UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

DIPLOMA PAPER

EXPECTED LOSS CALCULATION IN BANKING INDUSTRY: FACTORS DETERMINING LOSS GIVEN DEFAULT

Ljubljana, May 2007 JURE POLJŠAK

STATEMEMT

I, Jure Poljšak, student of the Faculty of Economics, state that I am the author of this diploma paper, whose mentor was dr. Marko Košak. I do not permit this paper to be published on the faculty's web pages.

In Ljubljana, May 7th 2007

Signature:

Table of content

List of Acronyms

CAD	capital adequacy directive
CRD	capital requirements directive
CRM	credit risk mitigation
DM	default mode models
EAD	exposure at default
GLM	generalised linear models
LGD	loss given default
MTM	mark-to-market models
NACE	Classification of Economic Activities in the European Community
PD	probability of default
PDF	probability density function
RDS	reference data set
RR	recovery rate
VAR	value-at-risk

1. INTRODUCTION

The subject of credit risk management has recently emerged as perhaps the principal and most challenging area of risk management in financial markets. This prominence has been motivated by a number of factors in the last decade, including the European Commission's introduction of and debate involving the new Basel II Capital Adequacy Directive (CAD III). This new accord on bank capital allows every bank to use its internal models to estimate loss given default (LGD). Banks are thus encouraged to assess the risk inherent of each individual credit agreement in more detail and scrutinise more closely the borrower's future ability to repay the debt. On the basis of this closer scrutiny, banks will take their loan decisions – to approve or reject an application and, if it approves it, at what price – in a more differentiated manner.

This diploma paper contributes to our understanding of bank loan credit risk by providing a framework to analyse the loss severity rate after a default event. This is then applied to a set of data obtained from bank x[1] concerning losses on loans to small and medium-size firms (SME) over the 2001-2005 period. It provides an understanding of the determinants of recovery rates and tests them empirically. Moreover, it gives information on the direct costs incurred by a bank in recoveries on bad and doubtful loans. Finally, to the best of the author's knowledge, this is the first empirical paper on bank loan LGD in Slovenia.

Accurate LGD estimates of defaulted facilities are important for provisioning reserves for credit losses, calculating risk capital and determining fair pricing for credit-risky obligations. The assessment of loss changes may lead to lower capital requirements in following years as far as that is supported by sound empirical evidence.

The structure of the paper is as follows. In Section 1 it is argued that with the new Basel II standards more quantitative approaches to estimating credit risk will come into force in the banking sector. This justifies the need and motivation for calculating LGD on bank loans. Section 2 exposes LGD, its methodology and the approaches to estimating it according to Basel II. In Section 3 a literature review, recent empirical studies, along with the correlation of PD and LGD are presented, adding some referential values of studies covering LGD on loans. Section 4 justifies the importance of data management with recommendations on which data to collect in data warehouses, pointing out collateral. The preparedness of Slovenian banks and their data availability, justified and proved in the case of a Slovenian bank, is presented in Section 5. The econometrical background required for these kinds of models is presented in Section 6, whereas Section 7 describes the research part of the paper with empirical results and the model. Section 8 reports estimates of direct costs incurred in recovery and is followed by the conclusion in the last section.

---

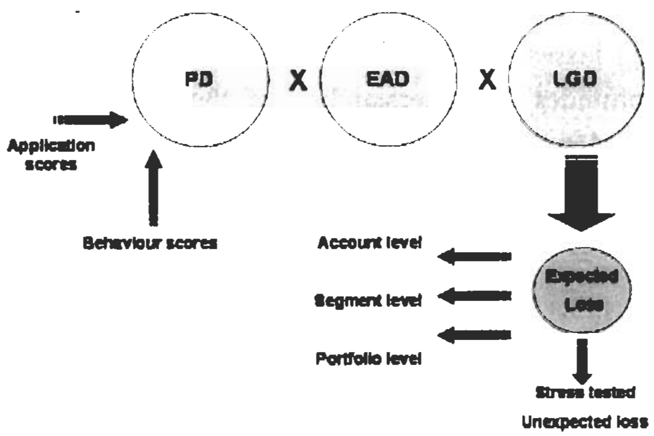
[1] Identity not disclosed: bank x

2. DETERMINATION OF LOSS GIVEN DEFAULT (LGD)

The three main variables affecting credit risk in the banking industry are: (1) exposure at default (EAD); (2) the probability of default (PD); and (3) the loss given default (LGD). The latter represents the main part of the research in this paper.

The following schematic (Figure 1) summarises the relationship between the model types required for Basel II. The biggest banks in the world have already implemented appropriate models to better manage the credit risk of their portfolios. By adopting the new Basel standards Slovenian banks have also focused their attention on measuring and managing risks stemming from credit exposure. This paper is focused on the most sophisticated part of the expected loss calculation which enables a bank to use Advanced IRB approaches, namely an LGD estimation for SMEs.

Figure 1: Models required for Basel II



Source: Strengthening the Czech banking sector – Application of Basel II, 2005, p.22.

A bank having all the required information in terms of PD, EAD and LGD estimates simply multiplies the three values and gets the expected loss either at the account, segment or portfolio level (see Figure 1). Increasing the accuracy of LGD and PD estimates improves the precision of the expected credit loss and consequently also the precision of both regulatory and economic capital allocation. In such a framework, provisions for credit losses should cover the expected losses, while economic capital is seen as a cushion for unexpected loss.

For the purpose of better understanding what loss given default indicates and what its characteristics are, some stylised facts about recoveries and losses from surveying the academic literature and practitioners should be pointed out (Schuermann, 2004):

- Most of the time, recovery as a percentage of exposure is either relatively high (around 70-80%) or low (around 20-30%). The recovery or loss distribution is said to be bimodal.
- The most important determinants of which mode a defaulted claim is likely to fall into are whether or not it is secured and its place in the capital structure of the obligor (bank loans typically have higher recovery rates than bonds).

- Recoveries are lower in recessions.
- The industry of the obligor seems to matter: tangible asset-intensive industries have higher recovery rates than service sector firms, with some exceptions such as high tech and telecom industries.
- There is no strong effect of the size of exposure on recoveries.

2.1 DEFINITIONS OF DEFAULT AND LOSS

Default and loss are the two main issues in the term loss given default (LGD) and therefore I provide their definitions as used within Basel II.

Typically a default occurs when any of the following conditions are met (Basel Committee on Banking Supervision, 2004, par. 453):

- a loan is placed on non-accrual;
- a charge-off has already occurred;
- the obligor is more than 90 days past due; and
- the obligor has filed for bankruptcy.

Paragraph 452 of the same document defines that default has occurred with regard to a particular obligor when either or both of the following events have taken place:

- the bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held); and
- the obligor is past due more than 90 days on any material credit obligations to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than what is currently outstanding.

If an obligor is in default, this affects all of their facilities.

When it comes to loss in estimating LGD, we talk about economic loss (Basel Committee on Banking Supervision, 2004, par. 460). Besides redemption, both the European Commission and the Basel Committee also include under economic loss material discount effects as well as the direct and indirect material costs associated with collecting on a collateral instrument. In other words, LGD includes three types of losses (Schuermann, 2004):

- the loss of principal;
- the carrying costs of non-performing loans, e.g., interest income foregone; and
- workout expenses (collections, legal etc).

Considering the aforementioned, the formula for the discounted recovery rate (RR), which can be defined also as 1 – LGD, can be represented as follows (Franks, Servigny, Davydenko, 2004, p. 29):

$$\text{Discounted recovery rate (RR)} = \frac{PV(CashFlow)}{EAD} = \frac{PV(R + M - C)}{P + I}$$

where:

- Cash Flow = proceeds of at-default principal, arrears of interest and fee payments;
- EAD = exposure at default;
- R = cash recovered from the customer, including any collateral and personal guarantees;
- M = market value of the received non-cash securities;
- C = direct costs attributable to recovery of the loan facility or a company's debt exposure by the bank, e.g. legal fees, costs of sale of collateral etc.;
- P = principal outstanding at the default date of debt exposure; and
- I = interest owing, but unpaid at the default date.

2.2 METHODOLOGY

There are two main classification criteria of the methods for obtaining the estimated LGD based on the type of input used (Basel Committee on Banking Supervision, 2005, p. 61). One possibility is to use *subjective methods*. These methods are based on the experience of experts and their judgement. Using this evaluation, the value of collateral, its prospective price in the market, the obligor's willingness and potential to repay as well as the estimated duration of the workout process are assessed. This method is useful for portfolios with almost no defaults or as an interim solution until internally developed objective models are fully implemented and tested. The second class of methods, so-called *objective methods*, uses empirical data on losses as the main input to calculate the expected LGD. They are based on more mathematical and statistical modelling. The objective methods for calculating LGD can be divided into four classes (Basel Committee on Banking Supervision, 2005, p. 62):

- **market LGD**: observed from market prices of defaulted bonds or marketable loans soon after the actual default event. It is used by most rating agencies' studies;
- **workout LGD**: the set of estimated cash flows resulting from the workout or collection process, properly discounted, and the estimated exposure;
- **implied market LGD**: LGDs derived from risky (but not defaulted) bond pricing using a theoretical asset-pricing model. This method is helpful in the case of portfolios such as sovereigns, banks or large corporates, for which market credit spreads are available; and
- **implied historical LGD** for retail portfolios.

From the point of view of the situation in the Slovenian banking sector, where loans still play the most important role in financing, the method workout LGD is the most applicable to SMEs (see Table 1). Attention needs to be paid to the timing of the cash flows from the distressed asset. Measuring this timing impacts on downstream estimates of the realised LGD. The cash flows should be discounted with an appropriate discount rate, which is a subject of the discussion and reported on in the section on the Advanced IRB Approach (see page 8).

Table 1: Classification of objective methods to obtain LGDs

Source	Measure	Type of facilities in the RDS		Most applicable to
		Defaulted facilities	Non-defaulted facilities	
Market values	Price differences	Market LGD		Large corporate, sovereigns, banks
	Credit spreads		Implied market LGD	Large corporate, sovereigns, banks
Recovery and cost experience	Discounted cash flows	Workout LGD		Retail, SMEs, large corporate
	Historical total losses and estimated PD	Implied historical LGD		Retail

Source: Basel Committee on Banking Supervision, 2005, p. 62.

The measure used for the level of recovery in default is ultimate recovery, which is the amount the debt holder will eventually recover after the default is resolved. Ultimate recoveries are usually discounted to reflect the time value of money (Franks, Servigny, Davydenko, 2004, p. 29).

2.3 *DEFINING LGD: THE BASEL II PERSPECTIVE*

The Basel Committee has designed three different approaches to LGD Measurement (Basel Committee on Banking Supervision, 2004). Classified from the simplest to the most complex and most flexible approaches, these are:

- the Standardised Approach;
- the Foundation IRB Approach; and
- the Advanced IRB Approach.

2.3.1 The Standardised Approach

The new Standardised Approach is comparable with the existing approach from 1989. The main difference lies in the greater distinction between risks for corporates, which brings better recovery expectations that help explain the most favourable treatment designed for some specific portfolios and collateral. It is based on a very compact set of predefined risk weights and does not require a bank to produce any explicit estimate of their recovery rates. LGD does not enter into capital computation in a direct and explicit way. The existing regulations predict a uniform weight of 100% for all corporate loans. The new Standardised Approach's weights depend on the ratings given to corporates by rating agencies. They range from 20% for the group with the best rating to 150% for the most risky corporates. Enterprises that have no external ratings, which represent the majority of European small and medium-sized enterprises, will have the existing weight of 100% for their unrated corporate loans.

The portfolios commanding lower risk weights comprise exposures secured by residential property and by commercial real estate. Loans collateralised by mortgages on residential property that is or will be occupied by the borrower or that is rented are risk-weighted at 35%. The risk weight for loans secured by commercial real estate may be as low as 50%. Also some widely used financial instruments, whose value can easily be marked to market, are accepted as collateral that can reduce total capital requirements. Eligible financial collaterals include: cash and deposits issued by a lending bank, bonds with a rating of at least BB, listed senior bonds issued by a bank, equities included in a main index or listed on a recognised stock exchange, mutual funds having a daily quoted price and investing in the mentioned instruments (Basel Committee on Banking Supervision, 2004, par. 145).

The impact of financial collateral on credit risk can be quantified with either the simple or comprehensive approach (Basel Committee on Banking Supervision, 2004, par. 129-138). In the simple approach, the portion of exposure covered by recognised collateral receives the risk weight applicable to the collateral itself, not to the original borrower, with the floor of 20%. In the comprehensive approach, no capital requirement is applied to the collateralised portion of the exposure but the value of the collateral (C) must be reduced by a haircut (Hc), reflecting the risk that the market value of the financial instrument may decrease before it is revaluated. The exposure amount after the risk mitigation, which is mainly driven by the level and type of collateralisation, is calculated as follows (Basel Committee on Banking Supervision, 2004, par. 147):

$$E^* = \max \{0, [E \times (1 + He) - C \times (1 - Hc - Hfx)]\}$$

where:

- E^* = the exposure value after risk mitigation;
- E = current value of the exposure;
- He = a haircut appropriate to the exposure –in cases where the loan is not issued in cash (e.g., in the case of securities lending) and its value may increase over time;
- C = the current value of the collateral received;
- Hc = a haircut appropriate to the collateral; and
- Hfx = a haircut appropriate for a currency mismatch between the collateral and exposure.

2.3.2 The Foundation IRB Approach

In the Foundation IRB Approach, LGD enters capital computations in a direct and explicit way. It enters as one of the four parameters (PD, LGD, EAD, maturity) on which capital requirements are based. While PDs should be estimated internally by banks, all other variables including LGD have been set by regulators.

The basic LGD is fixed at 45% for all senior unsecured exposures, all the exposures that are exclusively senior claims on corporate borrowers (loans, senior secured bonds etc). This value can be raised up to 75% for subordinated exposures depending on the collateral pledged to secure the loan. A set of rules that quantify the effect of financial and non-financial collaterals is introduced. The same haircut system as seen in the Standardised Approach is also widely used in the Foundation Approach of IRB. In this case the haircuts are not applied to the value of exposure, but directly to LGDs.

As far as financial instruments are concerned, the formula for computing the effective loss given default (LGD*) applicable to a collateralised transaction can be expressed as follows (Basel Committee on Banking Supervision, 2004, par. 291):

$$LGD^* = LGD \times (E^*/E)$$

where:

- LGD = basic LGD for all senior unsecured exposures before the recognition of collateral (45%);
- E = current value of the exposure; and
- E* = exposure value after risk mitigation as determined in the Standardised Approach.

Regarding non-financial assets, three different categories of collateral are accepted, namely receivables, real estate and other collateral (including physical capital, but excluding any assets acquired by the bank as a result of a loan default). The borrower's risk must not be directly dependent on the performance of the property serving as collateral. Moreover, the collateral has to be revaluated at least once a year, based on a market fair price. For the so-called IRB collaterals (see Table 2), the haircuts are replaced by a system of minimum and maximum thresholds (Tmin and Tmax) that help compute the adjusted LGD (LGD*) in the following way (Basel Committee on Banking Supervision, 2004):

$$LGD^* = \left\{ 45\% - \frac{\min\left(C/E, T\max\right)}{T\max}(45\% - LGD\min)\ ;\ \text{if } C/E \geq T\min \right\}$$

or

$$LGD^* = \{ \quad 45\% \quad ;\ \text{if } C/E \leq T\min \}$$

Where:

- LGDmin = minimum value that can be attained by the adjusted LGD, when $C/E \geq T\max$

- All other symbols retain the same meanings as described in the Standardised Approach (see page 6).

Table 2: Key parameters for the computation of LGD when IRB collateral is available

Type of IRB collateral	T_{min} (%)	T_{max} (%)	LGD_{min} (%)
Receivables	0	125	35
Commercial and residential real estate	30	140	35
Other collateral	30	140	40

Source: Basel Committee on Banking Supervision, 2004.

2.3.3 The Advanced IRB Approach

Banks adopting the Advanced IRB Approach will be allowed to use their own estimates of LGDs provided they can demonstrate to their supervisors that their internal models are conceptually sound and consistent with their historical experience. To demonstrate this, data on historical recovery rates must be collected and archived. This includes data on different components of the recoveries experienced on defaulted exposures; for instance amounts recovered, source of recovery (collateral and guarantees, type of liquidation and bankruptcy procedure), time period elapsed before the actual recovery and administrative costs (Resti, Sironi, 2005). All of the relevant information must be retained on a single-facility basis.

The Basel Committee also requires that the LGD estimates produced by banks are long-run estimates. This implies that such estimates cannot be lower than the long-run default-weighted average LGD rate based on the average economic loss of all observed defaults (Basel Committee on Banking Supervision, 2004, par. 468). This means we have to take into consideration the actual number of defaults each year along with the actual LGD rate of each year by calculating the average LGD.

The main difference between the two IRB approaches to LGD lies in the fact that in the Foundation Approach all computations are directly based on the current value of the collateral, while the Advanced Approach explicitly states that all collateral values must be evaluated in the light of historical recovery rates that are in turn dependent on the type of collateral, type of recovery method etc. Whereas in the Foundation Approach LGD reductions are only permitted when some eligible collateral is present, the internal models developed under the Advanced Approach are allowed to incorporate the effect of any LGD related variable provided it proves to be significant in explaining recovery rates.

DISCOUNT RATE

Which discount rate to use on cash received post-default is a question that is the subject of considerable disagreement amongst practitioners and banking supervisors. Table 26 in Annex 2 proves this variety and proposes different rates as suitable. For defaulted small and medium-sized enterprise (SME) bank loans, the mean discount rate is found to be similar, on average, to the contract rate pertaining at the time of default (Maclachlan, 2004).

Theoretically, the appropriate discount rate is the risk-appropriate rate. Practically, this rate is very difficult to infer from observable variables when there are no markets for these facilities.

According to Araten, Jacobs and Varshney (2004) banks can also use a 'vulture' discount rate to discount cash flows instead of the average interest rate, which would bring about lower recoveries.

However, all assumptions other than the risk free rate could be set by agreement by local supervisors to ensure initial consistency between banks. But for internal economic capital measurements banks could use their own estimates.

3. DEFAULT RECOVERY RATES

3.1 REVIEW OF THE LITERATURE AND DIFFERENT TYPES OF MODELS

Credit risk literature has devoted significant attention to estimating the first component (PD) of the three main variables (PD, LGD and EAD). Much less attention has been paid to estimations of the recovery rate (RR) that equals 1-LGD and to the relationship between PD and RR. This traditional focus on default analysis has been partly reversed by the recent increase in the number of studies dedicated to the subject of RR estimation and the relationship between PD and RR. Generally, recent evidence from many countries suggests that collateral values and RRs can be volatile; moreover, they tend to go down just when the number of defaults goes up in economic downturns (Schleifer, Vishny, 1992; Altman, 2001).

The credit risk models that have been developed over the last 30 years can be divided into two main groups, whereby the first group is further divided into two main approaches:

- credit pricing models
  - structural form approaches
  - reduced form models
- portfolio credit VAR models.

3.1.1 Structural Form Models

The basic intuition behind structural models is the following: a default occurs when the market value of a firm's assets is lower than that of its liabilities (Merton, 1974). Merton as the main representative of the *first generation* structural models established a theoretical framework using the principle of option pricing (Black, Scholes, 1973). The payment to debtholders at the maturity of the debt is the smaller of the quantities: either the face value of the debt or the market value of the firm's assets. The underlying logic is that when the value of a firm at maturity is greater than the face value of the debt, then the debtholder gets back the face value of the debt. On the other hand, if the value of the firm is smaller than the face value of the debt, the debtholder gets back the market value of the firm.

Under these types of model, the RR is therefore an endogenous variable as the creditors' payoff is a function of the residual value of the defaulted company's assets. PD and RR tend to be inversely related. For instance, if the volatility of the firm's assets increases its PD increases while the expected RR at default decreases since the possible asset values can be quite low relative to the liability levels (Altman et al., 2001).

The *second generation* of structural models still adopts the original Merton framework as far as the default process is concerned but removes one of the unrealistic assumptions of the Merton model, namely that a default can only occur upon the maturity of the debt when the firm's assets are no longer sufficient to cover its debt obligations. Instead, it is assumed that a default may occur at any time between the issuance and maturity of the debt and that a default is triggered when the value of the firm's assets reaches a lower threshold level (Longstaff, Schwartz, 1995).

The RR in the event of a default is exogenous and independent of the firm's assets according to the second generation of structural models. It is generally defined as a fixed ratio of the outstanding debt value and is therefore independent of the PD. For example, by looking at the history of defaults and the recovery ratios for various classes of debt of comparable firms one can form a reliable estimate of the RR (Longstaff, Schwartz, 1995) Despite some improvements, the second generation of the structural-form-based models, the same as the first generation models, still represents relatively poor empirical performance (Eom, Helwege, Huang, 2001).

3.1.2 Reduced Form Models

Unlike structural form models, reduced form models do not condition default on the value of the firm and parameters related to the firm's assets do not need to be estimated to implement them. In addition, reduced form models introduce separate explicit assumptions on the dynamics of both PD and RR. These variables are modelled independently of the structural features of the firm, its asset volatility and leverage. Reduced form models assume an exogenous RR that is independent of the PD (Duffie, Singleton, 1999).

Reduced form models differ fundamentally from typical structural form models in the degree of predictability of the default as they can accommodate defaults which come as a surprise. A default occurs when an exogenous random variable, which is assumed to be a driving factor for a default, undergoes a discrete shift in its level. The time at which the discrete shift will occur cannot be foretold on the basis of information available today (unpredictable Poisson events).

According to one of the models, it assumes an exogenous process for the expected loss at default, meaning that the RR does not depend on the value of the defaultable claim, it allows for a correlation between the default hazard-rate process and RR (Duffie, Singleton, 1999).

Most of the models examined are based on bonds, which are not necessarily applicable to bank loans and I therefore will not go into their details here.

3.1.3 Credit VAR Models

In the second half of the 1990s, banks and consultants started developing credit risk models aimed at measuring potential loss with a predetermined confidence level that a portfolio of credit exposures could suffer within a specified time horizon. These value-at-risk (VAR) models include JP Morgan's CreditMetrics (Gupton, Finger, Bhatia, 1997), Mckinsey's CreditPortfolioView (Wilson, 1998) and KMV's CreditPortfolioManager.

Credit VAR models can be divided into two main categories (Altman, Resti, Sironi, 2005):

- default mode (DM) models; and
- mark-to-market (MTM) models.

The two approaches basically differ in the amount of data necessary to feed them: limited in the case of DM models, and much wider in the case of MTM models.

The main output of a credit risk model is the probability density function (PDF) of the future losses in a credit portfolio. From the analysis of such a loss distribution, a financial institution can estimate both the expected loss and the unexpected loss in its credit portfolio.

Financial institutions typically apply credit risk models to evaluate the economic capital needed to face the risk associated with their credit portfolios. In such a framework, provisions for credit losses should cover the expected losses while economic capital is seen as a cushion for unexpected losses.

All credit VAR models treat RR and PD as two independent variables (Crouhy, Galai, Mark, 2000) similarly to reduced-form models where the RR is typically taken as an exogenous constant parameter or a stochastic variable.

3.2 LGD IN EMPIRICAL STUDIES

Rather than focusing on the theoretical issues pertaining to LGD modelling, most recent contributions have dealt with the estimation of RRs related to different types of credit assets. Since very few financial institutions have ample data on RR by asset type and by type of collateral, model builders and the analysts responsible for the Basel II IRB models have begun with estimates from public bond and private bank loan markets. On the other hand, some banks will research their own internal databases in order to conform to the requirements of the Advanced IRB Approach.

Bank loans are likely to have some characteristics that differ significantly from those of corporate bonds. Comparing the number of studies analysing bank loans or corporate bonds reveals that very few studies focus on the bank loan markets due to data unavailability.

3.2.1 Results of Studies on Bank Loans

Citibank's 24-year study examined 831 defaulted loans (Asarnow, Edwards, 1995). It reports an average cumulative recovery rate of 65%.

Moody's bank loans study comprised a sample of 58 bank loans (Carty, Lieberman, 1996). Based on secondary market prices for defaulted bank loans they reported an average defaulted bank loan price of 71%. They did not observe a bi-modal distribution, but reported a skewness toward the high end of the price scale. In the same study, the authors measured the recovery rate for a sample of 229 small and medium-sized loans in the US. They reported an average recovery rate of 79% based on the present value of cash flows. Again, the distribution was highly skewed towards the high end of the scale.

An average recovery rate of 68.2% was estimated for bank loans in Latin America (Hurt, Felsovalyi, 1998). This also shows that loan size is a contributory factor to loss rates, with large loan defaults exhibiting lower recovery rates.

Resuming the results of some recent works, Emery (2003) reports some referential values which are of interest to our research. The median RR on secured bank loans is 73.0% and 50.5% on senior unsecured bank loans. Several researches have also presented fairly high variance levels across industrial sectors (Verde, 2003). Schuermann (2004) recently highlighted the importance of the industry factor in determining LGD in a survey of academic and practitioner literature.

One of the largest and more recent studies that focuses on loans to small and medium-sized enterprises was made by Standard & Poor's Risk Solution Department (Franks, Servigny, Davydenko, 2004). It considers collateral as the key driver of recovery rates, which vary across banks within the same country and jurisdiction. Recovery rates also differ across countries where banks respond to different bankruptcy regimes and codes by adjusting different lending practices. In France, for instance, banks demand higher levels of collateral and target specific forms of collateral. The recovery rate in France differs significantly (it is lower) from recoveries in the UK and Germany.

Finally, I point out a recent analysis of the determinants of LGD rates using a portfolio of credits given by the largest private Portuguese bank, Banco Comercial Portugues (Dermine, Carvalho, 2005).[2] An average cumulative recovery estimate of 71% is calculated on a sample of 371 defaulted loans to SMEs. These loans were granted in the period from June 1995 to December 2000. The estimates of RRs are based on discounted cash flows recovered after the default event.

3.3 THE PD-RR RELATIONSHIP

During the last five years, new approaches have enriched the theoretical and empirical estimation of LGD by focusing on the relationship between PD and RR. Frye's model is based on assumptions that the same economic conditions that cause defaults to rise might cause RRs to decline (Frye, 2000). The intuition behind his theoretical model is the following: if a borrower defaults on a loan, the bank's recovery may depend on the value of the loan collateral. The value of the collateral, like the value of other assets, depends on economic conditions. If the economy experiences a recession, RRs may decrease just as default rates tend to increase.

This evidence indicates that recovery risk is a systematic risk component. As such, it should attract a risk-premium and should be adequately considered in credit risk management applications. This view also seems to have been shared by regulators: while in the early drafts of the new Basel Accord LGD it was treated as a fixed parameter, the final Accord now accepts and underlines that it is stochastic in nature and may jump to substantially higher levels as the credit cycle slows down (Basel Committee on Banking Supervision, 2004).

4. DATA MANAGEMENT RECOMMENDATIONS

4.1 IMPLICATION OF A DATA WAREHOUSE

The most important parts of determining the realised LGD and the expected LGD are data and data quality, which need to be regularly validated to ensure the consistency of the risk parameter and the accuracy, completeness and appropriateness of the input factors to estimate LGD.

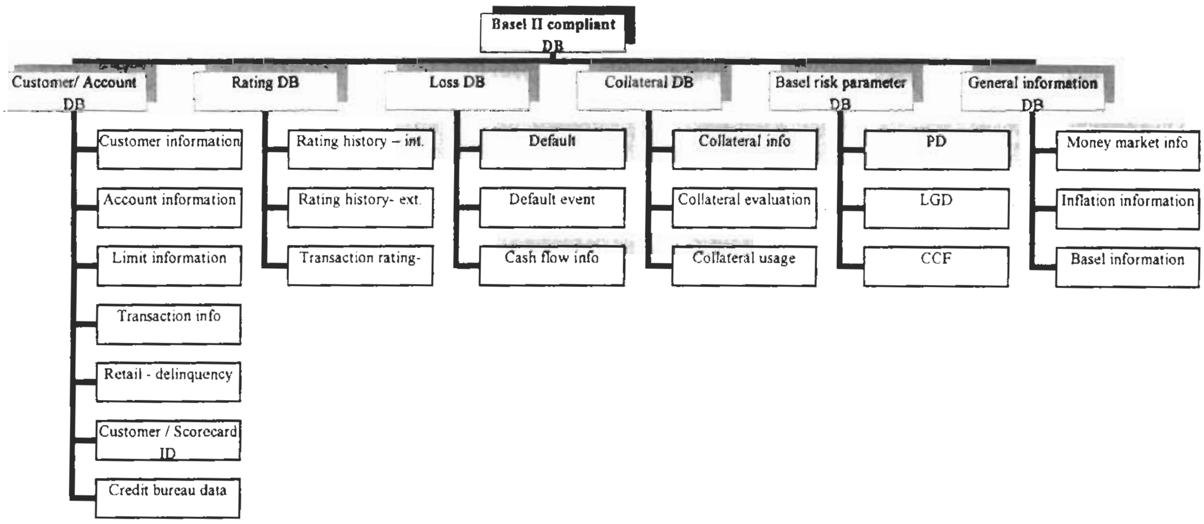
Decisions as to which data should be stored during the workout process with respect to the customer and the collateral are generally made by experienced experts in the bank. It is necessary that banks create a risk data warehouse (or similar risk database) where all information related to losses, ratings and other information related to Basel II, such as risk parameters, are stored and to ensure a system environment in which none of the records are lost and the relevant data fields can be easily used to estimate LGD.

[2] The results and referential values are presented in Annex 1.

The risk data warehouse includes all risk drivers essential for the estimation of all risk parameters, i.e. not only LGD. These risk drivers can vary for different risk parameters, so for loss estimation the considered risk drivers should at least relate to:

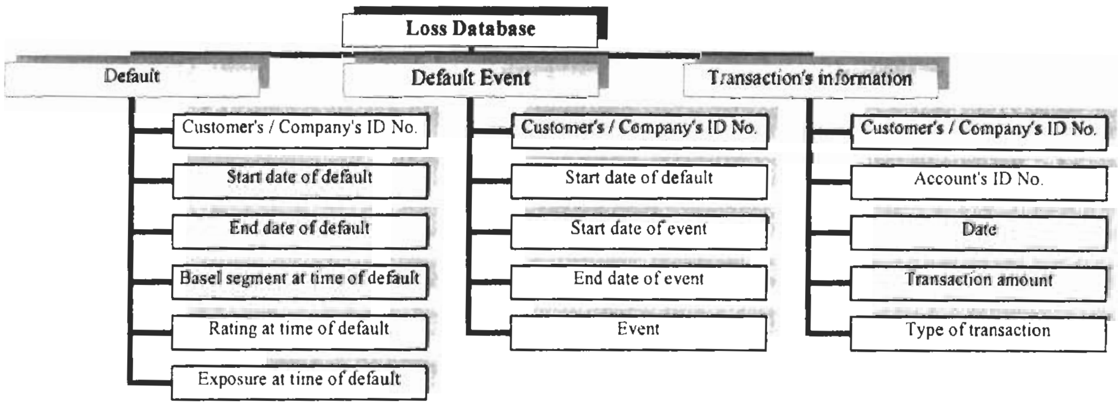
- Transaction – facility type, collateral, seasoning, recovery procedure;
- Borrower – asset class, geographic region, line of business;
- Institution – internal organisation and governance, specific departments or entities dealing with recoveries; and
- External - interest rate, legal framework.

Figure 2: Basel II-compliant database structure



Source: Strengthening the Czech banking sector – Application of Basel II, 2005, p. 91.

Figure 3: Loss database: illustrative data fields



Source: Strengthening the Czech banking sector – Application of Basel II, 2005, p. 92.

Figure 2 summarises the topics to be addressed for calculation in the risk data warehouse and for the demonstration of compliance with Basel II. It is not intended to be a fixed recommendation for the development of a data system structure.

A loss database, which forms part of the risk data warehouse, should among other things store information on the start and end dates of a default case (end of the workout process or restructuring), any facts about the workout process, costs and recoveries, default events as well as rating information and segmentation information about the defaulting borrower. Figure 3 presents the information that could be obtained in a loss database.

4.1.1 Collateral

Annex VII of the draft version of the Capital Requirements Directive (CRD) or CAD III deals with credit risk mitigation (CRM). In part 2 of this Annex, the minimum requirements for eligible collateral, netting agreements and guarantees are stated. Considering the requirements and data storage, I will highlight some of the recommendations for banks.

The best way for banks to store and monitor collateral information is to implement a collateral management system. This system should not only store information on the type of collateral or its assignment to individual borrowers and accounts. Other information such as the first date of evaluation or the realisation date of the collateral should also be captured. All of these data are vital for the loss calculation regardless of whether collateral is included directly or indirectly in the LGD estimation.

The realisation date is needed in order to check whether the collateral linked to a particular defaulted loan has actually been used for reducing the loss. The loan may have been overcollateralised and therefore additional realisations from a particular piece of collateral are not needed. The date of the first evaluation can be used not only for the first automated re-evaluation process but also for the purpose of monitoring the value change over time, especially with respect to mortgages where loan contracts can last up to 25 years or more.

Further, information on the validity period of the collateral is important especially if this does not match with the maturity of the loan it is related to. Therefore the 'start date' and the 'end date' of the period for which the collateral can be used as a credit risk mitigation technique should be stored in the collateral management system, in addition to any prior liens.

Another requirement of the draft directive is that the collateral needs to be regularly revalued. The reassessment of the value of the collateral and its storage also offers a good indicator for following market trends and using the information later on for risk assessment, pricing or limit-setting purposes.

As far as the realisation of collateral is concerned, it can be realised as a whole or partly depending on the type of collateral. We identified three main groups of collateral:

- collateral realised once and as a whole, like houses (mortgage) or cars (movables);
- collateral realised partly and/or in more than one instance, like securities, deposit accounts, assignment/cession of receivables; and
- collateral depending on the loan amount, like guarantees.

For the first group of collateral, the bank has to set rules so that the amount finally realised for a specific piece of collateral is stored in the data system. Many banks only store the amount needed to cover the defaulted loan.

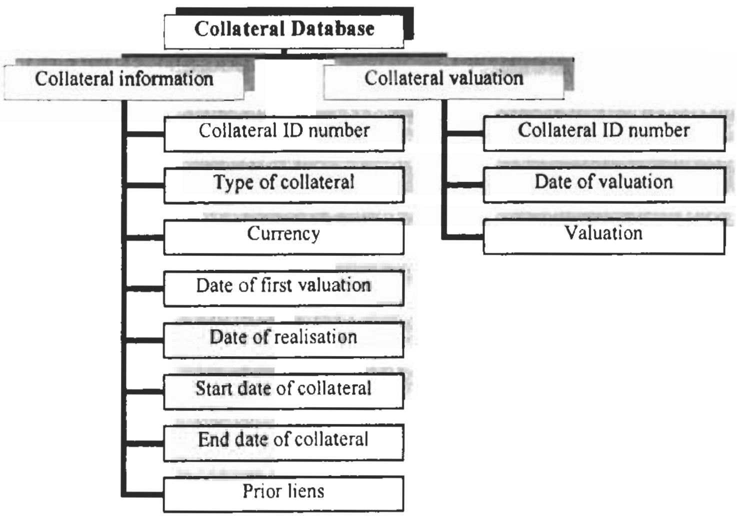
If the collateral is e.g. a deposit, the bank might only wish to realise that part of the portfolio needed to cover the defaulted loan or which might be easy to sell at the time. To overcome this problem, we recommend the following options for the second group of collateral:

- the use of supervisory haircuts;
- an evaluation using internal haircuts (e.g. VAR-models: especially for security deposits/portfolio); and
- a calculation of the collateral recovery rate and volatility, using information about the realisation as a whole (especially for the assignment of receivables).

The last group of collateral contains collateral such as guarantees and letters of comfort. The value of such a guarantee depends on the agreement (given in an absolute value or as a percentage of the loan) as well as the amount currently outstanding of the assigned loan. As in most cases the guarantor will only cover the maximum of the actual amount outstanding and any agreed-upon limit, the value of the collateral cannot be calculated before the default date. To estimate the recovery rate of this collateral group, the estimated value of the collateral is this maximum amount.

Finally, Figure 4 presents the minimum information a collateral management system or collateral database should store, as explained above.

Figure 4: Data requirements for collateral management



Source: Strengthening the Czech banking sector – Application of Basel II, 2005, p. 94.

5. PREPAREDNESS OF SLOVENIAN BANKS

In order to assess the preparedness of Slovenian banks with respect to LGD calculations, two interviews were conducted. The first was with a Slovenian bank which is targeting the Standardised Approach, at least in the short term, and another one at the Bank of Slovenia. It helped me gain some insights into Slovenia's preparations for the new Basel standards and the adoption of approaches to credit risk management.

5.1 GENERAL ISSUES REGARDING DATA STORAGE

Most banks store information about their collection process such as positive and negative cash flows (earnings and costs) coming from either the borrower himself or the collateral linked to the loan. In some cases banks cannot distinguish the source of the payment received, i.e. whether the cash flows relate to the borrower or the collateral. Also, payments cannot always be distinguished between the individual accounts they relate to and payments are applied pro-rata to all accounts of a customer, based on average levels of collateralisation or products. In extreme cases, payments cannot even be attached to individual customers but only to customer groups as they are recorded in 'summary' general ledger accounts (this is often the case with payments received after the final write-off of loans). Also, information on the reasons for the default or different default stages are not kept in databases or files.

5.2 DATA AVAILABILITY AND THE CURRENT STATUS OF DATA MANAGEMENT

Data availability, as well as other problems not discussed above, makes it hard for a bank to use more sophisticated mathematical and statistical methods in its LGD estimation process. To use eligible regression, potential 'loss' drivers, which have to capture statistically significant information, have to be identified. As this basic requirement can only rarely be expected in the environment described above, the available empirical data can be better used for an analysis of average losses.

In this section I briefly introduce and evaluate the data ('loss' drivers) I have managed to collect from bank x in order to define the necessary factors for the model calculating the recovery rate, their extensiveness and quality in compliance with the Basel requirements and the type of data that should be collected in the future to enable the Advanced IRB Approach (also see the previous chapter on data management recommendations).

Most of the data I have collected for the purpose of the model are yearly-based (the end of each year) because of the easier grouping of data that are not always fully coherent with other data sources.

The estimates of LGD must be based on a minimum data observation period that should cover at least one complete economic cycle. This observation should not be shorter than a period of seven years for at least one source (Basel Committee on Banking Supervision, 2004). We encountered problems in fulfilling these requirements as most of our data have only been explicitly tabled since 2001 when a collateral policy came into force and changed their storage within the bank. Having held consultations with an official from the Bank of Slovenia setting a pivot model for the period from 2001 to 2005 would be a good basis to work on and it should then be gradually updated with additional necessary information when available. We also have to take into account that there have been no shocks in the last economic cycle of seven years.

5.2.1 Dependent Variable (Recovery Rate)

For the recovery rate calculation we need all the data about repayments on each facility. It can be calculated from differences in exposure every year after a default from the database of exposures that are based on a number of facilities. Some adjustments are required.

There exists a database of repayments by year for each client in default along with its recovery method. It is obligor-based. The possible adjustments are therefore connected to the obligor-based repayments that are then equally contributed to each of the company's facilities of the client (the average recovery on one's obligor facility).

There is another data source in bank x that includes all exposures of the facilities. But there is a problem with the calculation of repayments from this data source since facility numbers change every year. Every change in the conditions of the facilities also causes a change in the facility number and therefore the facility numbers cannot be directly compared with the exposures.

5.2.2 Variables Describing the Collateralisation of Facilities

The data describing the collateral features of each facility are gathered from a data source that could be already treated as a data management model. It contains most of the required information:

- facility number;
- collateral number;
- type of collateral; and
- value of collateral.

We consider collateral to be the most important category. Nevertheless, it may not be statistically significant for all types of collateral.

Moreover, we have to take into account the high possibility that there are facilities with several items of collateral covering them but, on the other hand, there are also examples of the same collateral covering several facilities.

The problem occurring with the tabled collateral value is that there is mostly no market value of the available collateral (most often occurred with the expired assignment of receivables).

BANKRUPTCY CODE

The bankruptcy law and its influence on creditors' rights can seriously affect the priority of claims and the proceeds of collateral that accrue to the secured lender. Slovenia represents a special case with its own bankruptcy law and it therefore cannot be directly compared with other countries' laws, e.g.: the collateral level in France tends to be higher than the level in Germany or the UK (Franks, Servigny, Davydenko, 2004, p. 70). In particular, French banks respond to a creditor-unfriendly bankruptcy code by requiring more collateral than lenders elsewhere, and by relying on particular collateral forms that minimise the statutory dilution of their claims in bankruptcy (Davydenko, Franks, 2005, pp. 1-3).

5.2.3 Explanatory Variable Industry

From the observed data for bank x we can calculate the average recovery rate for industrial sectors and so we can see what is the volatility of recoveries among different sectors of business activity. Bank x has data available on the average probability of default for industries that were gathered from an internal PD model. Considering both predictors we can calculate the expected loss (see Section 7.6 on page 41).

5.2.4 Other Information Contributing to the Final Recovery

The differentiation between a short-term and long-term loan can represent an important issue when it comes to recoveries.

The final rating given to a defaulted company before closing the facility and recovery method can also represent information that helps us calculate recoveries. All the information is fully available in the bank's internal computer documents.

Regarding the method of recovery there are different costs of collateral realisation or debt reorganisation and liquidation. The bank can assume the standard restructuring/liquidation intensity for certain customers or product types.

5.2.5 Macroeconomic Factors

The Basel requirements add that for certain types of exposures loss severities may not exhibit cyclical variability and LGD may not differ from the long-run default-weighted average. However, for other exposures, this cyclical variability in loss severities may be important and banks will need to incorporate it into their LGD estimates (Basel Committee on Banking Supervision, 2004).

Since 2001 the economic conditions have been very much in favour of economic growth so we cannot capture a 'downgrade LGD'. The economy has not experienced any important shocks that could also have negatively impacted on the LGD results. On the contrary, the economy has been flourishing which has also resulted in lower losses. Therefore, macroeconomic factors do not provide important indicators in the case of Slovenia for the last few years.

5.2.6 General Requirements on Data Availability

The data required to develop LGD models will vary significantly according to the methodology for calculating LGD measures chosen by the bank. It will also depend on specific business practices of banks with respect to taking collateral, i.e. whether specific collateral is attached to a specific facility or whether collateral is kept at the customer level and is available to be allocated across all the customer's facilities.

In any event, banks will need to collect at least information on historical provisions and write-offs, the dates and amounts of recoveries, the sources of recovery (e.g. from repayments by the customer, realisation of collateral or payment under a guarantee), the values and types of collateral, the costs of the collection process (both in terms of the cost of internal departments tasked with workout and collections and in terms of external legal and administrative costs).

6. ECONOMETRIC METHODOLOGY

6.1 INTRODUCTION TO THE STATISTICAL ANALYSIS

In a data analysis we need to understand the variables we will be analysing in order to get the greatest benefit from our data and to be sure that our conclusions are justified. One of the most important features of a variable is its measurement level (scale). Variables differ in how well they can be measured or how much measurable information their measurement scales can provide. The usual linear regression methods are designed to model **scale variables** where we can assume that any particular case (record) can take any value within the range of the variable. However, in our case we have some important **categorical variables** that do not fit this profile.

6.1.1 Categorical Variables

Categorical variables have two primary types of scales. Variables having categories without a natural ordering are called **nominal.** For nominal variables the order of listing the categories is irrelevant. The statistical analysis does not depend on that ordering (Agresti, 2002, pp. 1-4).

Many categorical variables do have ordered categories. Such variables are called **ordinal.** They have ordered categories but the distances between categories are unknown. The methods for ordinal variables utilise category ordering (Agresti, 2002, pp. 1-4).

A variable's measurement scale determines which statistical methods are appropriate.

6.2 REGRESSION WITH AN ORDINAL OUTCOME VARIABLE

To handle ordinal outcome (dependent) variables special methods are necessary. One way to model ordinal variables is to use a scoring scheme where each category of the ordinal outcome variable is assigned a score on a continuous scale to account for the unequal distances between categories. The drawback of this method is that one has to know which score values to assign to the ordinal categories before starting. In many cases, the appropriate score values cannot be determined *a priori* (SPSS Advanced Models 10.0, 1999, pp. 242-243).

An alternative approach uses a generalisation of a linear regression called a **generalised linear model** to predict cumulative probabilities for the categories. With this method we can get a separate equation for each category of the ordinal dependent variable. Each equation gives a predicted probability of being in the corresponding category or any lower category. With no predictors in the model, predictions are only based on the overall probabilities of being in each category. The prediction for the last category is always 1.0 since all cases must be either in the last category or a lower category. As a result, the prediction for the last category is not needed (SPSS Advanced Models 10.0, 1999, p. 243).

6.3 GENERALISED LINEAR MODEL – ORDINAL REGRESSION

6.3.1 Generalised Linear Models – A Computational Approach

Generalised linear models (GLMs) are used to do regression modelling for non-normal distributed data with a minimum of extra complications compared with a normal linear regression. GLMs are flexible enough to include a wide range of common situations but at the same time allow most of the familiar ideas of normal linear regression to carry over.

The generalised linear model can be used to predict responses for both dependent variables with discrete distributions and for dependent variables which are nonlinearly related to the predictors.

To summarise the basic ideas, the generalised linear model differs from the general linear model (of which, for example, multiple regression is a special case) in two major respects (McCullagh, Nelder, 1989).

- first, the distribution of the dependent or response variable can be (explicitly) non-normal and does not have to be continuous. It can be binomial, multinomial or ordinal multinomial (i.e., contain information on ranks only); and
- second, the dependent variable values are predicted from a linear combination of predictor variables which are connected to the dependent variable via a **link function**.

The general linear model for a single dependent variable can be considered a special case of the generalised linear model. In the general linear model the dependent variable values are expected to follow the normal distribution and the link function is a simple identity function (i.e., the linear combination of values for the predictor variables is not transformed).

To illustrate, in the general linear model a response variable Y is linearly associated with values of the X variables by (McCullagh, Nelder, 1989):

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

where e stands for the error variability that cannot be accounted for by the predictors; note that the expected value of e is assumed to be 0, while the relationship in the generalised linear model is assumed to be:

$$Y = g\ (b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e)$$

where e is the error, and g(…) is a function. Formally, the inverse function of g(…), say f(…), is called the link function, so that:

$$f(mu_y) = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

where mu_y stands for the expected value of y.

Various link functions can be chosen depending on the assumed distribution of the *y* variable values (see Annex 4 – McCullagh, Nelder, 1989).

6.3.2 Estimation in the Generalised Linear Model

The values of the parameters (b_0 through b_k and the scale parameter) in the generalised linear model are obtained by a maximum likelihood (ML) estimation, which requires iterative computational procedures. There are many iterative methods for ML estimation in the generalised linear model, of which the Newton-Raphson and Fisher-Scoring methods are among the most efficient and widely used (Dobson, 2002). The Fisher-scoring (or iterative re-weighted least squares) method in particular provides a unified algorithm for all generalised linear models, as well as providing the expected variance-covariance matrix of parameter estimates as a by-product of its computations.

6.3.3 Statistical Significance Testing

Tests for the significance of effects in the model can be performed via the Wald statistic, the likelihood ratio (LR), or score statistic (McCullagh, Nelder, 1989). The Wald statistic, which is computed as the generalised inner product of the parameter estimates with the respective variance-covariance matrix, is an easily computed, efficient statistic for testing the significance of effects. The score statistic is obtained from the generalised inner product of the score vector with a Hessian matrix (the matrix of the second-order partial derivatives of the maximum likelihood parameter estimates). The likelihood ratio (LR) test requires the greatest computational effort (another iterative estimation procedure) and is thus not as fast as the first two methods; however, the LR test provides the most asymptotically efficient test known (Agresti, 1996; McCullagh, Nelder, 1989).

6.3.4 Diagnostics in the Generalised Linear Model

The two basic types of residuals are the so-called Pearson residuals and deviance residuals. Pearson residuals are based on the difference between observed responses and the predicted values; deviance residuals are based on the contribution of the observed responses to the log-likelihood statistic. In addition, leverage scores, studentised residuals, generalised Cook's D, and other observational statistics (statistics based on individual observations) can be computed (Hosmer, Lemeshow, 1989).

6.4 ORDINAL REGRESSION (USING SPSS SOFTWARE)

As far as the requirements for using ordinal regression as one of the GLM models are concerned, it is expected that the dependent variable is assumed to be ordinal and can be numeric or string. The ordering is determined by sorting the values of the dependent variable in ascending order. The lowest value defines the first category. Factor variables are assumed to be categorical. Covariate variables must be numeric. Ordinal regression is eligible in our case where most of the explanatory variables are categorical and the response variable is ordinal.

Generalised linear models are a very powerful class of models which can be used to answer a great number of statistical questions. The basic form of a generalised linear model can also be written in the following way:

$$\text{link}(\gamma_j) = \theta_j - (\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

where γ_j is the cumulative probability for the j-th category, θ_j is the threshold for the j-th category, $\beta_1 \ldots \beta_k$ are the regression coefficients, $x_1 \ldots x_k$ are the predictor variables, and k is the number of predictors. There are several important things to notice (SPSS Advanced Models 10.0, 1999, pp. 244-245):

- The model is based on the notion that there is some latent continuous outcome variable (such as the recovery) and that the manifest ordinal outcome variable arises from discretising the underlying continuum into j ordered groups. The cut-off values on this continuous distribution that define the categories are estimated by threshold θ_j. In some cases, there is a good theoretical justification for assuming such an underlying distribution. However, even where there is no theoretical concept that links to the latent variable the model can still perform quite well and give valid results. Also in the ordinal regression model, the thresholds are estimated as part of the model and need not be specified *a priori*.
- The thresholds or constants in the model θj (corresponding to the intercept in linear regression models) depend only on which category's probability is being predicted. Values of the predictor (independent) variables do not affect this part of the model.
- The prediction part of the model, $(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$, depends only on the predictors and is independent of the outcome category. These first two properties imply that the results will be a set of parallel lines, one for each category of the outcome variable.
- Rather than predicting the actual cumulative probabilities, the model predicts a function of those values. This function is called, as we have already mentioned, the link function, and in general we can choose the form of the link function when we build the model (see Annex 4). This allows us to choose a link function based on the problem under consideration to optimise our results.

There are three major components in an ordinal regression model (SPSS Advanced Models 10.0, 1999, p. 245):

- **Location component,** $(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$, includes the coefficients and predictor variables. It represents the main part of the model. It uses the predictor variables to calculate predicted probabilities of membership in the categories for each case.
- **Scale component** is an optional modification to the basic model to account for differences in variability for different values of the predictor variables. The model with a scale component follows the form:

$$\text{link}(\gamma j) = \frac{\theta j - (\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}{\exp(\tau_1 z_1 + \tau_2 z_2 + \ldots + \tau_m z_m)}$$

  Where $\tau_1 \ldots \tau_m$ are coefficients for the scale component and $z_1 \ldots z_m$ are m predictor variables for the scale component (scale from the same set of variables as the x's).

- **Link function** is a transformation of the cumulative probabilities that allows an estimation of the model. Five link functions are available in the ordinal regression procedure and are summarised in Table 3.

Table 3: Link functions in ordinal regression

Function	Form	Typical application
Logit	$\log(\gamma / 1-\gamma)$	evenly distributed categories
Complementary log-log	$\log(-\log(1-\gamma))$	higher categories more probable
Negative log-log	$\{-\log(-\log(\gamma))\}$	lower categories more probable
Probit	$\Phi^{-1}(\gamma)$	analyses with explicit normally distributed latent variable
Cauchit (inverse Cauchy)	$\tan(\pi(\gamma-0.5))$	outcome with many extreme values

Source: SPSS Advanced Models 10.0, 1999.

7. LGD MODEL SPECIFICATIONS: THE CASE OF BANK X

In the paper we have so far provided the main features, a short review of the literature and the general framework for measuring and understanding LGDs. A challenge that risk managers in Slovenia as well as in other countries around the globe are facing at the moment is how to implement a methodology to estimate the degree of recovery risk surrounding specific classes of exposure, e.g. bank loans. They need to adopt a methodology to estimate LGD and fair provisions on non-performing loans. LGD estimates are important inputs in the pricing of credit risk, the measurement of bank profitability and solvency. The banks using the IRB Approach must compare their total eligible provisions with their total expected losses (Basel Committee on Banking Supervision, 2004).

The model specifications in the case of bank x provide some empirical evidences on cumulative recovery rates, on their determinants and finally also on the direct costs incurred in recovery.

7.1 DATABASE STRUCTURE

The paper provides a pilot model for a particular Slovenian bank. The goal is to calculate the LGD for SMEs. We have included all the defaulted companies with an average exposure of more than SIT 1,000,000[3]. The population consists of 305 companies in default[4].

[3] Note: € 1 = SIT 239.64

[4] The data gathered for this study do not include any reference to the identity of clients or any other information that according to the Slovenian Banking Law cannot be disclosed.

The database with all the required information needed for the calculation of LGD consists of 124 companies that defaulted in the period from 2001 to 2004 and were closed down by the end of 2005. The reason we did not take data before 2001 into account was that the data stored on collateral management were insufficient.

All the data used in the sample were collected internally by different bank reports within the risk management division.

7.1.1 Default and Recovery Measurements

We consider a company as being in default when a payment has been missed for more than 90 days as is recommended by the Basel II perspective. Translating this perspective into Slovenian banking standards and ratings it means that every company which has a rating worse than or equal to C[5] is shifted into the recovery process.

Considering that no market data have been available on the price of loans as at the default date, defined most frequently as the trading price one month after the default, we estimate the present value of actual recovered cash flows. To calculate the present value we decided to take the appropriate discount rate. We did not have access to the interest rate charged on individual loans. We experienced the same problems as most other studies (Asarnow, Edwards, 1995; Carty, Lieberman, 1996). Instead of the actual interest rate we have taken the approximate average interest rate on SIT credits (as all exposures were tabled in SIT) for the 2001-2005 period as an alternative (see Table 27 in Annex 3). It equals 10%.

In order to measure the cash flows recovered after a default event we tracked at each end of the year the post-default credit balances. Capital recovery is a reduction of the total balance. The total cash flow recovered is this capital recovery plus the interest on the outstanding balance.

While, at first glance, the tracking of cash flows after a default event appears a relatively simple (but time-consuming) exercise, special cases did require some adjustment:

- upgraded companies that were shifted back to 'performing' companies were excluded from the model; and
- only those companies and facilities whose recovery proceedings have been closed down or finished are included.

7.1.2 Explanatory Variables

For the statistical model we have built as well as for the empirical analysis of recoveries we consider several determinant factors that are of great importance for determining the dependant variable RR.

---

[5] Rating system of the Bank of Slovenia.

We can group the variables we find important for our internal calculation of LGD in the following categories:

COLLATERAL

Collateral is the main indicator of recovery on a facility in the case of default. I have divided collateral types into five groups:

- financial collateral (bank deposits, securities, bonds);
- real estate collateral;
- physical collateral (movables);
- guarantees; and
- assignments of receivables.

In cases where there are several types of collateral on the same facility we take the primary collateral in terms of liquidity as the preferential collateral type.

LOAN TYPES

Loans differ by their maturity or securitisation. In the case of maturity we divide loans into two categories: long-term and short-term, which are either secured or unsecured.

INDUSTRY

Industry may be an important indicator of recovery rates. We calculate recovery averages by industrial sector, with reference to the Classification of Economic Activities in the European Community (NACE classification), to refine the historical estimates of LGD.

SIZE OF EXPOSURE

There is another factor, namely the size of a facility's exposure at the time of default (small, medium, large), whereby:

- Small: debt outstanding < SIT 10,000,000;
- Medium: SIT 10,000,000 < Debt outstanding < SIT 100,000,000; and
- Large: Debt outstanding > SIT 100,000,000.

LAST RATING

There might be different recovery rates according to the different last ratings given by credit analysts when a company was in default[6].

---

[6] Rating system of the Bank of Slovenia.

RECOVERY METHOD

The recovery level also depends on the method of the recovery procedure. The broad classification of methods which bank x uses in its internal documents is the following:

- sale of the credit to a third party;
- debt rescheduling;
- informal workout;
- non-judicial foreclosure or execution;
- judicial foreclosure (immoveable assets);
- judicial proceedings and execution (moveable assets);
- liquidation proceedings (bankruptcy);
- formal rehabilitation;
- conversion of debt to equity; and
- other.

7.2 DESCRIPTIVE STATISTICS AND ANALYSIS OF THE DATABASE

Tables 4 and 5 reproduce information on the number of defaults per year and on the amount of debt outstanding at the time of default. Table 6 shows the number of loans with a guarantee, collateral or those that are unsecured.

Table 4: Descriptive statistics for the sample of bad and doubtful loans

Year of default	Number of defaults per year	Percentage
2001	59	47.6
2002	22	17.7
2003	31	25.0
2004	12	9.7
Total	124	100.0

Source: own calculations.

Table 5: Descriptive statistics for the sample of bad and doubtful loans

Debt outstanding at the time of default (SIT)	Number of observations	Percentage
Large	12	9.7
Medium	37	29.8
Small	75	60.5
Total	124	100.0

Note: Small: Debt outstanding < SIT 10,000,000; Medium: SIT 10,000,000 < Debt outstanding < SIT 100,000,000; Large: Debt outstanding > SIT 100,000,000

Source: own calculations.

We observe that the series of 124 default cases is highly skewed towards 2001. 46% of the observed bad loans were defaulted in 2001. The distribution of the debt outstanding is highly skewed towards the low end (small exposures). 61% of the debt exposure involves amounts less than SIT 10,000,000.

Table 6: Descriptive statistics for the sample of bad and doubtful loans

Forms of collateral/guarantee	Number of observations	Percentage
Financial collateral (bank deposits, securities, bonds)	6	4.8
Guarantees	38	30.6
Physical collateral (movables, others)	6	4.8
Real estate collateral (mortgages)	54	43.5
Assignments of receivables	5	4.0
Unsecured	**15**	**12.1**
Total	124	100.0

Source: own calculations.

In Table 6 the various forms of collateral are reported. In 12% of the cases there is no guarantee or collateral which means that a very large proportion of bank loans are collateralised in comparison to the results of a study in Portugal (Dermine, Carvalho, 2005). The most frequent form of collateral used by SMEs is a mortgage (44%).

Looking at the cases where there is no collateral or guarantee (unsecured), we can see that most of the unsecured loans are short-term loans. 17% of the short-term loans were totally unsecured compared to only 5% of the long-term loans that are unsecured (see Table 7).

Table 7: Descriptive statistics for the sample of bad and doubtful loans

Type of loan	Secured	Frequency of secured	Unsecured	Frequency of unsecured	Total
Long-term	52	95%	3	5%	55
Short-term	57	83%	12	17%	69
Total	109		15		124

Source: own calculations.

Table 8 shows the concentration of different forms of collateral in the sample according to the amount of debt outstanding at the time of default. We can see that large exposures are also most frequently secured with one of the forms of collateral (excluding guarantees). Large exposures are also less commonly totally unsecured. Small exposures at default are most likely to be secured by a guarantee as there were 28 such cases out of 38, which was the total number of guarantees in the sample.

Table 9 reports the concentration of default cases in different business sectors and the use of guarantees/collaterals/unsecured loans across these sectors. Eleven business sectors have been created with reference to the NACE classification. Default cases are observed in all business sectors, with a concentration in manufacturing (16% of the default cases), wholesale and retail trade (45% of the default cases).

Table 8: Number of loans with collateral/a guarantee or being unsecured according to the size of debt outstanding at the time of default

Debt outstanding at the time of default	Collateral	Frequency of collateral	Guarantee	Frequency of guarantee	Unsecured	Frequency of unsecured	Total
Large	9	75.0%	2	16.7%	1	8.3%	12
Medium	24	64.9%	8	21.6%	5	13.5%	37
Small	38	50.7%	28	37.3%	9	12.0%	75
Total	71		38		15		124

Note: Small: Debt outstanding < SIT 10,000,000, Medium: SIT 10,000,000 < Debt outstanding < SIT 100,000,000, Large: Debt outstanding > SIT 100,000,000

Source: own calculations.

Table 9: Number of default cases by industrial sectors (NACE)

NACE economic activities	Number of defaults with collateral	Number of defaults with guarantees	Number of unsecured defaults	Number of defaults (total)	Frequency
C – Mining	1			1	0.8
D – Manufacturing	13	6	1	20	16.1
E - Electricity, gas and water supply		1		1	0.8
F – Construction	4	1	1	6	4.8
G – Wholesale and retail trade	36	16	4	56	45.2
H - Hotels and restaurants	1	6		7	5.6
I - Transport, storage and communication	3	1		4	3.2
J - Financial intermediation	3	1		4	3.2
K - Real estate	7	4	6	17	13.7
M – Education	1		2	3	2.4
O - Other service activities	2	2	1	5	4.0
Total	71	38	15	124	100

Source: own calculations.

A further aggregation, as used in the econometric tests, leads to four activity sectors: real sector (sectors C, F, H, K) manufacturing (sector D), trade (G) and services (E, I, J, M, O)[7]. A quite similar aggregation into the four activity sectors was made by the Portuguese model (Dermine, Carvalho, 2005).

The relative use of a guarantee and collateral seems quite uniformly spread across the four aggregated sectors (see Table 10). Because of the small number of unsecured defaults the distribution across the four aggregated sectors is more volatile compared to collateral and guarantees.

[7] The grouping is made in accordance with the information important for the bank and does not necessarily suit the name of the aggregated activity sector.

Table 10: Number of default cases by aggregated industrial sectors

Aggregated industrial sectors	Number of defaults with collateral	In %	Number of defaults with guarantees	in %	Number of unsecured defaults	in %	Total
Manufacturing	13	65.0	6	30.0	1	5.0	20
Real	13	41.9	11	35.5	7	22.6	31
Services	9	52.9	5	29.4	3	17.6	17
Trade	36	64.3	16	28.6	4	7.1	56
Total	71	57.3	38	30.6	15	12.1	124

Note: aggregated sectors consist of the following sectors according to European Union's economic activity codes (NACE): Real (sectors C, F, H, K); Manufacturing (sector D), Trade, (sector G), Services (sectors E, I, J, M, O).
Source: own calculations.

In Table 11 we report the 12-, 24-, 36- and 48-month cumulative recovery rates for the total sample. The method of the recovery calculation is presented in Annex 7. The mean cumulative recovery rate of 73% (48-month cumulative recovery) is of the same order of magnitude as those reported by Asarnow and Edwards (1995) and Hurt and Felsovalyi (1998) for Latin America. It is also of the same order of magnitude as reported by Dermine and Carvalho (2005).

Table 11: Sample unweighted cumulative recovery rates

	12-month cumulative recovery	24-month cumulative recovery	36-month cumulative recovery	48-month cumulative recovery
Mean	0.48	0.62	0.70	0.73
Median	0.50	0.88	0.91	0.91
Standard Deviation	0.41	0.40	0.36	0.35
Minimum	0	0	0	0
Maximum	1	1	1	1

Note: cumulative recovery rates are calculated for the total sample of 124 facilities taking them all at a time. We focus on the time factor of recoveries of the loans in the sample not taking into account the duration of resolution proceedings (see Annex 7). This information tells us that 12 months after a default occurs on average 48% of the sample's exposure at default was recovered, 24 months after a default on average 62% of the sample's facility exposure was recovered etc.
Source: own calculations.

Table 12: Univariate statistics on recovery rates

	Cumulative recovery (loans with collateral)	Cumulative recovery (loans with a guarantee)	Cumulative recovery (unsecured loans)
Mean	0.76	0.72	0.62
Median	0.91	0.91	0.87
Mode	1	1	1
Standard Deviation	0.33	0.37	0.39
Range	1	1	1
Minimum	0	0	0
Maximum	1	1	1
Count	71	38	15

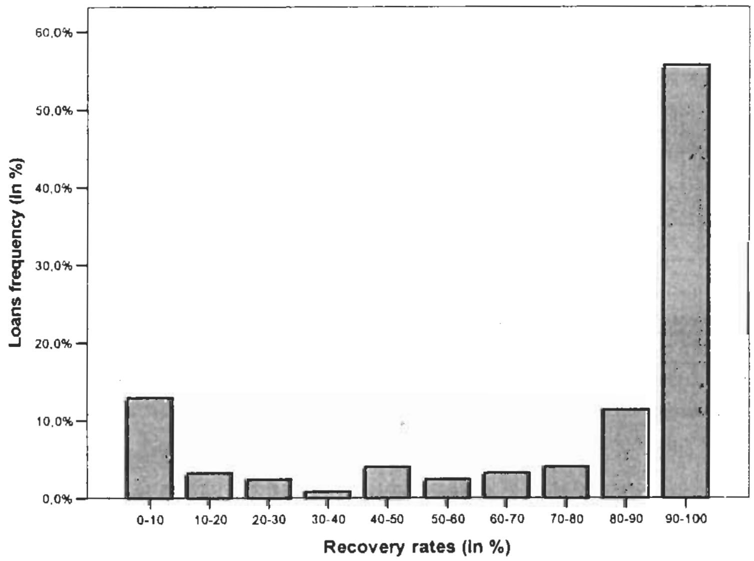
Source: own calculations.

The cumulative recovery on loans with collateral is 76%, on loans with guarantees 72%, while unsecured loans have an average cumulative recovery rate of 62% (as presented in Table 12). From these results we can conclude that collateral as such is a better source of securitisation than a guarantee. Either of the securitisations contributes to higher recoveries.

7.3 EMPIRICAL RESULTS ON CUMULATIVE RECOVERY RATES

First, it is of interest to analyse the distribution of cumulative recovery rates across the sample of loans. The distribution of cumulative recovery rates is reproduced in Figure 5. This figure shows a bi-modal distribution with many observations with a low recovery and many with an almost complete recovery (more than 50% of the cases having recoveries between 90% and 100%). These results are quite similar to those reported by Dermine and Carvalho for Portugal (2005), Asarnow and Edwards (1995) and Schuermann (2004) for the US, and Hurt and Felsovalyi (1998) for Latin America. All these studies present a bi-modal distribution of recoveries.

Figure 5: Sample distribution of cumulative recovery rates



Source: own calculations.

The loan portfolio models that incorporate a probability distribution for recovery rates should take this bi-modal distribution into account.

Table 13 indicates the value-weighting effect on cumulative recovery rates (also see Figure 11 in Annex 8). Considering formula LGD = 1 – RR we can observe that smaller exposures have on average higher losses (almost 30%) in comparison to large exposures (22%), which usually also means bigger companies. This leads to different conclusions to those made by Dermine and Carvalho (2005) on the Portuguese model. The reasons for this can be found in the worse collateralisation of small loans compared to large loans, which can also be seen in the tables in the previous section describing the sample database.

Table 13: Value-weighting effect on cumulative recovery rates

Debt outstanding at the time of default (SIT)	Average of LGD
Large	0.218
Medium	0.231
Small	0.298
Total	0.270

Source: own calculations.

7.4 ANALYSIS OF THE DETERMINANTS AFFECTING RECOVERY RATES

As we have already presented the statistical methodology and explanatory variables in previous chapters, we can continue with a statistical analysis and the practical background of the result as already started in the previous section with the contribution of the size of the exposure at default to cumulative recovery rates.

According to the literature mentioned herein, the most important determinant for calculating LGD is the collateralisation of each facility. Banks consider this when pricing a loan. Providing more valuable collateral may help reduce the interest one pays. If one's rating is relatively poor, collateral may help in getting a loan. It pays to inquire what types of collateral one's bank is willing to accept. Note that banks are very conservative in estimating the value of collateral as it is difficult to assess the actual recovery value in the case of default and since it requires a considerable effort by the bank to sell collateral to recover loan losses. The impact of collateral on reducing the risks of a loan depends on its type and liquidity (see Figure 10 in Annex 6 for a qualitative estimation).

Table 14: Average LGD rates by forms of collateral

Forms of collateral/guarantee	Average of LGD
Assignment of receivables	0.604
Guarantees	0.280
Real estate collateral	0.236
Financial collateral	0.139
Physical collateral**	0.093
Total	0.270

*** Note: it is noted physical collateral is unusually low*

Source: own calculations.

Table 14 presents the average LGD rates regarding the different securitisation types. The assignment of receivables as one of the important collateral types used by Slovenian banks has proven to be very inefficient collateral. After holding consultations with an expert in the fields within bank x, the results are not surprising since collectors manage to collect very few sources from the assignment of receivables and have more costs than profit arising from it. The contracts usually prove to be out-of-date and very ineffective. We can find reasons in bad collateral management in Slovenia especially in terms of the assignment of receivables. Financial collateral is, according to the results, expected to be the safest type of collateral, excluding the unusually low losses on physical collateral (also see Figure 12 in Annex 8).

The predictor last rating was included as a determinant as we can see considerable changes in losses depending on the last rating given to the obligor by credit analysts. Banks use ratings as the main input for calculating the expected loss implied by a given loan. In addition, the required share of the capital to be set aside to take into account the possibility that losses will be higher than expected will also depend on the rating. Thus, the rating is the key indicator of the cost a bank incurs for a given loan and it was taken as a predictor in LGD.

The losses ascend in accordance with worse ratings, ranging from a 15% average loss on those exposures rated C to a 44% average loss on those exposures having a rating E as we can see from Table 15 (also see Figure 13 in Annex 8).

Table 15: Average LGD rate by last rating

Last rating*	Average of LGD
C	0.146
D	0.223
E	0.441
Total	0.270

**Note: rating system of the Bank of Slovenia*
Source: own calculations.

Table 16: Average LGD rate by aggregated industrial sectors

Aggregated industrial sectors	Average of LGD
Trade	0.288
Real	0.282
Services	0.257
Manufacturing	0.211
Total	0.270

Note: aggregated sectors consist of the following sectors according to European Union's economic activity codes (NACE): Real (sectors C, F, H, K); Manufacturing (sector D), Trade (sector G), Services (sectors E, I, J, M, O).
Source: own calculations.

The results in Table 15 can be compared with the risk coverage in the portfolio of the Bank of Slovenia as well as in the case of bank x. We can observe that the loan loss provisions for ratings C, D and E, which are treated as signs of bad debt, are substantially higher than those observed in our results. Figure 14 in Annex 8 presents the difference between the average LGD rate by last rating and loan loss provisions by the rating required by the Bank of Slovenia.

The predictor aggregated industrial sector is another important factor of risk calculation. This predictor is a derivative of a fully-segmented factor of industrial activity by NACE. Table 16 reports (see Figure 15 in Annex 8 for the distribution of the average EAD by aggregate industrial sectors) that loans in the trade sector (Sector G) experience the highest risk of having a loss after facing a default (29%). Manufacturing (Sector D) seems to be the safest economic activity from the creditor perspective of the obligor's repayments after a default occurs (21%).

Another predictor that can explain to what extent the average loss on one's facility may be expected is the method by which the recovery process is executed. Table 17 presents the frequency of methods used by bank x and what are the average results in terms of loss by each of the methods (see Figure 16 in Annex 8). According to the sample, the most frequent method noted is the method 'others' (44%), following by a repayment arrangement with the client (13%) and regular repayments (12%).

Table 17: Average LGD rate by recovery method

RECOVERY METHOD (bank X classification)	Average of LGD	Frequency of recovery method
2B - Debt rescheduling along execution process	0.540	2%
7 - liquidation proceedings (bankruptcy)	0.455	6%
4 - Non-judicial foreclosure or execution	0.369	3%
10B – Others	0.354	44%
1A - Sale of claims to a third party	0.226	7%
10A - Regular repayments	0.196	12%
1B - Repayments arrangement with a discount for a client	0.124	6%
1C - Repayment arrangement with a client (mostly in one instalment)	0.116	13%
5A - Repayment from court auction sale	0.092	2%
2A – Debt rescheduling – not in execution process	0.058	3%
3 - Informal workout	0.055	2%
Total	0.272	100%

Note: Method 8a – Formal rehabilitation (reprogramming) was excluded due to the unrepresentative sample (just one case in the sample).
Source: own calculations.

From Table 17 we also observe the highest average losses by the methods debt rescheduling along execution process (54%) followed by liquidation proceedings (bankruptcy) where we can expect a loss of 45%. Informal workout and debt rescheduling without an execution process bring the most favourable results with the lowest average losses for the bank. Also repayment from a court auction sale, where mortgages as collateral are sold, testify that the collateral-type mortgage reflects a good quality securitisation.

The loan's maturity is another important factor in calculating the price of a loan. It is taken into account by almost all banks. Commonly, interest rates are lower for short-term loans than for long-term ones. This is because uncertainty rises with a longer loan maturity (European Commission, 2005, p. 24). What we can see from the empirical results in Table 18 is that there are higher average losses on short-term loans (30% on short-term against 23% on long-term ones). We can find the reasons for this in a more thorough review of the sample by type of loan (see Tables 19 and 20 and Figures 17 and 18 in Annex 8). In Table 18 we also see the average losses by secured loans (26%) and unsecured loans (38%).

Table 18: Average LGD rate by maturity of loan (type of loan) and by securitisation

Securitisation	Type of loan		Total
	Long-term	Short-term	
Secured	0.215	0.291	0.255
Unsecured	0.537	0.338	0.378
Total	0.233	0.299	0.270

Source: own calculations.

Considering the sample's average LGD rate by type of loan in terms of the amount of the debt outstanding at the time of default and the loan's maturity (short-term or long-term), both in relation to a facility's collateralisation, I highlight the following conclusions:

- Short-term loans experience higher losses due to worse collateral types (especially the assignment of receivables – 60% average LGD).
- Real estate collateral seems to be relatively effective collateral with the average recoveries being quite similar for either short-term or long-term loans (23% average LGD).
- Financial collateral as one of the primary collateral types (deposit or any kind of securities) is limited for large exposures where its value represents just a certain percentage of the exposure. Usually there is another type of collateral combined with financial collateral for the same facility. Large exposures are also often long-term ones and therefore we are witness to larger losses in the long-term where financial collateral is the primary collateral.
- Most unsecured facilities are short-term ones, which has also contributed to the higher average LGD of short-term loans.

Table 19: Average LGD rate by type of facility (exposure and securitisation)

Debt outstanding at the time of default (SIT)	Assignment of receivables	Unsecured	Guarantee	Real estate collateral	Financial collateral	Physical collateral	Total
Large		0.068	0.880	0.045	0.416	0.050	0.218
Medium		0.504	0.245	0.185	0.089	0.000	0.231
Small	0.604	0.342	0.247	0.328	0.079	0.127	0.298
Total	0.604	0.378	0.280	0.236	0.139	0.093	0.270

Source: own calculations.

Table 20: Average LGD by type of facility (maturity and securitisation)

Type of loan	Assignment of receivables	Financial collateral	Guarantee	Physical collateral	Real estate collateral	Unsecured	Total
Long-term		0.199	0.158	0.117	0.237	0.537	0.233
Short-term	0.604	0.078	0.308	0.045	0.232	0.338	0.299
Total	0.604	0.139	0.280	0.093	0.236	0.378	0.270

Source: own calculations.

7.5 RESULTS OF THE ECONOMETRICAL MODEL

IDENTIFYING THE OUTCOME VARIABLE AND PREDICTORS FOR THE LOCATION MODEL

After having identified the outcome variable of the model, which we divided into five classes ranging from 0 to 100% recovery, we also chose all the predictors for the location component of the model. A preliminary exploratory analysis was undertaken in order to identify the likely explanatory variables already presented in the paper. Another step included empirical considerations to evaluate the importance of each variable. All of the dependent variables in the model are entered as factors since they are all categorical variables, namely:

- primary collateral type;
- sector of industry;
- type of loan;
- last rating;
- size of the debt outstanding at the time of default; and
- recovery method.

We chose the location-only component assuming that the scale component will not be necessary.

7.5.1 Evaluating the Model

Due to the relatively small population of the available data we were constrained to make aggregations of the groups on some explanatory nominal variables. With more segmented variables we increase our determinant coefficients in the regression, but the model is only 'artificially' better as this coefficient is not corrected by the number of degrees of freedom (see Table 28 in Annex 9).

PREDICTIVE VALUE OF THE MODEL

To answer the question of whether the model gives adequate predictions we can examine the model-fitting information table (see Table 29 in Annex 9). Here we see the -2 log-likelihood values for the intercept-only (baseline) model and the final model (with the explanatory variables). While the -2 log likelihood statistics themselves are suspect due to the large number of empty cells in the model, the difference of log-likelihoods can usually still be interpreted as chi-square distributed statistics (McCullagh, Nelder, 1989). The significant chi-square statistics indicates that the model gives a significant improvement over the baseline intercept-only model. This basically tells us that the model gives better predictions than if we merely guessed on the basis of marginal probabilities for the outcome categories.

CHI-SQUARE-BASED FIT STATISTICS

The goodness-of-fit table (see Table 30 in Annex 9) contains Pearson's chi-square statistics for the model and another chi-square statistic based on the deviance. These statistics are intended to test whether the observed data are inconsistent with the fitted model. The significant Pearson's chi-square indicates that the data and the model predictions are similar and that we do have a good model. These statistics can be sensitive to empty cells so we tried to reduce the empty cells by grouping the explanatory variables as in the case of an aggregated industrial sector.

PSEUDO-R^2- MEASURES

These measures attempt to serve the same function as the coefficient of determination in linear regression models, namely to summarise the proportion of variance in the dependent variable associated with independent variables. In ordinal regression models, these measures are based on likelihood ratios rather than raw residuals. Three different methods are used to estimate the coefficient of determination (see Table 31 in Annex 9). Cox and Snell's R^2 is a well-known generalisation of the usual R^2 designed to be applied when a maximum likelihood estimation is used, as with ordinal regression. However, with categorical outcomes it has a theoretical maximum value of less than 1.0. For this reason, Nagelkerke (1991) proposed a modification that allows the index to take values in the full range from 0 to 1. McFadden's R^2 is another version that is based on the log-likelihood kernels for the intercept-only model and the full estimated model (SPSS Advanced Models 10.0, 1999, p. 252).

In our model, the Nagelkerke's Pseudo R-square is respectable with values around 0.36. It is expected for this type of analyses since there is a large variation in recovery rates and we are limited with our set of explanatory variables.

CLASSIFICATION TABLE

The next step in evaluating the model is to examine the predictions it generates. The model is based on predicting cumulative probabilities. However, what we are most interested in is how often the model can produce correct predicted categories based on the values of the predictor variables as it is important for calculating a credit's price. For these reasons we constructed a classification table (a confusion matrix) by cross-tabulating the predicted categories with the actual categories (see Table 32 in Annex 9).

The model seems to do a respectable job of predicting outcome categories, at least for the most frequent category (80–100% recovery). The model correctly classifies 93% of the last category cases. On the other hand, the model with generated predictive variables predicted relatively poorly according to the category we were most interested in, namely the worst recovery category (0–20%). 45% of the cases were assigned to this category, but 75% of the cases in the range of 20–40% recovery were predicted at a 0–20% recovery, which is a desirable result for predicting ordinal responses as the cases in category 2 are more likely to be classified as category 1 (0–20% recovery) than category 5 (80-100%).

TEST OF PARALLEL LINES

For location-only models, the test of parallel lines can help us assess whether the assumption that the parameters are the same for all categories is reasonable. This test compares the estimated model with one set of coefficient for all categories (see page 21) to a model with a separate set of coefficients for each category:

$\text{link}(\gamma_j) = \theta_j - (\beta_{1j}x_1 + \beta_{2j}x_2 + \ldots + \beta_{kj}x_k)$

We can see that the general model (with separate parameters for each category) gives a significant improvement in the model fit (see Table 33 in Annex 9). We cannot assume that the values of the location parameters are constant across the categories of the response.

PREDICTORS IN THE MODEL

The individual predictors in the model can be seen in the parameter estimates (see Table 34 in Annex 9). The threshold parameters are not particularly important from a theoretical standpoint. We focus on location parameters which relate the predictor variables to the cumulative recovery category probabilities.

In our case, the statistical tests showed that the recovery method did not have much value in the model and it was therefore excluded from the model so as to make it more parsimonious.

7.5.2 Revising the Model

First we consider it worth trying another link function. We estimated a new model with the complementary log-log link function as another of the link functions that may be appropriate to see whether the change increases the predictive utility of the model, keeping the same set of independent variables.

The model fitting statistics (see Table 35 in Annex 9) indicates that this model is once again better than simple guessing. The chi-square statistic, comparing the full model to the intercept-only model, has about the same magnitude. Considering the pseudo-R^2 measures (see Table 36 in Annex 9), we can see that changing the link function does not improve the model's ability to account for patterns in the outcome variable.

Looking at the classification table (see Table 37 in Annex 9) this model seems to be slightly worse at predicting the lower categories, which we consider to be an important indicator that we should retain the original model. The most important goal of the risk management division is to find and estimate those clients that might be risky in their recoveries.

When looking at the classification table, we observe very good results of predictions when considering all the predictors segmented to the official classifications of each predictor as tabled by the bank. But the underlying problem is the too small set of available data, thus a high level of singularities in the Fisher information matrix are encountered.

7.5.3 Interpreting the Model

With a model that gives an adequate fit to the data, as in the case of our original model, we can make some interpretations based on parameter estimates (see Table 34 in Annex 9).

According to the statistical tests, all our explanatory variables in the model seem to be important. Due to the minimisation of the number of explanatory variables, the aim of which is to have a statistically correct model, we excluded the recovery method from the model.

The first observation is that, as expected, the collateral variables (real estate, physical) have a statistically significant positive effect on the cumulative recovery at a significance of 0.05. Financial collateral also has a positive effect. As was the case of the univariate figures (although not statistically significant), the assignment of receivables has a negative effect on the cumulative recovery.

The aggregated industrial sector dummies are significant and positive in most cases and confirm the observation that recoveries in the aggregated sector 'trade' (the base case) are lower than the other four aggregated industrial sectors.

The next observation to confirm our empirical analysis of the cumulative recovery rates was the last rating. A better rating also brings higher recoveries.

The sign of coefficients for the explanatory variable 'the size of the debt outstanding at the time of default' gives us another important insight into the effects of the predictors in the model. Positive signs (although not statistically significant at $p=0.05$) indicate that recoveries are higher with larger exposures, confirming our observation in the empirical analysis and the sample univariate weighted and unweighted average cumulative recovery rates.

Finally, the model shows a statistically significant negative sign for a long-term loan. This indicates an inverse relationship in relation to the output variable. A long-term loan is expected to fall into a lower category than a short-term one, which means a lower level of recovery.

PREDICTIONS

For each case, the predicted outcome category is simply the category with the highest probability given the pattern of the predicted values for that case. For example, suppose we have a small (debt outstanding at the time of default) short-term (type) loan from a client in the industrial sector »trade«, having financial collateral as securitisation and being finally rated C, then the model gives the following individual category probabilities:

- recovery 0-20%: 0.06;
- recovery 20-40%: 0.00;
- recovery 40-60%: 0.01:
- recovery 60-80%: 0.01; and
- recovery 80-100%: 0.92.

Clearly, the last category with the highest recovery is the most likely category for this case according to the model, with a predicted probability of 0.92. We thus predict that this facility will be successfully recovered.

7.5.4 Robustness Tests

Two types of robustness tests have been conducted. The first test was a random selection of 90% of the observations. The results in terms of coefficient signs are consistent with those of the specifications in the original tests and so with those of the base specifications, confirming the statistical significance of the loan collateralisation, the last rating, the type of loan, and of the same significance level by the type of industry and size effect (see Table 38 in Annex 9).

In the second test we wanted to ensure that the size effect was not driven by high recoveries relative to a few large loans so we eliminated the 10% largest loans from the sample. The results were again consistent with the results of the original model and the base specifications. For the sake of space, the estimated parameters are not reported here.

7.6 FIRST ATTEMPT TO CALCULATE EXPECTED LOSS BY INDUSTRIAL SECTORS

Table 21 presents the results of a PD model and the empirical (historical) analysis of LGD. The average expected loss on a loan equals 1.4%, which defines the risk price of a loan for a bank. Figure 19 in Annex 10 shows which industrial sectors are expected to have the highest expected losses. According to the results the most unsafe industries are transport and construction. We observe higher LGDs in these two sectors (about 50%) in comparison to the bank's average (27%), while the PDs are not expected to deviate from the average (5%). The financial intermediation sector has very low LGDs, which contributes to the lowest expected loss.

Table 21: Calculation of expected loss by industrial activities

Sector of industrial activity	Average of LGD	Average of PD	Expected loss (in %)
D – Manufacturing	0.211	0.035	0.747
F – Construction	0.519	0.047	2.432
G - Wholesale and retail trade	0.288	0.045	1.304
H - Hotels and restaurants	0.191	0.058	1.098
I - Transport, storage and communication	0.505	0.049	2.492
J - Financial intermediation	0.046	0.055	0.250
K - Real estate	0.254	0.037	0.951
M – Education	0.275	0.058	1.605
O - Other service activities	0.254	0.064	1.634
Total	0.274	0.050	1.366

Note: sectors C and E were excluded due to the unrepresentative sample (just one case in the sector).
Source: own calculations.

8. WORKOUT COSTS INCURRED WITH RECOVERIES

So far, the analysis has been concerned with gross recoveries. For loan pricing, the calculation of LGD and capital requirements, or the calculation of the loan loss provision we also need to know the recoveries net of the costs incurred by a bank to recover these cash flows. Not much literature provides information on such data with the exception of the study on the Portuguese model (Dermine, Carvalho, 2005). I will report the figures specific to bank x concerning the costs of claims proceedings and recoveries.

At bank x there is one department that handles bad and doubtful loans, namely the Asset Recovery Management (ARM) department. This department refers to all collectors as well as internal lawyers and law courts to recover cash. I will report the data on direct costs incurred in recovery, collected for the 2002-2005 period (see Table 39 in Annex 11). The data do not cover the salaries of the department's employees.

The workout costs are divided into two units: individuals and corporates. Here I present just the costs of recoveries in the corporate portfolio. For reasons of confidentiality, all figures in Table 39 have been scaled by a common multiplicative factor. Therefore only the percentages are relevant. We observe that the average costs of proceeding per recovered cash amounts to 0.44%. These costs include the costs of external lawyers' fees and the costs of proceedings at law courts. According to an expert opinion, recovery costs on smaller loans in terms of the percentage of recovery are substantially higher than on large loans. The same was reported by Dermine and Carvalho (2005) in their study on the Portuguese model.

However, excluding the internal costs of employees' salaries within the ARM department, the ratio workout cost per fund recovered is relatively low. Therefore I will not go into further details of the workout cost structure within bank x. By adding the salaries of the bank's employees to these costs we can make an approximation of 1-2% of workout costs by exposure and apply these workout costs implicitly in the higher discount rate.

9. CONCLUSIONS

Banks have been – and will continue to be for the foreseeable future – the most important source of funding for SMEs. Under the new Basel II framework, the minimum amount of capital banks are required to set aside will (with some exceptions) no longer depend mostly on the size of the loan, but also and significantly on the risk of the loan. Generally speaking, this change will make 'riskier' lending somewhat more expensive (i.e. more capital consuming) for banks, while relatively safer lending will become less costly. The changes introduced by Basel II reflect the trend in the banking industry towards more quantitative and differentiated risk management. On the basis of closer scrutiny, presented in the paper in the case of bank x for SMEs, banks will take their loan decisions and offer a wider range of price and credit conditions.

Loan LGDs were estimated for a sample of 124 corporate loans of a Slovenian bank over the 2001-2005 period. The estimates were based on the discounted value of cash flows recovered after a default event. A univariate approach was applied to measure cumulative recovery on bad and doubtful loans. The average recovery estimate of 73% was in the same order as that obtained in the study of a Portuguese bank (Dermine, Carvalho, 2005). A multivariate approach was then applied to analyse the determinants of recovery rates. Three main conclusions can be drawn from this empirical case study. The first is that the frequency distribution of loan LGD appears bi-modal with many cases presenting recoveries close to 0% and a concentration of other cases presenting high recoveries from 80 to 100%. The second conclusion is that a multivariate analysis of the determinants of loan losses allows us to identify several statistically significant explanatory variables. These include collateral, industry sector, loan maturity and rating. Third, estimates of workout costs incurred by the bank in recovery are estimated at 1-2% of the exposure. A word of caution here is that this study, based on a relatively small available dataset of one single bank, might have captured some of the bank's idiosyncrasies.

The paper provides some crucial information and solutions proving that Slovenian banks could also calculate the risk in their portfolio more accurately. It is a step towards a better understanding of the determinants of bank loan LGD. It gives recommendations on how to store data on recovery over time, which should open the way to the development of a measure of loan loss provisioning. By adding an additional set of the available data into our LGD calculations, gathered from an efficiently run risk data warehouse, it could be validated and represent an important element of comparative advantage in the Slovenian financial market.

10. POVZETEK V SLOVENŠČINI

10.1 *Predmet proučevanja in struktura diplomskega dela*

Diplomsko delo zajema področje upravljanja s kreditnimi tveganji, in sicer nove, napredne notranje pristope na področju kreditnih tveganj, katerim banke namenjajo ali bodo namenjale vse več pozornosti. Novi sporazum s področja kapitalske ustreznosti (CAD III) omogoča vsaki banki uporabo internih modelov za izračun izgube ob neplačilu (LGD), kar spodbuja banke k bolj temeljiti oceni pripadajočih tveganj ob vsaki odobritvi kredita. Banke bodo s tem lahko bolj razločevale dobre in slabe naložbe in s tem ponudile tudi več različnih cen (obrestnih mer) posameznih kreditov.

Diplomsko delo je empirične narave in prispeva k razumevanju kreditnih tveganj za bančna posojila s predstavitvijo metodologije za analizo stopenj izgub po nastanku dogodka neplačila, ki jo uporabimo na vzorcu podatkov banke x o izgubah pri posojilih malih in srednjih podjetij v razdobju 2001-2005. Rezultati predstavljajo prvo tovrstno analizo na portfelju malih in srednjih podjetij kake izmed slovenskih bank. Predstavljene so pojasnjevalne spremenljivke izgub, njihova univariatna in multivariatna analiza kot tudi neposredni stroški, povezani s pridobivanjem poplačil.

Natančne stopnje poplačila posameznih izpostav so pomembne za pripravo popravkov vrednosti in rezervacij (angl. provisions) za kreditne izgube, za izračun tveganega kapitala kot tudi za določitev "pravične" vrednosti za tvegane kreditne obveznosti. Ocena spreminjanja izgub lahko vodi do nižjih kapitalskih zahtev za naslednja leta, v kolikor je ta ocena podprta s smiselnimi empiričnimi dokazi.

Vsebinsko je delo razdeljeno na 9 poglavij. Prvo opravičuje pomen te raziskave, drugo poglavje v ospredje postavi pojem izgube ob neplačilu, metodologijo in pristope izračuna le-tega glede na nove baselske standarde. V tretjem poglavju povzemam dosedanjo svetovno literaturo s tega področja, nedavne empirične študije, predstavljena je tudi povezava med verjetnostjo neplačila (PD) in samo izgubo ob neplačilu, dodanih pa je tudi nekaj referenčnih vrednosti rezultatov študij o izgubah na posojilih. Četrto poglavje zagovarja pomembnost efektivnega upravljanja s podatki z dodanimi predlogi shranjevanja podatkov v podatkovnih skladiščih. V petem poglavju se preselimo k slovenskim bankam, kjer proučujem njihovo pripravljenost na nove baselske standarde ocenjevanja kreditnih tveganj, prikazano bolj natančno na primeru banke x. Potrebno ekonometrično znanje, ki se uporablja za izračun tovrstnih modelov, predstavljam v šestem poglavju, medtem ko se v sedmem poglavju zatečemo k raziskovalnemu delu naloge in empiričnim rezultatom, preverjenimi tudi z modelom. Osmo poglavje na kratko predstavi direktne stroške, zajete pri doseganju poplačil, nato pa sledi zaključek v zadnjem poglavju.

10.2 Opredelitev izgube ob neplačilu

Na kreditno tveganje v bankah vplivajo tri glavne komponente: (1) izpostava ob neplačilu (EAD), (2) verjetnost neplačila (PD) in (3) izguba ob neplačilu (LGD). Zadnji kazalnik predstavlja osrednjo temo te diplomske naloge.

Neplačilo (angl. default) in izguba sta dva pomembna pojma znotraj termina izguba ob neplačilu. Ko dolžnik postane neplačnik, to velja za vse njegove obveznosti. Ta dogodek pa v splošnem nastane v primeru, ko je dolžnik več kot 90 dni v zaostanku s plačili na kateri izmed obveznosti do banke, ali je malo verjetno, da bo v celoti poplačal svoje obveznosti do banke. Izguba pri oceni LGD predstavlja ekonomsko izgubo, v katero Evropska komisija kot tudi Baselski odbor vključujeta izgubo glavnice, izgubljene opuščene obresti kot tudi stroške povezane s prenosom terjatve v izterjavo (angl. workout costs). Upoštevajoč te predpostavke lahko zapišemo formulo za izračun stopnje poplačila (angl. RR, ki je izračunan kot RR = 1 – LGD) na sledeči način:

$$\text{Diskontirana stopnja poplačila (RR)} = \frac{PV(CashFlow)}{EAD} = \frac{PV(R+M-C)}{P+I},$$

kjer števec predstavlja sedanjo vrednost (PV) dejanskih denarnih tokov (CashFlow) v denarni obliki s strani dolžnika, zavarovanja ali garancij (R), tržno vrednost ob prodaji vrednostnih papirjev in drugih nefinančnih zavarovanj (M), zmanjšano za sedanjo vrednost direktnih stroškov, povezanih s samo izterjavo (C). Imenovalec predstavlja izpostavo ob neplačilu (EAD), sestavljeno iz glavnice (P) in obresti (I), ki jih dolžnik dolguje, a jih ni plačal.

LGD kazalnik ocenjujemo lahko s pomočjo subjektivnih metod, ki bazirajo na izkušnjah strokovnjakov in njihovih ocenah, ali s tako imenovanimi objektivnimi metodami, kjer uporabljamo empirične podatke o izgubah. Slednje metode temeljijo na matematičnem ali statističnem modeliranju. Diplomska naloga se osredotoča na objektivno »workout« metodo, kjer je posebna pozornost posvečena časovnemu razporedu plačil, zaradi česar denarne tokove preračunamo na sedanjo vrednost z uporabo ustreznega diskontnega faktorja. Uporabo in višino le-tega lahko banka ob odobritvi nadzornikov določi tudi sama.

Baselski odbor za bančni nadzor loči tri pristope izračunavanja LGD-ja, in sicer od najbolj preprostega (standardizirani pristop) do dveh načinov notranjih pristopov bonitet (osnovni in napredni notranji pristop bonitet), kjer slednji prikazuje najbolj kompleksno in najbolj fleksibilno obliko pristopov. Banke, ki bodo sprejele napredne pristope, morajo zbrati podatke o preteklih poplačilih za obdobje najmanj sedmih let, in sicer morajo informacije obdržati na nivoju posamezne naložbe in ne dolžnika. Glavna razlika med notranjima pristopoma je v dejstvu, da so pri osnovnem pristopu vsi izračuni direktno bazirani na sedanji vrednosti zavarovanja, medtem ko morajo biti pri naprednem pristopu vse vrednosti zavarovanj ocenjene v luči preteklih stopenj poplačil, ki pa so poleg vrste zavarovanj odvisne tudi od drugih kazalnikov (npr. metode poplačila, sektorja gospodarske dejavnosti), ki značilno vplivajo na pojasnjevanje stopnje poplačila.

10.3 Stopnje poplačila slabih posojil

Literatura o kreditnem tveganju se je do nedavnega v večini primerov posvečala prvi izmed treh glavnih komponent kreditnega tveganja (PD, LGD in EAD), torej verjetnosti neplačila. V zadnjem času pa se je pojavil porast študij, posvečenih stopnji poplačila v primeru neplačila in njeni povezavi z verjetnostjo neplačila.

Modele kreditnih tveganj, ki so bili razviti v zadnjih 30 letih, ločimo na dve glavni skupini, in sicer na modele kreditnih ocen (angl. credit pricing models), ki jih delimo na strukturne modele (angl. structural form approaches) in pol-strukturne (angl. reduced form models). Osnova razumevanja strukturnih modelov je naslednja: neplačilo se pojavi, ko tržna vrednost sredstev podjetja pade pod nivo vrednosti obveznosti tega podjetja. Poplačilo je v tem primeru enako tržni vrednosti podjetja. Začetnik in glavni predstavnik te veje je bil Merton. Kasnejši predstavniki strukturnih modelov so odstranili prvotno nerealistično predpostavko, da se neplačilo lahko pojavi le ob dospelosti dolga. Pol-strukturni (angl. reduced form models) se v osnovi od tipičnih strukturnih modelov razlikujejo v stopnji možnosti napovedovanja neplačila, saj upoštevajo tudi nenadna presenečenja. PD in LGD so računani neodvisno od strukturnih značilnosti podjetja. Neplačilo se pojavi, ko neka zunanja naključna spremenljivka, ki predstavlja ključni faktor za nastanek neplačila, prestane diskretno spremembo v njeni vrednosti, čas te spremembe pa ne more biti napovedan. Večina modelov se nanaša na obveznice, kar ni vedno aplikativno tudi za bančna posojila.

V drugi polovici devetdesetih let so velike banke in svetovalci začeli razvijati modele kreditnih tveganj z namenom ocenjevanja potencialne izgube ob predhodno določeni stopnji zaupanja (JP Morgan's CreditMetrics, KMV's CreditPortfolioManager). Pri teh modelih tvegane vrednosti (VAR) ločimo dve vrsti, in sicer glede na količino podatkov, potrebnih za model: DM (default mode) modele in MTM (mark-to-market) modele. Druga, torej modeli trenutne tržne vrednosti, zahteva večjo količino podatkov. Glavni proizvod VAR modelov kreditnih tveganj je funkcija verjetnostne porazdelitve (PDF) prihodnjih izgub na kreditnem portfelju. S pomočjo take analize finančne institucije lahko ocenijo pričakovane in nepričakovane izgube na njihovem portfelju.

Analitiki v bankah so se začeli ukvarjati s svojimi internimi bazami podatkov z namenom prilagoditve na zahteve naprednih IRB pristopov. Ena boljših empiričnih raziskav o faktorjih ki vplivajo na LGD, ki lahko služi tudi kot primerjava za raziskave v Sloveniji, je bila študija avtorjev Dermine in Carvalho (2005) na portfelju malih in srednjih podjetij največje portugalske banke Banco Comercial Portugues.

10.4 Priporočila za upravljanje s podatki

Najpomembnejši del pri določanju realiziranega LGD-ja kot tudi pričakovanega LGD-ja so podatki in njihova kvaliteta. Banke vzpostavljajo skladišča podatkov za obvladovanje tveganj, v katerih naj bi se beležile vse potrebne informacije, povezane z baselskimi standardi. Če želimo vzpostaviti napredne IRB pristope, morajo biti informacije zbrane na ravni posamezne naložbe. Podatkovno skladišče za področje tveganj naj bi tako vsebovalo vse dejavnike tveganj, potrebne za oceno tako LGD-ja kot tudi ostalih parametrov tveganj. Ti dejavniki se razlikujejo glede na različne parametre, ki jih želimo izračunati. Dejavniki za oceno izgube naj bi se nanašali na: informacije o transakciji (vrsta naložbe, zavarovanje, metoda poplačila), posojilojemalca (vrsta sredstev, geografska regija, vrsta gospodarske dejavnosti), notranje informacije organiziranosti in nadzora procesa pridobivanja poplačila, kot tudi na zunanje informacije o obrestni meri in pravnem okvirju. Basel II ne predpisuje strukture podatkovnih sistemov.

Del osnutka direktive o kapitalskih zahtevah (CRD) se nanaša tudi na blaženje kreditnih tveganj (CRM), ki so v večini pogojena z vrsto zavarovanj, zato je smiselno povzeti nekaj priporočil njihovega skladiščenja podatkov. Najboljša pot za banke je vzpostavitev sistema upravljanja z zavarovanji. Ta sistem bi poleg splošnih informacij o vrsti zavarovanj na posameznih naložbah pri posameznih posojilojemalcih vseboval tudi datume vrednotenja oz. realizacije posojila, kajti tudi ti podatki igrajo pomembno vlogo pri računanju izgube. Informacija o časovni veljavnosti posojila je pomembna, saj se ta ne sklada vedno z dospelostjo posojila, na katerega se nanaša. Omeniti velja še redna ponovna vrednotenja posojil z namenom sledenja tržni vrednosti zavarovanja.

10.5 Pripravljenost slovenskih bank

Dostopnost podatkov kot tudi drugi problemi (npr. majhnost slovenskih bank) otežujejo bankam uporabo bolj sofisticiranih matematičnih in statističnih metod pri ocenjevanju LGD-ja. Za uporabo ustrezne regresije moramo identificirati potencialne dejavnike izgube, ki morajo predstavljati statistično značilno informacijo. V omenjenih oteževalnih okoliščinah je ta zahteva le redko dosežena, vendar empirične podatke lahko še vedno uporabimo za analizo povprečnih izgub.

Dejansko stanje pripravljenosti na napredne pristope predstavljam na primeru banke x, kjer sem se lotil pridobivanja podatkov z namenom določitve potrebnih dejavnikov za model ocenjevanja poplačil ob neplačilu za mala in srednja podjetja. Zbrani podatki so na letni osnovi (zadnji dan v letu). Problem se pojavi pri časovni komponenti zajemanja podatkov, saj naj bi glede na baselske predpise ta interval opazovanja ne bil krajši od sedmih let. Uspel sem pridobiti podatke za obdobje 2001-2005.

Največjo težavo pri pridobivanju podatkov predstavlja stopnja poplačila kot naša odvisna spremenljivka. Poplačila so namreč v bazah obeležena le na nivoju komitenta in ne na nivoju posamezne naložbe. Tu so zato potrebne določene prilagoditve. Pri podjetjih z več naložbami smo vzeli povprečno stopnjo poplačila, ki so jo ta podjetja dosegla glede na celotno izpostavo.

Podatki o lastnostih zavarovanj na posamezni naložbi, ki je pri računanju izgub ključna komponenta, so relativno dobro prikazani v enem izmed virov podatkov v banki x, kjer je večina potrebnih informacij zajeta. Tu gre omeniti tudi zakone in pravila v zvezi s pravicami kreditodajalca do kreditojemalca, ki vplivajo na postopke unovčevanja zavarovanj. Vsaka država ima svoje zakone o stečaju, ki jih včasih ne moremo neposredno primerjati med državami. Tako je npr. stopnja zavarovanj v Franciji v povprečju višja v primerjavi z Nemčijo in Veliko Britanijo, kar so raziskovali pri S&P (Franks, Servigny, Davydenko, 2004).

Podatki o gospodarski panogi, v kateri posluje podjetje, kot dejavniku izgube, so na voljo v internih bazah, prav tako pa ima banka x že dostopne podatke o povprečni verjetnosti neplačila po posameznih panogah.

O posamezni naložbi sem poleg že omenjenih dejavnikov zavarovanja in industrijske panoge uspel pridobiti tudi podatke o dospelosti posojila (kratkoročno, dolgoročno), zadnjo bonitetno oceno kreditnih analitikov in metodo, preko katere je bilo izvedeno poplačilo. Pridobljeni dejavniki za model v banki x, ki vplivajo na izgubo, so primerljivi s tistimi iz različnih raziskav v dosedanji literaturi.

Ciklična komponenta na primeru Slovenije v zadnjih letih ni igrala pomembne vloge, saj je gospodarstvo uspešno raslo in tako ni pričakovati izrazitih odstopanj ali padcev nivoja poplačil na račun krize v gospodarstvu (angl. downgrade LGD) ali kakega drugega makroekonomskega dejavnika.

Podatki, potrebni za pripravo LGD modelov, se bodo razlikovali glede na metodologijo, s katero bodo banke določale LGD. V vsakem primeru bodo banke morale zbirati informacije o oslabitvah (popravkih vrednosti in rezervacije), datumih in vrednostih odplačil, virih poplačila (iz naslova zavarovanja, garancije ali samega dolžnika), vrednosti in vrstah zavarovanj, stroških pridobivanja sredstev. Banka x večino od omenjenih informacij ustrezno beleži v svojih bazah, dobrodošla pa bi bila večja povezanost med njimi.

10.6 Statistična metodologija

Pred analizo podatkov moramo opredeliti vrsto spremenljivk, ki se pojavljajo v našem modelu. Poznamo dva osnovna tipa kategoričnih spremenljivk, in sicer nominalne in vrstilne (ordinalne) spremenljivke. Slednje imajo za razliko od nominalnih spremenljivk razrede razvrščene po velikosti.

Na podlagi ugotovitve vrst spremenljivk poiščemo najbolj primerno statistično metodo. Izbrali smo posplošitev linearne regresije, ki jo imenujemo tudi posplošeni linearni modeli (angl. generalised linear models ali GLM), za napoved kumulativnih verjetnosti posameznih razredov. Tako dobimo za vsak razred ordinalne odvisne spremenljivke svojo enačbo, ki nam napove verjetnost nahajanja v nanašajočem ali kakem nižjem razredu.

V statističnem orodju SPSS smo izbrali ordinalno regresijo kot eno izmed oblik GLM statistike, primerno za analizo našega modela. Namesto napovedovanja aktualnih kumulativnih verjetnosti model napoveduje funkcijo teh vrednosti, imenovano tudi povezovalna funkcija (angl. link function).

10.7 Specifikacija LGD modela na primeru banke x

LGD model predstavljen v diplomskem delu je pilotski model za eno od slovenskih bank, katerega cilj je ocena izgube ob neplačilu (LGD) za mala in srednja podjetja. V bazo podatkov smo zajeli vsa podjetja neplačnike s povprečno izpostavo več kot 1 mio SIT. Vse zahtevane informacije smo uspeli pridobiti za 124 podjetij oziroma njihovih naložb, ki so postala neplačniki v razdobju 2001-2004 in so imeli zaključene postopke poplačila do konca leta 2005.

Definicija, ki smo jo uporabili kot kriterij neplačila, je 90 dnevna zamuda s plačilom, kot je tudi predlagano s strani baselskega odbora za bančni nadzor. Ocenjevali smo neto sedanjo vrednost (NSV) dejanskih poplačil, ko je bilo podjetje ocenjeno kot neplačnik. Za diskontno stopnjo smo vzeli povprečno obrestno mero v razdobju 2001-2004, ki je znašala 10 %, saj nismo imeli podatka o dejanski obrestni meri vsakega od posojil. Potrebno je omeniti še dve prilagoditvi, ki smo jih upoštevali pri pridobivanju naše baze podatkov, in sicer smo iz našega vzorca izločili vsa podjetja, katerim so bile ukinjene rezervacije (angl. upgraded), poleg tega pa smo zajeli le podjetja in njihove izpostave, katerih postopki poplačil so bili končani do konca leta 2005.

Pojasnjevalne spremenljivke poplačil, ki smo jih uporabili za model kot tudi za empirično analizo, so naslednje:

- zavarovanje (finančno – bančni depozit, vrednostni papirji; nepremičnine – hipoteka; fizično – premičnine; poroštvo; odstop terjatev);
- tip posojila glede na dospelost (kratkoročno ali dolgoročno posojilo);
- gospodarska panoga glede na klasifikacijo gospodarske dejavnosti po NACE;
- velikost izpostave: majhna (< 10 mio SIT), srednja (10 mio – 100 mio SIT), velika (> 100 mio SIT);
- zadnja bonitetna ocena (C, D ali E po klasifikaciji Banke Slovenije); ter
- metoda poplačila.

Opisna statistika nam pokaže, da je bilo v našem vzorcu 46% posojil, ki so postala neplačniki v letu 2001. 61% posojil v vzorcu je imelo izpostavo ob neplačilu manjšo od 10 mio SIT. V 12% primerov posojilo ni imelo zavarovanja, najpogostejša oblika zavarovanja pa je bila hipoteka (44%). Večina nezavarovanih posojil je bila kratkoročnih. Primeri neplačil so se zgodili v vseh gospodarskih panogah, največ v trgovini (45%).

Gospodarske panoge smo za namen ekonometričnih testov združili v štiri agregatne sektorje aktivnosti[8]: realni sektor (C, F, H, K), predelovalne dejavnosti (D), trgovina (G) in storitve (E, I, J, M, O).

Povprečna vzorčna kumulativna stopnja poplačila je bila 73%, kar predstavlja dokaj podobne rezultate kot so jih poročali Asarnov in Edwards (1995), Hurt in Felsovalyi (1998) za Latinsko Ameriko ter Dermine, Carvalho (2005) za Portugalsko. Povprečna vzorčna kumulativna stopnja poplačila za posojila z zavarovanji je bila 76%, za posojila s poroštvom kot vrsto zavarovanja 72% in za nezavarovana posojila 62%. To potrjuje hipotezo, da vsako izmed zavarovanj doprinese k višjemu poplačilu.

Analiza porazdelitve kumulativne stopnje poplačila na vzorcu posojil pokaže na tipično bimodalno porazdelitev s koncentracijo dogodkov poplačil višjih od 90% izpostave ob neplačilu in zgostitvijo poplačil v razredu od 0-10%. Podobno porazdelite poplačil ugotavljajo Dermine in Carvalho (2005) za Portugalsko, Schuermann (2004) za Ameriko in Hurt, Felsovalyi (1998) za Latinsko Ameriko.

Vzorčne kumulativne stopnje poplačila, tehtane z vrednostjo izpostave, pokažejo pozitiven vpliv velikosti izpostave na stopnjo poplačila. Tako imajo posojila z majhno izpostavljenostjo v povprečju višjo izgubo (30%) v primerjavi s velikimi posojili in visokimi izpostavami ob neplačilu (22%). Razloge gre iskati v slabših zavarovanjih pri malih izpostavah v primerjavi z večjimi.

Pri analizi dejavnikov, ki vplivajo na poplačila, se najprej ustavimo pri zavarovanju, saj je sodeč po literaturi to tudi najpomembnejša determinanta poplačila, ki jo upoštevajo banke pri določanju obrestnih mer in dodelitvi posojil. Na podlagi analize vzorca ugotovimo, da je najslabša oblika zavarovanja odstop terjatev, saj so le-te v večini primerov ob nastopu neplačila že zapadle. Najboljša oblika zavarovanja je finančno zavarovanje v obliki depozita ali vrednostnih papirjev.

Zadnja bonitetna ocena igra pomembno vlogo pri oceni, kakšna bo končna stopnja poplačil. Banke uporabljajo bonitetne ocene kot glavni učinek vpliva na pričakovano izgubo pri danem posojilu. Zahtevani delež kapitala, namenjenega rezervacijam, upoštevajoč verjetnost da bo izguba večja od pričakovane, je tudi odvisen od bonitetne ocene. Izgube naraščajo z zniževanjem bonitetnih ocen, tako je bila izguba posojil z zadnjo bonitetno oceno C,

---

[8] Imena agregatnih sektorjev dejavnosti ne ustrezajo nujno sami razvrstitvi posamezne gospodarske dejavnosti v širšo skupino.

ugotovljena na podlagi našega vzorca, v povprečju 15%. Posojila z bonitetno oceno E so izkazovala 44% izgubo. Če upoštevamo oslabitve na naložbah, ki so zahtevane s strani Banke Slovenije, vidimo, da so te v povprečju približno dvakrat višje od dejanskih izgub.

Analiza izgub glede na štiri agregatne sektorje industrijske dejavnosti pokaže najvišjo povprečno izgubo v sektorju trgovskih dejavnosti (29%). Za sektor predelovalnih dejavnosti ugotovimo najnižjo povprečno stopnjo izgube (21%).

Naslednji dejavnik, katerega vpliv na končno poplačilo smo upoštevali, je bila metoda poplačila. V vzorcu se najpogosteje pojavi metoda »drugo« (44%), kateri sledi dogovorno poplačilo s stranko (13%) in pa redna poplačila (12%). V primeru reprograma ob tožbi so bile v povprečju največje izgube (54%), sledi jim stečaj ali likvidacija, kjer lahko v povprečju pričakujemo 45% izgubo. Prestrukturiranje in reprogram, ki še ni v tožbi, se izkažeta kot metodi, pri katerih smo bili priča najmanjšim izgubam.

Zapadlost posojila je bil zadnji izmed dejavnikov, ki smo jih opazovali. Večinoma so obrestne mere nižje pri kratkoročnih posojilih v primerjavi z dolgoročnimi, saj naj bi negotovost naraščala z daljšo dospelostjo posojila. Empirična analiza na podlagi vzorca pokaže ravno nasprotno. Kratkoročna posojila (povprečna izguba 30 %) se izkažejo za slabše poplačana v primeru neplačila kot dolgoročna posojia (23% povprečna stopnja izgube). Razloge za to gre iskati v natančnejšem pregledu posojil, zajetih v vzorcu. To pa privede do naslednjih ugotovitev:

- kratkoročna posojila dosegajo večjo izgubo zaradi slabšega zavarovanja (odstop terjatev, kjer je stopnja izgube kar 60%);
- hipoteka je dokaj zanesljivo zavarovanje s 23% povprečno stopnjo izgube na naložbah, zavarovanih s hipoteko;
- finančno zavarovanje kot najbolj likvidna vrsta zavarovanj ponavadi ne dosega vrednosti, ki bi zadoščala za pokritje izpostave, zaradi tega se v večini primerov poleg npr. depozita pojavlja še druga vrsta zavarovanj;
- večina nezavarovanih naložb je kratkoročnih, kar tudi doprinese k večji izgubi pri kratkoročnih posojilih.

Po opravljeni empirični analizi naloga prikazuje še rezultate modela, ki sem ga izdelal s pomočjo SPSS-a. Določili smo že omenjene pojasnjevalne spremenljivke, ki v našem modelu vplivajo na odvisno spremenljivko (stopnja poplačila). Glede na značilnosti porazdelitve naše odvisne spremenljivke smo izbrali povezovalno funkcijo 'Cauchit'.

Statistično značilna Hi-kvadrat statistika pokaže, da nam model da boljšo napoved, kot če bi le domnevali na podlagi mejnih verjetnosti za razrede odvisne spremenljivke. Napovedi modela se tudi statistično značilno prilagajajo podatkom, kar priča o dobrem modelu. Determinacijski koeficienti pokažejo, da je 36% variance pojasnjene z našimi pojasnjevalnimi spremenljivkami. Iz klasifikacijske tabele razberemo, da je model pravilno napovedal 93% primerov, ki so se dejansko nahajali v razredu s poplačili 80 do 100 %. 45% primerov, ki so se nahajali v kategoriji s poplačili le do 20% vrednosti izpostave, je bilo pravilno napovedanih. Slabše napovedi pri tej kategoriji malce popravi dejstvo, da je bilo 75% primerov poplačil iz drugega razreda (20-40% poplačilo) napovedano poplačilo 0-20%, kar je v skladu z ordinalnostjo (primeri iz drugega razreda so bližje prvemu kot petemu).

Pri interpretaciji modela smo ugotavljali predvsem statistično značilnost koeficientov ter njihov predznak. Pri večini koeficientov so se potrdili rezultati vplivov, pridobljenih na podlagi empirične analize kljub temu, da nekateri izmed njih niso bili statistično značilni pri stopnji 0,05. Model pa pokaže statistično značilen negativen vpliv dospelosti posojila na njegovo poplačilo, kar je v skladu s teoretičnimi predpostavkami, vendar nasprotno od empiričnih rezultatov, pridobljenih na podlagi našega vzorca. Za dolgoročno posojila se pričakuje, da se bodo nahajala v nižjem razredu kot kratkoročna posojila, kar posledično pomeni nižjo stopnjo poplačila.

Napravili smo tudi dva testa moči modela. V prvem smo naključno izbrali 90% opazovanih primerov, pri drugem pa smo izključili 10% posojil z največjo izpostavo z namenom izključitve možnosti, da bi bil vpliv velikosti izpostave ob neplačilu pogojen s strani visokih poplačil na nekaj velikih izpostavah. Rezultati so bili skladni s prvotnim modelom in prvotnimi specifikacijami.

10.8 Prvi poskus izračuna pričakovane izgube po industrijskih sektorjih

Z izračunanim LGD-jem po posameznih industrijskih sektorjih smo pridobili pomembno manjkajočo komponento pri računanju pričakovane izgube. Ob upoštevanju rezultatov modela verjetnosti nastanka neplačila (PD), ki ga v banki x že uporabljajo, za banko x sedaj lahko izračunamo tudi končno pričakovano izgubo po posameznem industrijskem sektorju. Povprečna pričakovana izguba je 1,4%, kar nam pove dejansko povprečno ceno tveganja za posojilo. Rezultati po posameznih gospodarskih dejavnostih pokažejo, da je pričakovana izguba največja pri panogah transport in gradbeništvo, večinoma na račun večjega LGD-ja (okoli 50%), najnižja pa v sektorju finančnega posredništva, in sicer na račun zelo nizkega LGD-ja.

10.9 Stroški izterjave poplačil

Za določanje cen posojil, izračun LGD-ja in kapitalskih zahtev, kot tudi za izračun oslabitev na naložbah, moramo pri poplačilih upoštevati tudi stroške, ki so nastali s procesom izterjave. V banki x se s tem področjem ukvarja sektor za razreševanje naložb. Podatki, prikazani v nalogi, ne vsebujejo stroškov plač zaposlenih v tem oddelku. V povprečju so stroški postopka izterjave v letih 2002-2005 predstavljali 0,44% poplačane vsote, zajemali pa so stroške zunanjih odvetnikov in sodne stroške. Stroški izterjave, izraženi v odstotkih od poplačila, so višji pri manjših posojilih kot pri večjih. Če bi upoštevali tudi plače zaposlenih, lahko vzamemo približek teh stroškov med 1-2%, ki jih lahko implicitno upoštevamo pri določanju diskontne stopnje.

10.10 Sklep

Banke so bile in bodo še naprej najpomembnejši vir financiranja malim in srednjim podjetjem. V okviru Basla II minimalni zahtevani kapital ne bo več v večini odvisen samo od velikosti posojila, ampak tudi od tveganosti posojila. Bolj tvegana posojila bodo tako postala tudi večji potrošniki kapitala. Uvedene spremembe s strani Basla II odsevajo trend k vse bolj kvantitativnemu in diferenciranemu upravljanju tveganj. Na podlagi bolj natančnih pregledov in analiz, kar je prikazala ta diplomska naloga za banko x, bodo banke sprejemale odločitve o posojilih in nudile tudi širšo cenovno lestvico kreditov in kreditnih pogojev.

Izgube ob neplačilih so bile ocenjene na vzorcu 124 posojil malih in srednjih podjetij pri banki x za obdobje 2001-2005. Ocene so bazirale na diskontiranih vrednostih dejanskih poplačil po nastopu neplačila. Univariatna analiza je bila uporabljena za merjenje kumulativnih stopenj poplačil na »slabih« posojilih. Ocena povprečne stopnje poplačila znaša 73% in je podobna tisti na primeru portugalske banke (Dermine, Carvalho, 2005). S pomočjo multivariatne analize smo preverjali pojasnjevalne spremenljivke poplačil. Na podlagi empirične študije primera lahko povzamemo tri pomembnejše ugotovitve:

- frekvenčna porazdelitev izgube bimodalna s koncentracijo visokih poplačil (med 90 in 100%) in nizkih poplačil (med 0 in 10 %);
- multivariatna analiza identificira nekaj statistično značilnih pojasnjevalnih spremenljivk (zavarovanje, industrijska panoga, dospelost posojila, bonitetna ocena posojila);
- ocena stroškov povezanih s prenosom terjatve v izterjavo 1-2% od vrednosti izpostave.

Naloga prikaže nekaj pomembnih informacij in rešitev, ki dokazujejo, da bi tudi slovenske banke lahko bolj natančno ocenjevale tveganja na posameznem portfelju. Študija pripomore k boljšem razumevanju determinant LGD-ja pri bančnih posojilih in z nekaj napotki o skladiščenju podatkov poplačil odpre vrata k razvoju mere ocenjevanja oslabitev posojil. S pomočjo dodajanja novih podatkov in širjenja vzorca bi model v prihodnosti lahko tudi potrdili in ga uporabljali, kar bi predstavljajo pomembno primerjalno prednost v slovenskem finančnem prostoru.

SLOVARČEK

Angleški izraz	**Slovenski prevod**
Capital Adequacy Directive (CAD)	direktiva o kapitalski ustreznosti
Credit pricing models	modeli kreditnih ocen
Credit risk mitigation	blaženje kreditnih tveganj
Default	neplačilo
Default mode models	modeli načinov nastopa neplačila
Exposure at default (EAD)	izpostava ob neplačilu
Generalized linear models (GLM)	posplošeni linearni modeli
Internal-Rating-Based (IRB)	pristop notranjih bonitet
Loss given default (LGD)	izguba ob neplačilu
Mark-to-market (MTM) models	modeli trenutne tržne vrednosti
Probability density function	funkcija verjetnostne porazdelitve
Probability of default (PD)	verjetnost neplačila
Provisions	popravki vrednosti in rezervacij
Recovery rate (RR)	stopnja poplačila
Reduced form models	pol-strukturni modeli
Structural form models	strukturni modeli
Value-at-risk (VAR) models	modeli tvegane vrednosti
Workout costs	stroški pri prenosu terjatve v izterjavo

BIBLIOGRAPHY

1. Agresti A.: Categorical Data Analysis. Hoboken (New Jersey) : John Wiley & Sons, Inc., 2002. 742 p.

2. Altman E. I., Resti A., Sironi A.: Recovery Risk: The next Challenge in Credit Risk Management. London : Risk Books, 2005. 364 p.

3. Altman E. I.: Altman High Yield Bond and Default Study. US Fixed Income High Yield Report. W.p.: Salomon Smith Barney, 2001.

4. Altman E. I. et al.: Analyzing and Explaining Default Recovery Rates. Research Report. London : ISDA, 2001. 97 p.

5. Araten M., Jacobs M., Varshney P.: Measuring LGD on Commercial Loans: An 18-Year Internal Study. The RMA Journal, Philadelphia, May, 2004, pp. 28-35.

6. Asarnow E., Edwards D.: Measuring Loss on Defaulted Bank Loans: a 24-Year Study. The Journal of Commercial Lending, W.p., 77 (1995), 7, pp. 11-23.

7. Black F., Scholes M.: The Pricing of Options and Corporate Liabilities. Journal of Political Economics, Chicago, May, 1973, pp. 637-659.

8. Carty L., Lieberman D.: Defaulted Bank Loan Recoveries. Moody's Investors Service, November, 1996.

9. Crouhy M., Galai D., Mark R.: A Comparative Analysis of Current Credit Risk Models. Journal of Banking & Finance, W.p., 24 (2000), pp. 59-117.

10. Davydenko S., Franks J.: Do bankruptcy codes matter? A study of defaults in France, Germany and the UK. London : London Business School, 2005. 47 p.

11. Dermine J., Carvalho C.: Bank Loan Losses-Given Default. W.p.: A Case Study, 2005. 40 p.

12. Dobson A. J.: An Introduction to Generalized Linear Models. Boca Raton : Chapman & Hall/CRC, 2002. 242 p.

13. Duffie D., Singleton K. J.: Modelling the Term Structures of Defaultable Bonds. Review of Financial Studies, Oxford, 12 (1999), 4, pp. 687-720.

14. Emery K.: Moody's Loan Default Database as of November 2003. Moody's Investors Service, December, 2003.

15. Eom Y. H., Helwege J., Huang J.: Structural Models of Corporate Bond Pricing: An Empirical Analysis. Mimeo, 2001. 50 p.

16. Franks J., Servigny A., Davydenko S.: A comparative analysis of the recovery process and recovery rates for private companies in the U.K., France, and Germany. Standard & Poor's Risk Solutions Report. New York : Standard & Poor's Risk Solutions, 2004. 93 p.

17. Frye J.: Depressing Recoveries. Risk, London, November, 2000, pp. 108-113.

18. Gupton G., Finger C., Bhatia M.: CreditMetrics Technical Document. New York : JP Morgan & Co., 1997. 212 p.

19. Hosmer D.W., Lemeshow S.: Applied Logistic Regression. New York : John Wiley & Sons, Inc., 1989. 376 p.

20. Hurt L., Felsovalyi A.: Measuring Loss on Latin American Defaulted Bank Loans, a 27-Year Study of 27 Countries. The Journal of Lending and Credit Risk Management, W.p., 80 (1998), pp. 41-46.

21. Longstaff F. A., Schwartz E. S.: A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. Journal of Finance, Oxford, 50 (1995), 3, pp. 789-819.

22. Maclachlan I.: Choosing the Discount Factor for Estimating Economic LGD. Melbourne : Australian and New Zealand Banking Group Ltd, 2004. 24 p.

23. McCullagh P., Nelder J.A.: Generalized Linear Models. Monographs on Statistics and Applied Probability 37, Boca Raton : Chapman & Hall/CRC, 1989. 540 p.

24. Merton R. C.: On the Pricing of Corporate Debt: The Risk Structure of Recovery Rates of Interest Rates. Journal of Finance, Oxford, 29 (1974), 2, pp. 449-471.

25. Resti A., Sironi A.: Defining LGD: The Basel II Perspective. Milan : Bocconi University, 2005. 15 p.

26. Schleifer A., Vishny R.: Liquidation Values and Debt Capacity: a Market Equilibrium Approach. Journal of Finance, Oxford, 47 (1992), 4, pp. 1343-1366.

27. Schuermann T.: What Do We Know About Loss Given Default?. New York : Federal Reserve Bank of New York, 2004. 32 p.

28. Verde M.: Recovery Rates Return to Historic Norms. New York : FITCH Ratings, September, 2003.

29. Wilson T. C.: Portfolio Credit Risk. Economic Policy Review, New York, Federal Reserve Bank of New York, October, 1998, pp. 71-82.

SOURCES

1. Basel Committee on Banking Supervision: International Convergence of Capital Measurement and Capital Standards. Basel : Bank for International Settlement, 2004. 157 p.

2. Basel Committee on Banking Supervision: Studies on the Validation of Internal Rating Systems. Working Paper No. 14. Basel : Bank for International Settlement, 2005. 112 p.

3. Engineering Statistic Handbook. [URL: http://www.itl.nist.gov/div898/handbook/index.htm], 17.6.2006.

4. European Commission: How to deal with the new rating culture. A practical guide to loan financing for small and medium-sized enterprises. Brussels : European Commission, 2005. 52 p.

5. Generalized Linear Models. [URL: http://www.statsoft.com/textbook/stglz.html], 2006

6. Internal documents SKB, 2001-2006.

7. Internal documents Société Générale, 2001-2006.

8. SPSS Advanced Models 10.0. Chicago: SPSS Inc., 1999. 325 p.

9. Strengthening the Czech Banking Sector - Application of Basel II. Phase 5 Report: Generic Rating/Scoring Methodology. W.p., PricewaterhouseCoopers, 2005. 137 p.

Annexes

ANNEX 1: Empirical Results of Portuguese Model (Dermine, Carvalho, 2005)

Table 22: Univariate Statistics on Recovery Rates

Sample Unweighted Cumulative Recovery Rates, a Pool-based Approach					
	Mean	Median	Minimum	Maximum	Standard Deviation
12-month cumulative recovery	52%	49%	0%	100%	44%
24-month cumulative recovery	65%	91%	0%	100%	41%
36-month cumulative recovery	69%	.95%	0%	100%	38%
48-month cumulative recovery (total sample)	71%	95%	0%	100%	37%[21]
48-month cumulative recovery (loans with no guarantee/collateral)	76%	92%	0%	100%	33%
48-month cumulative recovery (loans with personal guarantee only)	64%	93%	0%	100%	42%
48-month cumulative recovery (loans with collateral)	92%	98%	55%	100%	12%

Note: The pool-based approach includes for each horizon (12 months, 24 months, 36 months, 48 months) the subset of loans with recovery data available for that horizon.

Source: Dermine, Carvalho, 2005

Table 23: Log.log estimates of Cumulative Recoveries

Explanatory variable	12-month cumulative recovery (p-value)	24-month cumulative recovery (p-value)	36-month cumulative recovery (p-value)	48-month cumulative recovery (p-value)
Constant	1.65 (0.03*)	1.62 (0.02*)	3.24 (0.00*)	4.57 (0.00*)
Loan Size	- 0.58 (0.01*)	-0.86 (0.00*)	-1.18 (0.00*)	-1.26 (0.00)
Personal Guarantee	- 0.16 (0.31)	-0.21 (0.29)	-0.05 (0.83)	-0.35 (0.23)
Real Estate Collateral	0.28 (0.4)	0.29 (0.59)	0.67 (0.28)	1.98 (0.00)
Physical Collateral	- 0.36 (0.5)	0.58 (0.60)	-0.02 (0.99)	2.93 (0.00)
Financial Collateral	0.43 (0.30)	0.38 (0.42)	0.39 (0.5)	2.09 (0.02)
Year 1996	0.34 (0.10)	0.44 (0.06)	0.41 (0.1)	0.34 (0.20)
Year 1997	0.69 (0.01*)	0.54 (0.06)	0.52 (0.08)	
Year 1998	0.09 (0.69)	0.11 (0.68)		
Year 1999	- 0.47 (0.04*)			
1.Agriculture/Fishing	- 0.68 (0.43)	0.16 (0.84)	- 0.78 (0.49)	-0.56 (0.64)
2.Mining	- 2.19 (0.01*)	- 1.49 (0.06)	-1.49 (0.19)	-2.09 (0.02*)
3.Construction	-1.11 (0.16)	-0.82 (0.24)	-2.15 (0.04*)	-2.58 (0.00*)
5.Real Estate	-0.59 (0.53)	0.32 (0.76)	-1.70 (0.17)	-3.21 (0.00*)
6.Food/Bbeverages	-1.13 (0.21)	0.14 (0.89)	-0.35 (0.78)	-1.73 (0.08)
7.Textiles	-1.31 (0.11)	-1.14 (0.12)	- 2.53 (0.02*)	-3.57 (0.00*)
8.Chemicals	- 1.72 (0.04*)	-0.41 (0.71)	-1.65 (0.21)	-2.38 (0.04*)
9.Machinery	-1.09 (0.18)	-0.89 (0.23)	-2.62 (0.02)	-4.05 (0.00*)
10.Paper/Printing	-1.10 (0.16)	-1.51 (0.08)	-1.74 (0.28)	-4.32 (0.00*)
11.Other Non-mineral	-1.10 (0.18)	-0.84 (0.26)	-2.23 (0.05*)	-3.46 (0.00*)
12.Wholesale Trade	-1.21 (0.11)	-0.91 (0.16)	-2.60 (0.01*)	-3.65 (0.00*)
13.Retail Trade	-1.48 (0.05*)	-1.08 (0.12)	-2.83 (0.006*)	-4.00 (0.00*)
14.Transport	-1.81 (0.02*)	-1.51 (0.03*)	-2.69 (0.01*)	-3.62 (0.00*)
15.Other Services	-1.27 (0.10)	- 0.41 (0.57)	-1.42 (0.21)	-2.47 (0.01*)
Wald (Qui-squared) Test (p-value)	55.98 (0.00*)	34.7 (0.042*)	46.44 (0.00*)	390.8 (0.00*)
Reset Test (p-value)	0.58 (0.56)	-0.21 (0.83)	-1.03 (0.30	-2.58 (0.01)
Pseudo R^2	0.13	0.10	0.13	0.20
Number of Observations	317	270	213	154

* Represents significance at the 5% level.

Note: The table presents the estimation of the log-log regression for the cumulative recovery rates at four horizons, respectiverly 12-, 24-, 36- and 48-months. Cumulative recoveries are measured in cents per euro. The loan size is 1 million EUR. Collateral, year, and industry sectors are represented by dummies.

Source: Dermine, Carvalho, 2005

Table 24: Log-log estimates of Cumulative Recoveries

Explanatory variable	12-month cumulative recovery (p-value)	24-month cumulative recovery (p-value)	36-month cumulative recovery (p-value)	48-month cumulative recovery (p-value)
Constant	0.49 (0.04*)	0.72 (0.01*)	1.02 (0.00*)	1.79 (0.00*)
Loan Size	-0.66 (0.00*)	-0.76 (0.00*)	-0.84 (0.00*)	-0.77 (0.00*)
Personal Guarantee	-0.17 (0.26)	-0.27 (0.16)	-0.14 (0.53)	-0.38 (0.15)
Collateral	0.31 (0.18)	0.61 (0.07)	0.69 (0.08)	1.72 (0.00*)
Year 1996	0.23 (0.26)	0.34 (0.13)	0.35 (0.16)	0.31 (0.24)
Year 1997	0.63 (0.01*)	0.49 (0.07)	0.41 (0.15)	
Year 1998	0.10 (0.67)	0.12 (0.63)		
Year 1999	-0.48 (0.04*)			
II.Manufacturing Sector	-0.30 (0.21)	-0.36 (0.19)	-0.33 (0.35)	-1.20 (0.02*)
III.Trade Sector	-0.25 (0.25)	-0.25 (0.33)	-0.66 (0.04*)	-1.25 (0.01*)
IV.Services Sector	-0.29 (0.26)	-0.00 (0.99)	0.09 (0.83)	-0.42 (0.48)
Age of firm	0.01 (0.03*)	0.01 (0.02*)	0.01 (0.04*)	0.01 (0.04*)
Wald (Qui-squared) test (p-value)	37.6 (0.00*)	31.31 (0.00*)	38.48 (0.00*)	43.43 (0.00*)
Reset Test (p-value)	1.59 (0.11)	0.25 (0.80)	0.87 (0.39)	0.23 (0.82)
Pseudo R^2	0.11	0.08	0.11	0.18
Number of Observations	316	269	212	153

* Represents significance at the 5% level.

Note: The table presents the estimation of the log-log regression for the cumulative recovery rates at four horizons, respectively 12, 24-, 36- and 48-months. Cumulative recoveries are measured in cents per euro. The loan size is 1 million EUR. Collateral, year, and industry sectors are represented by dummies. The age of the firm is in number of months.

Source: Dermine, Carvalho, 2005

Table 25: Workout costs incurred in Recovery (2002)

Internal Recovery Department			
	Standardized Unit	Specialized Unit	Total
Total internal costs	296	727	1023
Amount Recovered/Restructured During the Year	7252	78 .000	85.252
Internal Recovery Cost per Euro Recovered (%)	4.1%	0.9%	1.2%

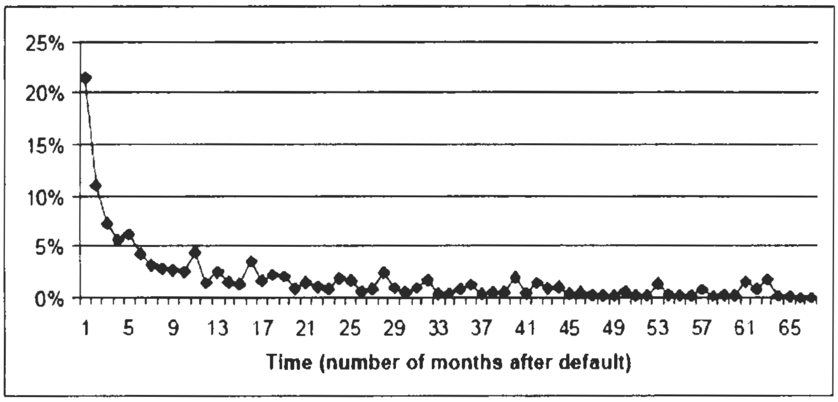
Contentious Department	
Internal Contentious Cost	278
External Lawyers and Court Expenses	1257
Total Internal and External Cost	1535
Cash Flows Recovered	14748
Contentious Recovery Cost per euro (%)	10.4%

Total Direct Cost (Internal and Contentious)	
Total Internal and External Cost	2.558
Total Amount Recovered	100.000
Average Recovery Cost per Euro (%)	2.6%

Note: This table reports the workout direct cost incurred in recovery by Banco Comercial Portugues in 2002. For reasons of confidentiality, the absolute figures have been scaled by a common factor. Only percentage figures are relevant. The standardized unit deals with loans with a value below 75,000 EUR, and the specialized unit deals with larger loans. The contentious department refers the cases to external lawyers or law courts.

Source: Dermine, Carvalho, 2005

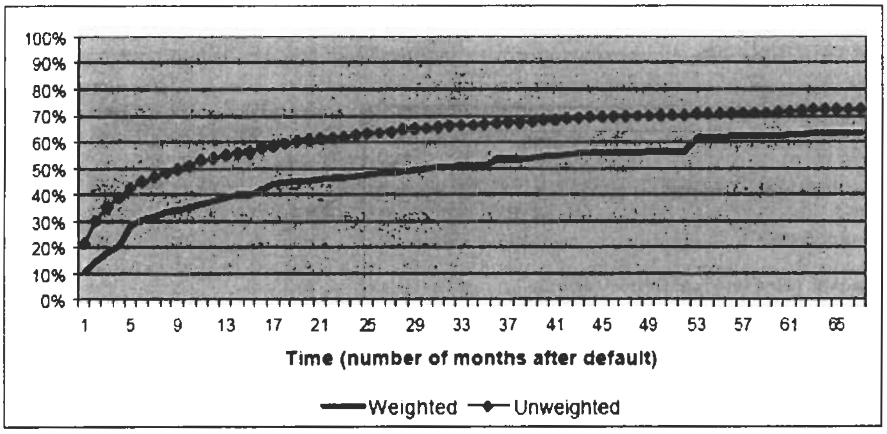
Figure 6: Sample unweighted Marginal Recovery Rate at time t+n ($SMRR_{t+n}$)



Note: This figure presents the marginal recovery n-months after default. The mortality-based approach is used to calculate the marginal recoveries.

Source: Dermine, Carvalho, 2005

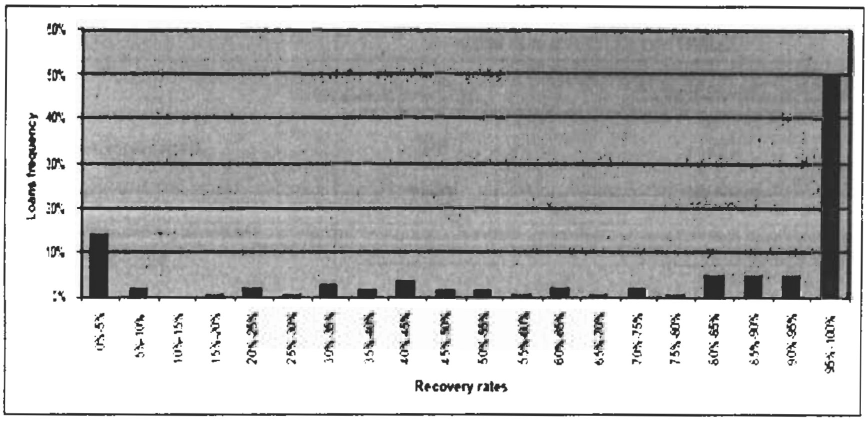
Figure 7: Sample Unweighted and Weighted Cumulative Recovery Rate at time t+n (SCRRt+n)



Note: The figure presents the cumulative weighted and unweighted recovery rates n-months after default. They have been calculated with the mortality-based approach.

Source: Dermine, Carvalho, 2005

Figure 8: Distribution of Cumulative Recovery Rates 48 Months after Default



Note: The figure presents the frequency of cumulative recovery rates on individual loans. Due to data limitation (five years), the cumulative recovery is calculated up to 48 months after default. This does not seem restrictive as Figure 6 indicates that the most of the recovery is achieved 48 months after default.

Source: Dermine, Carvalho, 2005

ANNEX 2: Alternative Discount Rate Methods

Table 26: Alternative discount rate methods

Source	Method
Edwards and Asarnow (1995)	Contractual loan rate including penalty
Eales and Bosworth (1998)	Lender's cost of equity
Friedman and Sandow (2003)	Coupon rate
Gupton and Stein (2002)	Market value one month post-default
Carty, Hamilton, Keenan, Moss, Mulvaney, Marshella, and Subhas (1998)	Contractual loan rate
Araten (2004)	15% flat rate justified by ex post realised returns on the Moody's Bankruptcy Bond Index (Hamilton and Berthault (2000))
OCC, the Board of Governors, FDIC, and OTS (2003)	The discount rate must be no less than the contract interest rate on new originations of a type similar to the transaction in question, for the lowest-quality grade in which a bank originates such transactions. Where possible, the rate should reflect the fixed rate on newly originated exposures with term corresponding to the average resolution period of defaulting assets (paragraph 134)
FSA (2003)	Firms should use the same rate as that used for an asset of similar risk. They should not use the risk free rate or the firms hurdle rate (unless the firm only invests in risky assets such as defaulted debt instruments) (page 68, Annex 3)
IAS 39 (2003)	Effective original loan rate (the rate that exactly discounts expected future cash payments or receipts through the expected life of the financial instrument)

Source: Maclachlan, 2004, p. 17

ANNEX 3: Average Interest Rates on Loans in bank x

Table 27: Average interest rates on loans in bank x (2001-2005)

Type of loan/Date	dec.01	dec.02	dec.03	dec.04	AVERAGE
Short-term credit (SIT)	0,13	0,12	0,07	0,06	**0,10**
Long-term credit (SIT)	0,14	0,12	0,07	0,05	**0,10**

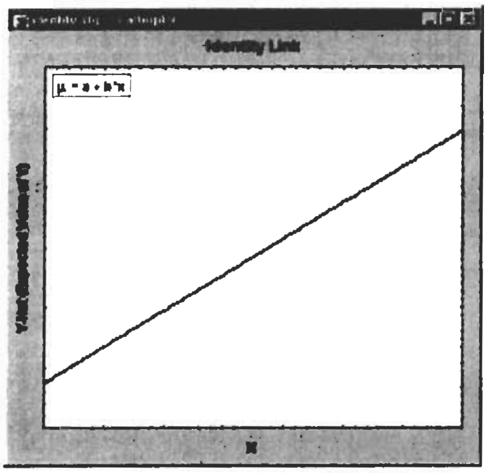
Source: own calculations.

ANNEX 4: Link Functions and Distributions

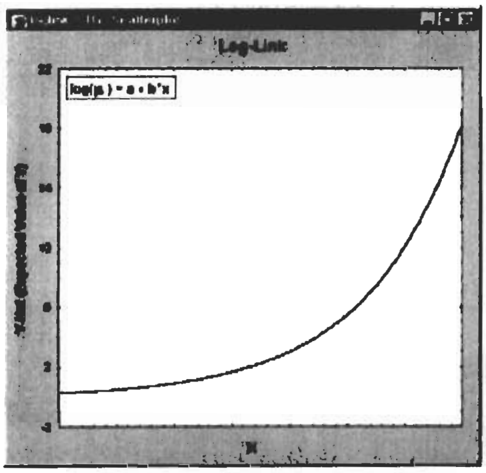
Various link functions can be chosen, depending on the assumed distribution of the *y* variable values (McCullagh, Nelder, 1989):

<u>*Normal, Gamma, Inverse normal, and Poisson distributions:*</u>

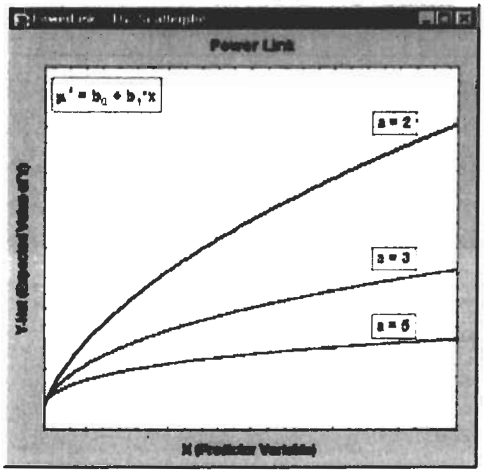
- Identity link: **f(z) = z**



- Log link: f(z) = log(z)



- Power link: **f(z) = z$^{\mathbf{a}}$**, for given a

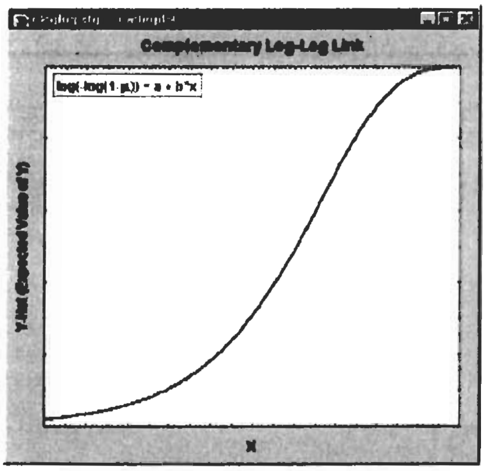


<u>*Binomial, and Ordinal Multinomial distributions:*</u>

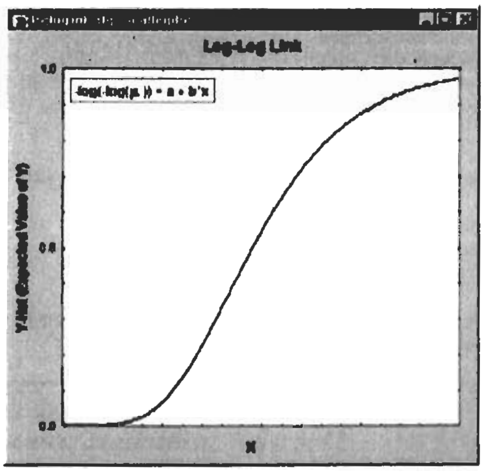
- Logit link: **f(z)=log(z/(1-z))**
- Probit link: **f(z)=invnorm(z)**

where *invnorm* is the inverse of the standard normal cumulative distribution function

- Complementary log-log link: **f(z)=log(-log(1-z))**



- Log-log link: **f(z)=-log(-log(z))**



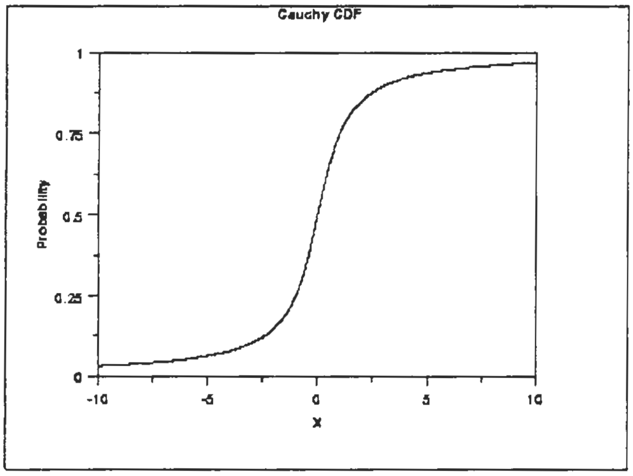
Multinomial distribution:

- Generalized logit link: **(f(z1|z2,…,zc)=log(x1/(1-z1-…-zc))**

where the model has **c+1** categories.

ANNEX 5: Cauchy Cumulative Distribution Function

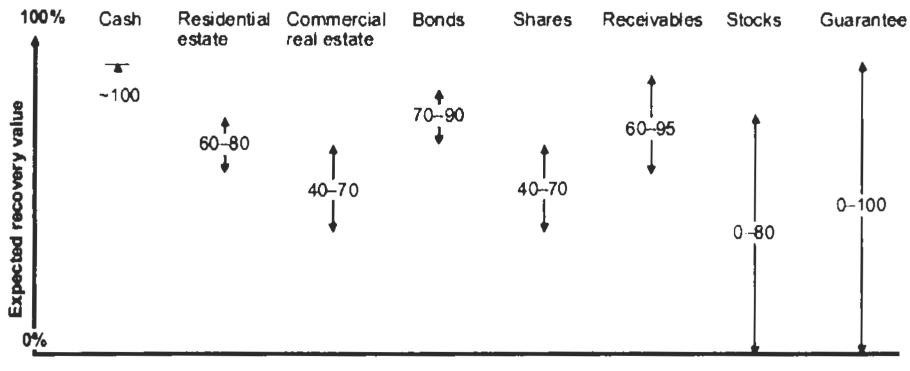
Figure 9: Cauchy cumulative distribution function



Source: Engineering Statistics Handbook, 2006.

ANNEX 6: Expected Recovery Value According to European Experts (Qualitative Estimate)

Figure 10: Expected recovery value by different collateral types



Source: European Commission, p. 24.

ANNEX 7: A Method to Analyse Recovery Rates and Loan-loss Provisioning (Provided by LGD Estimation)

The application of the method used in our analyses was done with some references to the mortality-based approach that was firstly introduced in the study on Portuguese bank loans LGDs by Dermine and Carvalho (2005). It examines the percentage of a bad and doubtful loan which is recovered n years after the default date. This methodology is appropriate because the population sample is changing over time.

To define the concepts of the method to analyse recovery rates, used to measure loan recovery rate, it is, for expository reasons, useful to refer to a simple example. Consider a loan of €100 that enters the 'default' category in December 2000. We track the subsequent payments on this loan, assuming, for expository convenience, that all payments take place at the end of the year. The interest/discount rate is 10%.

	dec.01	dec.02	dec.03	dec.04
Loan outstanding	100*			
Cash payment	0	50	26	14

**Note: Exposure at default*
Source: own calculations.

Let us define the *Marginal Recovery Rate* at December 2001, MRR_1, as the proportion of the outstanding loan in December 2001 that is being paid, one period (in the example, one year) after default:

MRR_1 = Cash flow paid $_1$ / Loan outstanding at the time of default

= 50/110 = (50 / 1.10) / 100 = 0.454

The marginal recovery rate can also be interpreted as the percentage repayment on the loan outstanding at the time of default, in present value terms.

Similarly, one can define the *Marginal Recovery Rate* at December 2002 as:

MRR_2 = Cash flow paid $_2$ / Loan outstanding at the time of default

= 26/121 = (25 / 1.21) / 100 =

= 0.214

and the *Cumulative Recovery Rate* in December 2002, CRR_2, is defined as:

$CRR_2 = (MRR_1 + MRR_2)$

= 0.669

Similarly to the calculation of CRR_2 we can calculate also CRR_3. The final result would be $CRR_3 = 77.5\%$

CRR_T, represents the proportion of the initial default loan that has been repaid (in present value terms), T periods after default.

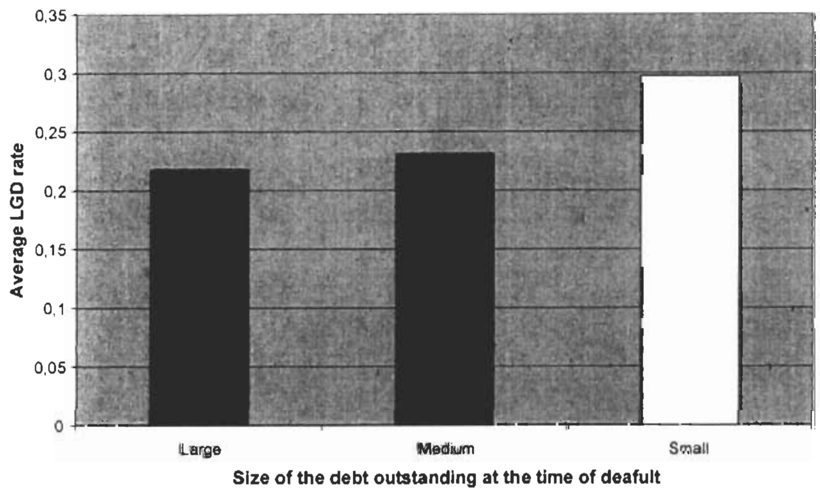
Finally, the Loan Loss Provision, LPP, which is provided by LGD calculation, on a loan balance outstanding at the default date, December 2000, is defined as

$LLP = LGD = 1 - CRR_3 = 1 - 0.775 = 22.5\%$

This figure, 22.5%, represents the percentage of the loan that will not be recovered (interest included) in the future. With perfect foresight, it would serve as a base for provisioning on the loan at the time of default.

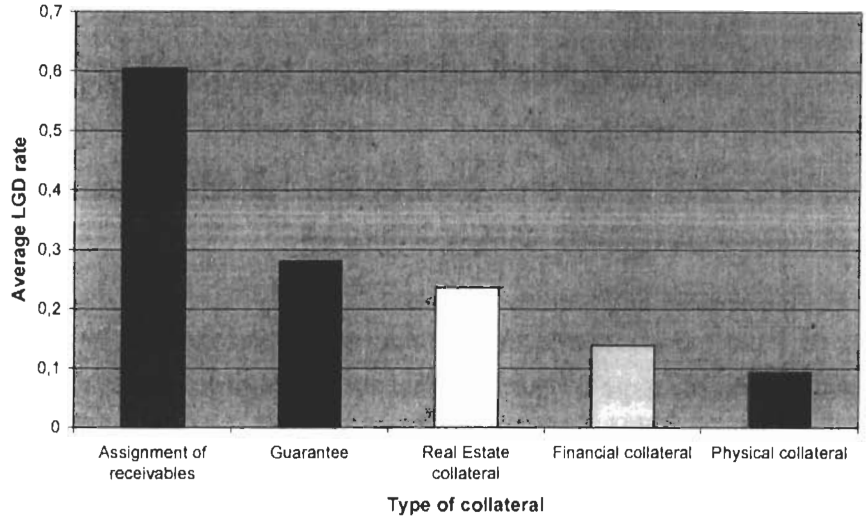
ANNEX 8: Empirical Results - Figures

Figure 11: Average LGD rate by the size of the debt outstanding at the time of default



Source: own calculations.

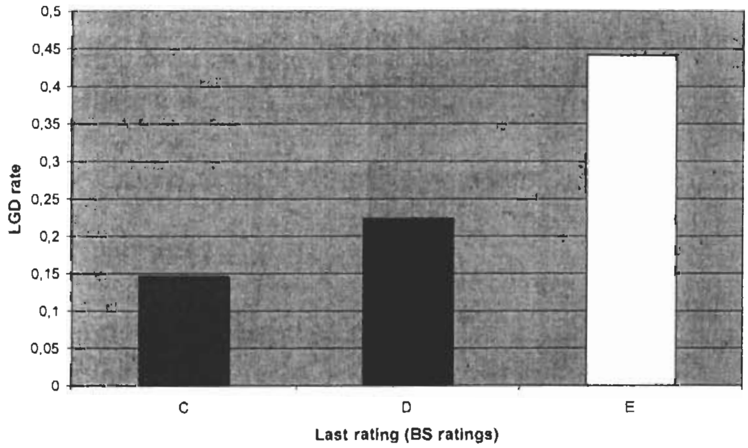
Figure 12: Average LGD rate by type of collateral



Note: Physical collateral noted to be unusually low

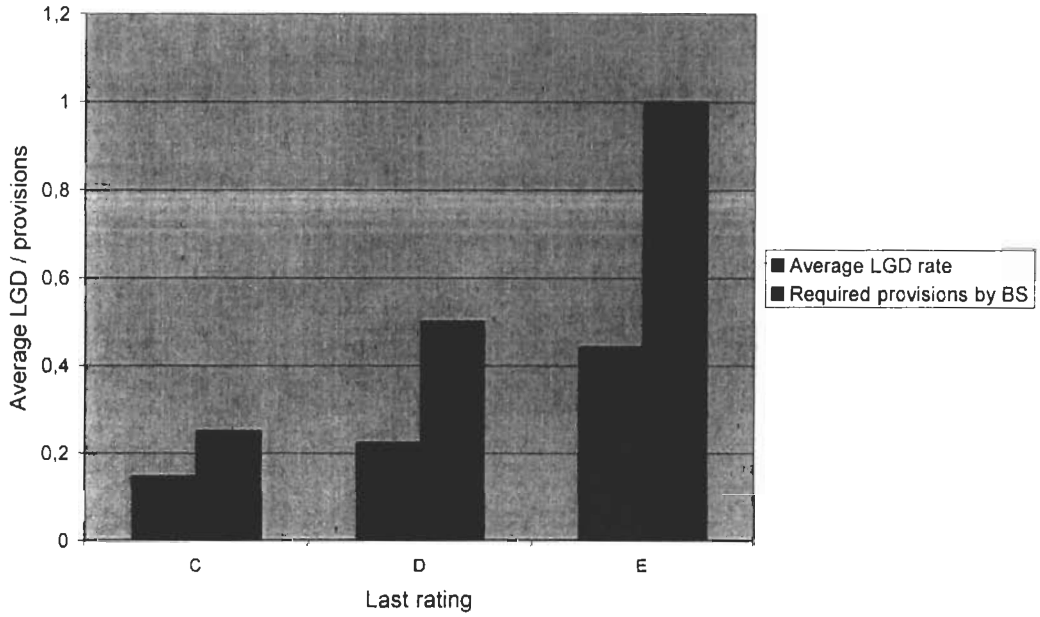
Source: own calculations.

Figure 13: Average LGD rate by last rating



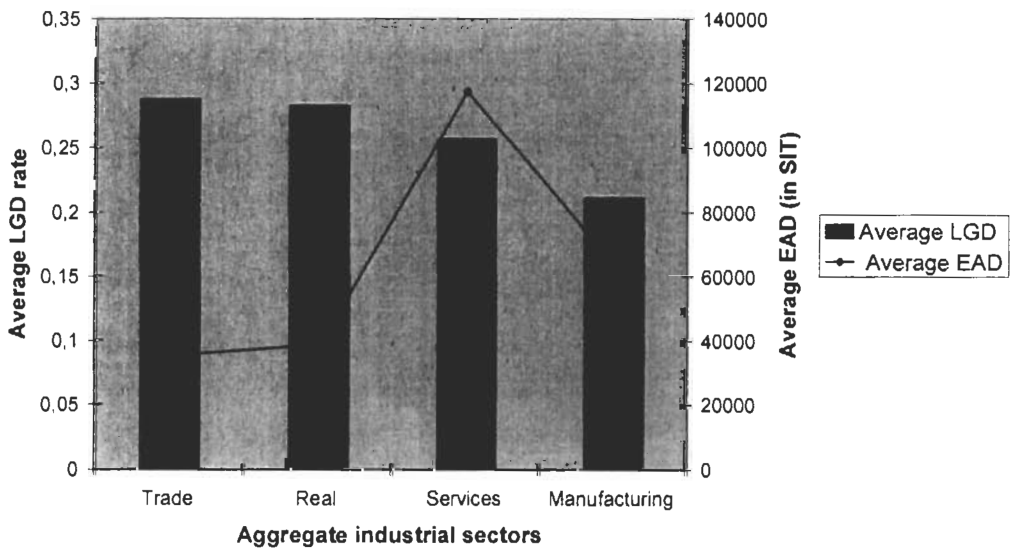
Source: own calculations.

Figure 14: Average LGD rates by last rating and loss provisions by rating required from Bank of Slovenia



Source: own calculations.

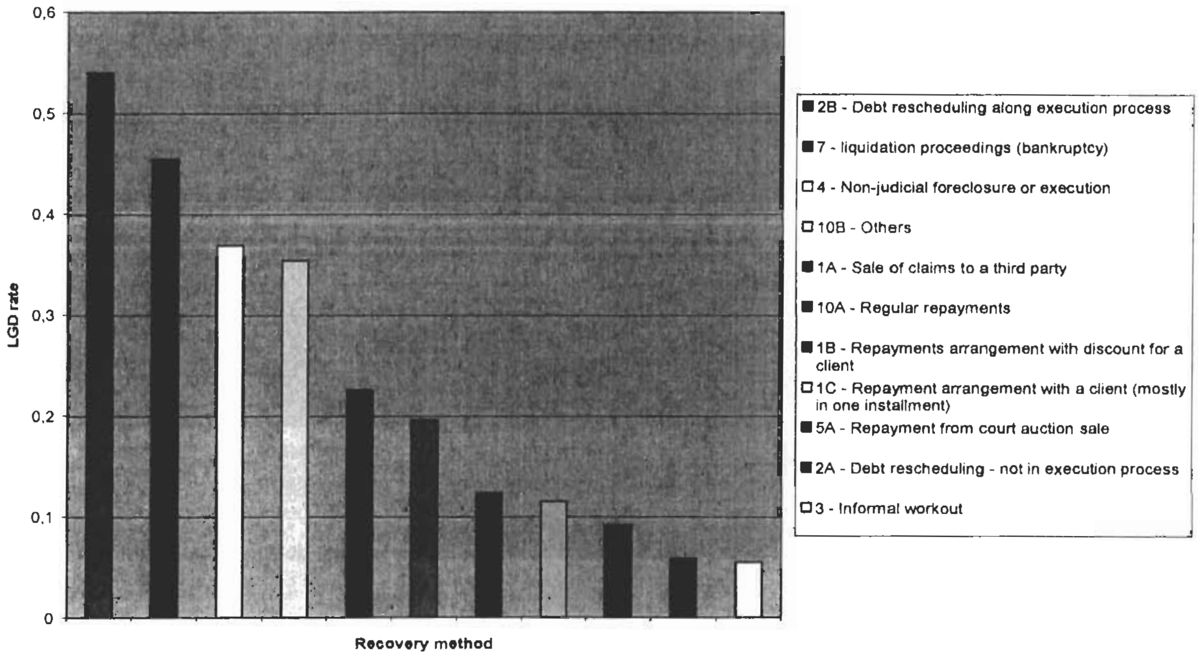
Figure 15: Average LGD rate by aggregated industrial sectors



Note: Aggregated sectors consists of the following sectors according to European Union economic activity codes (NACE): Real (sectors C, F, H, K); Manufacturing (sector D), Trade (sector G), Services (sectors E, I, J, M, O).

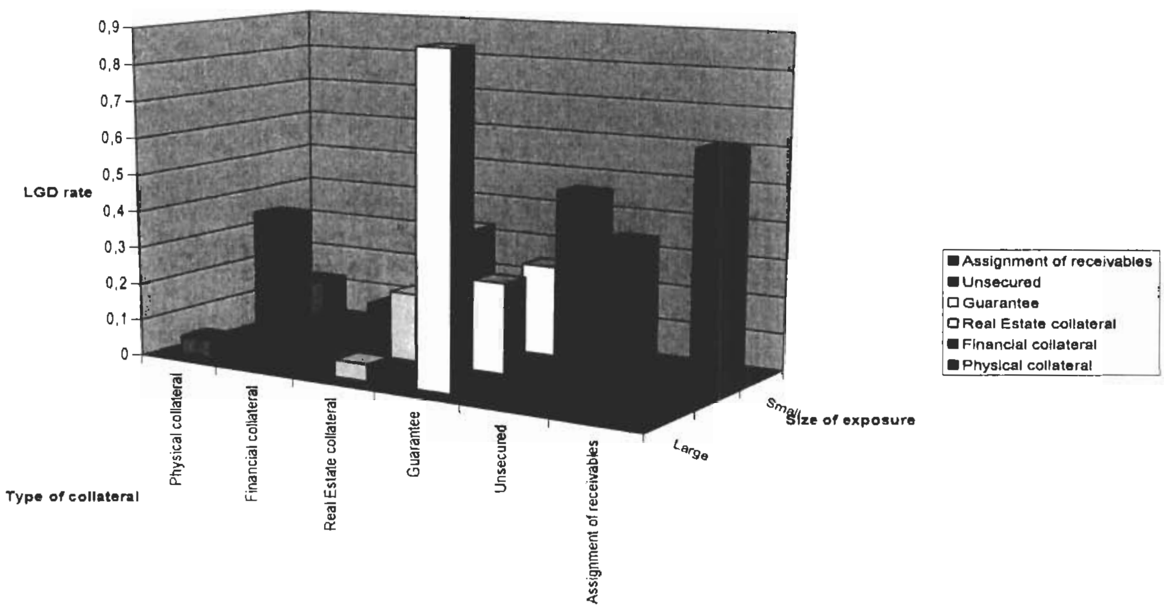
Source: own calculations.

Figure 16: Average LGD rate by type of recovery method



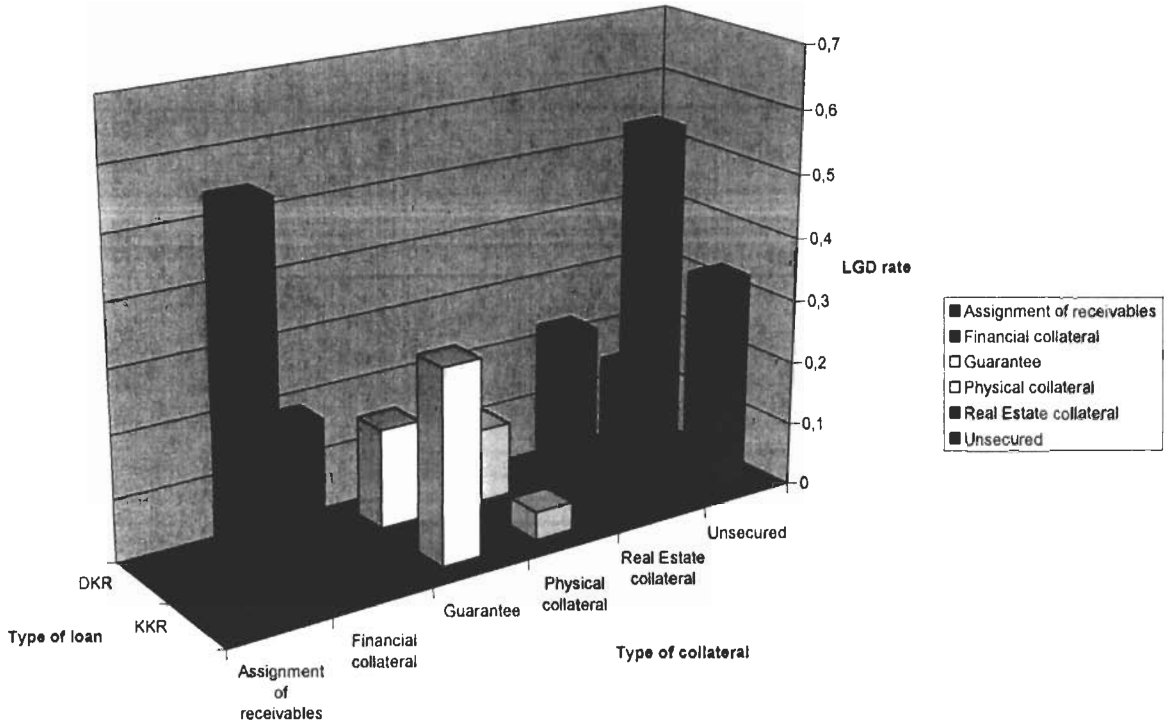
Note: Method 8a – Formal rehabilitation (reprogram) was excluded due to unrepresentative sample.
Source: own calculations.

Figure 17: Average LGD by type of loan



Source: own calculations.

Figure 18: Average LGD by type of loan



Source: own calculations.

ANNEX 9: Output of the Econometric Analysis

- **Evaluating the model**

Table 28: Warnings about empty cells

There are 285 (75,0%) cells (i.e., dependent variable levels by combinations of predictor variable values) with zero frequencies.

Source: own calculations.

Table 29: Model-fitting information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	218,095			
Final	170,696	47,399	13	,000

Link function: Cauchit.
Source: own calculations.

Table 30: Goodness-of-fit table

	Chi-Square	df	Sig.
Pearson	333,194	287	,031
Deviance	141,081	287	1,000

Link function: Cauchit.
Source: own calculations.

Table 31: Pseudo- R^2 measures

Cox and Snell	,318
Nagelkerke	,363
McFadden	,184

Link function: Cauchit.
Source: own calculations.

Table 32: Classification table for the original model

			Predicted Response Category		Total
			0-20	80-100	
RR class (in %)	0-20	Count	9	11	20
		% within RR class (in %)	45	55	100
	20-40	Count	3	1	4
		% within RR class (in %)	75	25	100
	40-60	Count	2	6	8
		% within RR class (in %)	25	75	100
	60-80	Count	0	9	9
		% within RR class (in %)	0	100	100
	80-100	Count	6	77	83
		% within RR class (in %)	7	93	100
Total		Count	20	104	124
		% within RR class (in %)	16	84	100

Source: own calculations.

Table 33: Test of parallel lines

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	170,696			
General	81,846(a)	88,850(b)	39	,000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a The log-likelihood value cannot be further increased after maximum number of step-halving.

b The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

c Link function: Cauchit.

Source: own calculations.

Table 34: Paramenter estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[RRclassin = 0-20]	1,180	,821	2,069	1	,150	-,428	2,789
	[RRclassin = 20-40]	1,608	,831	3,743	1	,053	-,021	3,237
	[RRclassin = 40-60]	2,334	,897	6,769	1	,009	,576	4,093
	[RRclassin = 60-80]	3,001	,979	9,390	1	,002	1,081	4,920
Location	[Formsofcollguarantee =Assignment of receivables]	-,854	1,180	,524	1	,469	-3,166	1,459
	[Formsofcollguarantee =Financial collateral]	2,820	1,729	2,659	1	,103	-,570	6,210
	[Formsofcollguarantee =Personal guarantee]	,729	,643	1,288	1	,256	-,530	1,989
	[Formsofcollguarantee =Physical collateral]	4,621	1,658	7,770	1	,005	1,372	7,871
	[Formsofcollguarantee =Real Estate collateral]	3,856	1,213	10,109	1	,001	1,479	6,233
	[Formsofcollguarantee =Unsecured]	0(a)	.	.	0	.	.	.
	[Aggregateindustrialse ctors=Manufacturing]	1,475	,699	4,458	1	,035	,106	2,845
	[Aggregateindustrialse ctors=Real]	1,806	,752	5,777	1	,016	,333	3,279
	[Aggregateindustrialse ctors=Service]	1,188	,777	2,339	1	,126	-,334	2,710
	[Aggregateindustrialse ctors=Trade]	0(a)	.	.	0	.	.	.
	[Typeofloan=DKR]	-2,637	,920	8,225	1	,004	-4,440	-,835
	[Typeofloan=KKR]	0(a)	.	.	0	.	.	.
	[ZADNJIRATING=C]	3,856	,983	15,379	1	,000	1,929	5,784
	[ZADNJIRATING=D]	2,190	,886	6,106	1	,013	,453	3,928
	[ZADNJIRATING=E]	0(a)	.	.	0	.	.	.
	[SizeofEAD=Large]	,605	,844	,514	1	,473	-1,049	2,259
	[SizeofEAD=Medium]	,929	,555	2,801	1	,094	-,159	2,017
	[SizeofEAD=Small]	0(a)	.	.	0	.	.	.

Link function: Cauchit.

a This parameter is set to zero because it is redundant.

Source: own calculations.

- **Revising the model**

Table 35: Model-fitting information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	218,095			
Final	172,690	45,405	13	,000

Link function: Complementary Log-log.
Source: own calculations.

Table 36: Pseudo- R^2 measures

Cox and Snell	,307
Nagelkerke	,350
McFadden	,176

Link function: Complementary Log-log.
Source: own calculations.

Table 37: Classification table

			Predicted Response Category		Total
			0-20	80-100	
RR class (in %)	0-20	Count	8	12	20
		% within RR class (in %)	40	60	100
	20-40	Count	2	2	4
		% within RR class (in %)	50	50	100
	40-60	Count	3	5	8
		% within RR class (in %)	38	62	100
	60-80	Count	3	6	9
		% within RR class (in %)	33	67	100
	80-100	Count	3	80	83
		% within RR class (in %)	4	96	100
Total		Count	19	105	124
		% within RR class (in %)	15	85	100

Source: own calculations.

- **Robustness test (random selection of 90% of the observations)**

Table 38: Parameter estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[RRclassin = 0-20]	1,094	,882	1,538	1	,215	-,635	2,822
	[RRclassin = 20-40]	1,645	,896	3,373	1	,066	-,110	3,401
	[RRclassin = 40-60]	2,519	,995	6,406	1	,011	,568	4,470
	[RRclassin = 60-80]	3,204	1,088	8,668	1	,003	1,071	5,336
Location	[Formsofcollguarantee =Assignment of receivables]	-1,036	1,422	,531	1	,466	-3,823	1,751
	[Formsofcollguarantee =Financial collateral]	4,326	1,937	4,987	1	,026	,529	8,123
	[Formsofcollguarantee =Guarantee]	1,031	,693	2,216	1	,137	-,326	2,389
	[Formsofcollguarantee =Physical collateral]	5,877	1,989	8,732	1	,003	1,979	9,775
	[Formsofcollguarantee =Real Estate collateral]	4,982	1,595	9,760	1	,002	1,857	8,108
	[Formsofcollguarantee =Unsecured]	0(a)	.	.	0	.	.	.
	[Aggregateindustrialsectors=Manufacturing]	1,260	,712	3,136	1	,077	-,135	2,655
	[Aggregateindustrialsectors=Real]	2,047	,836	5,996	1	,014	,408	3,685
	[Aggregateindustrialsectors=Services]	1,645	,921	3,189	1	,074	-,160	3,451
	[Aggregateindustrialsectors=Trade]	0(a)	.	.	0	.	.	.
	[Typeofloan=DKR]	-3,671	1,321	7,718	1	,005	-6,261	-1,081
	[Typeofloan=KKR]	0(a)	.	.	0	.	.	.
	[ZADNJIRATING=C]	4,551	1,317	11,930	1	,001	1,968	7,133
	[ZADNJIRATING=D]	2,066	,917	5,077	1	,024	,269	3,863
	[ZADNJIRATING=E]	0(a)	.	.	0	.	.	.
	[SizeofEAD=Large]	,455	,828	,301	1	,583	-1,169	2,078
	[SizeofEAD=Medium]	1,077	,611	3,103	1	,078	-,121	2,274
	[SizeofEAD=Small]	0(a)	.	.	0	.	.	.

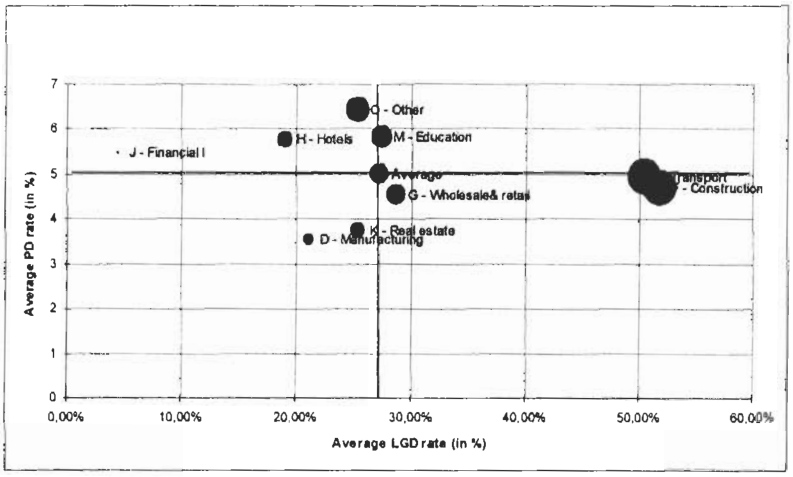
Link function: Cauchit.

a This parameter is set to zero because it is redundant.

Source: own calculations.

ANNEX 10: Calculation of Expected Loss Estimation by Industrial Sector

Figure 19: Expected loss by industrial activities



*Note: * Sectors C and E were excluded from the chart due to unrepresentative sample*
Source: own presentation.

ANNEX 11: Workout Costs Incurred in Recovery

Table 39: Workout costs incurred in recovery (2002 – 2005)

	Cost of proceedings				
CORPORATES AND SP	FEES	COST OF PROCEEDINGS AND EXPERTISE	TOTAL	FUNDS RECOVERED	RATIO COST/RECOVERY (in %)
Total 2002	231.517.533,00	93.587.712,00	244.800.474,00	42.590.322.000,00	0,57
Total 2003	172.597.233,00	- 10.326.481,14	162.270.751,86	41.073.390.000,00	0,40
Total 2004	98.760.648,00	13.609.842,00	112.370.490,00	33.048.453.000,00	0,34
Total 2005	88.297.155,75	5.066.180,67	93.363.336,42	19.490.703.000,00	0,48
AVERAGE	167.625.138,00	32.290.357,62	173.147.238,62	38.904.055.000,00	0,44

Source: internal calculations (bank x).

388