“Analysis of loss given default”

AUTHORS	Markus Höchstötter Abdolreza Nazemi

ARTICLE INFO	Markus Höchstötter and Abdolreza Nazemi (2013). Analysis of loss given default. *Investment Management and Financial Innovations*, *10*(4)
RELEASED ON	Monday, 09 December 2013
JOURNAL	"Investment Management and Financial Innovations"
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
0	0	0

© The author(s) 2026. This publication is an open access article.

Markus Höchstötter (Germany), Abdolreza Nazemi (Germany)

Analysis of loss given default

Abstract

As of now, there exist a vast variety of approaches quantifying the recovery of defaulted debt or, alternatively, the loss given default (LGD). This article endeavors to give a comprehensive account of the existing models of the recovery rate. Furthermore, it gives a detailed listing of the different types of debt that evoke different recovery processes and, hence, necessitate different definitions of recovery rate. It becomes obvious that there is a multitude of approaches rendering it almost infeasible to compare the different results directly.

Keywords: loss given default, recovery rate, GLM, beta distribution, beta kernels, SVM, credit default swaps.
JEL Classification: C13, C35, C52.

Introduction

The terms LGD and recovery rate have become widely used as of lately. In light of the recent crisis, sufficient amount of energy has to be committed to the assessment of these quantities. According to Standard & Poor's, between 2007 and 2011, a total of 496 rated institutions mostly from the US have defaulted representing over one trillion in debt outstanding and dwarfing anything seen so far. 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.

The fear of any repetition of the current crisis in the future may be founded in light of the seemingly increasing occurrence of financial crises as brought to attention by Stiglitz (1998) already one decade ago. Support to this hypothesis is provided by the fact that in the USA alone, public sector as well as corporate and consumer debt have reached dizzying levels. According to the Board of Governors of 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.

The loss related quantities such as the probability of a default, the amount of potential loss, as well as the amount recovered in case of a default are stated by the Basel II accords. Even though not mandatory for all lenders, the terminology and definitions are used widely in the context of credit risk[1]. However, the definitions in particular with respect to recovery are not unique. Here, we will present a variety of studies that refer to different interpretations of the definition of the recovery rate such as, for example, in the context of market recovery versus ultimate recovery[2]. We endeavor to provide an extension of the well known references such as Schuermann (2004) or Altman et al. (2005).

The paper is organized as follows. We divide the article into two sections. In section 1, the more extensive of the two, the individual subsections present the different methods to estimate the recovery rate. In detail, we subdivide the models into regressions, distributional approaches, alternative approaches such as neural networks and support vector machines, and stochastic recovery models. Section 2 lists the analyses related to debt type subdivided into the subsections on bonds and corporate debt as well as consumer debt such as bank loans and non-bank credit. Finally, pointing out possible directions for further research, the last section concludes the paper.

1. 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. A summary of the methods in this section can be found in Table 1.

1.1. Regression. We begin with the literature on data mining methods that we subsumed under the term regression. The simplest model in this context is ordinary least squares (OLS) regression $y = x^T\beta$ for modeling LGD as, for example, in Bellotti and Crook (2008). Bellotti and Crook (2012) also regress LGD on macro variables. An interesting early work is provided by Livingstone and Lunt (1992) who consider a multiple regression on social, economic and psychological factors related to debt. The most common generalized linear models (GLM) are given by logistic regression

© Markus Höchstötter, Abdolreza Nazemi, 2013.
Markus Höchstötter, Dr., Department of Econometrics and Statistics, School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany.
Abdolreza Nazemi, Dr., Department of Econometrics and Statistics, School of Economics and Business Engineering, Karlsruhe Institute of Technology, Germany.
We are very grateful to Yun (Ellen) Zhu, Marliese Uhrig-Homburg and Ser-Huang Poon for their valuable comments.

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

[2] One could also just as well consider the diverging definitions as individual definitions, themselves.

$$G(x^T\beta) = \exp(x^T\beta),$$

such as in Chen and Chen (2010) and probit

$$G(x^T\beta) = \Phi(x^T\beta),$$

where Φ denotes the standard normal cumulative distribution function. Chava et al. (2011) use fundamental variables in the context of OLS, logit, and probit, alternatively. Dermine and Neto De Carvalho (2006) as well as Grunert and Weber (2009) perform OLS regression also including macro variables and transforming through a logistic link. Bellotti and Crook (2008) also transform $G(x^T\beta)$ into a beta distributed LGD through quantile matching. Jacobs and Karagozoglu (2011) basically follow this approach but instead use a mixture of beta distributions as link function. Qi and Zhao (2011) model LGD based on logistic and inverse-Gaussian regression followed by a beta transformation. In the context of GLM, Bastos (2010a) uses the log-log link function

$$G(x^T\beta) = \exp(-\exp(-x^T\beta)),$$

instead. Belotti and Crook (2008) also apply Tobit regression by censoring the LGD according to

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

with the linear model $x^T\beta$ superimposed by some normally distributed noise u.

1.2. Distributional methods. 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[1]. Gupton and Stein (2002) state that recovery rates should be modeled by a beta distribution with density function,

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

for $y \in [0,1]$ and $\alpha, \beta > 0$, where $\Gamma(\cdot)$ denotes the gamma function[2].

Chen (1999) introduces a beta kernel estimator much of the same spirit as the well-known Gaussian kernel. Given n observations Y_i within $\in [0,1]$, its design is given by

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

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 $f(\cdot; c, d)$. 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 one rescales the original estimator (1) 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

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

Calabrese and Zenga (2010) introduce an alternative beta kernel estimator

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

to overcome the well-known boundary problem of the original beta kernel estimator. As pointed out by Calabrese (2010), however, 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.

1.3. Alternative methods. As the last category of modeling techniques, we present a collection of different approaches that have not found widespread use in contrast to the ones listed in the two subsections before. Qi and Zhao (2011) succesfully introduce neural networks as a non-linear approach to model LGD. This is also done by Bastos (2010b). Common to any design are an 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

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

[1] 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.

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

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.

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 hyper plane threshold with perpendicular vector w maximizing the minimal distance of each of the two groups from the threshold. With the optimal hyper plane, the training data keep a minimum distance of b from the hyper plane to guarantee generality of the model. The optimization problem using all n observations (y_i, x_i), $x_j \in R^d$ is thus given

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

or in the dual form

$$\max_a \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1,j}^{n} a_i a_j y_i y_j < x_i, x_j >,$$
$$s.t. \ \sum_{i=1}^{n} a_i y_i = 0 \quad (3)$$

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 (2) and (3). To this end, the original data vector $x_j \in R^d$ 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. Common kernel functions are, for example, polynomial $k(x_i, x) = < x_i, x >^p$ or radial basis $k(x_i, x) = \exp(-\| x_i - x \|^2 / c)$. Loterman et al. (2011) apply for the modeling of LGD non-linear techniques such as Classification and Regression Trees (CART), 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

$$\max_s \ \Delta i(s,t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R), \quad (4)$$

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. They also apply 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 (4). MARS approaches non-linearity 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 $y = \phi(x)^t b + \varepsilon$ with the original data x mapped into a higher feature space by ϕ to obtain a higher degree of linearity. Matuszyk et al. (2010) introduce LGD modeling based on a decision tree using a weights-of-evidence (WOE) approach for coarse-classification of continuous exogenous variables to determine the most significant characteristics for the prediction of high and low LGD. This is repeated in Thomas at el. (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.

1.4. Stochastic recovery. This section distinguishes itself from the subsection 1.2 in that the stochastic recovery rate presented here is embedded into more complex credit risk models rather than being individually modeled and fitted in an inferential sense.

We begin our discussion with some research on stochastic recovery based on Wiener processes. In the spirit of Merton (1974), Peura and Jokivuolle (2005) define default as when asset value A falls below default level D. Given that event, the expected loss equals the amount by which the firm falls short of meeting its obligation B at maturity of the debt, i.e.,

$$ELGD = \frac{1}{B}[\max(0, B - V_T) \,|\, A_T < D].$$

That is, as long as there is more collateral than mature debt, the loss is zero. Thus, the recovery is implicitly included in the collateral term. In Düllmann and Trapp (2005), the recovery of firm i is modeled as the logit

$$R_i = \frac{\exp(Y_i(X))}{1 + \exp(Y_i(X))}$$

with linear driver

$$Y_i(X) = \mu + \sigma\sqrt{\omega}\, X + \sigma\sqrt{1-\omega}\, Z_j$$

composed of the common factor X and idiosyncratic random variables Z_i that are *i.i.d.* standard normal. The results of the analysis are based on the S&P Credit Pro data base bonds and loans between January 1982 and December 1999 with market recovery at default as well as ultimate recovery.

Chabaane et al. (2005) model the correlation between some default event and the corresponding recovery using multivariate normal latent variables $\Psi_j,\ j = 1,\ldots,n$, where n denotes the number of firms, the correlation is given by $\rho = \mathrm{cov}(\Psi_i, \Psi_j)$, and the normally distributed common factor is Ψ. Recoveries are modeled as random variables $R_j = \exp(\mu + \sigma\xi_j)$ where the linear driver evolves itself as

$$\xi_j = \sqrt{\beta}\,\xi + \sqrt{1-\beta}\,\bar{\xi}_j$$

correlated with the common factors through

$$\nu = \mathrm{cov}(\Psi, \xi)$$
$$\gamma = \mathrm{cov}(\bar{\Psi}_j, \bar{\xi}_j)$$

In the following paragraphs, we will present the more recent and more sophisticated models that were developed in the context of credit derivative pricing. Andersen and Sidenius (2004) who provide the model of some of the literature thereafter use the Gaussian copula for default. The corresponding loss is given by

$$l_i = l^{\max}(1 - c_i(\mu_i + b_i \cdot Z + \xi_i))$$

for some given maximum loss $l_i^{\max}$ with constant μ_i, systematic factor terms $b_i \cdot Z$, and idiosyncratic factors ε_i and ξ_i independent of any other factors. The Z, ε_i, and ξ_i are assumed Gaussian with the mapping function c_i given by the cumulative normal distribution function, i.e., $c_i = \Phi$, for all i. The unconditional as well as conditional distribution with first two moments of the stochastic recovery R_i is capable of producing the same shapes of the density function as the beta distribution. The extensions to a Student's-*t* copula are straight-forward. Numerical results show that the random recovery model is not capable of handling the observed correlation skews, but of reproducing heavy-tailed losses prevailing in reality[1]. Krekel (2008) also use the standard Gaussian copula model for the random default variable X_i of obligor i. The recovery R^i if default occurs no later than payment date T_k has a discrete distribution of J states

$$R^i = \begin{cases} r_1^i & \text{with probability} \quad p_1^i \\ r_2^i & \text{with probability} \quad p_2^i \\ \vdots & \vdots \qquad\qquad\qquad \vdots \\ r_J^i & \text{with probability} \quad p_J^i \end{cases} \tag{5}$$

The authors suggest a simple unconditional probability distribution such as

$$R^i = \begin{cases} 60\% & \text{with probability} \quad p_1^i = 40\% \\ 40\% & \text{with probability} \quad p_2^i = 30\% \\ 20\% & \text{with probability} \quad p_3^i = 20\% \\ 0\% & \text{with probability} \quad p_4^i = 10\% \end{cases}$$

The authors partition the real line by additional thresholds $c_{k,J}^i = -\infty < c_{k,J-1}^i < \ldots < c_{k,0}^i = c_k^i$, where each compartment $[c_{k,J}^i, c_{k,J-1}^i]$ corresponds to the respective recovery rate value r_j^i from (5). Thus, given a certain value x of the market factor, X, the conditional probability distribution of the recovery rate depends on the correlation ρ of the default triggering random variables X_i. So, by construction, it is

$$p_{k,j}^i(x) = \begin{cases} \dfrac{\Phi\left(\dfrac{c_{k,j}^i - \sqrt{\rho}}{\sqrt{1-\rho}}\right) - \Phi\left(\dfrac{c_{k,j-1}^i - \sqrt{\rho}}{\sqrt{1-\rho}}\right)}{\Phi\left(\dfrac{c_k^i - \sqrt{\rho}}{\sqrt{1-\rho}}\right)} & \text{if } 0 < \rho < 1 \\ p_j^i & \text{if } \rho = 0 \\ \left.\begin{matrix} 1 & c_{k,j}^i < x < c_{k,j-1}^i \\ 0 & \text{else} \end{matrix}\right\} & \text{if } \rho = 1 \end{cases}$$

[1] A correlation skew is obtained from calibrating the correlation parameter to each portfolio loss tranche. The typical interpolation curve depicts a convex upward sloping graph.

Amraoui and Hitier (2008) state that the standard recovery is set equal to $\overline{R} = 40\%$. Hence, the loss is bounded from above. Thus, to match market prices, a mark-down from $\overline{R}$ to some lower, stochastic $\tilde{R}$ is introduced. It is made possible through the following relationship of $\tilde{R}$ to the recovery rate R_i of issuer i through the common factor X

$$(1-R_i(X)) = \left(1-\tilde{R}\right)\frac{g_\rho(\tilde{p}_i, X)}{g_\rho(p_i, X)},$$

where p_i and $\tilde{p}_i$ denote the unconditional probabilities of default of issue i in the scenarios $\overline{R}$ and $\tilde{R}$, respectively, while the function $g(\cdot, X)$ is the conditional probability of default conditional on the common factor X. For consistency reasons, one has to guarantee that $E^{\mathbf{Q}}[R_i \mid \tau_i < T] = \overline{R}$ with risk neutral measure Q, for any maturity T. That is, conditional on default, the average recovery rate has to equal the stripping recovery.

Prampolini and Dinnis (2009) provide a discussion of the method introduced by Amraoui and Hitier (2008) in a single tranche CDO (STCDO) pricing context, i.e., the pricing of insurance against default of a certain percentage range of the credit portfolio. The base tranche is defined as the loss interval [0, d], where d is the so-called detachment of the tranche, i.e., its upper bound. The loss at time t incurred by the protector of the base tranche [0, d] is defined as

$$L_d(t) = \min\{L_P(t), d\},$$

where L_P denotes the portfolio loss at time. The remaining tranche percentage is given by $N_d = 1 - - L_d(t) - R_d(t)$. Here, $\overline{R}_i$ denotes the deterministic recovery and $\tilde{R}_i$ is the lower bound of the stochastic recovery to be determined by the modeler[1]. The authors claim $\tilde{R}_i$ to be the proper choice since positive values will bound the loss from attaining 100% ever, almost surely.

Bennani and Maetz (2009) introduce the spot recovery rate given default at time t denoted as $r(t) \mid \tau = t$. In the factor model, the conditional spot recovery is given by $r(t, X) = E[r(\tau) \mid \tau = t, X]$. The recovery to maturity if default happens before time t is denoted as $R(t) = r(\tau) \mid \tau \le t$ while in the factor model the conditional recovery to maturity is given by $R(t, X) = E[r(\tau) \mid \tau \le t, X]$. Furthermore, the conditional default probability for issuer i is $p_i(t,X) = P(\tau_i \le t \mid X)$.

Consequently, the recovery until maturity of obligor i conditional on X is given by

$$R_i(t, X) = \int_0^t r_i(s, X)\frac{dp_i(s, X)}{\int_0^t dp_i(s, X)}$$

which is equivalent to

$$r_i(t, X) = \frac{\partial_t R_i(t, X)\, p(t, X)}{\partial_t p(t, X)}.$$

Defining the loss at time t as $l_t = (1 - r_\tau)1_{\tau \le t}$, we obtain the expected loss condition (EL)

$$E[r_\tau \mid \tau = t] = R^{Mkt}.$$

In a factor model, the recovery depends only on default time τ and common factor X, such that

$$r_\tau = r(\tau, X) = r(\Phi^{-1}(p_\tau), X$$

$$r_\tau 1 \sim r(\sqrt{\rho}\, X + \sqrt{1-\rho} Y, X) 1_{Y \le \frac{\Phi^{-1}p(t) - \sqrt{\rho}\, X}{\sqrt{1-\rho}}}$$

$$E[r(\tau, X) \mid \tau = t] = E\left[r(t, \sqrt{\rho}\Phi^{-1}(p_t) + \sqrt{1-\rho}\, X\right] = R^{Mkt},$$

where Φ denotes the standard normal cumulative distribution function. Then, the conditional normally distributed spot recovery rate is

$$r_i(t, X) = \Phi(\alpha_\rho X \times \beta_i(t))$$

with α_ρ determining the dependence structure ρ from the Gaussian copula. With this notation, constant recovery can be modeled through $\alpha_\rho = 0$. Then, two spot recovery models are used. The first assumes the same functional form of the spot recoveries for all $t \ge 0$. The second one is equivalent to the first one with a different recovery at time $t = 0$, however, i.e., $r_i(0^+, x) = R^{Mkt}$.

A bottom-up dynamic correlation modeling framework is intrduced by Li (2010) with consistent stochastic recovery. The recovery rate conditional on default between t_1 and t_2, i.e., $\tau \in (t_1,t_2)$, is given by $r(t_1,t_2) \in [0,1]$ with mean $\mu(t_1,t_2) = \mathrm{E}[r(t_1,t_2)] \in [0,1]$. The spot recovery rate conditional on exact default between t and $t + dt$, i.e., $\tau \in (t, t + dt)$, is $r(t,t)$. The model permits a bounded variance $\sigma^2(t_1,t_2) \in [0, \mu(t_1,t_2)(1 - \mu(t_1,t_2))]$. For consistent CDO pricing, it is shown that only $\mu(p,p)$ and $\sigma_i(p,p)$ need to be specified. The relationship between the spot mean recovery $\mu(t, t)$ and the term mean recovery $\mu(0,t)$ can be expressed by

$$\mu(0,t) = \frac{1}{p(t)}\int_0^t \mu(s,s)\, p(s) = \int_0^{p(t)} \mu(p, p) ds\,.$$

Now, the unconditional term recovery rate at time t for issuer i is obtained through integration over all possible market factor values x as

[1] In Amraoui and Hitier (2008), it is set equal to 40%.

$$R_i(0,t) = \frac{1}{p_i(t)} \int \mu(0, p_i(x,t)) p_i(x,t) f(x,t) dx.$$

Schneider et al. (2010) pursue the identification of implied LGD from credit default swaps (CDS) spreads. Theory requires that at inception, the value of the fix premia $s_t(T)$ that the protection seller is expecting to receive throughout the live of the derivative, i.e. $T_n \in \{t, t+1, \ldots, T\}$, given the information at time t[1]

$$L_t^{\text{pre}}(T) = s_t(T) \left(\sum_{n=1}^{T} (T_n - T_{n-1}) \tilde{P}(t, T_n) + E_t^{\mathbb{Q}} \left[\int_t^T (u - T_{I(u)}) \eta_u \exp\left\{ -\int_t^u \tilde{r}_v dv \right\} du \right] \right)$$

has to equal the expected discounted payments at default of the protection buyer at time *t*.

$$L_t^{\text{pro}}(T) = E_t^{\mathbb{Q}} \left[\int_t^T L\tilde{G}D_u \, \eta_u \exp\left\{ -\int_t^u \tilde{r}_v dv \right\} du \right].$$

$L\tilde{G}D_u$ denotes the expected LGD in *u*. From this equality given the observed premia, the unknown expected LGD is singled out which is assumed constant over time. In their model, the default intensities subdivided into short- and long-term components are driven by a latent Wiener process in combination with a Poisson jump process (Cox process). The cross section of the LDG is then regressed on the industry sectors and credit rating.

Das and Hanouna (2009) model the recovery rate as implied from CDS spreads. For the derivation of the latter, they use the equality of premium and protection leg as shown in Schneider et al. (2010). In their model, the recovery rate in period *j* is denoted as ϕ_j in the event of default between T_{j-1} at T_j. The recovery rates are conceived as dynamic, stochastic, and endogenous. Moreover, Das and Hanouna (2009) model the entire term structures of the recovery rate. The stock price *S* is the only state variable. With default intensity $\xi[i,j] = \frac{1}{S[i,j]^b}$, the probability of default is given as $\gamma[i,j] = 1 - \exp(-\xi[i,j]h)$. The constant *h* is the coupon frequency. Then, in period *j*, the recovery rate is given by the probit model $\phi[i,j] = g(a_0 + a_1\lambda)\,[i,j]$, where the link function *g* is the normal cumulative distribution function. Consequently, the entire model uses only three parameters a_0, a_1, and *b*.

Table 1. Summary by model

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

[1] $I(u)$ denotes the last premium payment date before default at u, η_u the intensity of default, and $\tilde{r} = r + \eta$ the risk adjusted short rate. Expectation is computed with respect to the risk neutral measure Q.

2. 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.

2.1. Bonds and corporate debt. Schuermann (2004) finds that seniority of debt has an enormous impact on the distribution of the recovery rate. Recovery on junior debt is predominantly low compared to senior debt. Also, the industry type has significant impact on recovery. Moreover, bonds behave differently from loans due to different control rights with the latter generally yielding higher recovery rates. Similar results are obtained by Felsovalyi and Hurt (1998) who analyze the recovery on loans issued to commercial industrial borrowers. Altman (2008) studied the recovery process on defaulted bonds and came to the same conclusion regarding seniority. Also, recovery is found to be lower than for loans. They additionally report high variation across industries. Jacobs and Karagozoglu (2011) find that macro-economic factors play as import role as industry conditions, equity returns, and the price of tradable debt at default in addition to the debtor related characteristics. 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. Asarnow and Edwards (1995) compare corporate and industrial loans to structured loans concluding that structuring seems to have a very positive effect on recovery. Renault and Scaillet (2004), Bastos (2010b), Bastos (2010a), and Qi and Zhao (2011) analyze corporate loans with respect to modeling techniques. They find that non-parametric models outperform parametric models. The results of this subsection are summarized in Table 2.

Table 2. Summary by debt type (corporate)

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&L 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

2.2. 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.

2.2.1. Bank loans. Calabrese (2010) reports a high concentration of recovery at zero and one. Grunert and Weber (2009) state that the inclusion of macro variables does not improve model quality. In contrast, Caselli et al. (2008) find that macro-economic factors are important. However, the recovery rate hinges more on the loan-to-value ratio atdefault. Zhang and Thomas (2012) find exposure at default as the single most important determinant. Avery et al. (2004) argue that situational circumstances matter immensely with respect to recovery. Livingstone and Lunt (1992) conclude that socio-demographic factors play a relatively minor role in personal debt and debt repayment. Attitudinal factors are found to be important predictors of debt and debt repayments. Hao et al. (2009) obtain that loan-specific characteristics are significant for loan recovery discrimination. 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 during the whole life and last 12 months, as well as the time until default. Zhang (2009) examines the influence of loan covenants on recovery rates which are found to be highly signidicant predictors. With respect to modeling, Grunert and Weber (2009) state that the beta distribution is not useful for modeling recovery rates. Loterman et al. (2011) report that SVM and non-linear neural networks have better predictive ability than parametric or regression methods.

2.2.2. Non-bank credit. Bellotti and Crook (2008) and Bellotti (2010) study credit card debt with respect to correlation of default and LGD denying its influence on portfolio Value-at-Risk. Chen and Chen (2010) find that social, demographic, and economic factors are relevant in the explanation of recoveryof residential mortgages. Bellotti and Crook (2012) find that bank interest rates and unemployment rates significantly predict recovery. Qi and Yang (2009) analyze LGD of insured mortgages and find that the current loan-to-value (CLTV) as well as the initial loan-to-value (LTV) are positively correlated. Thomas et al. (2012) analyze the success of varying debt collection processes.

A summary of the findings of the following two subsections is provided in Table 3.

Table 3. Summary by debt type (consumer)

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
Chenand Chen	Mortgage loan	1,880	1987-2007	-	-	Taiwan	2010
Hao et al.	Loss metric 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
Qiand 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

Conclusion

This article presented the current state of research on recovery rates. In section 1, the different models and methods were given. Analyses sorted by type of debt followed in section 2. Of the many different approaches, those considering exogenous variables for the prediction such as, for example, statistical ones appear to be more powerful than those in the context of pricing due to fewer degrees of freedom of the latter. However, it might be diffcult to integrate the different approaches without excessive inflation of model complexity. Also, study in this field has been insufficient with respect to generality of findings due to a lack of easily accessible recovery data as a result of the preponderance of bank data compared to other debt.

References

1. 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*, Cambridge: Cambridge University Press, pp. 175-206.
2. 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*, Risk Books, London, pp. 41-59.
3. Altman, E.I., A. Resti, and A. Sironi (2005b). *Recovery risk: The next challenge in credit risk management*, Risk Books, London.
4. Amraoui, S. and S. Hitier (2008). Optimal Stochastic Recovery for Base Correlation, Working Paper, BNP Paribas.
5. Andersen, L. and J. Sidenius (2004). Extensions to the Gaussian copula: random recovery and random factor loadings, *Journal of Credit Risk*, 1 (1), pp. 29-70.

6. Asarnow, E. and D. Edwards (1995). Measuring loss on defaulted bank loans. A 24-year study, *Journal of Commercial Lending*, 77 (7), November 23.
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, pp. 835-856.
8. Bastos, J. (2010a). Forecasting bank loans loss-given default, *Journal of Banking and Finance*, 34 (10), pp. 2510-2517.
9. Bastos, J. (2010b). *Predicting bank loan recovery rates with neural networks*, Technical report.
10. BCBS (2005). International convergence of capital measurement and capital standards. a revised framework, bank for international settlements. Available at http://www.bis.org.
11. BCBS (2011). Basel III: A global regulatory framework for more resilient banks and banking systems, bank for international settlements. Available at http://www.bis.org.
12. Bellotti, T. (2010). A simulation study of Basel II expected loss distribution for a portfolio of credit cards, *Journal of Financial Services Marketing*, 14, pp. 268-277.
13. Bellotti, T. and J. Crook (2008). *Modelling and predicting loss given default for credit cards*, Technical report.
14. Bellotti, T. and J. Crook (2012). Loss given default models incorporating macroeconomic variables for credit cards, *International Journal of Forecasting*, 28 (1), pp. 171-182.
15. Bennani, N. and J. Maetz (2009). A Spot Stochastic Recovery Extension of the Gaussian Copula, Credit Derivatives Working Paper.
16. Board of Governors of the Federal Reserve System (2011). Federal Reserve statistical release. Available at http://www.federal-reserve.gov.
17. Boettger, M., A. Guthoff, and T. Heidorn (2008). Loss given default modelle zur schaetzung von recovery rates, Frankfurt School Working Paper Series (96).
18. Calabrese, R. (2010). Regression for recovery rates with both continuous and discrete characteristics, Proceedings of the 45th Scientific Meeting of the Italian Statistical Society (SIS), Italy.
19. Calabrese, R. and M. Zenga (2010). Bank loan recovery rates: Measuring and nonparametric density estimation, *Journal of Banking and Finance*, 35 (5), pp. 903-911.
20. 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), pp. 1-34.
21. Chabaane, A., J.-P. Laurent, and J. Salomon (2005). Credit Risk Assessment and Stochastic LGD: An Investigation of Correlation Effects. In E. Altman, A. Resti, and A. Sironi (eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*, Chapter 15, Risk Books.
22. Chava, S., C. Stefanescu, and S. Turnbull (2011). Modeling the Loss Distribution, *Management Science*, 57 (7), pp. 1267-1287.
23. Chen, S. (1999). Beta kernel estimators for density functions, *Computational Statistics and Data Analysis*, 31, pp. 131-145.
24. Chen, T.H. and C.W. Chen (2010). Application of data mining to the spatial heterogeneity of foreclosed mortgages, *Expert Systems with Application*, 37 (2), pp. 993-997.
25. Crook, J., D. Edelman, and L. Thomas (2007). Recent developments in consumer credit risk assessment, *European Journal of Operations Research*, 183 (3), pp. 1447-1465.
26. Das, S.R. and P. Hanouna (2009). Implied recovery, *Journal of Economic Dynamic & Control*, 33, pp. 1837-1857.
27. Dermine, J. and C. Neto De Carvalho (2006). Bank loan losses-given-default: A case study, *Journal of Banking and Finance*, 30 (4), pp. 1219-1243.
28. Duellmann, K. and M. Trapp (2005). Systematic Risk in Recovery Rates of US Corporate Credit Exposure. In E. Altman, A. Resti, and A. Sironi (eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*, Chapter 13, London: Risk Books.
29. Engelmann, B. and R. Rauhmeier (2006). *The Basel II risk parameters: Estimation, validation, and stress testing*, Springer, Berlin.
30. Felsovalyi, A. and L. Hurt (1998). Measuring loss on Latin American defaulted bank loans: A 27 year study of 27 countries, *Journal of Lending and Credit Risk Management*, 81 (2), pp. 41-46.
31. Filho, A., L.C. Thomas, and C. Mues (2010). Optimizing the collection process in consumer credit, *Production and Operations Management*, 19 (6), pp. 698-708.
32. Gourieroux, C. and A. Monfort (2006). (Non) consistency of the beta kernel estimator for recovery rate distribution, Working Paper.
33. Grunert, J. and M. Weber (2009). Recovery rates of commercial lending: Empirical evidence for German companies, *Journal of Banking and Finance*, 33, pp. 505-513.
34. Gupton, G.M. and R.M. Stein (2002). *Losscalc: Model for predicting loss given default*, Moody's February.
35. Hao, C., M. Yu-Chao, C. Mu-zi, T. Yue, W.W Bo, C. Min, and Y. Xiao-Guang (2009). Recovery discrimination based on optimized − variables support vector machine for nonperforming loans, *Systems Engineering − Theory and Practice*, 29 (12), pp. 23-30.
36. Jacobs, M. and A. Karagozoglu (2011). Modeling ultimate loss-given-default on corporate debt, *Journal of Fixed Income*, 21 (1), pp. 6-20.
37. Kaneko, T. and H. Nakagawa (2008). A bank loan pricing model based on recovery rate distribution, *International Journal of Innovative Computing*, 4 (1), pp. 101-108.
38. Krekel, M. (2008). Pricing distressed CDOs with Base Correlation and Stochastic Recovery, Working Paper.

39. Li, Y. (2010, April). A Dynamic Correlation Modelling Framework with Consistent Stochastic Recovery, Technical report, Barclays Capital.
40. Livingstone, S.M. and P.K. Lunt (1992). Predicting personal debt and debt repayment: Psychological, social and economic determinants, *Journal of Economic Psychology*, 13, pp. 111-134.
41. Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2012). Benchmarking regression algorithms for loss given default modeling, *International Journal of Forecasting*, 28 (1), pp. 161-170.
42. Matuszyk, A., C. Mues, and L. Thomas (2010). Modelling LGD for unsecured personal loans: Decision tree approach, *Journal of the Operational Research Society*, 61 (3), pp. 393-398.
43. Peter, C. (2006). Estimating loss given default: Experience from banking practice. In B. Engelmann and R. Rauhmeier (eds.), The Basel II Risk Parameters, Springer-Verlag, Heidelberg, pp. 143-174.
44. Peura, S. and E. Jokivuolle (2005). LGD in a Structural Model of Default. In E. Altman, A. Resti, and A. Sironi (eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*, Chapter 11, Risk Books.
45. Porto, R. (2011). A brief note on implied historical lgd, *The Journal of Credit Risk*, 7 (2), pp. 73-81.
46. Prampolini, A. and M. Dinnis (2009). CDO mapping with stochastic recovery, Working paper, HSH Nordbank AG.
47. Qi, M. and X. Yang (2009). Loss given default of high loan-to-value residential mortgages, *Journal of Banking and Finance*, 33 (5), pp. 788-799.
48. Qi, M. and X. Zhao (2011). Comparison of modeling methods for loss given default, *Journal of Banking and Finance*, 35 (11), pp. 2842-2855.
49. 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*, 28 (12), pp. 2915-2931.
50. Robert C. Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance*, 29, pp. 449-470.
51. Schneider, P., L. Soegner, and T. Veya (2010). The Economic Role of Jumps and Recovery Rates in the Market for Corporate Default Risk, *Journal of Financial and Quantitative Analysis*, 45 (6), pp. 1517-1547.
52. Schuermann, T. (2004). What do we know about loss given default? In D. Shimko (ed.), Credit Risk Models and Management (2 ed.), Risk Books, London, pp. 249-274.
53. Standard & Poor's (2011). Standard & Poor's Financial Services LLC. http://www.standardandpoors.com.
54. Stiglitz, J. (1998). Must financial crises be this frequent and this painful? Technical report, McKay Lecture, Pittsburgh, Pennsylvania.
55. Thomas, L.C. (2009). *Consumer Credit Models*, Oxford, UK: Oxford University Press.
56. Thomas, L.C., A. Matuszyk, and A. Moore (2012). Comparing debt characteristics and LGD models for different collection policies, *International Journal of Forecasting*, 28 (1), pp. 196-203.
57. Zhang, J. and L.C. Thomas (2012). Comparison of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD, *International Journal of Forecasting*, 28 (1), pp. 204-215.
58. Zhang, Z. (2009). Recovery Rates and Macroeconomic Conditions: The Role of Loan Covenants, AFA 2010 Atlanta Meetings Paper.