's Investor Service, June
KENNEDY, P. [1992]: A Guide to Econometrics. Cambridge, MA, The MIT Press
KIM, I. J. – RAMASWAMY, K. – SUNDARESAN, S. [1993]: Does Default Risk in Coupons Affect the Valuation of Corporate Bonds?: A Contingent Claims Model. *Financial Management*, Vol. 22, No. 3, pp. 117-31.
KIM, J. O. – CURRY, J. [1977]: The treatment of missing data in multivariate analysis. *Sociological Methods Research*, Vol. 6, Issue 2, pp. 215-240.
KLEINBAUM, D. G. – KUPPER, L. L. – MULLER, K. E. – NIZAM, A. [1998]: Applied Regression Analyis and Other Multivariable Methods. 3rd Edition, Brooks/Cole Publishing Company, Duxbury Press, pp. 656-686.
KLUGMAN, S. A. – PANJER, H. H. – WILLMOT, G. E. [2008]: Loss Models. From Data to Decisions. 3rd Edition, John Wiley & Sons, Inc., New Jersey
KPMG [2010]: Bázel III és a szabályozási keretrendszer változása. Tények és feladatok, koncepciók és dilemmák. (The Changes of Basel III and the Regulatory Frameworks. Facts and Tasks, Concepts and Dilemmas) KPMG Tanácsadó Kft. URL: http://www.kpmg.com/HU/hu/IssuesAndInsights/ArticlesPublications/Documents/Bázel%20III%20és%20a%20szabályozási%20keretrendszer%20változása_2010_web.pdf (downloaded: 30.07.2011.)
KRÖPFL, B. – PESCHEK, W. – SCHNEIDER, E. – SCHÖNLIEB, A. [2000]: Alkalmazott statisztika. (Applied Statistics) Műszaki Könyvkiadó, Budapest
LANDO, D. [2004]: Credit Risk Modeling: Theory and Applications. Princeton University Press
LEE, S.-P. – LIU, D.-Y. [2002]: Determinants of Defaults in Residential Mortgage Payments: A Statistical Analysis. *The International Journal of Management*, Vol. 49, No. 2, June, pp. 377-389.
LEI [2010]: An assessment of the long-term economic impact of stronger capital and liquidity requirements. Basel Committee on Banking Supervision, Long-term Economic Impact Workgroup, August 2010. URL: http://www.bis.org/publ/bcbs173.pdf (downloaded: 30.07.2011.)
LEKKAS, V. – QUIGLEY, J. M. – VAN ORDER, R. [1993]: Loan Loss Severity and Optimal Mortgage Default. *Journal of the American Real Estate and Urban Economics Association*, Vol. 21, No. 4, pp. 353-371.
LEVY, A. – HU, Z. [2006]: Incorporating Systematic Risk in Recovery: Theory and Evidence. Working Paper, Moody's KMV Company, Moody's Investor Service
LINDLEY, D. V. [1957]: A Statistical Paradox. *Biometrika*, Vol. 44, No. 1-2, pp. 187-192.
LITTLE, R. J. A. – RUBIN, D. B. [2002]: Statistical Analysis with Missing Data. 2nd Edition, John Wiley & Sons, New Jersey
LONG, J. S. [1997]: Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences, Series 7, SAGE Publications, Inc.
LONGSTAFF, F. A. – SCHWARTZ, E. S. [1995]: A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *Journal of Finance*, Vol. 50, pp. 789-819.
LUCAS, A. [2006]: Basel II Problem Solving. QFRMC Workshop and Conference on Basel II & Credit Risk Modelling in Consumer Lending, Southampton
MACKINNON, J. G. [1992]: Model Specification Tests and Artificial Regression. *Journal of Economic Literature*, Vol. 30, pp. 102-145.
MACLACHLAN, I. [2005]: Choosing the Discount Factor for Estimating Economic LGD. In: ALTMAN, E. I. – RESTI, A. – SIRONI, A. (eds.): Recovery Risk. The Next Challenge in Credit Risk Management. Risk Books, London, pp. 285-305.
MADDALA, G. S. [1983]: Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, Cambridge University Press, Section 2, Discrete Regression Models
MADDALA, G. S. [2004]: Bevezetés az ökonometriába (Introduction to Econometrics). Nemzetközi Tankönyvkiadó, Budapest. Source: MADDALA, G. S. [2001]: Introduction to Econometrics, John Wiley & Sons, Ltd.
MAG [2010]: Assessing the macroeconomic impact of the transition to stronger capital and liquidity requirements – Interim Report. Basel Committee on Banking Supervision, Macroeconomic Assessment Group, December 2010. URL: http://www.bis.org/publ/othp12.htm (downloaded: 30.07.2011.)
MANSKI, C. F. [1991]: Regression. *Journal of Economic Literature*, American Economic Association, Vol. 29, No. 1, March, pp. 34-50.
MAYS, E. (ed.) [1998]: Credit Risk Modeling. Desing and Application. Glenlake Publishing Company, Ltd., Amacom, American Management Association, New York

MCNABB, H. – WYNN, A. [2000]: Principles and Practice of Consumer Credit Risk Management. Institute of Financial Services, CIB Publishing
MEDVEGYEV, P. [2002]: Valószínűségszámítás. Fejezetek a matematikai analízisből és a valószínűségszámításból. (Calculus of Probability. Chapters of the Mathematical Analysis and the Calculus of Probability) Vol. 1, Aula Kiadó, Budapest
MELNICK, E. L. – EVERITT, B. S. (eds.) [2008]: Encyclopedia of Quantitative Risk Analysis and Assessment. John Wiley & Sons, Inc.
MERTON, R. C. [1973]: Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, Vol. 4, pp. 41-83.
MERTON, R. C. [1974]: On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, Vol. 29, pp. 449-471.
MIU, P. – OZDEMIR, B. [2006]: Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD Correlations. *Journal of Credit Risk*, Summer, Vol. 2, No. 2, pp. 43-68.
MOOD, A. M. – GRAYBILL, F. A. – BOES, D. C. [1974]: Introduction to the Theory of Statistics. McGraw-Hill, New York
MORAL, G. – GARCÍA-BAENA, R. [2002]: LGD Estimates in a Mortgage Portfolio. Estabilidad Financiera, Banco de España, Vol. 3, pp. 127–164. (Financial Stability Review)
MORAL, G. – OROZ, M. [2002]: Interest Rates and LGD Estimates. Manuscript
MORRISON, D. F. [1967]: Multivariate Statistical Methods. McGraw-Hill, New York
MYERS, R. H. [1990]: Classical and Modern Regression with Applications. PWS-KENT Publishing Co., Boston
NELDER, J. A. – WEDDERBURN, R. W. [1972]: Generalized linear models. *Journal of the Royal Statistical Society*, Series A, Vol. 135, No. 3, pp. 370-384.
NEROV, M. – PRESS, S. J. [1973]: Univariate and Multivariate Log-Linear and Logistic Models. Report R-1306-EDA/NIH, Rand Corporation, Santa Monica, CA, December
NETER, J. – WASSERMAN, W. – KUTNER, M. H. [1990]: Applied Linear Statistical Models. IRWIN Inc., Boston
OENB. [2004]: Guidelines on Credit Risk Management. Rating Models and Validation. Oesterreichische Nationalbank. URL: http://www.oenb.at/en/img/rating_models_tcm16-22933.pdf (downloaded: 02.08.2010.)
O'SHEA, S. – BONELLI, S. – GROSSMAN, R. [2001]: Bank Loan and Bond Recovery Study: 1997-2000. Fitch Loan products special report, IBCA, March
PAPKE, L. E. – WOOLDRIDGE, J. M. [1996]: Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. *Journal of Applied Econometrics*, Vol. 11, No. 6, pp. 619-632.
PAULOVICS, O. [2005]: LGD modellezés elméletben és gyakorlatban. (LGD Modelling in Theory and in Practice) *Hitelintézeti Szemle*, Vol. 4, No. 5-6, pp. 63-83.
PAULOVICS, O. [2006]: Hitelezési-veszteség modellek. (Credit Loss Models) *Development and Finance*. Vol. 3, No. 3, pp. 74-81.
PETER, C. [2006]: Estimating Loss Given Default – Experiences from Banking Practise. In: ENGELMANN, B. – RAUHMEIER, R. (eds.): The Basel II Risk Parameters: Estimation, Validation, and Stress Testing. Springerlink, Budapest, pp. 143-175.
PLACKETT, R. L. [1981]: The Analysis of Categorical Data. Griffin, London
POLÍVKA, J. [2008]: LGD Parameter Scoring using Beta Regression Model. Conference Paper, Ostrava, 11-12.
PYKHTIN, M. [2003]: Unexpected Recovery Risk. *Risk Magazine*, August, Vol. 16, No. 8, pp. 74-78.
PWC [2004]: PricewaterhouseCoopers study on the consequences of the draft proposed new capital requirements for banks and investment firms in the European Union. PricewaterhouseCoopers, 8th April 2004. URL: http://ec.europa.eu/internal_market/bank/docs/regcapital/studies/2004-04-basel-impact-study_en.pdf (downloaded: 30.07.2011.)
QI, M. [2005]: Survey of Research on Downturn LGD. Basel Accord Implementation Group Work Paper
QI, M. – YANG, X. [2009]: Loss Given Default of High Loan-to-value Residential Mortgages. Risk Management Research Report, *Journal of Banking & Finance*, 33, pp. 788-799. (QI, M. – YANG, X. [2007]: Loss Given Default of High Loan-to-value Residential Mortgages. Economics and Policy Analysis working Paper, No. 4, August)
QUERCI, F. [2005]: Loss Given Default on a medium-sized Italian bank's loans: an empirical exercise. European Financial Management Association, Milan, Italy
RAMANATHAN, R. [1993]: Statistical Methods in Econometrics. Academic Press, San Diego
RAMANATHAN, R. [2003]: Bevezetés az ökonometriába, alkalmazásokkal. (Introductory Econometrics with Applications) Panem Kiadó, Budapest. (Source: RAMANATHAN, R. [2002]: Introductory Econometrics with Applications. 5th Edition, Harcourt College Publishers, New York)
RAMSEY, J. B. [1969]: Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of Royal Statistical Society*, Series B, Vol. 31, pp. 350-371.

RENAULT, O. – SCAILLET, O. [2004]: On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities. *Journal of Banking and Finance*, Vol. 28, No. 12, pp. 2915-2931.
ROCHE, J. – BRENNAN, W. – MCGIRT, D. – VERDE, M. [1998]: Bank Loan Ratings. In: FABOZZI, F. J. [1998]: Bank Loans: Secondary Market and Portfolio Management. Frank J. Fabozzi Associates, New Hope, Pennsylvania, pp. 57-70.
ROSENBERG, E. – GLEIT, A. [1994]: Quantitative Methods in Credit Management: A Survey. *Operations Research*, Vol. 42, pp. 589-613.
ROTH, P. L. – SWITZER, F. S. [1995]: A Monte Carlo Analysis of Missing Data Techniques in a HRM Setting. *Journal of Management*, Vol. 21, No. 5, pp. 1003-1023.
RUBIN, D. B. [1976]: Inference and Missing Data. *Biometrica*, Vol. 63, pp. 581-593.
SABATO, G. – SCHMID, M. M. [2008]: Estimating Conservative Loss Given Default. 10th September. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1136762 (downloaded: 03.09.2010.)
SACHS, L. [1982]: Applied Statistics. A Handbook of Techniques. Springer Verlag, New York – Heidelberg – Berlin
SCHILLER, R. – WEIS, A. [1997]: Evaluating Real Estate Valuation Systems. Paper presented at the American Real Estate and Urban Economics Meeting, January
SCHLAFER, J. L. [1997]: Analysis of Incomplete Multivariate Data. Chapman & Hall, London
SCHUERMANN, T. [2005]: What Do We Know About Loss Given Default? In: ALTMAN, E. – RESTI, A. – SIRONI, A. (eds.): Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books, London, pp. 3-24.
SOÓS, J. [2011]: Az új európai pénzügyi felügyeleti struktúra. (The New Structure of the European Financial Supervision) *Európai Tükör*, No. 2, pp. 126-140.
SHARMA, S. [1996]: Applied Multivariate Techniques. John Wiley & Sons, Inc., New York
SHIBATA, R. [1981]: An Optimal Selection of Regression Variables. *Biometrica*, Vol. 68, pp. 45-54.
SILVEY, D. S. [1959]: The Lagrange Multiplier Test. *Annals of Mathematical Statistics*, Vol. 30, pp. 389-407.
SMITH, L. – SANCHEZ, S. – LAWRENCE, E. [1999]: A comprehensive model for managing credit risk on home mortgage portfolios. *Decision Sciences*, Vol. 27, No. 2, pp. 291-317.
SPANOS, A. [1999]: Probability Theory and Statistical Inference. Econometric Modeling with Observational Data. Cambridge University Press, pp. 77-190.
SZÁJER, J. [2010]: Az Európai Parlament megerősített szerepe Lisszabon után. Átruházott és végrehajtási jogkörök – újabb lépés egy demokratikusabb Európai Unió felé. (The Strengthened Role of the European Parliament after Lisbon. Delegated and Executive Powers – a Further Step Towards a More Democratic European Union) Európai Tükör, Vol. 10, No. 9, pp. 27-31.
SZOMBATI, A. [2010]: Bázel III. rendszerszintű hatásai itthon és Európában. (The System Level Impacts of Basel III in Hungary and in Europe) MNB-szemle, Magyar Nemzeti Bank, December, pp. 33-42.
TAJTI, ZS. [2010]: An application of historical internal and external data for loss given default calculation. 7th International Conference of PhD Students, University of Miskolc, 8-12th August, pp. 131-136.
TAJTI, ZS. [2011]: A bázeli ajánlások és a tőkemegfelelési direktíva (CRD) formálódása. (Changes of the Basel Recommendations and of the Capital Adequacy Directive (CRD) *Hitelintézeti Szemle*, Vol. 10, No. 5, pp. 499-519.
TERTÁK, E. DR. [2010]: Változások a bankszabályozásban. (Changes in the Bank Regulation) 48. Közgazdász vándorgyűlés, Szeged, 1st October. URL: http://www.mkt.hu/docs/2010-10-03-11-09-09-Tertak_Elemer.ppt (downloaded: 31.07.2011.)
THEIL, H. [1971]: Principles of Econometrics. John Wiley & Sons, Inc., New York
THOMAS, L. C. – MATUSZYK, A. – MOORE, A. [2007a]: Collections policy comparison in LGD modelling. 3rd September, URL: http://www.management.soton.ac.uk/research/publications/CRR-09-03.pdf (downloaded: 09.08.2010.)
THOMAS, L. C. – MUES, C. – MATUSZYK, A. [2007b]: Modelling LGD for unsecured personal loans: Decision tree approach. CORMSIS WP 07/07, School of Management, University of Southampton. *Journal of Operational Research Society*, Vol. 61, March, pp. 393-398.
THORNBURN, K. [2000]: Bankruptcy auctions: costs, debt recovery and firm survival. *Journal of Financial Economics*, Vol. 58, No. 3, pp. 337-368.
TOBIN, J. [1958]: Estimation of Relationships for Limited Dependent Variables. *Econometrica*, Vol. 26, pp. 24-36.
TUKEY, J. W. [1977]: Explanatory Data Analysis. Addison-Wesley, Tukey
UNAL, H. – MADAN, D. – GÜNTAY, L. [2001]: Pricing the Risk of Recovery in Default with APR violation. University of Maryland, August
VERDE, M. [2003]: Recovery Rates Return to Historic Norms. Fitch Ratings, September
WEBSTER, A. [1992]: Applied Statistics for Business and Economics. Broadley University, Richard D. Irwin, Inc., Homewood IL 60430, Boston MA 02116
WEISBERG, S. [1985]: Applied Linear Regression. 2nd Edition, John Wiley & Sons, Inc., New York

WILSON, T. C. [1998]: Portfolio Credit Risk. Federal Reserve Board of New York, Economic Policy Review, October, pp. 71-82.
WITTEN, I. H. – FRANK, E. [2005]: Data Mining. Practical Machine Learning Tools and Techniques. 2nd Edition, Elsevier, Morgan Kaufmann Publishers, Inc.
WOOLDRIDGE, J. M. [2009]: Introductory Econometrics: A modern approach. 4th Edition (International student), South-Western. Multiple Regression Analysis: Estimation. Chapter 3, pp. 68-113.; More on Specification and Data Issues, Chapter 9, pp. 300-337.; Limited Dependent Variable Models and Sample Selection Corrections, Chapter 17, pp. 574-622.
ZHANG, Z. [2009a]: Recovery Rates and Macroeconomic Conditions: The Role of Loan Covenants. MPRA, Munich Personal RePEc Archive, Paper No. 17521, Boston College, 2nd September 2009
ZHANG, J. – THOMAS, L. C. [2009b]: Comparison of single distribution and mixture distribution models for modelling LGD. Quantitative Financial Risk Management Centre, School of Management, University of Southampton, CRR-09-04

Laws

Act CXII of 1996 on Credit Institutions and Financial Enterprises *(Hpt. – 1996. évi CXII. törvény a hitelintézetekről és a pénzügyi vállalkozásokról)*
Government Decree No. 244/2000 on the Rules for Specifying the Capital Requirement Necessary as Collateral for Trading Book Positions and Risks and Currency Exchange Rate Risks and on the Detailed Rules for Maintaining a Trading Book *(Kkr. – 244/2000. (XII.24.) Kormányrendelet a kereskedési könyvben nyilvántartott pozíciók, kockázatvállalások, a devizaárfolyam kockázat és nagykockázatok fedezetéhez szükséges tőkekövetelmény megállapításának szabályairól és a kereskedési könyv vezetésének részletes szabályairól)*
Government Decree No. 250/2000 on the Characteristics of the Annual Reporting and Bookkeeping Obligations of Credit Institutions and Financial Enterprises *(250/2000. (XII.24.) Kormányrendelet a hitelintézetek és a pénzügyi vállalkozások beszámoló készítési és könyvvezetési kötelezettségeinek sajátosságairól)*
Government Decree No. 381/2007 on the Management of Credit Institution Counterparty Risk *(381/2007. (XII.23.) Kormányrendelet a hitelintézet partnerkockázatának kezeléséről)*
Government Decree No. 196/2007 on the Management and Capital Requirement of Credit Risk *(Hkr. – 196/2007. (VII.30.) Kormányrendelet a hitelezési kockázat kezeléséről és tőkekövetelményéről)*
Government Decree No. 200/2007 on the Management and Capital Requirement of Operational Risk *(Mkr. – 200/2007. (VII.30.) Kormányrendelet a működési kockázat kezeléséről és tőkekövetelményéről)*