

Chalupka, Radovan; Kopecsni, Juraj

Working Paper

Modelling bank loan LGD of corporate and SME segments: A case study

IES Working Paper, No. 27/2008

Provided in Cooperation with:

Charles University, Institute of Economic Studies (IES)

Suggested Citation: Chalupka, Radovan; Kopecsni, Juraj (2008) : Modelling bank loan LGD of corporate and SME segments: A case study, IES Working Paper, No. 27/2008, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:

<https://hdl.handle.net/10419/83329>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

Modelling Bank Loan LGD of Corporate and SME Segments A Case Study

**Radovan Chalupka
Juraj Kopecsni**

IES Working Paper: 27/2008



Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
<http://ies.fsv.cuni.cz>

Disclaimer: The IES Working Papers is an online paper series for works by the faculty and students of the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. The papers are peer reviewed, but they are *not* edited or formatted by the editors. The views expressed in documents served by this site do not reflect the views of the IES or any other Charles University Department. They are the sole property of the respective authors. Additional info at: ies@fsv.cuni.cz

Copyright Notice: Although all documents published by the IES are provided without charge, they are licensed for personal, academic or educational use. All rights are reserved by the authors.

Citations: All references to documents served by this site must be appropriately cited.

Bibliographic information:

Chalupka, R., Kopecsni, J. (2008). “ Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study ” IES Working Paper 27/2008. IES FSV. Charles University.

This paper can be downloaded at: <http://ies.fsv.cuni.cz>

Modelling Bank Loan LGD of Corporate and SME Segments A Case Study

Radovan Chalupka*

Juraj Kopecsni#

*IES, Charles University Prague,
E-mail: chalupka@fsv.cuni.cz

IES, Charles University Prague,
E-mail: kopecsni@centrum.cz

November 2008

Abstract:

The aim of this paper is to propose a methodology to estimate loss given default (LGD) and apply it to a set of micro-data of loans to SME and corporations of an anonymous commercial bank from Central Europe. LGD estimates are important inputs in the pricing of credit risk and the measurement of bank profitability and solvency. Basel II Advance IRB Approach requires internally estimates of LGD to calculate risk-weighted assets and to estimate expected loss.

We analyse the recovery rate dynamically over time and identify the efficient recovery period of a workout department. Moreover, we focus on the appropriate choice of a discount factor by introducing risk premium based on a risk level of collaterals. We apply statistical methods to estimate LGD and test empirically its determinants.

Particularly, we analyse generalised linear models using symmetric logit and asymmetric log-log link functions for ordinal responses as well as for fractional responses. For fractional responses we employ two alternatives, a beta inflated distribution and a quasi-maximum likelihood estimator.

We find out that the main drivers of LGD are a relative value of collateral, a loan size as well as a year of the loan origination. Different models provided similar results. As for the different links in more complex models, log-log models in some cases perform better, implying an asymmetric response of the dependent variable.

Keywords: credit risk, bank loan, loss given default, LGD, recovery rate, fractional responses, ordinal regression, quasi-maximum likelihood estimator

JEL: G21, G28.

Acknowledgements

This research was supported by the Grant Agency of Charles University grant no. 131707/2007 A-EK and by the Grant Agency of the Czech Republic grant no. 402/05/H510.

1. Introduction

The New Basel Capital Accord (Basel Committee on Banking Supervision, 2006) has been created with an objective to better adjust regulatory capital with the underlying risk in a bank's credit portfolio. The new Accord requires international banks to develop and use internal risk models for calculating credit risk capital requirement. It allows banks to compute their regulatory capital in two ways: (1) using a revised standardised approach based on the 1998 Capital Accord which uses regulatory ratings for risk weighting assets or (2) using an internal rating based (IRB) approach where banks are permitted to develop and use their own internal risk ratings.

The IRB approach is based on three key parameters used to estimate credit risk: PD – the probability of default of a borrower over a one-year horizon, LGD – the loss given default, the credit loss incurred if a counterparty of the bank defaults and EAD – exposure at default. These parameters are used to estimate the expected loss, which is a product of PD, LGD and EAD. There are two variants of IRB available to banks, the foundation and the advanced approach. The difference is in estimation of the parameters. In the foundation approach only PD is estimated internally, LGD and EAD are based on supervisory values. In the advanced approach all parameters are determined by the bank.

Most of the banks are prepared to use the foundation approach, since they have already built internal models to estimate PD. However, many banks are not ready to implement fully the advanced IRB approach for the non-retail segment. This is because to move from the foundation approach to the advanced approach requires banks also to model and determine LGD.

Banks need to understand LGD, its components and various associated issues. This research contributes to propose a methodology to estimate loss given default and then apply it to a set

of micro-data of loans to small and medium sized enterprises (SMEs) and corporations. The data were provided by an anonymous commercial bank from Central Europe (the “Bank”). The access to a unique database of loans enables us to show empirically a timing of defaulted loans recovery, cumulative recovery rates and economic determinants of LGD.

The first contribution of our paper is hence the proposition of a methodology for the advanced IRB modelling of corporation and SME LGD. The majority of banks still do not possess advanced models to assess riskiness of their credit portfolios, primarily due to lack of quality data. Additionally, important assumptions about costs, discount factors, downturn aspects and regulatory requirements have to be made. The paper presents answers to majority of these issues; it gathers currently used methodologies, assesses their performance and suggests necessary details. The devised advanced IRB methodology with possible options is useful for banks to measure more accurately the riskiness of their credit portfolio.

Specifically, we have focused our attention on the choice of an appropriate discount rate, which has a critical impact on the ex-post observed LGD. There is no agreement about which rate to choose. We discuss possible alternatives and state the most appropriate options. Additionally, we have performed a dynamic analysis of workout LGD in order to understand the timing and the process of distressed loans recoveries. This information helps to increase the bank’s internal workout process efficiency what leads to a lower LGD. We are going one step further beyond the requirements of Basel II and we analyse not only the ultimate result of the recovery process but also how this process evolves in time.

The second purpose of our paper is an empirical study based on the set of micro-data received from the Bank. Based on the literature, we have proposed and applied three different statistical modelling techniques in order to estimate determinants of the LGD — (1) generalised linear models using symmetric logit and asymmetric log-log link functions for ordinal responses as well as (2) for fractional responses using beta inflated distribution and (3) quasi-maximum likelihood estimator. Moreover, several ways how to measure predictive performance are suggested.

Our paper is organised as follows; the second section is a brief literature review, the third section discusses different issues regarding LGD mostly from a regulatory perspective, the fourth section tackles the issue of an appropriate discount rate, the fifth section focuses on characteristics of LGD from a modelling perspective and on a description of data, the next section depicts the methodology used, while the last three sections provide results, goodness-of-fit performance measures and conclusions, respectively.

2. Literature Review

Banks using the advanced IRB approach need to consider common characteristics of losses and recoveries. These basic characteristics are bimodality, seniority and type of collateral, business cycles, industry and size of loan. The following studies explored these characteristics for both bonds and loans.

Recovery rates¹, defined as a percentage of recovered exposure during the workout process tend to have a bimodal distribution. Bimodality implies that most of the loans have recovery close to 100% (full repayment) or there is no recovery at all (bankruptcy). Bimodality makes parametric modelling of recovery difficult and requires a non-parametric approach (Renault and Scaillet, 2004).

The second important issue is collateral of defaulted claims and their place in capital structure. Bank loans are typically at the top of the capital structure implying generally higher recovery rates than bonds. Recovery rate tends to be higher (i.e. LGD tends to be lower) when the claim is secured by a collateral with high quality. Asarnov and Edwards (1995), Carey (1998) and Gupton et al. (2000) confirmed that seniority and collateral matter. They used primarily data from Citibank and Moody's.

There is strong evidence that recoveries in recessions are lower than during expansions, for instance according to Carey (1998) and Frye (2000). Employing Moody's data they showed that during recessions recoveries are lower by one third.

Other studies by Grossman et al. (2001) and Acharya et al. (2003) argue that industry is another important determinant of LGD. Results of Altman and Kishore (1996) provide evidence that some industries such as utilities (70% average recovery) do better than others (e.g. manufacturing 42%).

The most ambiguous key characteristic is the size of a loan. Asarnov and Edwards (1995) and Carty and Lieberman (1996) found no relationship between LGD and size of loan on the U.S. market. Thornburn (2000) obtained similar negative result for Swedish business bankruptcies. However, Hurt and Felsovalyi (1998) show that large loan default exhibit lower recovery rates. They attribute it to the fact that large loans are often unsecured, and they are provided to economic groups that are family owned.

Currently, the bank loan LGD is not explored well by theoretical and empirical literature. Although several empirical academic studies have analysed credit risk on corporate bonds, only few studies have been applied to bank loans. The reason for this is that since bank loans are private instruments, few data are publicly available.

¹ The same applies to LGD defined as 100% minus a recovery rate percentage.

Altman (1989) and Altman and Suggitt (2000) applied actuarial analysis to study mortality rates of U.S. corporate bonds. This was followed by studies on recovery rates in the bond market, on corporate bonds reported information, on the probability of default over time for different bond ratings, on recovery rates based on market prices at the time of default, on estimates of rating transition matrices and on the degree of correlation between default frequencies and recovery rates. For instance we can mention papers by Frye (2000 and 2003), Nickell et al. (2000), Allen and Saunders (2003), Altman et al. (2003), and Acharya et al. (2003). Altman et al. (2003) report that recovery rates on defaulted bonds are negatively affected by the supply of defaulted bonds. Acharya et al. (2003) found that recoveries on individual bonds are affected not only by seniority and collateral, but also by the industry conditions at the time of default. These two empirical studies validate the theoretical study by Shleifer and Vishny (1992) who examined the impact of industry conditions on liquidation values.

The most important studies focusing on the bank loan markets are the following. Asarnow and Edwards (1995) analysed 831 defaulted loans at Citibank over the period 1970-1993 and show that the distribution of recovery rates is bimodal, with concentration of recovery rates on either the low or high end of the distribution. Their average recovery rate is 65%. Carty and Lieberman (1996) measured the recovery rate on a sample of 58 bank loans for the period 1989-1996 and report skewness toward the high end of price scale with the average recovery of 71%. Gupton et al. (2000) report higher recovery rate of 70% for senior secured loans than for unsecured loans (52%) based on 1989-2000 data sample consisting of 181 observations. The above studies focused on the U.S. market. Hurt and Felsovalyi (1998) who analysed 1,149 bank loan losses in Latin America over 1970–1996 find average recovery rate of 68%. Another study by Franks et al. (2004) calculate recovery rates of 2,280 defaulted companies whose data was taken from 10 banks in three countries over the period 1984-2003. They find country specific bankruptcy regime, which indicates significantly different recovery rate. Average recovery rates are 53% for France, 61% for Germany and 75% for UK. Summary of studies are reported in table in *Appendix*.

None of the above studies provide information on the timing of recoveries. Paper written by Dermine and Neto de Carvalho (2006) is the first study in which authors take into consideration the timing as an important factor. Furthermore, this paper is the first study to apply the workout LGD methodology on a micro-data set from Europe. They estimate LGD for a sample of 374 corporate loans over period 1995–2000. The estimates are based on the discounted value of cash flows recovered after the default event and the estimated average

recovery is 71%. They find that beta distribution does not capture the bimodality of data and using multivariate analysis they identify several significant explanatory variables.

3. *Characteristics and Regulatory Issues of LGD*

The key issues of LGD (Schuermann, 2004) are the following, (1) definition and measurement, (2) key drivers and (3) modelling and estimation approaches. In this section we describe characteristics of LGD focusing on the regulatory issues.

LGD is usually defined as a ratio of losses to an exposure at default. There are three classes of LGD for an instrument. These are market, workout and implied market LGD. Market LGD is observed from market prices of defaulted bonds or marketable loans soon after the actual default event. Workout LGD is derived from a set of estimated cash flows resulting from a workout and collection process, properly discounted to a date of default. Thirdly, implied market LGD is derived from risky but not defaulted bond prices using a theoretical asset pricing model. In this paper only workout LGD is considered.

Definition of a default

There is no standard definition of a default. Different definitions are used for different purposes. Even the international rating agencies, like S&P, Moody's and Fitch, use different default definitions. However, the measured loss at the default depends on the definition of default, so it is important to make clear the definition that is used.

According to the Bank for International Settlements (BIS) a default is a situation when an obligor is unlikely to pay its credit obligations or the obligor is past due more than 90 days on any material credit obligation. We follow the second part of this definition.

Measurement of LGD

There are four ways to measure long-term average LGD on a portfolio level, default weighted averaging vs. time weighted averaging and default count averaging vs. exposure-weighted averaging.

	Default count averaging	Exposure weighted averaging
Default weighted averaging	$LGD = \frac{\sum_{y=1}^m \sum_{i=1}^{n_y} LR_{i,y}}{\sum_{y=1}^m n_y}$	$LGD = \frac{\sum_{y=1}^m \sum_{i=1}^{n_y} EAD_{i,y} \times LR_{i,y}}{\sum_{y=1}^m \sum_{i=1}^{n_y} EAD_{i,y}}$
Time weighted averaging	$LGD = \frac{\sum_{y=1}^m \left(\frac{\sum_{i=1}^{n_y} LR_{i,y}}{n_y} \right)}{m}$	$LGD = \frac{\sum_{y=1}^m \left(\frac{\sum_{i=1}^{n_y} EAD_{i,y} \times LR_{i,y}}{\sum_{i=1}^{n_y} EAD_{i,y}} \right)}{m}$

Table 1 – Different measurement of LGD on a portfolio level

Table 1 shows these four options, i is a default observation, y is the year of default, there are n_y defaults in each year and a total of m years of observations, LR is the loss rate or LGD for each observation.

The time weighted averaging is less desirable as it smoothes out high LGD years with low ones and can therefore underestimate LGD. Therefore, in practice the default weighted averaging is used. For the non-retail segment the default count averaging is recommended and we use it for the analysis of LGD in time. On the other hand, the exposure weighted averaging is frequently used for retail portfolios.

Economic Loss

The definition of loss used in LGD estimation for regulatory purposes is the economic loss. When measuring the economic loss, all relevant factors should be taken into account, such as material discount effects and material direct and indirect costs associated with collection of the exposure². Direct (external) costs include fees to an insolvency practitioner, costs of selling assets, costs of running a business and other professional fees. Indirect (internal) costs are costs incurred by a bank for recovery in the form of intensive care and workout department costs. Economic loss should also consider the cost of holding the non-performing assets (funding costs) over a workout period. Funding costs should be reflected in an appropriate discount rate, which includes a risk premium of the underlying assets³. Moreover, it is also important to understand the effectiveness of the workout process in time, particularly to make appropriate assumptions for modelling LGD⁴.

² Taking into account these factors distinguishes an economic loss from an accounting loss.

³ The issue of an appropriate discount rate is discussed in the *Section 4*.

⁴ The analysis of a workout period length and time distribution of recoveries is presented in the next part.

To estimate internal costs several methods are possible. Aggregate workout costs or costs of intensive care of the workout department could be related to the (1) aggregated amount of exposure, (2) aggregate recovery amount or (3) to the number of defaults in a given period. The reasoning for the first alternative is that more costs are allocated to events with larger exposure. However, the amount recovered is even more important, so the option where higher costs are related to the higher recoveries could be more appropriate.⁵ The third case suggests that workout costs are more or less constant, regardless of the size of an exposure or recovery of a particular file. In this paper internal cost of the workout process are estimated as 2% relative to the recovered amount based on past experience. Actual external costs were available for each default case, so they are used in our analysis.

Length of a workout period, Cumulative Recovery Rates and Time Distribution

The estimation of a workout period length and analysis of recoveries in time is important from both regulatory and modelling perspective. The recovery period starts when the client defaults or when the workout department undertakes the file. The recovery period ends when the file is officially written-off or when the counterparty is cured and gets back to the portfolio of performing loans. Nowadays, most of the issues in Central European commercial banks are non-closed because of relatively short period since the transition to market economy and increase in defaults. Some of the non-closed files can be included in the sample of closed files as the estimated amount to be recovered is not significant. For these cases we can consider their length of the workout period:

- until non-recovered value is less than 5% of EAD
- one year after default (mainly used in retail)
- +25% upper percentile from the distribution of length of workout period
- until effective recovery period (useful for non-retail).

In this study, the last option was used; the effective workout period was estimated based on the cumulative recovery rate analysis. Cumulative recovery rate was calculated in order to show a dynamic evolution of recovery rates over time⁶, i.e. the time distribution of recovery rates. This enables to analyse the evolution of recovered amount and identify the reasons of non-efficient workout process. The population sample is changing over time, so this approach is appropriate, because it takes into the consideration the time as an important factor. Count

⁵ For these two alternatives it is preferable to set a floor and a ceiling for minimum and maximum internal costs.

⁶ The methodology is based on a univariate mortality-based approach, which was applied in Dermine and Neto de Carvalho (2006); calculation does not include internal and external costs.

weighted and exposure weighted average cumulative and marginal recovery rates are calculated quarterly after the default date (*Figure 1* and *Figure 2*).

Firstly, there is clear evidence about the size effect; counterparties with higher exposure at default have significantly lower recovery rates. Secondly, within the first year of recovery the count weighted average recovery rate is 37%, at the end of the third year it is 60% and after that average recovery rate increases only slightly. At the end of the seventh year recovery rate achieves 67%⁷ and after that it is almost unchanged. Counterparties with a short recovery period (less than 1 year) have different behaviour than those who are subject to a long recovery process⁸.

Average marginal recovery rate is calculated as the average recovery rate on the remaining exposure at the end of the particular period and it is shown in *Figure 2*.

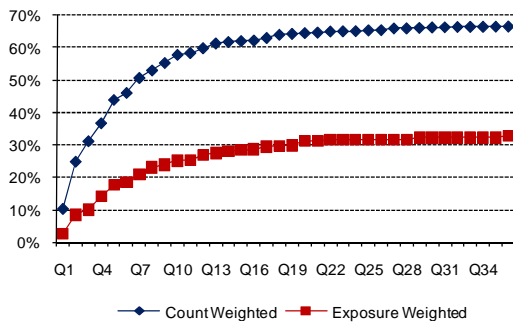


Figure 1 – Average cumulative recovery rate

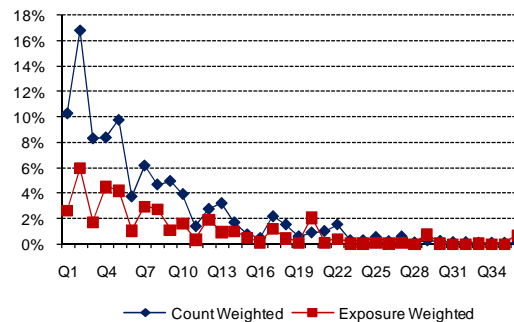


Figure 2 – Average marginal recovery rate

The figure clearly shows that the highest marginal recovery rate is at the beginning of the recovery period. Based on this analysis we can conclude that the workout process is effective until the end of the third year. After the third year of recovery process there are only minor recovered amounts, mainly due to earlier defaulted counterparties with a long recovery period. Workout process has different evolution for counterparties with large exposure (less volatile) than for small loans. Moreover, counterparties with a high recovery in the first two quarters, tend to have low LGDs with a short recovery period. To conclude, cumulative and marginal recovery rates with time distribution should be considered in any LGD model for a defaulted portfolio to better estimate the remaining amount to be recovered.

Downturn LGD

The last important regulatory issue we would like to discuss is downturn LGD. Basel II requires reflecting economic downturn conditions when estimating LGD. This LGD cannot be

⁷ Exposure weighted average recovery rates after 1, 3 and 7 years are 14%, 27% and 32% respectively

⁸ For comparison, see the figures in *Appendix*.

less than a long-run default-weighted average. To estimate downturn LGD based on own historical data, banks need to have at least seven years long period dataset. Currently in most of Central European commercial banks this condition is not met. For this reason to achieve an indication of downturn LGD several options are available:

- use a different discount factor
- work with default weighted LGD instead of exposure weighted or time weighted LGD
- take into the consideration the non-closed files, where the recovery is lower
- use macroeconomic factors within several stress scenarios
- choose 5 worst years out of last 7 years.

To estimate downturn LGD non-closed files are recommended to be included in the model, until there are enough long periods available. In this study, also non-closed files are included, hence the estimated LGD can be considered as an indication of downturn LGD.

4. Discount rate

In order to calculate LGD for a particular client ex-post realised cash-flows have to be discounted back to the time of default. There is no agreement about which rate to choose hence in this section we discuss the possible alternatives and state those options which we believe are the most appropriate and are used in our calculation of LGD.

A *pre-default* required rate (k)⁹ (a contract rate) to discount a stream of cash-flows such as interest and loan repayments can be decomposed into three components:

- a risk-free rate (r_f),
- a default premium (δ_{dp}),
- a risk-premium (δ_{rp}).

A risk-free rate represents a risk-neutral measure of a time value of money and is typically represented by a yield of government security such as a Treasury bill or a Treasury bond. A default premium included in the contract rate is a compensation for expected reduction in received cash-flows due to expected default and less than full recovery of payments from some clients. For a specific pool of clients with similar risk characteristics a bank estimates a probability of default (π) and a recovery rate (rr) to arrive at a default premium. A bank receives expected cash-flows which under risk-neutrality are discounted by the risk-free rate to arrive at the present value. A risk-averse bank, however, demands compensation for the volatility of actual cash-flows from expected ones, hence a risk-premium is added to the

⁹ ‘Discount rate’, ‘required rate’ and ‘expected rate’ are usually used interchangeably and this is the case also in this paper.

discount rate. While the default premium ensures that the return is at the level of risk-free rate *on average*, adding the risk-premium provides an additional compensation (above the risk-free rate) for the fact that the return may be lower in *an individual case*. The intuition behind the different risk components can be formalised by a single-period case (e.g. Jorion 2007). Using the notation already defined and assuming a loan with a single cash-flow (full repayment) in one year, the present value equals:

$$PV = \frac{\$100}{1+k} = \frac{\$100}{1+r_f+d_{rp}+d_{dp}} = \frac{\$100}{1+r_f+d_{rp}} \times (1-p) + \frac{rr \times \$100}{1+r_f+d_{rp}} p \quad (1)$$

The present value of the loan is simply a probability-weighted average of non-default and default cash-flows. Please note that the full cash-flow is discounted by the full discount rate (including the default premium), while the reduced (expected) cash-flows discount is without the default premium as the default risk is already reflected in the parameters π and rr . With 0% default probability or 100% recovery rate, there is no default risk and the default premium term cancels out.

By rearranging the equation (1) and dropping-out the second order terms, the full discount rate can be alternatively defined as the sum of the risk-free rate, a default probability multiplied by a loss given default¹⁰ and the risk premium:

$$k \approx r_f + d_{dp} + d_{rp} = r_f + p(1-rr) + d_{rp} \quad (2)$$

While the risk-free rate is directly observable, the other parameters have to be inferred. Risk premium is of particular interest, as it is needed to calculate LGD (1-rr) from ex-post realised cash-flows. It is determined by the level the risk-aversion of investors, i.e. how much they require for a specific level of risk. Maclachlan (2005) lists various proposals of which discount rate to use to calculate LGD from ex-post realised cash flows¹¹, we briefly summarise the pros and cons of the most promising alternatives:

Original contractual loan rate – It is argued that as this rate reflects the opportunity cost of losing future payments and the risk of the client, hence it ought to be used. The problem with this approach is threefold. Firstly, the risk (as reflected in δ_{rp}) typically changes from the date of loan origination to the point of default and the original premium might no longer be representative. Secondly, if the expected inflation reflected in the risk-free rate is significantly

¹⁰ As already defined, loss given default = 1 – recovery rate.

¹¹ For a large enough sample, under rational expectations, ex-post realised cash-flows are on average a good approximation of ex-ante expected cash-flows (e.g. Brady, et al. 2007), hence these cash-flows are to be discounted by ex-ante expected rate.

different, using contractual rate is not appropriate. Thirdly, the contractual rate includes the default premium (δ_{dp}) which must not be used (as we have already discussed) to discount cash flows that have been already reduced by the realisation of a default risk.

Lender's cost of equity – Cost of equity of a bank is a sum of a risk-free rate and a risk premium so as such it meets the definition of the rate in the equation (1) to discount expected risky payments. However, in general it is defined as one number representing overall required rate of the bank averaging out the risk of all future cash-flows. To measure the LGD more reliably, we have to distinguish between the risks and hence have different rates to discount cash-flows with different risks.

Ex post defaulted bond and loan returns – Brady et al. (2007) conducted a study using recovery information of 1,139 defaulted bonds and loans from 1987 through the second quarter of 2005. The database they used contained the information about market prices of defaulted instruments, information about recovered cash-flows and various characteristics of instruments such as presence of collateral, S&P rating, industry code, debt structure and instrument type. By equating 30 day average price of the defaulted debt with recovery values, they calculated “the most likely estimates” of discount rates. The factors found to be determining the risk premium¹² were obligor's initial rating, whether or not the industry is in a stressed condition at the time of default, relative seniority to other debt and an instrument type. Regarding the instrument types, point estimate of risk premium for bank debt was 9.4%, in between of the range of other types (senior secured bonds 4.1%, senior unsecured bonds 23.1, senior subordinated bonds -1.1%, subordinated bonds 0.6%). Bank debt sub-divided based on other factors considered was not found significantly different from the overall figure. However, there is an important limitation of the data used; secured debt was not distinguished based on the actual level of collateral (full or partial). The premium can serve as a useful benchmark in the calculation of appropriate discount rate and it is also used in this paper.

Systematic asset risk class – The next option (Maclachlan 2005) proposes to use a systematic risk of the asset class under risk. If the defaulted debt is secured by a collateral independent of the company, the systematic risk of the collateral is to be used to determine the discount rate. If the debt is unsecured, the overall risk of company assets is to be used. For calculation of the required premium standard CAPM¹³ is used. This enables to distinguish between various risks

¹² Risk premium is the difference between the computed most likely discount rate and the average risk-free rate for that period.

¹³ CAPM – Capital Asset Pricing Model.

(discount rates) based on different sources of net cash flows. Five levels of discount premiums were recognised, 0 basis points (bps) for liquidation of cash collateral, 240 bps for liquidation of residential mortgage, 420 bps for liquidation of small SME, 480 bps for liquidation of large SME, 600 bps for liquidation of high-volatility commercial real estate (HVCRE), continuation of the original contract and re-negotiation of contract and the highest premium of 990 bps for a guarantee payment. We are going to use these premiums to calculate different discount rates for the calculation of LGD. Compared to the previous approach with flat 940 bps premium, this approach seems much less conservative as only the last category has a higher risk premium. As one client generally has more than one type of collateral, we weight the risk-premiums based on the percentage of particular collateral out of the exposition at default (EAD) to arrive at a composite discount rate for a particular client.

In our calculations we tested flat LGD premiums of 0-9% (each time increased by 1%) and the 9.4% premium. Increasing the premium by 1% resulted in an increase of LGD by approximately the same percentage point. This relatively small effect is due to relatively short average workout period and significant portion of payments received in early years of workout periods. Additionally, LGD was calculated using different premiums for each asset class of collaterals. The resulting average LGD is similar to the flat premium of 5% which is in the range of currently accepted equity risk premium¹⁴ reflecting average risk premium required by investors. As we consider using asset class premiums as the most plausible approach, LGD calculated by this approach was used in further calculations. The effect of application of this discount rate is shown in *Figure 3* below.

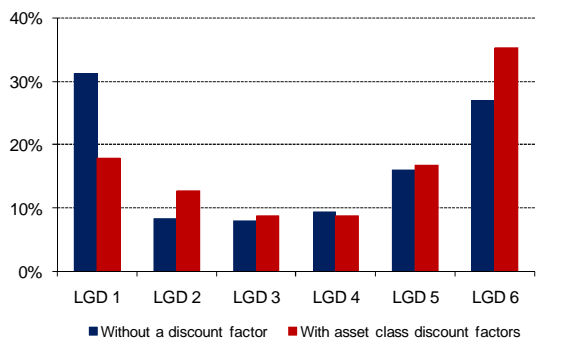


Figure 3 – *The effect of a discount factor on LGD*¹⁵

5. Determinants and Modelling Issues of LGD

In this part, we describe the portfolio that is analysed in the paper and discuss determinants and modelling issues of LGD.

¹⁴ Equity risk premium is the difference between the return of stocks and risk-free government bonds.

¹⁵ LGD grades 1 to 6 are based on Moody's grades and described in the next chapter.

Data sample

The original data sample is based on all available historical closed files for 1989–2007¹⁶ and all open defaulted issues. In the first step we used all closed files. Secondly, we decided to enhance the dataset and we included those non-closed files whose recovery period was currently longer than the effective recovery period. As analysed in the *Section 3*, after twelve quarters of the workout process recovery increases only slightly. Hence, in these cases we do not expect significant increase of recovery rate in the remaining part of the workout process. We are aware of the fact, that our estimation of LGD could be overestimated¹⁷ as a result of inclusion of non-closed counterparties.

Additionally, we decided to split the sample into two parts; the first subsample includes the cases closed within a year whereas the second part contains defaults with longer recovery period. Observations with a very short workout period likely represent special cases that are different from a normal workout process. These might be either “technical defaults” when a client falls in the definition of default for temporarily having past due obligations (LGD close to 0%) or the cases of frauds with LGD close to 100%. Possibly different determinants of LGD might be important for each subsample, so we will analyse the whole sample and each of the subsamples separately. The overall LGD is 52%, for files closed within a year the figure is 16%, while for the second subsample LGD amounts to 60%¹⁸. Our results are in line with other empirical studies reported in *Appendix*.

The observations are aggregated at the level of a client; altogether there are several hundred data points¹⁹. For each default case the amount of cash flows received from the workout process²⁰ and their timing are available together with other data collected by the workout department such as exposure at default, type and amount of collateral, type of loan, a year of loan origination, etc.

Bimodal distribution

The LGD modelling techniques significantly differ from the techniques used in PD modelling. LGD does not have a normal distribution like PD, but bimodal. While a PD model estimates a likelihood of default, which is a binomial event, LGD model has to estimate a loss severity,

¹⁶ In the early years of this period, however, not all defaults were recorded and some of the information was missing. Moreover, recent defaults are not closed and the workout period is short, so the data are not included in our dataset. The majority of quality data is for the period 1995–2004.

¹⁷ On the other hand, as we have noted, employing this approach is an indication of downturn LGD.

¹⁸ For comparison with some of the studies which included only closed files, the second subsample can be further divided into closed files (LGD of 34%) and open files (LGD of 67%).

¹⁹ A more exact number of observations is not presented to preserve confidentiality of the Bank.

²⁰ The cash flows from the workout process equal recovered amount minus direct costs of recoveries.

which is a continuous variable. Thirdly, PD is based on a fixed event while LGD is a dynamic issue taking several years. These features make LGD modelling more challenging. More detailed data is to be collected and more advanced techniques are required to be applied. In general, there are more factors having impact on LGD, however, at the same time less data is available to modellers.

One possibility how to account for bimodality is to map continuous LGD to a number of LGD grades. In each of these classes, data are more normally distributed than overall LGD. We use LGD grades based on Moody's²¹

LGD1	0 % <= LGD < 10 %	LGD4	50 % <= LGD < 70%
LGD2	10 % <= LGD < 30 %	LGD5	70 % <= LGD < 90%
LGD3	30 % <= LGD < 50%	LGD6	90 % <= LGD < 100%.

The frequency counts based on these grades as already depicted in *Figure 3*, it reveals the binomial pattern of the LGD distribution.

LGD Censoring between 0 and 1

LGD can be less than 0%, implying that the bank ultimately recovers more than 100%. This is possible due to the fact that LGD is expressed as a percentage of an EAD. After default the claim on the borrower rises due to interest accruals, fees and fines. If the client ultimately pays the full claim, which contains the interest accruals and fines, the recovery can exceed the original EAD resulting in a negative LGD. However, there are two reasons, why LGD needs to be cut off from below. From a practical point of view, the bank recovers the full amount of original EAD and the final LGD is 0%. Secondly to avoid distortions from the smallest files (for which the ultimate recovery might reach levels of 1000% or more), LGD needs to be cut off from above.

LGD can be more than 100% e.g. as a result of high discount rates and workout costs, which exceed recoveries. Again, LGD needs to be cut off to avoid distortions. The appropriate cut off rate depends on the sample and needs to be decided based on the number of observations that is cut off and the effect of the cut off on distribution and average of LGD. In this paper the LGD is censored between 0 and 1, similarly to many other publications.

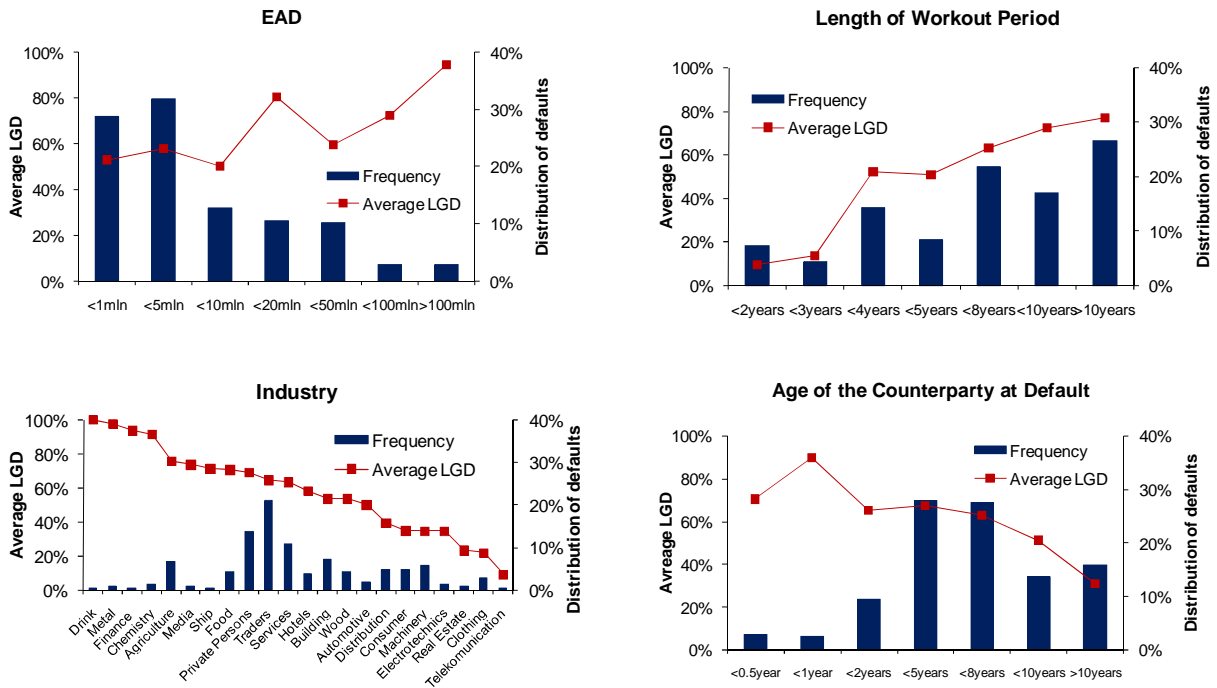
Typical Risk Drivers

The following charts (*Figure 4*) show the distribution of the portfolio and average LGD according to factors typically discussed in the literature – EAD, length of workout process, industry, age of the counterparty at the moment of default and other. Columns marked on the right axis indicate frequencies and the lines are the average LGDs with values shown on the

²¹ Alternatives are the other major rating agencies such as S&P and Fitch with similar LGD grades.

left axis. In general, features of our data sample are consistent with the characteristics described in the literature²².

There is a signal that counterparties with larger EAD and a longer workout period have higher LGD. Counterparties operated in a particular industry sector have lower LGD; for instance in the machinery sector there is two times lower LGD than in agriculture sector. The graphs also indicate that more experienced counterparties result in a higher recovery rate and there is strong evidence that the Bank has lower LGD on more secured counterparties. Interestingly, the length of the performing loan period has a negative effect on the bank recovery rate. Finally, counterparties originated and defaulted in the early years of the sample have significantly higher LGD than the more recent defaults. More recently, a defaulted counterparty has a shorter recovery period as it is shown in the figure, indicating that workout process is getting more efficient.



²² The sample consists also non-closed files whose recovery period is currently longer than effective recovery period. For this reason graphs do not show the definite figures.

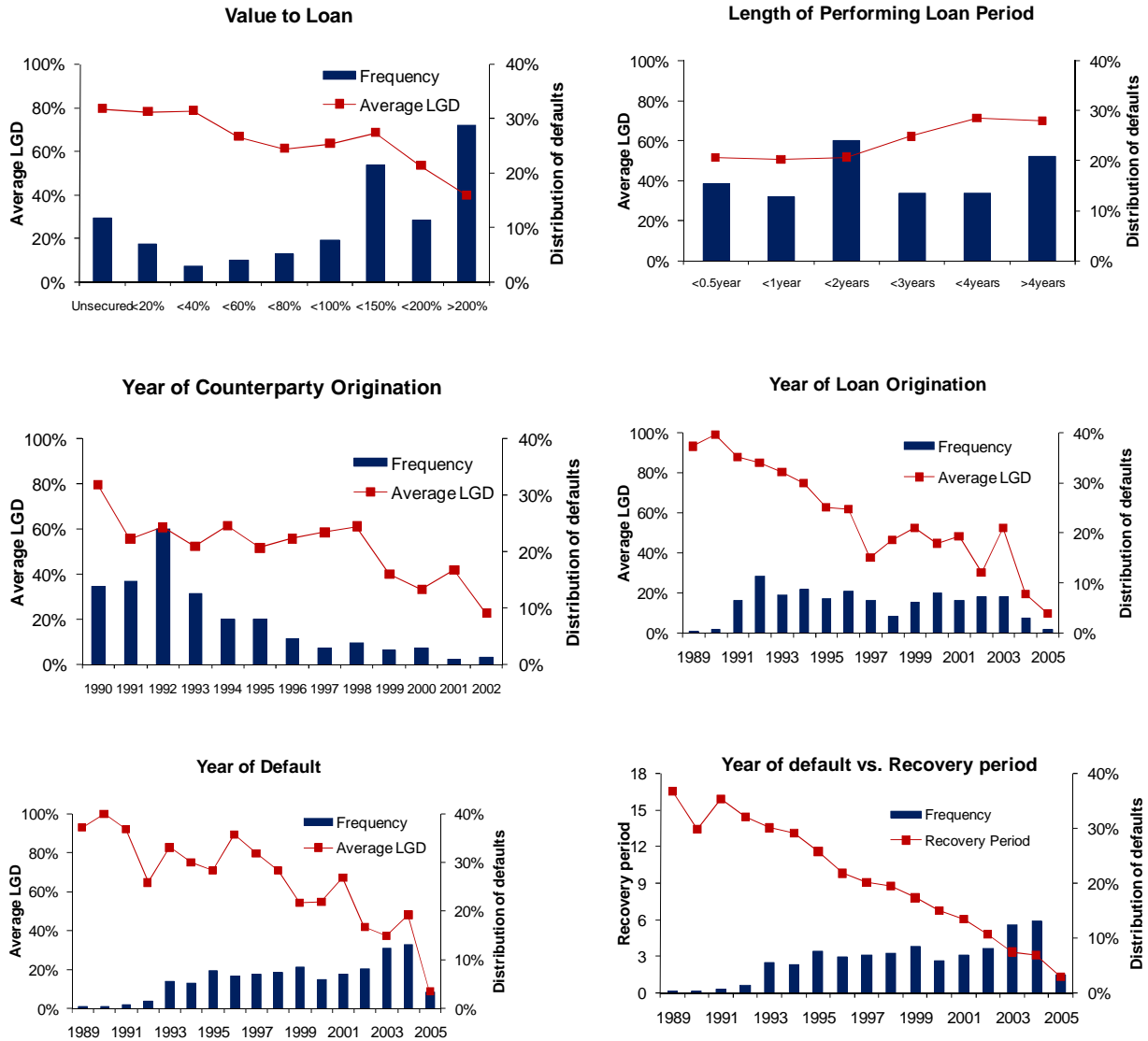


Figure 4 – Characteristics of typical risk drivers

6. Methodology

The following paragraphs describe how the data were processed before carrying out the regression models. Missing data are handled in the following ways:

- Observations with missing data are excluded from the dataset. This option was used in the cases when data necessary for modelling is missing, such as a collateral value.
- Missing data are added, replaced by an average or median value of the portfolio, replaced by a lower or higher cut-off. The age of counterparty is an example.
- Missing data are not replaced neither those observation are excluded. These data currently are not essential for modelling and was kept as unchanged, illustrative factor is related to

the industry where the missing industry was coded as one level along with data where the industry information was available.

Outliers are detected based on the distribution of a factor and an expert judgment. A specific issue is the age of the firm for private persons. In the case of outliers, the appropriate conservative cut-off value is applied, which is determined based on the median, quantiles and power statistics²³ of the factor.

Different types of data need different transformation and adjustment in order to receive a more powerful model. For continuous factors normalising is applied after the elimination of outliers. This is useful when in the model variables like EAD are included (with a wide range of 0 to hundreds of millions in currency units) and factors like age of counterparty (with a narrow range of 0-30 in years).

For categorical factors transforming into dummy variables is carried out, such as a year of default, year of origination, number of collaterals. As an alternative for factors, collateral type or industry, grouping similar categories into one class is employed. The model should distinguish basic types of collateral and industry.

We have used four collateral type classes based on the risk aspect of the collateral, similar to the classes used in the calculation of the discount rate:

Class A: low risk – cash, land and residential real estate

Class B: lower average risk – movables and receivables

Class C: upper average risk – commercial real estate

Class D: high risk – securities and guarantees

In the datasets there are 30 industry groups, we grouped them into fewer categories based on two classifications in *Table 2*:

Standard Industry Codes (SIC)		Alternative industry classification	
A	Agriculture, Forestry, And Fishing	A	Aviation and Transport Services
B	Mining	B	Business Services
C	Construction	C	Consumer Business
D	Manufacturing	D	Energy and Resources
E	Transportation, Communications, Electric, Gas and Sanitary Services	E	Financial Services
F	Wholesale Trade	F	Life Sciences and Health Care
G	Retail Trade	G	Manufacturing
H	Finance, Insurance and Real Estate	H	Public Sector
I	Services	I	Real Estate
J	Administration	J	Technology, Media and Telecommunications

Table 2 – Different industry classifications

Additionally, we “compressed” the alternative industry classification even further by having only two groups, the first one containing the “new industries” (Financial Services, Life

²³ The power statistic is measured as accuracy ratio defined in Sobehart and Keenan (2007).

Sciences and Health Care, Technology, Media and Telecommunications and Business and Consumer Services) and the rest being the “traditional industries”.

Explanatory variables used

From a statistical modelling point of view, factors are divided into continuous factors (can be of any value), categorical factors (can be of only certain number of values) and dummy factors (can be of two values – zero and one). However, from a practical point of view factors are divided into four main categories. We list the variables that are available for our analysis and in *Table 3* we show those determinants of recovery which are actually used in the models. Counterparty related factors²⁴: industry classification, age of the company at the default, year of default, year of company origination, year of loan origination, and length of business connection at the default.

Contract related factors²⁵: type of the contract, exposure at default, interest rate on the loan, tenure, and number of different type of contracts.

Collateral related factors: collateral type, collateral value by type, aggregate collateral value, collateral value relative to the EAD, collateral value as a percentage of aggregate collateral value, number of collaterals, and diversification as a number of different collaterals.

Macroeconomic factors²⁶ are not analysed, because the dataset is relatively short.

Recovery rate determinants	Type	Correlation
Counterparty related factors		
Age of a counterparty	Continuous	Positive
Length of business connection	Continuous	?
Year of default before 1995	Dummy	Negative
Year of loan origination before 1995	Dummy	Negative
New industries	Dummy	?
Industry not specified	Dummy	?
Contract related factors		
Exposure at default	Continuous	Negative
Number of loans	Categorical	?
Investment type of loan	Dummy	?
Overdraft type of loan	Dummy	?
Revolving type of loan	Dummy	?
Purpose type of loan	Dummy	?
Collateral related factors		
Collateral value of A relative to EAD	Continuous	Positive
Collateral value of B relative to EAD	Continuous	Positive
Collateral value of C relative to EAD	Continuous	Positive
Collateral value of D relative to EAD	Continuous	Positive
Number of different collaterals	Categorical	Positive

Table 3 – Recovery rate determinants used in the models (type of variable and expected correlation with recovery rate)

²⁴ Other possible counterparty related factors are a legal form of the company, size of the company, probability of default one year before default, length of time spent in default, intensity of business connection as distance from the domicile, financial indicators such as profitability, liquidity, solvency, capital market ratio, structure of the balance sheet, stock return volatility.

²⁵ Other possible contract related factors are seniority of the loan, and size of the loan.

²⁶ Possible macroeconomic factors are default rates, interest rate, GDP growth, inflation rate, industry concentration.

Multivariate analysis

Three different generalised linear models are applied in order to estimate determinants of the LGD — the first one uses ordinal responses of dependent variable, the other two employ fractional responses either assuming beta inflated distribution or a more general model estimated by the quasi-maximum likelihood estimator. In all three cases logit and log-log link functions are used. As a benchmark, firstly classical linear regression model was used to fit the data.

Models with fractional responses using quasi-maximum likelihood estimator

Since LGD is a continuous variable typically bounded within the interval $[0, 1]$, we need to map the limited interval of LGD onto potentially unlimited interval of LGD scores ($\beta'x$). For this procedure a Generalised Linear Models (GLM) with an appropriate link function can be used (McCullagh and Nelder, 1989). Several link functions are possible. We have applied the logit and log-log links, which are the most common and enable us to capture both, a symmetric (logit) and an asymmetric case (log-log). The quasi-maximum likelihood estimator (QML) described below does not assume a particular distribution and it is hence more flexible to fit the data than a model using a particular distribution.

If we denote the transformation function as $G(\cdot)$, the logit link using the logistic function is

$$G(a + \beta'x) = \frac{\exp(a + \beta'x)}{1 + \exp(a + \beta'x)},$$

the log-log link using the extreme value distribution for dependent variable (the standard Gumbel's case) is

$$G(a + \beta'x) = e^{-e^{-a+\beta'x}},$$

and the complementary log-log link is

$$G(a + \beta'x) = 1 - e^{-e^{a+\beta'x}}.$$

To estimate this GLM we use the non-linear estimation procedure which maximises a Bernoulli log-likelihood function²⁷

$$L_i(a, \mathbf{b}) = y_i [\log G(a + \mathbf{b}'\mathbf{x}_i)] + (1 - y_i) \log [1 - G(a + \mathbf{b}'\mathbf{x}_i)].$$

where a and \mathbf{b} are an estimated value of a and \mathbf{b} .

²⁷ For further technical details and practical applications see Papke and Wooldridge (1996).

Models with fractional responses using a beta distribution

A beta distribution has also been used to model LGD, for example in the commercially available application LossCalc by Moody's (Gupton and Stein, 2002). This approach assumes that LGD has a beta distribution. As the values of the distribution itself are bounded within the range [0, 1] a link function has to be used to map LGD scores into this interval. Again, we have used the logit link and the log-log links. As LGD of 0% or 100% are values which are normally observable and have hence non-zero probabilities p_0 and p_1 , we have used the inflated beta distribution²⁸ with the location, scale and two shape parameters, μ , σ , ν , and τ respectively that allows for 0 and 1 defined by

$$f_Y(y | \mathbf{m}, \mathbf{s}, \mathbf{n}, \mathbf{t}) = \begin{cases} p_0 & \text{if } y = 0 \\ (1 - p_0 - p_1) \frac{1}{B(\mathbf{a}, \mathbf{b})} y^{\mathbf{a}-1} (1 - y)^{\mathbf{b}-1} & \text{if } 0 < y < 1, \\ p_1 & \text{if } y = 1 \end{cases}$$

for $0 \leq y \leq 1$, where $\alpha = \mu(1 - \sigma^2) / \sigma^2$, $\beta = (1 - \mu)(1 - \sigma^2) / \sigma^2$, $p_0 = \nu(1 + \nu + \tau)^{-1}$, $p_1 = \tau(1 + \nu + \tau)^{-1}$ so $\alpha > 0$, $\beta > 0$, $0 < p_0 < 1$, $0 < p_1 < 1 - p_0$.

The estimates of β of this GLM model were produced using maximum likelihood.

Models with ordinal responses

As an alternative technique, we have modelled 6 discrete LGD grades defined earlier using ordinal regression instead of continuous dependent variable used in the previous models²⁹. These models might be more appropriate if we expect default cases to be homogenous *within* a LGD grade but being different *between* grades, either by having a different response to factors (different β) or a different likelihood of a default case to fall into a particular grade (a different intercept). The ordinary regression model using cumulative logit link function is defined as:

$$\begin{aligned} \text{logit}[P(Y \leq j | \mathbf{x})] &= \log \frac{P(Y \leq j | \mathbf{x})}{1 - P(Y \leq j | \mathbf{x})} \\ &= \log \frac{p_1(\mathbf{x}) + \mathbf{K} + p_j(\mathbf{x})}{p_{j+1}(\mathbf{x}) + \mathbf{K} + p_j(\mathbf{x})}, \quad j = 1, \mathbf{K}, J - 1 \end{aligned}$$

Each cumulative logit uses all J response categories. A model for $\text{logit}[P(Y \leq j)]$ alone is an ordinary logit model for a binary response in which categories 1 to j form one outcome and

²⁸ This definition of beta inflated distribution is based on (Stasinopoulos D. M. et al, 2008).

²⁹ Since LGD grades are a dependent variable in ordinal regressions, we have used recovery rates (not LGD) as a fractional response to have the same sign of estimated coefficients in both cases.

categories $j + 1$ to J form the second. A model that simultaneously uses all cumulative logits is

$$\text{logit}[P(Y \leq j | \mathbf{x})] = \mathbf{a}_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, \mathbf{K}, J - 1$$

Each cumulative logit has its own intercept. The $\{\alpha_j\}$ are increasing in j , since $P(Y \leq j | \mathbf{x})$ increases in j for fixed x , and the logit is an increasing function of this probability. This model has the same effects $\boldsymbol{\beta}$ for each logit and we have used it since we consider the same effects in each grade as appropriate; we allow for different intercepts.

To fit this special case of the GLM, let (y_{i1}, \dots, y_{iJ}) be binary indicators of the response of the response for subject i . The likelihood function (e.g. Agresti (2002)) is

$$\begin{aligned} \prod_{i=1}^n \left[\prod_{j=1}^J p_j(\mathbf{x}_i)^{y_{ij}} \right] &= \prod_{i=1}^n \left[\prod_{j=1}^J (P(Y \leq j | \mathbf{x}_i) - P(Y \leq j-1 | \mathbf{x}_i))^{y_{ij}} \right] \\ &= \prod_{i=1}^n \left[\prod_{j=1}^J \left(\frac{\exp(\mathbf{a}_j + \boldsymbol{\beta}'\mathbf{x}_i)}{1 + \exp(\mathbf{a}_j + \boldsymbol{\beta}'\mathbf{x}_i)} - \frac{\exp(\mathbf{a}_{j-1} + \boldsymbol{\beta}'\mathbf{x}_i)}{1 + \exp(\mathbf{a}_{j-1} + \boldsymbol{\beta}'\mathbf{x}_i)} \right)^{y_{ij}} \right]. \end{aligned}$$

It is minimised as a function of different intercepts α_j and common slope coefficients $\boldsymbol{\beta}$ for each LGD grade.

The complementary log-log link for ordinal regression model is defined as

$$\log\{-\log[1 - P(Y \leq j | \mathbf{x})]\} = \mathbf{a}_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, \mathbf{K}, J - 1.$$

With this link, $P(Y \leq j)$ approaches 1 at a faster rate than it approaches 0.

The log-log link

$$\log\{-\log[P(Y \leq j | \mathbf{x})]\} = \mathbf{a}_j + \boldsymbol{\beta}'\mathbf{x}, \quad j = 1, \mathbf{K}, J - 1$$

is appropriate when the complementary log-log link holds for the categories listed in reverse order.

Selecting the appropriate model

In order to select the most appropriate model, some commonly used procedures are followed. Continuous variables are plotted against LGD (and against LGD grades for ordinal responses) to get “a feel” of the underlying relationship. Similarly, categorical variables are tabulated to form an expectation of a potential relationship. Moreover, a frequency table provides information whether there are enough counts for each cell to estimate reliably the effect³⁰. Thirdly, univariate regressions using each explanatory variable separately is performed to see the effect of each variable independent of the other effects. Then all potentially plausible

³⁰ This is important for ordinal regression as we have six grades and we have to have enough observations for each explanatory variable in each grade.

variables are put together in a regression model. Afterwards, variables not contributing significantly to the explanatory power of a model are gradually eliminated from the model (backward elimination) based on Akaike (AIC) and Schwarz information criteria (SIC). For models where continuous dependent variable is estimated and a specific distribution is assumed, worm plot for residuals (van Buuren et al., 2001), and QQ-plots³¹ were utilised to have a visual indication of normality of residuals.

7. Results

The results presented in the following sections using more conservative SIC³² reveals interesting findings. As expected, collaterals of class A and C have a positive and strong effect on recovery rates. These collaterals represent land, residential real estate, cash and commercial real estate; hence there is no surprise to find a strong positive relationship, higher proportion of collateral as a % of EAD increases recovery or likelihood of recovery. On the other hand, a year of loan origination has a negative effect. There are two categories of loans, the loans originated before 1995 and the loans that started later. It was much more likely to encounter high LGD before 1995 than afterwards. Negative sign for loans originating before 1995 indicate that either the workout process is getting more efficient or recoveries improve in time per se. During 1990s Central European economies experienced economic transition and stabilisation in the second half of the decade and later could be a factor for this recovery improvement. EAD is the next variable significant in almost all the models. The correlation with recoveries is negative, for higher loans recoveries tend to be lower. The first explanation from the literature could be a weaker link between the management and company results in the case of big companies having high bank loans. This contradicts the assumption that a bank intensifies the enquiry of the creditworthiness and the monitoring of the borrower for high loans. The second explanation could be high leverage of big companies and violation of the absolute priority rule. If a big company defaults, there are many creditors competing for the company's assets so the recovery for a bank can be small. The effect of other variables is not so unambiguous and the results are different for different models, the specifics are discussed for each class of models.

There are different determinants of LGD for the two subsamples and the whole sample. The main reason is that subsamples contain different type of default cases so the important factors and the estimated coefficients are different. Subsample A (defaults with longer recovery

³¹ In the QQ-plots sample values are plotted against theoretical values predicted by a distribution.

³² AIC yields similar results but allows more variables to be included in the models, we report only results based on SIC.

period than 1 year) has a relatively high concentration of high LGD, the most important factors are EAD, year of loan origination, and collaterals of A and C class. On the other hand, subsample B (default cases closed within a year) observations concentrate close to zero LGD and there is a smaller number of significant explanatory variables. These are EAD and a number of different collateral types. Specific collateral values are not so important for these cases, because these are mainly better clients with temporarily problems and the problems are usually cured so the collateral is not realised. However, the presence of collateral is important. For the whole sample, not all determinants of LGD from each subsample are significant. This can be explained by the fact that the whole sample has a bimodal distribution³³ so some of the LGD determinants from each subsample offset each other in the whole sample. The best fit for the whole sample is achieved for models with ordinal responses, which are able to capture the bimodality.³⁴ For subsample A the best results are obtained by the linear model, although the difference in performance measured by the power statistic compared to the other models is not significant. For subsample B, the quasi-maximum likelihood estimator with the log-log link function provides the best results due to the ability to capture the asymmetric response.

Classical linear regression model

The linear regression model as the benchmark is the simplest case in which a continuous recovery rate variable is regressed on a linear combination of explanatory variables. The major drawback of this method is that the predicted values can be outside the range [0, 1].

Rather surprisingly, the simple linear model is able “to remove” bimodality of the whole sample as shown by residuals which are inside the bounds of confidence intervals of the worm plot, close to normal quantile in the QQ-plot as shown in *Figure 5* below.

For the whole sample and subsample A there are two more significant variables apart from the common factors discussed in the previous subsection (*Table 4*). Length of business relationship has a strong negative effect on recovery rate. For clients with long relationship lower recoveries can be explained by a less prudent attitude to “familiar” clients. On the other hand, a number of different collateral classes of a client (a proxy for diversification) has a positive impact, but the effect is weaker. For subsample B, a number of loans and different

³³ Subsamples have rather unimodal concentration either on the left side with zero LGD (subsample B) or on the right side with high LGD (subsample A).

³⁴ Alternatively, instead of using the whole sample to estimate the parameters, a model for the whole sample could be constructed as a combination of two models based on two subsamples. However, for such model it is necessary to estimate probability of counterparty belonging into one of the subsamples. Such estimation is beyond the scope of our paper.

loan types are the additional determinants, but compared to EAD their quantitative impact is rather weak. In the whole sample only the overdraft type of loan remains significant.

Recovery rate determinants	Subsample A (>1 year)			Subsample B (<1 year)			Whole sample		
	Value	Std. error	P-value	Value	Std. error	P-value	Value	Std. error	P-value
Exposure at default –EAD	-0.230	0.091	0.012	-2.250	0.518	0.000	-0.330	0.089	0.000
Collateral class A as % of EAD	0.411	0.095	0.000				0.359	0.091	0.000
Collateral class C as % of EAD	0.395	0.097	0.000				0.329	0.082	0.000
Length of business relationship	-0.197	0.062	0.002				-0.179	0.060	0.003
Number of different collateral classes	0.079	0.028	0.005	0.211	0.046	0.000	0.103	0.026	0.000
Year of loan origination before 1995	-0.283	0.040	0.000	-0.369	0.162	0.027	-0.298	0.039	0.000
Number of loans				-0.097	0.032	0.004			
Investment type of loan				0.154	0.074	0.042			
Overdraft type of loan				0.125	0.045	0.007	0.186	0.045	0.000
Purpose type of loan				0.143	0.064	0.031			

Table 4 – Results of the classical linear regression model

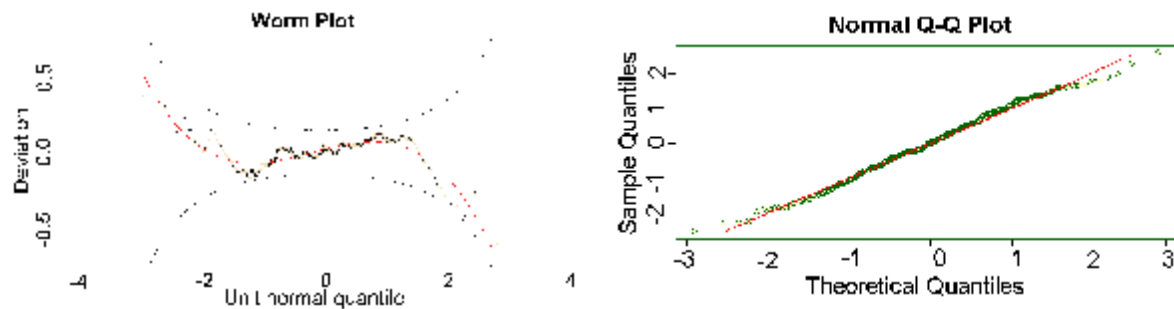


Figure 5 – Tests of normality of residuals for the linear model (the whole sample)

Models with fractional responses using quasi-maximum likelihood estimator

Applying the asymmetric log-log link function³⁵ (Table 5) to regress continuous recovery rates estimated by quasi-maximum likelihood yields better results than the logit link function especially for subsample B, where there is a high concentration on zero LGD.

For subsample A, collateral of class A and C along with a year of loan origination are the major determinants of LGD. EAD which is strongly negatively correlated to the collateral classes appears not to be a significant factor³⁶. In the subsample B, factor EAD with number of different collateral type are the main drivers. Strong negative effect of EAD for loans closed within a year can be explained by the fact that defaults on high exposures indicate a real problem (possibly a fraud with a very low recovery) whereas small exposures are rather technical defaults which are cured with low LGD.

³⁵ Determinants of recovery rates for the logit and complementary log-log function are shown in the summary table in Appendix.

³⁶ However, a larger sample would enable to distinguish the effects of highly correlated factors better and we expect EAD to have an impact.

Recovery rate determinants	Subsample A (>1 year)			Subsample B (<1 year)			Whole sample		
	Value	Std. error	P-value	Value	Std. error	P-value	Value	Std. error	P-value
Exposure at default –EAD				-15.950	3.261	0.000	-1.128	0.271	0.000
Collateral class A as % of EAD	1.802	0.562	0.001				1.491	0.546	0.006
Collateral class C as % of EAD	1.599	0.359	0.000				1.612	0.358	0.000
Number of different collateral classes				1.589	0.282	0.000			
Year of loan origination before 1995	-1.032	0.107	0.000				-1.128	0.112	0.000
Overdraft type of loan							0.825	0.194	0.000

Table 5 – Results of fractional responses (quasi-maximum likelihood estimator, log-log link)

Models with fractional responses using a beta distribution

The fit employing the inflated beta distribution and the logit link is very similar to the log-log link³⁷. Logit reasonably fits the data (*Figure 6*), although the fit is worse for higher recovery rates.

Compared to the other models, collateral of class A (residential real estate) is not significant for any subsample. The model for subsample B (*Table 6*), identify as the additional important factors collateral of class B (movables and receivables), age of counterparty and length of business connection, as well as different type of loans. Moreover, collateral of class C has a negative impact, offsetting the positive influence from subsample A and hence making it insignificant in the whole sample. These singularities can be explained by the assumption of beta distributed errors. The other factors are in line with the previous results.

Recovery rate determinants	Subsample A (>1 year)			Subsample B (<1 year)			Whole sample		
	Value	Std. error	P-value	Value	Std. error	P-value	Value	Std. error	P-value
Exposure at default –EAD	-1.426	0.348	0.000	-24.382	2.530	0.000	-1.946	0.354	0.000
Collateral class B as % of EAD				1.854	0.532	0.001			
Collateral class C as % of EAD	0.963	0.405	0.018	-2.227	0.744	0.005			
Age of a counterparty				1.329	0.550	0.020			
Length of business relationship				1.900	0.862	0.033			
Number of different collateral classes							0.311	0.111	0.005
Year of loan origination before 1995	-1.364	0.169	0.000				-1.390	0.168	0.000
Industry not specified	-0.725	0.182	0.000				-0.695	0.166	0.000
Investment type of loan				2.984	0.625	0.000			
Revolving type of loan				1.718	0.386	0.000			
Overdraft type of loan				0.909	0.376	0.020	0.845	0.188	0.000
Purpose type of loan				1.443	0.444	0.002			

Table 6 – Results of fractional responses using beta distribution (logit link)

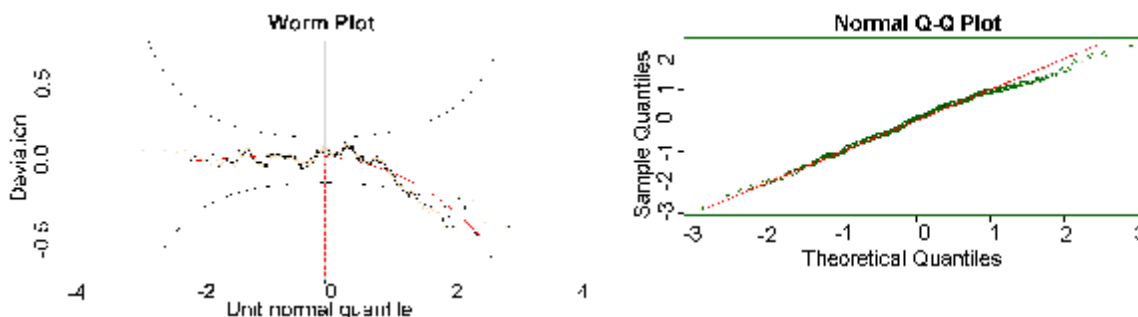


Figure 6 – Tests of normality of residuals for beta distribution (logit link, the whole sample)

³⁷ The log-log and complementary log-log links are again shown in the summary table in *Appendix*.

Models with ordinal responses

This kind of models was not applied on the subsample B due not sufficient number of observations for each LGD grade³⁸. *Table 7* shows the results for the cumulative logit link model. Increasing the length of business relationship measured as a period between the date of bank account opening and a date of default decreases the likelihood of low LGD, similarly to the classical linear model. Also consistent to previous results, a higher number of different collateral types and older counterparty increase the probability of low LGD. Loans types (apart from overdrafts) and other variables proved not to be statistically significant determinants of LGD.

Recovery rate determinants	Subsample A (>1 year)			Subsample B (<1 year)			Whole sample		
	Value	Std. error	P-value	Value	Std. error	P-value	Value	Std. error	P-value
Exposure at default –EAD	-2.500	0.744	-3.358	n/a	n/a	n/a	-3.471	0.795	-4.367
Collateral class A as % of EAD	2.799	0.650	4.309	n/a	n/a	n/a	2.242	0.634	3.538
Collateral class C as % of EAD	2.338	0.648	3.606	n/a	n/a	n/a	1.802	0.591	3.051
Age of a counterparty				n/a	n/a	n/a	1.202	0.464	2.589
Length of business relationship	-1.208	0.409	-2.956	n/a	n/a	n/a	-1.348	0.422	-3.195
Number of different collateral classes	0.581	0.191	3.049	n/a	n/a	n/a	0.652	0.187	3.491
Year of loan origination before 1995	-1.769	0.275	-6.442	n/a	n/a	n/a	-1.796	0.291	-6.175
Industry not specified				n/a	n/a	n/a	-0.811	0.271	-2.987
Overdraft type of loan				n/a	n/a	n/a	1.133	0.310	3.653

Table 7 – Results of ordinal regression model (logit link)

8. Comparing goodness-of-fit of the models

Goodness-of-fit summary measures offer an overall indication of a model fit. Out of parametric performance measures, mean square error (MSE), mean absolute error (MAD) and correlation between observed and modelled LGD have been evaluated to compare suggested models predictive power. The outcomes from correlation are listed in *Table 8*.

Model	Correlation		
	Subsample A	Subsample B	Whole sample
Linear model	0.603	0.841	0.602
Fractional response Logit link	0.580	0.846	0.536
Fractional response Log-log link	0.557	0.829	0.574
Fractional response Complementary Log-log link	0.573	0.820	0.534
Fractional response Beta - Logit Link	0.540	0.755	0.550
Fractional response Beta - Log-log link	0.541	0.784	0.543
Fractional response Beta - Complementary Log-log link	0.511	0.647	0.550
Ordinal response Logit link	0.548	n/a	0.610
Ordinal response Complementary Log-log link	0.563	n/a	0.605

Table 8 – Parametric performance measure - correlation

MSE, MAD and correlation coefficient measure model performance parametrically and are sensitive to a model calibration. In contrast, a power statistic is a non-parametric measure that focuses on the ability to discriminate “good” from “bad” outcomes without being sensitive to the calibration. It indicates a model’s power and ranges from zero to one. It provides information about different aspects of model performance not registered by the above mentioned measures.

³⁸ Moreover, also results of log-log link model are not available as only differently skewed complementary log-log link could be estimated. The complementary log-log link is again shown in the summary table in *Appendix*.

The Power statistic³⁹ is commonly used for PD models, where there are only two possibilities – a default or a non-default. However, in LGD models dependent variable is continuous. In order to apply power statistic, it is necessary to defined what is considered as “good” and what as “bad”. In this paper we propose three different alternatives. In the first, “bad” are those observations where the observed LGD is higher than the average LGD. Secondly, “bad” are those observations where the observed LGD is higher than the 75th percentile of the LGD. Finally, “bad” are those observations where the observed LGD is higher than the 25th percentile of the LGD. Results for the first case are reported in *Table 9*⁴⁰.

Model	Bad = above average LGD					
	Subsample A (> 1 year)		Subsample B (< 1 year)		Whole Sample	
	Power Statistic	SE	Power Statistic	SE	Power Statistic	SE
Linear model	62.4%	22.5%	61.6%	23.6%	72.0%	26.3%
Fractional response Logit link	57.8%	21.8%	68.3%	22.6%	69.3%	24.9%
Fractional response Log-log link	55.2%	21.3%	69.2%	20.1%	67.7%	25.0%
Fractional response Complementary Log-log link	57.2%	21.1%	68.3%	20.7%	69.0%	25.0%
Fractional response Beta - Logit Link	56.4%	16.8%	49.7%	20.3%	68.4%	20.9%
Fractional response Beta - Log-log link	59.9%	15.6%	56.8%	21.2%	71.0%	19.5%
Fractional response Beta - Complementary Log-log link	53.7%	16.7%	52.1%	14.9%	68.5%	21.0%

Table 9 – Binary power statistic (Accuracy ratio) of models, bad – above average LGD

Alternatively, ordinal power statistic can also be applied for LGD models. This statistic (*Table 10*) measures the ability of a certain model to differentiate between any numbers of rating categories in the correct order; hence it is particularly for the ordinal response models. The results are similar as in the previous case.

Model	Subsample A (> 1 year)		Subsample B (< 1 year)		Whole Sample	
	Power Statistic	SE	Power Statistic	SE	Power Statistic	SE
Linear model	64.4%	4.0%	87.2%	8.8%	70.8%	3.6%
Fractional response Logit link	60.4%	4.3%	88.5%	5.7%	66.5%	3.9%
Fractional response Log-log link	57.7%	5.0%	89.5%	6.4%	65.5%	3.5%
Fractional response Complementary Log-log link	59.3%	4.0%	88.7%	7.8%	66.6%	4.2%
Fractional response Beta - Logit Link	58.6%	4.9%	70.9%	16.8%	67.1%	3.6%
Fractional response Beta - Log-log link	58.5%	5.4%	69.0%	19.8%	66.8%	3.8%
Fractional response Beta - Complementary Log-log link	55.7%	4.7%	55.7%	21.5%	67.4%	3.6%
Ordinal response Logit link	58.3%	4.5%	n/a	n/a	72.0%	2.8%
Ordinal response Complementary Log-log link	61.1%	3.8%	n/a	n/a	71.8%	3.8%

Table 10 – Ordinary power statistic of models

For both versions of power statistics, the linear model, fractional models using quasi-maximum likelihood and ordinal response models perform similarly and relatively well in absolute terms. Power statistic is generally higher for the subsample B, however, in the case of beta distribution the difference compared to the whole sample is not so significant and it is even negative for complementary log-log link.

³⁹ The Power statistic is described for instance in Gupton and Stein (2002)

⁴⁰ The tables for the other two cases are presented in *Appendix*.

9. Conclusions

In this paper, we analysed several aspects of the economic loss. Particularly, we focus on an appropriate discount factor and timing of recovery rates to identify the efficient recovery period. Various statistical models are applied to test empirically the determinants of recovery rates. We find out that the main drivers are a relative value of collateral, a loan size and a year of the loan origination. Different models provide similar results. As for the different links in more complex models, log-log models in some cases perform better, implying an asymmetric response of the dependent variable. All the models performed relatively well when the overall fit of the different models is assessed. However, models with the commonly assumed beta distribution achieve slightly worse results and hence are not deemed optimal for our data.

From a policy perspective, our paper provides evidence that workout LGD is a viable option in credit risk estimation despite various methodological difficulties. In this study, we try to provide a reasonable detail of various issues to be tackled and proposed methodological alternatives how cope with these issues.

To account for different determinants of LGD for different workout period we split the sample into two parts, the first one being the defaults closed within a year while the second subsample are the defaults with longer recovery periods. We show that different determinants are important for each subsample and the effects are then aggregated in the whole sample. For short recovery periods, exposure at default is the most important factor supplemented by existence of a number of collaterals classes. For the longer recovery periods, the year of loan origination and values of real estate collaterals are the major determinants of LGD. Generally, other factors such as the length of business connection or overdraft type of loan appear in some of the models but their effect is weaker.

In this study we answer majority of issues in LGD modelling of advance IRB approach. Nonetheless, there are several ways in which our research can be improved. Firstly, a similar study can be done on a larger sample of data and hence some of the effects could be estimated more precisely. Secondly, correlation of recovery rate and probability of default, effects of macroeconomic factors and downturn LGD should be thoroughly analysed for a complete LGD model.

10. Literature

- Acharya et al. (2003): “Understanding the Recovery Rate of Defaulted Securities”, Working Paper
- Agresti, A. (2002): “Categorical Data Analysis, 2nd Edition”, Wiley Series in Probability and Statistics.
- Allen, L. and A. Saunders (2003): “A Survey of Cyclical Effects in Credit Risk Management Models”, BIS, Working papers 126, pp. 1-32
- Altman, E. (1989): “Measuring Corporate Bond Mortality and Performance”, Journal of Finance, 44, pp. 909-922
- Altman et al (2003): “The Link between Default and Recovery Rates: Implication for Credit Risk Models and Procyclicality”, Mimeo, pp. 1-46
- Altman, E. and Kishore (1996): “Almost everything you want to know about Recoveries on Defaulted Bonds”, Financial Analyst Journal, pp. 57-64
- Altman, E. and H. Suggitt (2000): “Default Rates in the Syndicated Bank Loan Market: A Mortality Analysis”, Journal of Banking & Finance, 24 (1, 2), pp. 229-253
- Asarnov E. and D. Edwards (1995): “Measuring Loss on Defaulted Bank Loans: A 24 year study”, Journal of Commercial Lending, 77(7), pp. 11-23
- Basel Committee on Banking Supervision (2006): “International Convergence of Capital Measurement and Capital Standards”, Bank for International Settlements, Basel
- Brady et al. (2007): “Discount Rate for Workout Recovery: An Empirical Study“. Available at SSRN: <http://ssrn.com/abstract=907073>
- Carey M. (1998): “Credit Risk in Private Debt Portfolios”, Journal of Finance, 53(4), pp. 1363-1387
- Carty, L. and D. Lieberman (1996): “Defaulted Bank Loan Recoveries”, Moody’s special comment
- Dermine J. and C. Neto de Carvalho (2006): “Bank Loan Losses-Given-Default, A case study”, Journal of Banking and Finance, 30, pp. 1219–1243.
- Franks et al. (2004): “A comparative analysis of the recovery process and recovery rates for private companies in the UK, France and Germany”, Standard& Poor’s Risk Solutions
- Frye (2000): “Depressing recoveries”, Chicago Fed working paper, Risk, November, pp. 108-111
- Frye (2003): “A false sense of Security”, Risk, August, pp. 63-67
- Grossman et al. (2001): “Bank Loan and Bond Recovery Study: 1997-2000”, Fitch Loan products special report
- Grunert J. and M. Weber (2005): “Recovery Rates of Bank Loans: Empirical Evidence for Germany”
- Gupton et al. (2000): “Bank Loan Loss Given Default”, Moody’s special comment.
- Gupton, G.M. and R. M. Stein (2002): “LossCalc: Model for predicting loss given default (LGD), Moody’s KMV”.
- Hurt, L. and A. Felsovaly (1998): “Measuring loss on Latin American defaulted bank loans, a 27-year study of 27 countries”, The Journal of Lending and Credit Risk Management, 80, October, pp. 41-46
- Jorion, P. (2007): “Financial Risk Manager Handbook“. Wiley Finance
- Maclachlan, I. (2005): “Choosing the Discount Factor for Estimating Economic LGD“, Recovery Risk, The Next Challenge in Credit Risk Management, Edited by Altman E., Resti A. and A. Sironi, Risk Books, pp. 285-305

- McCullagh, P. and J. A. Nelder (1989): “Generalized Linear Models”, Second edition, London: Chapman and Hall
- Nickell et al (2000): “Stability of Rating Transitions, in Special Issue on Credit Risk Modelling and Regulatory Issues”, *Journal of Banking and Finance*, 24 (1/2), pp. 203-228
- Papke, L.E. and J. M. Wooldridge (1996): “Econometric Method for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates”, *Journal of Applied Econometrics*, 11, pp. 619-632
- Renault, O. and O. Scaillet (2004): “On the Way to Recovery: A Non-parametric bias free estimation of recovery rate intensities”, *Journal of Banking and Finance*, 28, pp. 2915-31
- Schuermann, T. (2004): “What do we know about Loss Given Default?”, *Credit Risk Models and Management*, London, UK: Risk Books
- Shleifer, A. and R. W. Vishny (1992): “Liquidation Values and Debt Capacity: A Market Equilibrium Approach”, *Journal of Finance*, XLVII (4), pp. 1343-1366
- Sobehart, J.R and S.C. Keenan (2007): “Understanding performance measures for validationg default risk models: a review of performance metrics“, *Journal of Risk Model Validation*, 1, pp. 61-79
- Stasinopoulos, D. M. et al. (2008). Instructions on how to use the GAMLSS package in R
- Thornburn K. (2000): “Bankruptcy auctions: costs, debt recovery and firm survival”, *Journal of Financial Economics*, 58, pp. 337-368
- van Buuren, S. – Fredriks, M. (2001): “Worm plot: a simple diagnostic device for modelling growth reference curves”. *Statistics in Medicine*, 20: 1259–1277.

11. Appendix

Authors	Country	Years	Recovery	Observations
Asarnow and Edwards (1995)	USA	1970-1993	65%	831
Carty and Lieberman (1996)	USA	1989-1996	71%	58
Grossman et al. (1997)	USA	1991-1997	82%	60
Grossman et al. (1997)	United Kingdom	1991-1997	68%	14
Felsovalyi and Hurt (1998)	Latin America	1970-1996	68%	1149
Eales and Bosworth (1998)	Australia	1992-1995	69%	5782
Carty (1998)	USA	1986-1997	87%	200
Hamilton and Carty (1999)	USA	1982-1997	84%	195
van de Castle et al. (1999)	USA	1987-1997	85%	258
Bartlett (2000)	United Kingdom	1996-2000	77%	55
Gupton et al. (2000)	USA	1989-2000	70%	181
van de Castle et al. (2000)	USA	1987-1996	84%	264
Kabance (2001)	Mexico	1995-2001	40%	70
O'Shea et al. (2001)	USA	1997-2000	73%	35
Hamilton et al. (2002)	USA	1982-2001	71%	n/a
Bos et al. (2002)	USA	1988-2001	84%	528
Hamilton et al. (2004)	USA	2003	86%	21
Keisman (2003)	USA	1988-2003	79%	750
Araten et al. (2004)	USA	1982-1999	60%	3761
Franks et al. (2004)	United Kingdom	1984-2003	75%	1418
Franks et al. (2004)	France	1984-2003	53%	586
Franks et al. (2004)	Germany	1984-2003	61%	276

Table 11 – Empirical studies concerning the recovery rate of loans, Source: Grunert and Weber (2005)

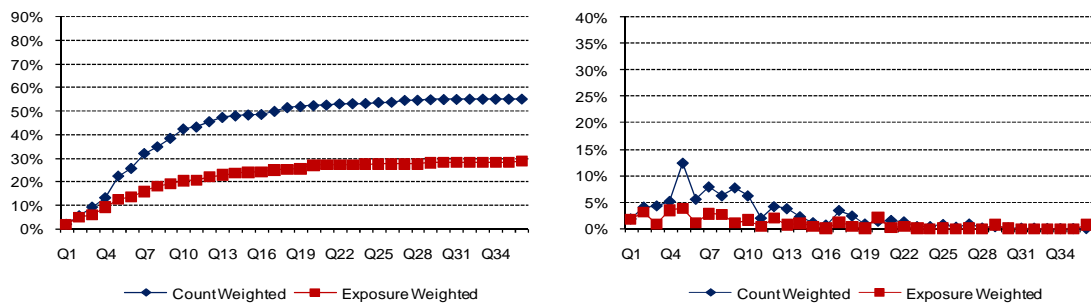


Figure 7 – Average cumulative and marginal recovery rate for closed files with longer than 1 year recovery period

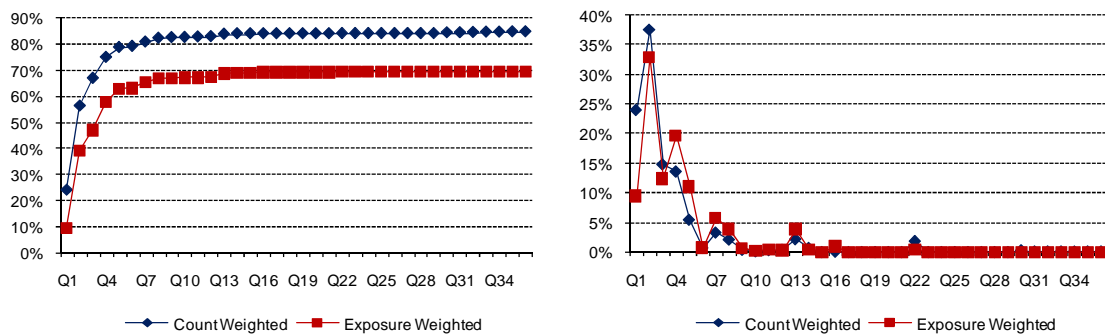


Figure 8 – Average cumulative and marginal recovery rate for closed files with shorter than 1 year recovery period

Model	Sample	Exposure at default - EAD	Collateral class A as % of EAD	Collateral class B as % of EAD	Collateral class C as % of EAD	Collateral class D as % of EAD	Age of a counterparty	Length of business relationship	Number of different collateral classes	Year of default before 1995	Year of loan origination before 1995	New industries	Industry not specified	Number of loans	Investment type of loan	Overdraft type of loan	Revolving type of loan	Purpose type of loan	
Linear model	A	-0.230	0.411		0.395			-0.197	0.079		-0.283								
Fractional response Logit link	A		2.552		2.565			-1.225			-1.622								
Fractional response Log-log link	A		1.802		1.599						-1.032								
Fractional response Complementary Log-log link	A		1.607		1.679			-0.939			-1.254								
Fractional response Beta - Logit Link	A	-1.426			0.963						-1.364		-0.725						
Fractional response Beta - Log-log link	A	-0.730			0.716						-0.797		-0.424						
Fractional response Beta - Complementary Log-log link	A	-1.230									-1.121		-0.611						
Ordinal response Logit link	A	-2.500	2.799		2.338			-1.208	0.581		-1.769								
Ordinal response Complementary Log-log link	A	-1.648	1.329		1.382		0.724	-0.943	0.367		-0.980		-0.507						
Linear model	B	-2.250							0.211		-0.369			-0.097	0.154	0.125			0.143
Fractional response Logit link	B	-27.240							2.149										
Fractional response Log-log link	B	-15.950							1.589										
Fractional response Complementary Log-log link	B	-13.096							0.936										
Fractional response Beta - Logit Link	B	-24.382		1.854	-2.227		1.329	1.900							2.984	1.718	0.909		1.443
Fractional response Beta - Log-log link	B	-20.540		1.766	-2.318		1.074	2.014				0.814			2.418	1.702	1.099		1.527
Fractional response Beta - Complementary Log-log link	B	-9.435		0.456											0.581				
Linear model	A+B	-0.330	0.359		0.329			-0.179	0.103		-0.298					0.186			
Fractional response Logit link	A+B	-2.873							0.666		-1.567					1.008			
Fractional response Log-log link	A+B	-1.128	1.491		1.612						-1.128					0.825			
Fractional response Complementary Log-log link	A+B	-2.254							0.471		-1.247					0.591			
Fractional response Beta - Logit Link	A+B	-1.946							0.311		-1.390		-0.695			0.845			
Fractional response Beta - Log-log link	A+B	-0.989							0.191		-0.830		-0.445			0.636			
Fractional response Beta - Complementary Log-log link	A+B	-1.504							0.237		-1.083		-0.454			0.500			
Ordinal response Logit link	A+B	-3.471	2.242		1.802		1.202	-1.348	0.652		-1.796		-0.811			1.133			
Ordinal response Complementary Log-log link	A+B	-2.218	1.144		1.050		0.725	-0.833	0.437		-1.002		-0.469			0.632			

Table 12 – Summary of significant determinants for all models and samples

Model	Bad = above 75 percentile LGD					
	Subsample A (> 1 year)		Subsample B (> 1 year)		Whole Sample	
	Power Statistic	SE	Power Statistic	SE	Power Statistic	SE
Linear model	52.0%	22.5%	47.4%	23.6%	58.5%	26.3%
Fractional response Logit link	49.2%	21.8%	45.4%	22.6%	52.9%	24.9%
Fractional response Log-log link	48.5%	21.3%	46.1%	20.1%	55.6%	25.0%
Fractional response Complementary Log-log link	47.6%	21.1%	45.4%	20.7%	52.6%	25.0%
Fractional response Beta - Logit Link	45.8%	16.8%	42.8%	20.3%	53.3%	20.9%
Fractional response Beta - Log-log link	48.0%	15.6%	51.4%	21.2%	54.3%	19.5%
Fractional response Beta - Complementary Log-log link	41.0%	16.7%	47.5%	14.9%	53.7%	21.0%

Table 13 – Binary power statistic (Accuracy ratio) of models, bad – above 75 percentile LGD

Model	Bad = above 25 percentile LGD					
	Subsample A (> 1 year)		Subsample B (> 1 year)		Whole Sample	
	Power Statistic	SE	Power Statistic	SE	Power Statistic	SE
Linear model	63.4%	22.5%	9.7%	23.4%	69.7%	26.3%
Fractional response Logit link	60.6%	21.8%	4.7%	22.5%	64.0%	24.9%
Fractional response Log-log link	59.9%	21.3%	5.4%	20.0%	64.5%	25.0%
Fractional response Complementary Log-log link	59.0%	21.1%	4.7%	20.6%	64.1%	25.0%
Fractional response Beta - Logit Link	55.9%	16.8%	39.2%	20.2%	65.5%	20.9%
Fractional response Beta - Log-log link	62.5%	15.6%	55.5%	21.0%	69.0%	19.5%
Fractional response Beta - Complementary Log-log link	50.3%	16.7%	22.2%	14.8%	65.6%	21.0%

Table 14 – Binary power statistic (Accuracy ratio) of models, bad – above 25% percentile LGD

IES Working Paper Series

2008

1. Irena Jindrichovska, Pavel Körner : *Determinants of corporate financing decisions: a survey evidence from Czech firms*
2. Petr Jakubík, Jaroslav Heřmánek : *Stress testing of the Czech banking sector*
3. Adam Geršl : *Performance and financing of the corporate sector: the role of foreign direct investment*
4. Jiří Witzany : *Valuation of Convexity Related Derivatives*
5. Tomáš Richter : *Použití (mikro)ekonomické metodologie při tvorbě a interpretaci soukromého práva*
6. František Turnovec : *Duality of Power in the European Parliament*
7. Natalie Svarciva, Petr Svarc : *Technology adoption and herding behavior in complex social networks*
8. Tomáš Havránek, Zuzana Iršová : *Intra-Industry Spillovers from Inward FDI: A Meta-Regression Analysis*
9. Libor Dušek, Juraj Kopecsni : *Policy Risk in Action: Pension Reforms and Social Security Wealth in Hungary, Czech Republic, and Slovakia*
10. Alexandr Kuchynka : *Volatility extraction using the Kalman filter*
11. Petr Kadeřábek, Aleš Slabý, Josef Vodička : *Stress Testing of Probability of Default of Individuals*
12. Karel Janda : *Which Government Interventions Are Good in Alleviating Credit Market Failures?*
13. Pavel Štika : *Možnosti analytického uchopení reciprocity v sociálních interakcích*
14. Michal Bauer, Julie Chytilová: *A Model of Human Capital, Time Discounting and Economic Growth*
15. Milan Rippel, Petr Teplý : *Operational Risk – Scenario Analysis*
16. Martin Gregor : *The Strategic Euro Laggards*
17. Radovan Chalupka, Petr Teplý : *Operational Risk Management and Implications for Bank's Economic Capital – a Case Study*
18. Vít Bubák : *Value-at-Risk on Central and Eastern European Stock Markets: An Empirical Investigation Using GARCH Models*
19. Petr Jakubík, Petr Teplý : *The Prediction of Corporate Bankruptcy and Czech Economy's Financial Stability through Logit Analysis*

20. Elisa Gaelotti : *Do domestic firms benefit from geographic proximity with FDI? Evidence from the privatization of the Czech glass industry*
21. Roman Horváth, Marek Rusnák : *How Important Are Foreign Shocks in Small Open Economy? The Case of Slovakia*
22. Ondřej Schneider : *Voting in the European Union - Central Europe's lost voice*
23. Fabricio Coricelli, Roman Horváth : *Price Setting and Market Structure: An Empirical Analysis of Micro Data*
24. Roman Horváth, Kamila Koprnická : *Inflation Differentials in EU New Member States: An Empirical Evidence*
25. Michal Franta, Branislav Saxa, Kateřina Šmídková : *Inflation Persistence: Is It Similar in the New EU Member States and the Euro Area Members?*
26. Jakub Seidler : *Implied Market Loss Given Default: structural-model approach*
27. Radovan Chalupka, Juraj Kopecsni : *Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study*

All papers can be downloaded at: <http://ies.fsv.cuni.cz>

