# Reflection on Recovery and Loss Given Default. What Is and What Is Amiss.

Markus Höchstötter[a,1], Abdolreza Nazemi[a,2], Svetlozar Rachev[a,b,3]

[a]*School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany*
[b]*Department of Applied Mathematics at Stony Brook University, New York, USA.*

**Abstract**

As of now, there exist a vast variety of approaches quantifying the recovery of defaulted debt or, alternatively, the loss-given default (LGD). However, to our knowledge, literature is still short of a contribution providing a comprehensive presentation of the different definitions of recovery and a comparative summary of the related models. For this reason, with this article, we endeavor to give a concise yet sufficient characterization of debt in general, a subdivision into corporate and personal debt, as well as a further distinction of the latter with respect to bank loans and non-bank loans. We claim to provide a comprehensive account of the literature on recovery rates and LGD focusing on the respective models, in particular.

*Key words:*
Loss-given-default, recovery rate, GLM, beta distribution, beta kernels, SVM.

*Email addresses:* (Markus Höchstötter), (Abdolreza Nazemi), (Svetlozar Rachev)

[1]Markus Höchstötter is a post doctoral researcher at the Chair of Statistics, Econometrics and Mathematical Finance at the School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany.

[2]Abdolreza Nazemi is a doctorate student at the Chair of Statistics, Econometrics and Mathematical Finance at the School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany.

[3]Svetlozar Rachev is Professor in the Department of Applied Mathematics at Stony Brook University in New York and at the Chair of Statistics, Econometrics and Mathematical Finance at the at the School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany.

*September 21, 2011*

## 1. Introduction

The latest crisis, the Subprime crisis of 2007, and its devastating effects can still be felt not only in the financial industry but essentially across all industries worldwide. On Wall Street as well as in Europe, some financial power houses fell victim to it lethally while others came to be in need of injections from their respective governments. For example, in 2008, Lehman Brothers collapsed under its debt load of 144 billion dollars while the German Sachsen LB needed resuscitation in form of close to three billion euros in 2007 after it failed to refinance maturing short-term debt of its subsidiary in Dublin. A good account of the collective damage can be found in Standard&Poor's (2011). Between 2007 and 2011, a total of 496 rated mostly US institutions have defaulted representing over one trillion in debt outstanding and dwarfing anything seen so far. The corresponding numbers from the entire time between 1981 and 2006 are only 1519 defaults and 622 million dollars in debt outstanding, respectively. Moreover, due to the current crisis, even entire countries found themselves on the brink of bankruptcy which still provides for daily headlines.

But the impact has not only been felt across the corporate and sovereign world but also on the consumer side as many employees were laid off by tilting or at least insecure employers. Almost like an overreaction, these gloomy conditions led banks to suppress lending on a large scale which resulted in the credit crunch. Their behavior evoked fear that the current problems might be on a par with those of the world financial crisis following 1929 leading to the great depression. Bailing out large corporations as well as enforcing the rights of individual borrowers, and imposing the most intrusive government regulation on the financial industry since those days were the paths followed by the governments even though sometimes considered highly contentious by market participants and economists.

Question may arise as to why all this happened. But maybe more important than the detection of one potential cause of the crisis is to design devises that could prevent such agony from spreading further or even impede any repetition of it in the future. The fear of the latter may be founded in light of the seemingly increasing occurrence of financial crises as brought to attention by Stiglitz (1998) already one decade ago. But in addition to crises that usually catch the majority of the world off-guard, another peril that mostly goes unnoticed is given by the increasing amount of debt potentially doing harm in a more continuous way and, in the worst case, resulting in evermore powerful shocks.

In the USA alone, public sector as well as corporate and consumer debt have reached dizzying levels. According to the Board of Governors of the Federal Reserve System (2011), the public sector debt amounts to over 14 trillion dollars while US corporates and privates have both accumulated similarly shocking amounts. This trend is by no means unique to the USA, however. Thomas (2009) presents equivalent tendencies for Europe, especially in the private sector. Hence, if this trend is not reversed, the need for the expansion of lending will remain a pressing issue. To guarantee a functioning lending system, it is imperative to reign in the possible damage that can result from a borrower's default on repaying the debt in the predetermined manner. For this, strong and capable methods have to be introduced and reinforced to direct the lender's attention to the inherent risks of his debt positions.

Even though not mandatory for all lenders, the terminology and definitions given in the Basel II accords are used widely in the context of credit risk.[4] In particular, the definitions of loss, expected loss, the loss given default conditional on default as well as recovery rate, and related loss thresholds such as the value-at-risk (VaR) are applied when financial institutions that do not necessarily qualify as banks invest in debt.[5]

The various literature on defaulted debt that we will cover in this paper and the respective references not only emphasize the difficulties arising from diverging definitions such as, for example, that of the recovery rate, in some cases, terms are not clearly defined, at all. For example, ultimate recovery is a very misleading quantity since the time-interval for the related work-out process is commonly selected arbitrarily rendering its interpretation more ambiguous. It is also revealed that the performance of the members of an ever growing population of predominantly quantitative methods for estimation and prediction is sensitive to the type of debt. Most of all, the methodical side suffers from a lack of freely accessible data to produce reliable results of general value.

The sequel of the paper is organized as follows. Section 2 is devoted to the introduction of the definitions of default, recovery rate, and loss given default as well as related terms. The different methods to estimate the respective quantities will be presented in section 3. A literature review on the results of analyses sorted by type of debt will follow in section 4. And finally, a summary will conclude in section 5 pointing out possible directions for further research.

## 2. Definitions

As stated in Thomas (2009), a financial institution generally offers five types of lending or, in other words, there exist five types of financial debt. Accordingly, financial debt can be coarsely classified as sovereign, corporate, retail such as consumer lending, bank, or equity. As can be found in BCBS (2005), the Basel accords set regulatory standards for financial institutions in handling the risks inherent in the different types of lending by demanding an adequate capital base. For the quantification of the risks, the Basel II framework introduces parameters related

[4] For details concerning the regulatory framework, the reader is kindly referred to BCBS (2005) and BCBS (2011).

[5] Before the Basel accords had been established, lenders in the retail sector resorted to a widely used tool, the credit scorecard. This device helps to assess the probability of a new customer to default. The actual score, usually a three digit number computed from the set of consumer characteristics, separates between potentially good and bad borrowers. While is has proven to be reliable for the purpose it was designed for, namely the assessment of the creditworthiness of the borrower before the inception of the relationship, it has failed to produce good results for the loss prediction once the borrower has defaulted. As reference, we recommend Hand (2001). However, we will not grant it any particular attention here.

to the default in the lending process and the resulting loss hereafter. It is the objective of this section to present these parameters in the context of their definitions.

### 2.1. Credit Risk and Capital Requirements

The most important parameters determined by the Basel II accords are the exposure at default, the loss-given-default, the probability of default and as parameters of location of the loss distribution, the expected loss as well as the value-at-risk. Crook et al. (2007) provide a good overview. The use of such parameters for the credit risk capital requirements requires knowledge of their values. As mentioned in Bellotti (2010), the Basel II accords allow banks to estimate them according to two approaches. The first is the standard approach and the second is the internal ratings based approach (IRB). In the standard approach, the estimates rely on ratings generated by external agencies. The IRB, on the other hand, permits banks to implement their own internal risk models and estimation methods.

### 2.2. Exposure at Default

While default itself is not uniquely defined in the International Convergence of Capital Measurement and Capital Standards, the regulation introduced by BCBS (2005) requires a bank to refer to a reference definition of default constituted for its internal use. For example, Standard&Poor's (2011) set default as taking place "on the first occurrence of a payment default on any financial obligation rated or unrated, other than a financial obligation subject to a bona fide commercial dispute." In general, it can be considered either as the worsening of the borrower's conditions such that he will most likely be unable to meet his obligations or as some sort of delay in the scheduled repayment process beyond some threshold on the account of the borrower.[6] The exposure at default (EAD) then is the remained of the original debt that is still owed by the borrower. The complexity of its calculation can be very much dependent on the type of debt.

Altman et al. (2005a) add that the considered exposure is significantly influenced by financial collateral. Two approaches have to be considered in this context, the simple and the comprehensive one. The simple approach allows the complete acknowledgement of collateral for the reduction of exposure while the comprehensive approach diminishes the reduction by some haircut.

### 2.3. Loss-given-Default and Recovery Rate

As stated in Altman et al. (2005a) and Altman et al. (2005b), the loss-given-default (LGD) is generally the outstanding amount owed after default has been recorded and consequently turns into the credit that is lost by a financial institution when a borrower defaults. Hence, it is close to its general understanding by researchers and practitioners. According to the regulatory framework, it has to be measured as a percentage of the EAD. The recovery rate (RR) is defined as its

[6] Often in the context of credit risk, the term obligor is used in lieu of borrower.

complement with respect to the full, i.e., one minus LGD. As stated in Peter (2006), this parameter has many different definitions resulting from the variety of definitions of EAD on the one hand and on different options of inclusion of recoveries on the other hand. Moreover, the type of seniority of the credit plays an important role in the determination of LGD.

Any amount that counts as recovery reduces the loss. This includes proceeds from facility or collateral sale. Guarantees, assets from the borrower, and restructured or cured exposures can also be considered. The regulation explicitly demands including material direct and indirect costs arising from the handling of defaulted exposure excluding internal costs as stated in Peter (2006). Generally, there are two different alternatives to define the recovery process. The first is the market approach where immediate recovery is achieved by the sale of the non-performing debt on a secondary market. The second is the ultimate recovery resulting from a longer lasting workout process. Engelmann and Rauhmeier (2006) distinguish between as many as three different recovery approaches by including the market implied method that prices the defaulted debt according to comparable debt on the market usually up to one month after default. Porto (2011) even categorize recovery processes into four groups, i.e., two market-based groups and two based on internal data. The latter two comprise the realized recoveries obtained after a completed workout process and the implied historical recoveries (or, equivalently, LGD) based on internal estimates, respectively.[7]

As Peter (2006) points out, the LGD has to be based on economic loss incurred by entity. In addition to the borrower related characteristics, it needs to reflect macro-economic conditions such as GDP, unemployment rate etc. In the IRB, LGD estimates are acceptable for retail exposure only while in the advanced internal rating based approach (IRBA), estimates can also be used for corporate, sovereign, and bank exposures. The LGD estimation follows subjective and objective methods, while objective methods can be further subdivided into explicit or implicit. For example, the market LGD approach is explicit. Financial institutions subject to the Basel regulation are held to perform regular comparisons between realized and estimated LGD to detect deviations and correspond appropriately. While the LGD parameter is an elementary component of the Basel regulation it is of interest for banks also in accounting and internal risk reporting.

### *2.4. Probability of Default*

Though not the focus of our paper, this as well as the following parameters is stated for the sake of completeness. In the regulation, the probability of default (PD) is defined as the "likelihood that a loan will not be repaid and fall into default. This PD will be calculated for each company who has a loan." Furthermore, the bank has to consider the credit history of the counterparty. Also, any characteristic of the investment has to be taken into account for the

[7]Porto (2011) show that both realized as well as implied historical recovery processes lead to the same figures for a portfolio of debt if the average EAD of the defaulted obligors equals that of the non-defaulted obligors.

calculation of the PD. General economic conditions will also influence the likelihood of default. It has been found that LGD is positively correlated with PD as stated in, for example, Bellotti (2010).

### *2.5. Parameters of Location of the Loss Distribution*

According to the Basel regulation, the expected loss (EL) from a default is composed of PD, LGD, and EAD in the following way

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

Bellotti (2010) mentions four risk elements in the computation of EL. One is the uncertainty at the account level. That is the ineptitude of any model to predict default for each individual account. The others are model estimation uncertainty, uncertainty with respect to the distribution of EAD, and the uncertainty arising from ignorance of the exact distribution of LGD exacerbated by its correlation with PD. The key parameters that determine the credit risk of financial assets, i.e., PD, LGD, and EAD yielding the EL have to be observed over a one-year horizon.

The value-at-risk (VaR) quantifies a threshold loss that will not be exceeded at a given confidence level defined as 99 percent. So, in calculating the value-at-risk, the Basel accords request the 99th percentile, one-tailed confidence interval of the loss distribution.

An efficient allocation of regulatory and economic capital calls for accurate estimates of these parameters. Moreover, this will also be paramount for the pricing of the credit risk in the context of debt instruments and credit derivatives.

## 3. Methods

In this section, we present literature using different parametric as well as non-parametric models and methods with applications in modeling LGD or, equivalently, recovery rate for various types of debt. In our opinion, an excellent overview of this topic is provided by Schuermann (2004). We add to this by including more recent advances and setting the focus on the methodology and modeling aspects. A summary of the methods in this section can be found in Table 1.

### *3.1. Regression*

We begin with the literature on data mining methods that we subsumed under the term regression. Bastos (2010a) uses fractional response on homogenous subsamples of the data. That is, the response variable recovery rate $r \in [0, 1]$ is regressed on the data vector $x \in R^d$ through some link function $E[r|x] = G(x^T\beta)$. The author resorted to the log-log link function

$$G(x^T\beta) = \exp(-\exp(-x^T\beta)) \tag{1}$$

The coefficient parameters $\beta$ are estimated through a Bernoulli quasi-maximum log-likelihood process

$$l(\beta) = \sum_{i=1}^{n} r_i \ln G(x_i^T\beta) + (1 - r_i) \ln\left(1 - G(x_i^T\beta)\right)$$

This procedure yields mixed results with respect to accuracy over different time horizons compared to alternatives such as the regression tree which successively splits the data into groups of nearly homogenous recovery rates based on some impurity measure $i$. More specifically, at each node $t$, the optimal split $s$ leads to the maximum decrease in impurity. That is, the objective is

$$\max_{s} \quad \Delta i(s,t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R) \tag{2}$$

where $p_L$ and $p_R$ denote the percentage of observations of node $t$ that are assigned to its child notes $t_L$ and $t_R$, respectively. In Bellotti and Crook (2008), for modeling LGD, ordinary least squares (OLS) regression $y = x^T\beta$ on covariates $x$ is applied as well as decision trees and Tobit regression. The latter treats the LGD as a censoring transformation

$$y = \begin{cases} x^T\beta + u & x^T\beta > 0 \\ 0 & x^T\beta \leq 0 \end{cases}$$

of the linear model $x^T\beta$ superimposed by some normally distributed noise $u$. It is found that OLS is the better choice over the alternatives. In their regression model, PD enters as significant covariate due to the joint dependence on macro-economic conditions. The data are then transformed through fractional logit

$$G(x^T\beta) = \exp(x^T\beta)), \tag{3}$$

and probit

$$G(x^T\beta) = \Phi(x^T\beta)), \tag{4}$$

where $\Phi$ denotes the standard normal cumulative distribution function, to finally transform the such obtained $G(x^T\beta)$ into a beta distributed LGD through quantile matching.[8] Jacobs and Karagozoglu (2011) basically follow this approach but instead use a mixture of beta distributions as link function. Bellotti and Crook (2011) model LGD based on OLS with macro variables additionally added. The resulting transform is performed through link functions such as in (1) and (3), beta quantile matching, and probit functions (4), as before. Chen and Chen (2010) use simple logistic regression for LGD while Dermine and Neto De Carvalho (2006) perform OLS regression also including macro variables and transforming through a logistic link to obtain a quantity in the LGD scale, i.e. bounded within $[0, 1]$. This is also found in Grunert and Weber (2009) for the modeling of the recovery rate. Qi and Zhao (2011) model LGD based on logistic and inverse-Gaussian regression followed by a beta transformation.[9] An interesting early work is provided by Livingstone and Lunt (1992). The authors consider in detail several social, economic and psychological factors related to debt. Linear discriminant function analysis is used to discriminate debtors from non-debtors using significant person-related covariates $x$. That is, a combination of the linear thresholds for each covariate classifies the person. In an additional multiple regression analysis based on these variables, it is analyzed how far people might get into debt and how much of their debts people repay.

### 3.2. *Distributional*

The methods collectively presented in this subsection provide either parametric distributions or related non-parametric approaches to model the distribution of LGD and the recovery rate.[10] Gupton and Stein (2002) state that beta LGD or recovery rates have a density that theoretically should be best described by a beta distribution since the support is limited to the interval between zero and one with various shapes governed by two parameters. For some $y \in [0, 1]$, the density function of the beta distribution is given by

$$f(y; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (1 - y)^{\beta - 1} \, y^{\alpha - 1}$$

where $\Gamma(\cdot)$ denotes the gamma function and the parameters $\alpha, \beta > 0$.[11] They apply to the distribution a correction factor, however, to allow for values slightly above 1.0 which are sometimes

[8] For the density function of the beta distribution, we refer the reader to subsection 3.2.

[9] The inverse Gaussian is a right-skewed distribution with support on the positive real line and density function $f(y; \delta, \nu) = (2\pi y^3 \nu)^{-0.5} \exp(-(1 - \delta y)^2/(2\nu y))$. In inverse-Gaussian regression, the linear term $x^T\beta$ enters the distribution as drift parameter $\delta$. Covariates $x$ and coefficients $\beta$ have to be such that $\delta > 0$.

[10] In the context of the Basel accords, the recovery rate is merely important for the computation of the appropriate capitalization of a bank. However, for the secondary market, a reliable prediction of the recovery rate is essential in pricing models of loans. Kaneko and Nakagawa (2008) apply a dynamic stochastic model to Japanese bank loans.

[11] Depending on the parameter values, the density can be symmetric with horizontal, U-shaped, or cup-shaped graph, or asymmetric.

observed in the context of bonds and also loans depending on the definition of recovery. Since this parametric approach appeared soon too limited, Chen (1999) introduced a beta kernel estimator much of the same spirit as the well-known Gaussian kernel. Given $n$ observations $Y_i$ within $[0, 1]$, its design is given by

$$\hat{f}_n(y, b) = \frac{\sum_{i=1}^{n} K_{\left(\frac{y}{b}+1, \frac{1-y}{b}+1\right)}(Y_i)}{n} \tag{5}$$

for some $y \in [0, 1]$ and parameter $b$ responsible for smoothing. The kernel function $K_{(c,d)}(\cdot)$ is given by the beta density function $f(\cdot; c, d)$. However, the observed sample data do not enter the beta density functions as parameter values as would have been the case with the Gaussian kernel, but instead enter as arguments. This was also done by Renault and Scaillet (2004). But the kernel density estimator (5) does not converge uniformly on $[0, 1]$ to the true density function because of the boundary problems at 0 and 1 due to its design. As modification of this estimator, Gourieroux and Monfort (2006) introduced new beta kernels which they referred to as *macro* and *micro* density estimators. The first rescales the original estimator (5) by the estimated total mass, i.e.,

$$\hat{f}_n^1(y, b) = \frac{\hat{f}_n(y, b)}{\int_0^1 \hat{f}_n(y, b) dy}$$

while the latter rescales at each observation according to

$$\hat{f}_n^{(1)}(y, b) = \frac{1}{n} \sum_{i=1}^{n} \frac{K\left(Y_i, y/b + 1, (1 - y)/b + 1\right)}{\int_0^1 K\left(Y_i, y/b + 1, (1 - y)/b + 1\right)}$$

As pointed out by Calabrese (2010), to truly copy the behavior of recovery rates, one has to model based on a discrete-continuous hybrid distribution where the continuous part $(0, 1)$ is given by a beta mixture and point mass is assigned to the values 0 and 1, respectively. For estimation purposes, the beta distributions are reparameterized using as new parameters the mean and the variance as expressions of the original parameters, i.e., $\mu = \alpha/(\alpha+\beta)$ and $\sigma^2 = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)] = \alpha(1-\alpha)/(\phi+1)$, respectively. So, which had been neglected before, parameters can be estimated jointly. In the literature prior to that, the second parameter, $\phi = \alpha+\beta$, interpreted as a nuisance parameter was not considered essential enough to be estimated. Finally, Calabrese and Zenga (2010) in addition to the already presented discrete-continuous mixed distribution of the recovery rate, introduced an alternative beta kernel estimator for the continuous part that had the observed sample data enter as parameter values in the fashion of the Gaussian kernel estimator so that it overcomes the boundary problem of the original beta kernel estimator that failed to represent probability mass one when integrated.

$$\hat{f}_M(y, b) = \frac{\sum_{i=1}^{n} K_{\left(\frac{Y_i}{b}+1, \frac{1-Y_i}{b}+1\right)}(y)}{n}$$

*3.3. Alternative Methods*

As the last category of modeling techniques, we present a collection of different approaches that have not found wide-spread use in contrast to the ones listed in the two subsections before. Qi and Zhao (2011) introduce neural networks as a non-linear approach to model LGD. This is also done by Bastos (2010b) for the recovery rate since neural networks supported by bootstrap display superior predictive performance over parametric regression models they are tested against. The design of the neural network is especially appealing because of its several layers of perceptrons. Common to any design are a input layer, one or more hidden layers of neurons, and an output layer. In the simplest version of only one hidden layer, input data consisting of observations $x_j$ of $j = 1, 2, \ldots, d$ variables enters neuron $i$ of the hidden layer to be transformed there into a weighted functional output

$$h_i = f^{(1)}\left(b_i + \sum_{j=1}^{d} w_{i,j} x_j\right)$$

with weights $w_{i,j}$ and neuron-specific constant $b_i$. Output from all $n_h$ hidden neurons is then turned into network output

$$y = f^{(2)}\left(b^{(2)} + \sum_{i=1}^{n_h} v_i h_i\right)$$

with neuron weights $v_i$. The neural network allows for a flexible yet sometimes unintuitive design. This technique is particularly apt in separating samples with respect to objective functions such as, for example, zero or full recovery.

Hao et al. (2009) model recovery rates for homogenous classes obtained through stepwise application of support vector machines (SVM). The SVM are used to separate debtors into two categories ($y = -1$ or $y = 1$) based on some hyperplane threshold with perpendicular vector $w$ maximizing the minimal distance of each of the two groups from the threshold. With the optimal hyperplane, the training data keep a minimum distance of $b$ from the hyperplane to guarantee generality of the model. The optimization problem using all $n$ observations $(y_i, x_i)$, $x_i \in R^d$ is thus given

$$\min_{w,b} \|w\|_2^2, \qquad s.t.\ \ y_i(<w, x_i> + b) \geq 1,\ \ i = 1, 2, \ldots, n \tag{6}$$

or in the dual form

$$\max_{a} \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1,j}^{n} a_i a_j y_i y_j < x_i, x_j >, \qquad s.t.\ \ \sum_{i=1}^{n} a_i y_i = 0 \tag{7}$$

where $< \cdot, \cdot >$ denotes the inner product. The separating rule is then given by $f(x) = sign(<w, x> + b)$ or, equivalently, $f(x) = sign(\sum_{i=1}^{n} a_i y_i < x_i, x > + b)$. A problem occurs if the data are not linearly separable as required by (6) and (7). To this end, the original data vector $x \in R^d$ is mapped into a higher dimensional ($k > d$) feature space with a non-linear function $\phi : R^d \to R^k, x \mapsto \phi(x)$. To circumvent the calculations of the inner products and associated dot products in the higher dimension, the so called kernel-trick is applied, requiring computation of $k(x_i, x_j) =< \phi(x_i), \phi(x_j) >$ for the dot products. Thus, the transformation into the dimension can be actually avoided. The resulting separating function is now $f(x) = sign(\sum_{i=1}^{n} a_i y_i k(x_i, x) + b)$. Common kernel functions are, for example, polynomial $k(x_i, x) =< x_i, x >^p$ or radial basis $k(x_i, x) = \exp(-\|x_i - x\|^2/c)$. The authors state that the advantages are given by the use of key observations only for the sake of speed, the translation of the discrimination problem into a quadratic problem, and the projection of the original problem onto a higher dimensional space to apply a linear discrimination function facilitating the original problem. They begin the modeling with a stepwise selection process of the variables most powerful to separate the data set into homogenous subsets. Loterman et al. (2011) apply for the modeling of LGD non-linear techniques such as Classification And Regression Trees (CART), Multivariate Adaptive Regression Splines (MARS), Least Squares Support Vector Machines (LSSVM), and Artificial Neural Networks (ANN) since their performance, respectively, is proven to exceed that of linear models. CART is as described by (2). MARS approaches non-linearities by representing the dependent variable as a linear composition $y = \sum_{j=1}^{k} b_j B_k(x)$ of $k$ basis functions. Basis functions are added and discarded in a two-step procedure. LSSVM is a version of SVM to conduct a linear regression of the form $y = \phi(x)^t b + \epsilon$ with the original data $x$ mapped into a higher feature space by $\phi$ to obtain a higher degree of linearity. Using a kernel $K(x, x_i) = \phi(x)^T \phi(x_i)$ simplifies the optimization in the preferred dual form $y = \sum_{i=1}^{n} a_i \phi(x)^T \phi(x_i) + e$. Matuszyk et al. (2010) Introduce LGD modeling based on a decision tree using a weights-of-evidence (WOE) approach to determine the most significant characteristics for the prediction of high and low LGD. Here, the observations are coarsely classified into $K$ classes $C_1, C_2, \ldots, C_K$. For each class $C_i$, the log-ratio

$$w_i = \log\left(\frac{g_i/b_i}{n_G/n_B}\right)$$

is computed where $g_i$ and $b_i$ denote the number of goods and bads in class $C_i$, respectively, while $n_G$ and $n_B$ are the number of goods and bads, respectively, of the entire population. In this manner, the importance of each class with respect to predicting good or bad recovery is evaluated. This is basically repeated in Thomas et al. (2011) augmented by a beta or normal function transformation. Filho et al. (2010) express the effect of the respective collection processes in predicting LGD. In this context, they use text mining methods to detect steps in the collecting process that are most helpful for obtaining a higher recovery rate.

[TABLE 1 HERE]

## 4. Results by Borrower Type

In the following, we will list research coarsely sorted by the two types of borrowers of most interest, i.e., corporate or retail borrowers.

### *4.1. Bonds and Corporate Debt*

As stated in Schuermann (2004), seniority of debt has an enormous impact on the distribution of the recovery rate. While for more junior levels, recovery is predominantly low, this is generally not the case with senior debt. Also, senior recovery rates display on average bimodality in their distributions which is not often the case with junior debt where much of the probability mass is located in the upper end. We will not present results in such detail as suggest by Schuermann (2004) who suggests a significant impact on LGD from the industry type. Defined junior debt, Schuermann (2004) bonds are different from loans due to different control rights of the bondholders which manifest itself particularly during the default. Loans to corporations, however, are equipped with higher seniority than bonds and, hence, should result in higher recovery rates.

Jacobs and Karagozoglu (2011) find in their analysis of 871 corporate bankruptcies occurring between 1985 and 2008 that macroeconomic factors play as import a role as industry conditions, equity returns, and the price of tradable debt at default in addition to the debtor related characteristics. Felsovalyi and Hurt (1998) analyze the recovery on 1,149 Citibank loans issued to commercial industrial borrowers between 1970 and 1996. They report an average LGD of 31.8% or, equivalently, a recovery rate of roughly 68%. Altman (2008) studied the recovery process on 2,071 defaulted bonds from Moody's Default Risk data base. Dependent on the seniority level of the bonds and securing, recovery is found to be between roughly 30% and 62%, so, on average, lower than the previous figure for loans. For a different survey conducted during 1982-2002, the authors find that the recovery rate is strongly dependent on the debt type. They additionally report high variation across industries. Moreover, they mention that individual factors best explain recovery rates. Asarnow and Edwards (1995) analyze 831 corporate and industrial loans plus 89 structured loans between 1970 and 1993. The resulting average LGD is 34.8% for the corporates and industrials while it is 12.8% for the structured loans. Thus, structuring seems to have a very positive effect on recovery.

With respect to methodology and modeling, Qi and Zhao (2011) analyze 3,751 US corporate loans in the period 1985-2008. As expected they find that non-parametric models outperform parametric models. With respect to the distribution, they argue that it should not be imperative to force a bi-modal candidate on the empirical data since bi-modality does not always appear to be detected. Moreover, the regression tree that came to use in their analysis has very high predictive accuracy. Their results showed that fractional response is superior to the often praised OLS regression. In their analysis of 623 bonds from the Standard & Poor's/PMD database during 1981-1999, Renault and Scaillet (2004) find in Monte Carlo simulations that beta-kernel estimator outperforms any alternative non-parametric estimators most often used in the LGD context. Bastos (2010b) analyze 374 Portuguese loans to small and medium size enterprises in the period 1995-2000 from Dermine and Neto de Carvalho (2006). The out-of-sample predictions based on neural networks prove better than any parametric regression. For the same 374 Portuguese bank loans, Bastos (2010a) detect bimodality in the recovery rate distribution. They find that the predictive accuracy of the regression tree, and even the historic averages, is higher than for the typical linear regression models. Moreover, they report similar recovery rate distributions over different time horizons with mean recovery rates of between 50%-85%. In their study on 10,000 short-term loans to small and midsize companies in Portugal between 1995-2000, Dermine and Neto De Carvalho (2006) detect that almost all available customer related variables bear significant explanatory power. Böttger et al. (2008) conclude that corporate debt is mainly driven by the six factors: seniority, securitization, jurisdiction, industry, economic cycle, and expected liquidity of the secondary market for the debt type. The results of this subsection are summarized in Table 2.

[TABLE 2 HERE]

### *4.2. Consumer Debt*

In the sequel, we differentiate between bank loans in the narrow sense and any other retail credit even if their characters might appear similar. A summary of the findings of the following two subsections is provided by Table 3.

#### *4.2.1. Bank loans*

Calabrese (2010) analyze 149,378 Italian bank loans in distress between 1998-99. They conclude that the capitalized recovery amount significantly influences the subsequent recovery rate. As is often the case with personal loans, they also report a high concentration of recovery at zero and one. In their study, Grunert and Weber (2009) observe 120 German bank loans during the period 1992-2003. They report an average recovery rate of 72.5% and median of 91.8%. The beta distribution is not useful as is often the case. Instead, an uni-modal left-skewed distribution is deemed better. The inclusion of macro variables does not improve model quality. A negative correlation between recovery rate and the creditworthiness of borrower is apparent while EAD is a significant covariate in the regression of recovery rates. Caselli et al. (2008) analyze 11,649 Italian bank loans advanced to private persons as well as SMEs. They find that while macro-economic factors such as GDP growth, employment, and total annual production are important, for personal loans, the recovery rate hinges more on the loan-to-value-at-default ratio. In the next study, Avery et al. (2004) argue that situational circumstances matter immensely with respect to recovery. Finally, Livingstone and Lunt (1992) find that socio-demographic factors play a relatively minor role in personal debt and debt repayment. Disposable income does not differ between those in debt and not in debt, although it predicts how far people would be in debt and is most important in determining debt repayment. Attitudinal factors (being pro-credit rather than anti-debt, or seeing credit as useful but problematic) are found to be important predictors of debt and debt repayments. Further psychological factors, focusing on economic attributions, locus of control, coping strategies and consumer pleasure are found important. In China, Hao et al. (2009) investigate 1,115 loans with 131 variables from LossMetrics. Several loan-specific characteristics are significant for loan recovery discrimination. They report accuracies at 95.7% (in-sample) and 95.4% (out-of-sample), respectively. For a data set of 50,000 defaulted personal loans in the UK between 1984-2004, Matuszyk et al. (2010) detect as the five most significant characteristics as predictors for LGD the loan amount, the application score, the number of months in arrears (1) during the whole life and (2) during last 12 months, as well as the time until default. Loterman et al. (2011) have five bank loan data sets. They report for all popular models goodness-of-fit of $4\% < R^2 < 43\%$. Of the methods, SVM and non-linear neural networks have better predictive performance which, as they state, is contrary to results in PD modeling. And Zhang and Thomas (2010) perform analysis on 27,278 UK personal bank loans defaulted sometime between 1987-1999 and in recovery until 2003. They report an average recovery rate of 42% while the most significant OLS regression variable is EAD. According to goodness-of-fit measures, mixture models are not better than regular linear regression.

*4.2.2. Non-bank credit*

Bellotti (2010) study 50,000 Brazilian credit cards. Their most important finding is that the correlation of PD and LGD does not increase the VaR of their portfolio loss which is in stark contrast to common belief. Bellotti and Crook (2008) have UK credit card loans from four different institutions from 1998-2004 including a rich set of borrower related variables. They also come to the conclusion that the correlation of PD and LGD does not increase their VaR. Bellotti and Crook (2011) analyze 55,000 UK retail credit cards between 1995-2005. they find that bank interest rates and unemployment rates up to lag six significantly predict recovery. The inclusion of all application variables works best if any interaction terms are excluded, however. Of all different regression methods, the best-performing on is OLS. In their study of 1880 individual residential foreclosed mortgages from the period 1987-2007, Chen and Chen (2010) find that of the 14 variables considered, the property location is strongly correlated with social, demographic, economic factors and thus is relevant in the explanation of recovery. Qi and Yang (2009) have 241,293 US high-loan-to-value and insured mortgages under analysis stemming from the period 1990-2003. Here, $29.2\% < LGD < 31.7\%$. Moreover, LGD and the current loan-to-value (CLTV) as well as the initial loan-to-value (LTV) are positively correlated. LGD and loan size, however, are negatively correlated while LGD and age of loan are positively correlated. LGD in general depends on further loan characteristics. The goodness-of-fit is stated as $R^2 = 61.2\%$. CLTV is given as the single most important determinant of LGD and, hence, an LTV update should be included in LGD models regularly.

[TABLE 3 HERE]

## 5. Summary

In section 2, we gave the definitions of the terms default, recovery rate, and loss given default as generally agreed to by regulators, academics, and practitioners. In section 3, the models and methods for recovery rates and LGD were given. And the literature review on the results of analyses sorted by type of debt followed in section 4. We saw that sometimes, such as in the case of default or EAD, for example, unique definitions were not provided by the banking supervision and much leeway existed potentially making it difficult to compare the quantities from different institutions. Since a proper assessment of the expected loss and LGD, in particular, are imperative for adequate capital measures and also in the debt pricing process, reliable and highly standardized models have to be available. However, as we saw, reality is different. While linear regression seems to be the most favorite choice, it is not necessarily the soundest one. Other more complex methods generated good predictive results. However, their structure is not always accessible to intuition and easily interpretable economically. Non-parametric methods such as the kernel estimator are reasonable since they take into account the distribution of LGD or the recovery rate. However, in contrast to the data mining methods, none of them, to our knowledge, regard macro-economic variables as requested by the Basel accords. Also, we could not find a model considering the dynamics of the debtor repayment patterns which might represent a significant factor for ultimate recovery. Furthermore, the impact of the collection process on recovery should deserve more attention. For the future, we suggest to direct research into any of these areas.

| Author or authors | Models | Year |
|---|---|---|
| Bastos | Fractional response regression,Regression Tree | 2010a |
| Bastos | Fractional Regression, Neural Network | 2010b |
| Bellotti and Crook | models incorporating macroeconomic variables | 2011 |
| Bellotti and Crook | OLS regression, Decision Tree and Tobit regression | 2008 |
| Calabrese and Zenga | beta kernel, mixed random variable | 2010 |
| Calabrese | beta regression, Mixed random variable, distributional | 2010 |
| Caselli et al. | Regression, Multivariate analysis | 2008 |
| Chen and Chen | Logistic Regression | 2010 |
| Chen | Beta kernel estimators (it is not on LGD) | 1999 |
| Dermine and Neto De Carvalho | OLS regression, Logistic Regression | 2006 |
| Filho et al. | Optimization, Text Mining | 2010 |
| Gourieroux and Monfort | Beta Kernel | 2006 |
| Grunert and Weber | Distribution of RR, regression | 2009 |
| Gupton and Stein | Distributional | 2002 |
| Hao et al. | support vector machine, discriminant analysis | 2009 |
| Jacobs Jr. and Karagozoglu | beta-link generalized linear model | 2011 |
| Loterman et al. | OLS, Ridge regression, Beta regression, logistic regression, CART, MARS, LSSVM, and ANN | 2011 |
| Qi and Yang | Regression | 2009 |
| Qi and Zhao | Regression tree, Neural network, Fractional response regression, Inverse Gaussian regression | 2011 |
| Renault and Scaillet | Kernel estimation, nonparametric estimators, Monte Carlo | 2004 |
| Thomas et al. | Box-Cox, linear Regression, Beta Distribution, Log Normal Transformation | 2009 |
| Thomas et al. | Modelling LGD for unsecured personal loans: Decision tree approach | |
| Yeh and Lien | Data Mining Techniques in PD | 2009 |
| Zhang and Thomas | Linear regression, Survival analysis, Mixture distribution | 2010 |

Table 1: Summary by model.

| Author or authors | Data | Sample Size | Sample Period | Mean of RR | Median of RR | Country | Year |
|---|---|---|---|---|---|---|---|
| Asarnow and Edwards | Bank | 89 | 1970-1993 | 0.873 | - | US | 1995 |
| Asarnow and Edwards | C&I loans | 831 | 1970-1993 | 0.652 | - | US | 1995 |
| Bastos | SMEs | 374 | Jun. 1995-Dec. 2000 | 0.694 | 0.946 | Portugal | 2010a |
| Caselli et al. | SME | 11,649 | 1990-2004 | 0.540 | 0.560 | Italy | 2008 |
| Caselli et al. | SMEs | 1,814 | Jan. 1990-Aug. 2004 | 0.54 | 0.63 | Italy | 2008 |
| Caselli et al. | SMEs | 1,925 | Jan. 1990-Aug. 2004 | 0.50 | 0.39 | Italy | 2008 |
| Caselli et al. | SMEs | 2,169 | Jan. 1990-Aug. 2004 | 0.53 | 0.56 | Italy | 2008 |
| Caselli et al. | SMEs | 2,423 | Jan. 1990-Aug. 2004 | 0.54 | 0.47 | Italy | 2008 |
| Caselli et al. | SMEs | 3,318 | Jan. 1990-Aug. 2004 | 0.58 | 0.64 | Italy | 2008 |
| Dermine and Neto de Carvalho | SMEs | 10,000 | Jun. 1995-Dec. 2000 | 0.71 | 0.95 | Portugal | 2006 |
| Felsovalyi and Hurt | Citibank Loans | 1,149 | 1970-1996 | 0.68 | - | LA | 1998 |
| Grunert and Weber | SME | 120 | 1992-2003 | 0.725 | 0.918 | Germany | 2009 |
| Jacobs Jr. and Karagozoglu | US Corporate | 3,902 | 1985-2008 | 0.6104 | 0.6841 | US | 2011 |
| Jones and Hensher (Altman) | Bank Loans | 1,324 | 1988-2006 | 0.772 | - | US | 2008 |
| Jones and Hensher (Altman) | Bonds | 2,071 | 1988-2006 | 0.30-0.62 | - | US | 2008 |
| Qi and Zhao | US Corporate | 3,751 | 1985-2008 | 0.4423 | 0.4529 | US | 2011 |
| Renault and Scaillet | Standard & Poor’s/PMD | 623 | 1981-1999 | 0.4215 | - | US | 2004 |
| Schuermann | Bonds | 282 | 1970-2003 | 0.4952 | 0.4475 | US | 2006 |

Table 2: Summary by debt type (corporate).

| Author or authors | Data | Sample Size | Sample Period | Mean of RR | Median of RR | Country | Year |
|---|---|---|---|---|---|---|---|
| Bellotti | Credit Card | 50,000 | 2003-2004 | - | - | Brazil | 2010 |
| Bellotti and Crook | Credit Card | 55,500 | 1998-2004 | - | - | UK | 2008 |
| Bellotti and Crook | Credit Card | 55,000 | 1999-2005 | - | - | UK | 2011 |
| Calabrese | Personal loan | 149,378 | 1998-1999 | 0.384 | 0.340 | Italy | 2010 |
| Caselli et al. | Personal loan | 11,649 | 1990-2004 | 0.540 | 0.560 | Italy | 2008 |
| Chen and Chen | Mortgage Loan | 1,880 | 1987-2007 | - | - | Taiwan | 2010 |
| Hao et al. | lossMetric database | 1115 | - | - | - | China | 2009 |
| Livingstone and Lunt | | | | | | | 1992 |
| Loterman et al. | Credit Card | 7,889 | - | - | - | - | 2011 |
| Loterman et al. | Mortgage Loan | 119,211 | - | - | - | - | 2011 |
| Loterman et al. | Mortgage Loan | 3,351 | - | - | - | - | 2011 |
| Loterman et al. | Mortgage Loan | 4,097 | - | - | - | - | 2011 |
| Loterman et al. | Personal loan | 47,853 | - | - | - | - | 2011 |
| Qi and Yang | Mortgage Insurance | 241,293 | 1990-2003 | Max. 0.568 | - | US and other | 2009 |
| Schuermann | Bank loans | 151 | 1970-2003 | 0.631 | 0.655 | | 2006 |
| Thomas et al. | Personal loan | 50,000 | 1989 - 2004 | - | - | UK | 2010 |
| Zhang and Thomas | Personal Loan | 27,278 | 1987-1999 | 0.420 | - | UK | 2010 |

Table 3: Summary by debt type (consumer).

**References**


Altman, E. I. (2008). Default recovery rates and lgd in credit risk modelling and practice: An updated review of the literature and empirical evidence. In S. Jones and D. A. Hensher (Eds.), *Advances in credit risk modeling and corporate bankruptcy prediction*, Chapter 7, pp. 175–206. Cambridge: Cambridge University Press.

Altman, E. I., A. Resti, and A. Sironi (2005a). Loss given default; a review of the literature in recovery risk. In E. I. Altman, A. Resti, and A. Sironi (Eds.), *Recovery risk: The next challenge in credit risk management*, Chapter 1, pp. 41–59. Risk Books, London.

Altman, E. I., A. Resti, and A. Sironi (Eds.) (2005b). *Recovery risk: The next challenge in credit risk management*. Risk Books, London.

Asarnow, E. and D. Edwards (1995, Nov 23). Measuring loss on defaulted bank loans. A 24-year study. *Journal of Commercial Lending Vol. 77*(7).

Avery, R. B., P. U. Calem, and G. B. Canner (2004). Consumer credit scoring: Do situational circumstances matter. *Journal of Banking and Finance 28*, 835–856.

Bastos, J. (2010a). Forecasting bank loans loss-given default. *Jourrnal of Banking and Finance Vol. 34*(10), 2510–2517.

Bastos, J. (2010b). Predicting bank loan recovery rates with neural networks. Technical report.

BCBS (2005). International convergence of capital measurement and capital standards. a revised framework, bank for international settlements. http://www.bis.org.

BCBS (2011). Basel III: A global regulatory framework for more resilient banks and banking systems, bank for international settlements. http://www.bis.org.

Bellotti, T. (2010). A simulation study of basel ii expected loss distribution for a portfolio of credit cards. *Journal of Financial Services Marketing Vol. 14*, 268–277.

Bellotti, T. and J. Crook (2008). Modelling and predicting loss given default for credit cards. Technical report.

Bellotti, T. and J. Crook (2011). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting (forthcoming)*.

Board of Governors of the Federal Reserve System (2011). Federal reserve statistical release. http://www.federalreserve.gov.

Böttger, M., A. Guthoff, and T. Heidorn (2008). Loss given default modelle zur schätzung von recovery rates. *Frankfurt School - Working Paper Series* (96).

Calabrese, R. (2010). Regression for recovery rates with both continuous and discrete characteristics, Proccedings of the 45th Scientific Meeting of the Italian Statitistical Society (SIS), Italy.

Calabrese, R. and M. Zenga (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking and Finance Vol. 35*(5), 903–911.

Caselli, S., S. Gatti, and F. Querci (2008). The sensitivity of the loss given default rate of systematic risk: new empirical evidence on bank loans. *Journal of Financial Services Research 34*(1), 1–34.

Chen, S. (1999). Beta kernel estimators for density functions. *Computational Statistics and Data Analysis 31*, 131–145.

Chen, T. H. and C. W. Chen (2010, March). Application of data mining to the spatial heterogeneity of foreclosed mortgages. *Expert Systems with Application Vol. 37*(2), 993–997.

Crook, J., D. Edelman, and L. Thomas (2007). Recent developments in consumer credit risk assessment. *European Journal of Operations Research Vol. 183*(3), 1447–1465.

Dermine, J. and C. Neto De Carvalho (2006). Bank loan losses -given-default: A case study. *Journal of Banking and Finance Vol. 30*(4), 1219–1243.

Engelmann, B. and R. Rauhmeier (2006). *The Basel II risk parameters: Estimation, validation, and stress testing*. Springer, Berlin.

Felsovalyi, A. and L. Hurt (1998). Measuring loss on latin american defaulted abnk loans: A 27 year study of 27 countries. *Journal of Lending and Credit Risk Management Vol. 81*(2), 41–46.

Filho, A., L. C. Thomas, and C. Mues (2010). Optimizing the collection process in consumer credit. *Production and Operations Management Vol. 19*(6), 698–708.

Gourieroux, C. and A. Monfort (2006). (non) consistency of the beta kernel estimator for recovery rate distribution. Technical report, Working Paper.

Grunert, J. and M. Weber (2009). Recovery rates of commerical lending: Empirical evidence for german companies. *Journal of Banking and Finance Vol. 33*, 505–513.

Gupton, G. M. and R. M. Stein (2002). Losscalc: Model for predicting loss given default. *Moody's February*.

Hand, D. (2001). Modeling consumer credit risk. *Journal of Management Mathematics Vol. 12*(2), 139–155.

Hao, C., M. Yu-Chao, C. Mu-zi, T. Yue, W. W Bo, C. Min, and Y. Xiao-Guang (2009, Dec). Recovery discrimination based on optimized -variables support vector machine for nonpeforming loans. *Systems Engineering - Theory and Practice Vol. 29*(12), 23–30.

Jacobs, M. and A. Karagozoglu (2011). Modeling ultimate loss-given-default on corporate debt. *Journal of Fixed Income 21*(1), 6–20.

Kaneko, T. and H. Nakagawa (2008). A bank loan pricing model based on recovery rate distribution. *Inetrnational Journal of Innovative Computing Vol. 4*(1), 101–108.

Livingstone, S. M. and P. K. Lunt (1992). Predicting personal debt and debt repayment: Psychological, social and economic determinants. *Journal of Economic Psychology 13*, 111–134.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2011). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting (in press)*.

Matuszyk, A., C. Mues, and L. Thomas (2010). Modelling LGD for unsecured personal loans: Decision tree approach. *Journal of the Operational Research Society Vol. 61*(3), 393–398.

Peter, C. (2006). Estimating loss given default: Experience from banking practice. In B. Engelmann and R. Rauhmeier (Eds.), *The Basel II Risk Parameters*, Chapter 8, pp. 143–174. Springer-Verlag, Heidelberg.

Porto, R. (2011). A brief note on implied historical lgd. *The Journal of Credit Risk 7*(2), 73–81.

Qi, M. and X. Yang (2009, May). Loss given default of high loan-to-value residential mortgages. *Journal of Banking and Finance Vol. 33*(5), 788–799.

Qi, M. and X. Zhao (2011). Comparison of modeling methods for loss given default. *Journal of Banking and Finance (in press)*.

Renault, O. and O. Scaillet (2004). On the way to recovery: A nonparametric bias-free estimation of recovery rates densities. *Journal of Banking and Finance Vol. 28*(12), 2915–2931.

Schuermann, T. (2004). What do we know about loss given default? In D. Shimko (Ed.), *Credit Risk Models and Management (2nd Ed.), 249-274*, Chapter 9. Risk Books, London.

Standard&Poor's (2011). Standard & poor's financial services llc. http://www.standardandpoors.com.

Stiglitz, J. (1998, 23 September). Must financial crises be this frequent and this painful? Technical report, McKay Lecture, Pittsburgh, Pennsylvania.

Thomas, L. C. (2009). *Consumer Credit Models*. Oxford, UK: Oxford University Press.

Thomas, L. C., A. Matuszyk, and A. Moore (2011). Comparing debt characteristics and LGD models for different collection policies. *International Journal of Forecasting (in press)*.

Zhang, J. and L. C. Thomas (2010, September). Comparison of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*.