“Analysis of loss given default”

 AUTHORS
 Markus Höchstötter
 Abdolreza Nazemi

 ARTICLE INFO

 JOURNAL
 "Investment Management and Financial Innovations"

 FOUNDER
 LLC “Consulting Publishing Company “Business Perspectives”

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

 © The author(s) 2018. This publication is an open access article.
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 all lenders, the terminology and definitions are used widely in the context of credit risk. 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. 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 Livingston 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.
\[ G(x^T \beta) = \exp(x^T \beta), \]
such as in Chen and Chen (2010) and probit \[ G(x^T \beta) = \Phi(x^T \beta), \]
where \( \Phi \) 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 distribution function. Chava et al. (2011) use \( \bar{G} \) and parameter \( b \) responsible for smoothing.

The kernel function \( K(c,d) \) is given by the beta density function \( f(c; d) \). 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.,

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

while the latter rescales at each observation according to

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

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

\[
\hat{f}_n(y,b) = \frac{\sum_{i=1}^n K(Y_i, y/b+1, (1-y)/b+1)}{n}
\]

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) successfully 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 enter neuron \( i \) of the hidden layer to be transformed there into a weighted functional output

\[
h_i = f^{(i)}(h_i + \sum_{j=1}^d w_{i,j} x_j)
\]

1 In the context of the Basle 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 \( w_{ij} \) 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 SVMs are used to separate debtors into two categories \((y = -1 \text{ or } y = 1)\) 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 \((y_i, x_i), x_i \in \mathbb{R}^d\) is thus given

\[
\min_{w,b} \left\| w \right\|_2^2, \quad \text{s.t.} \quad y_i (w, x_i > +b) \geq 1, i = 1, 2, \ldots, n \tag{2}
\]

or in the dual form

\[
\max \sum_{i=1}^n \left( \sum_{j=1}^n a_{ij} y_i y_j x_i x_j \right),
\]

\[
\text{s.t.} \quad \sum_{i=1}^n a_{ij} y_i = 0 \tag{3}
\]

where \( \langle \cdot, \cdot \rangle \) denotes the inner product. The separating rule is then given by \( f(x) = \text{sign}(\langle w, x \rangle + b) \) or, equivalently, \( f(x) = \text{sign}(\sum_{i=1}^n a_i y_i x_i x_j > +b) \).

A problem occurs if the data are not linearly separable as required by (2) and (3). To this end, the original data vector \( x_i \in \mathbb{R}^d \) is mapped into a higher dimensional \((k > d)\) feature space with a non-linear function \( \phi : \mathbb{R}^d \rightarrow \mathbb{R}^k, x \mapsto \phi(x) \). 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 \( k(x, x) = (x, x)^p \) or radial basis \( k(x, x) = \exp(-\left\| x - x \right\|^2 / \ell) \). Lotterman 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

\[
\max_s \Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R), \tag{4}
\]

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 \( y = \sum_{j=1}^k b_j B_j(x) \) 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 \( y = \phi(x)^T \beta + \epsilon \) with the original data \( x \) mapped into a higher feature space by \( \phi \) 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 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.

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.,

\[
\text{ELGD} = \frac{1}{B} \left[ \max(0, B - V_x) \right] A_x < D \}
\]

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
\[ Y_j(X) = \mu + \sigma \sqrt{\omega} X + \sigma \sqrt{1-\omega} Z_j \]
composed of the common factor \( X \) and idiosyncratic random variables \( Z_j \) 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 \( \rho = \text{cov}(\Psi_i, \Psi_j) \), and the normally distributed common factor is \( \Psi_i \). Recoveries are modeled as random variables \( R_j = \exp(\mu + \sigma \xi_j) \) where the linear driver evolves itself as
\[ \xi_j = \sqrt{\beta} \xi + \sqrt{1-\beta} \xi_j \]
correlated with the common factors through
\[ \nu = \text{cov}(\Psi_i, \xi) \]
\[ \gamma = \text{cov}(\Psi_j, \xi_j) \]

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 standard Gaussian copula for default. The corresponding loss is given by
\[ l_i = l_{\text{max}}(1-c_i (\mu_i + b_1 \cdot Z + \xi_i)) \]
for some given maximum loss \( l_{\text{max}} \) with constant \( \mu_i \), systematic factor terms \( b_1 \cdot Z \), and idiosyncratic factors \( \xi_i \) independent of any other factors. The \( Z \), \( \xi_i \), and \( \xi_j \) 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 straightforward. 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. Krekel (2008) also use the standard Gaussian copula model for the random default variable \( X_i \) for obligor \( i \). The recovery \( R_i \) if default occurs no later than payment date \( T_i \) has a discrete distribution of \( J \) states
\[
R_i = \begin{cases} 
1 \text{ with probability } p_i^1 \\
2 \text{ with probability } p_i^2 \\
\vdots & \vdots \\
J \text{ with probability } p_i^J 
\end{cases} 
\tag{5}
\]
The authors suggest a simple unconditional probability distribution such as
\[
R_i = \begin{cases} 
60\% \text{ with probability } p_i^1 = 40\% \\
40\% \text{ with probability } p_i^2 = 30\% \\
20\% \text{ with probability } p_i^3 = 20\% \\
0\% \text{ with probability } p_i^4 = 10\% 
\end{cases} 
\]
The authors partition the real line by additional thresholds \( c_i^1 < c_i^2 < \ldots < c_i^J = c_i \), where each compartment \([c_i^1, c_i^2, \ldots, c_i^J] \) corresponds to the respective recovery rate value \( r_j \) 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 \( \rho \) of the default triggering random variables \( X \). So, by construction, it is

\begin{align*}
\Phi \left( \frac{c_i^j - \sqrt{\rho}}{\sqrt{1-\rho}} \right) - \Phi \left( \frac{c_i^{j-1} - \sqrt{\rho}}{\sqrt{1-\rho}} \right) & \quad \text{if } 0 < \rho < 1 \\
\Phi \left( \frac{c_i^j - \sqrt{\rho}}{\sqrt{1-\rho}} \right) & \quad \text{if } \rho = 0 \\
1 & \quad \text{if } \rho = 1 \\
0 \quad \text{else} & \\
\end{align*}

\[ c_i^1 \leq x \leq c_i^{J-1} \]

\[ 0 \leq x \leq c_i \]

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 $R = 40\%$. Hence, the loss is bounded from above. Thus, to match market prices, a mark-down from $R$ 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 $X$

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

where $p_i$ and $\tilde{p}_i$ denote the unconditional probabilities of default of issue $i$ in the scenarios $R$ and $\tilde{R}$, respectively, while the function $g(-, X)$ is the conditional probability of default conditional on the common factor $X$. For consistency reasons, one has to guarantee that $E[\tilde{R} \mid \tilde{\tau} < T] = \tilde{R}$ 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, $R_d$ denotes the deterministic recovery and $\tilde{R}$ is the lower bound of the stochastic recovery to be determined by the modeler. The authors claim $\tilde{R}$ 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 $r(t, X) = E[r(t) \mid \tau = t, X]$. The recovery to maturity if default happens before time $t$ is denoted as $R(t) = r(t) \mid t \leq t$ while in the factor model the conditional recovery to maturity is given by $R(t, X) = E[r(t) \mid \tau \leq t, X]$. Furthermore, the conditional default probability for issuer $i$ is $p_i(t, X) = P(t_i \leq t \mid X)$.

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

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

which is equivalent to

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

Defining the loss at time $t$ as $l_i = (1 - r_i)1_{\tau \leq s}$, we obtain the expected loss condition (EL)

$$E[r_i \mid \tau = t] = R_i^{\text{Mkt}}.$$  

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

$$r_i = r(\tau, X) = r(\Phi^{-1}(p_i), X)$$

for $\tau \sim r(\sqrt{\rho} X + \sqrt{1 - \rho} Y, X)$ with risk neutral measure $Q$, for any maturity $T$. The second one is equivalent to the first one with a different recovery at time $t = 0$, however, i.e., $r_0(\Phi, x) = R_i^{\text{Mkt}}$.

A bottom-up dynamic correlation modeling framework is introduced 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 $r(t_1, t_2) \mid \tau \in [0, 1]$ with mean $\mu(t_1, t_2) = E[r(t_1, t_2)] \in [0, 1]$. The spot recovery rate conditional on exact default between $t$ and $t + dt$, i.e., $\tau \in (t, t + dt)$, is $r(t, t + dt)$.

The model imposes 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 $\sigma(p, p)$ 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

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

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%.
The risk adjusted short rate. Expectation is equal to the expected discounted payments at time $t$.

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

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 $s_i(T)$ that the protection seller is expecting to receive throughout the live of the derivative, i.e. $T_{i,a} \in \{ t_i, t + 1, \ldots, T \}$, given the information at time $t$.

 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 $\phi_i$ in the event of default between $T_{i,j-1}$ and $T_{i,j}$. 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 $\lambda[i,j] = \frac{1}{S[i,j]}$, 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

<table>
<thead>
<tr>
<th>Author or authors</th>
<th>Models</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bastos</td>
<td>Fractional response regression, regression tree</td>
<td>2010a</td>
</tr>
<tr>
<td>Bastos</td>
<td>Fractional Regression, neural network</td>
<td>2010b</td>
</tr>
<tr>
<td>Bellotti and crock</td>
<td>Models incorporating macroeconomic variables</td>
<td>2011</td>
</tr>
<tr>
<td>Bellotti and crock</td>
<td>OLS regression, Decision Tree and Tobit regression</td>
<td>2006</td>
</tr>
<tr>
<td>Calabrese and Zenga</td>
<td>Beta kernel, mixed random variable</td>
<td>2010</td>
</tr>
<tr>
<td>Calabrese</td>
<td>Beta regression, mixed random variable, distributional</td>
<td>2010</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>Regression, multivariate analysis</td>
<td>2008</td>
</tr>
<tr>
<td>Chen and Chen</td>
<td>Logistic regression</td>
<td>2010</td>
</tr>
<tr>
<td>Chen</td>
<td>Beta kernel estimators (it is not on LGD)</td>
<td>1999</td>
</tr>
<tr>
<td>Dernine and Neto De Carvalho</td>
<td>OLS regression, logistic regression</td>
<td>2006</td>
</tr>
<tr>
<td>Filho et al.</td>
<td>Optimization, text mining</td>
<td>2010</td>
</tr>
<tr>
<td>Gourieroux and Monfort</td>
<td>Beta kernel</td>
<td>2006</td>
</tr>
<tr>
<td>Grunert and Weber</td>
<td>Distribution of RR, regression</td>
<td>2009</td>
</tr>
<tr>
<td>Gupton and Stein</td>
<td>Distributional</td>
<td>2002</td>
</tr>
<tr>
<td>Hao et al.</td>
<td>Support vector machine, discriminant analysis</td>
<td>2009</td>
</tr>
<tr>
<td>Jacobs Jr. and Karagozoglu</td>
<td>Beta-link generalized linear model</td>
<td>2011</td>
</tr>
<tr>
<td>Loterman et al.</td>
<td>OLS, Ridge regression, beta regression, logistic regression, CART, MARS, LSSVM, and ANN</td>
<td>2011</td>
</tr>
<tr>
<td>Qi and Yang</td>
<td>Regression</td>
<td>2009</td>
</tr>
<tr>
<td>Qi and Zhao</td>
<td>Regression tree, neural network, fractional response regression, inverse Gaussian regression</td>
<td>2011</td>
</tr>
<tr>
<td>Renault and Scaillet</td>
<td>Kernel estimation, nonparametric estimators, Monte Carlo</td>
<td>2004</td>
</tr>
<tr>
<td>Thomas et al.</td>
<td>Box-Cox, linear regression, beta distribution, log normal transformation</td>
<td>2009</td>
</tr>
<tr>
<td>Thomas et al.</td>
<td>Modelling LGD for unsecured personal loans: decision tree approach</td>
<td>2009</td>
</tr>
<tr>
<td>Yeh and Lien</td>
<td>Data mining techniques in PD</td>
<td>2009</td>
</tr>
<tr>
<td>Zhang and Thomas</td>
<td>Linear regression, survival analysis, mixture distribution</td>
<td>2010</td>
</tr>
</tbody>
</table>

---

1 $T_{i,a}(u)$ denotes the last premium payment date before default at $u$, $\eta_u$ the intensity of default, and $\bar{r} = r + \eta$ the risk adjusted short rate. Expectation is computed with respect to the risk neutral measure $Q$.  

---

75
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.

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 at default. 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

<table>
<thead>
<tr>
<th>Author or authors</th>
<th>Data</th>
<th>Sample size</th>
<th>Sample period</th>
<th>Mean of RR</th>
<th>Median of RR</th>
<th>Country</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bastos</td>
<td>SMEs</td>
<td>374</td>
<td>Jun. 1995-Dec. 2000</td>
<td>0.694</td>
<td>0.946</td>
<td>Portugal</td>
<td>2010a</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SME</td>
<td>11,649</td>
<td>1990-2004</td>
<td>0.540</td>
<td>0.560</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SMEs</td>
<td>1,814</td>
<td>Jan. 1990-Aug. 2004</td>
<td>0.54</td>
<td>0.63</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SMEs</td>
<td>1,925</td>
<td>Jan. 1990-Aug. 2004</td>
<td>0.50</td>
<td>0.39</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SMEs</td>
<td>2,189</td>
<td>Jan. 1990-Aug. 2004</td>
<td>0.53</td>
<td>0.56</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SMEs</td>
<td>2,423</td>
<td>Jan. 1990-Aug. 2004</td>
<td>0.54</td>
<td>0.47</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Caselli et al.</td>
<td>SMEs</td>
<td>3,318</td>
<td>Jan. 1990-Aug. 2004</td>
<td>0.58</td>
<td>0.64</td>
<td>Italy</td>
<td>2008</td>
</tr>
<tr>
<td>Dermine and Neto de Carvalho</td>
<td>SMEs</td>
<td>10,000</td>
<td>Jun. 1995-Dec. 2000</td>
<td>0.71</td>
<td>0.95</td>
<td>Portugal</td>
<td>2006</td>
</tr>
<tr>
<td>Felsovalyi and Hurt</td>
<td>Citibank loans</td>
<td>1,149</td>
<td>1970-1996</td>
<td>0.68</td>
<td>-</td>
<td>LA</td>
<td>1998</td>
</tr>
<tr>
<td>Grunert and Weber</td>
<td>SME</td>
<td>120</td>
<td>1992-2003</td>
<td>0.725</td>
<td>0.918</td>
<td>Germany</td>
<td>2009</td>
</tr>
<tr>
<td>Jacobs Jr. and Karagozoglu</td>
<td>US Corporate</td>
<td>3,902</td>
<td>1985-2008</td>
<td>0.6104</td>
<td>0.6841</td>
<td>US</td>
<td>2011</td>
</tr>
<tr>
<td>Jones and Hensher (Altman)</td>
<td>Bank loans</td>
<td>1,324</td>
<td>1988-2006</td>
<td>0.772</td>
<td>-</td>
<td>US</td>
<td>2008</td>
</tr>
<tr>
<td>Jones and Hensher (Altman)</td>
<td>Bonds</td>
<td>2,071</td>
<td>1988-2006</td>
<td>0.30-0.62</td>
<td>-</td>
<td>US</td>
<td>2008</td>
</tr>
<tr>
<td>Qi and Zhao</td>
<td>US Corporate</td>
<td>3,751</td>
<td>1985-2008</td>
<td>0.4423</td>
<td>0.4529</td>
<td>US</td>
<td>2011</td>
</tr>
<tr>
<td>Schuermann</td>
<td>Bonds</td>
<td>282</td>
<td>1970-2003</td>
<td>0.4962</td>
<td>0.4475</td>
<td>US</td>
<td>2006</td>
</tr>
</tbody>
</table>
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 significant 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 recovery of 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>
<thead>
<tr>
<th>Author or authors</th>
<th>Data</th>
<th>Sample size</th>
<th>Sample period</th>
<th>Mean of RR</th>
<th>Median of RR</th>
<th>Country</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bellotti</td>
<td>Credit Card</td>
<td>50,000</td>
<td>2003-2004</td>
<td>-</td>
<td>-</td>
<td>Brazil</td>
<td>2010</td>
</tr>
<tr>
<td>Bellotti and Crook</td>
<td>Credit Card</td>
<td>55,500</td>
<td>1998-2004</td>
<td>-</td>
<td>-</td>
<td>UK</td>
<td>2008</td>
</tr>
<tr>
<td>Calabrese</td>
<td>Personal loan</td>
<td>149,378</td>
<td>1998-1999</td>
<td>0.384</td>
<td>0.340</td>
<td>Italy</td>
<td>2011</td>
</tr>
<tr>
<td>Chen and Chen</td>
<td>Mortgage loan</td>
<td>1,880</td>
<td>1987-2007</td>
<td>-</td>
<td>-</td>
<td>Taiwan</td>
<td>2010</td>
</tr>
<tr>
<td>Hao et al.</td>
<td>Loss metric database</td>
<td>1115</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>China</td>
<td>2009</td>
</tr>
<tr>
<td>Livingstone and Lunt</td>
<td>Credit card</td>
<td>7,889</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1992</td>
</tr>
<tr>
<td>Loterman et al.</td>
<td>Credit card</td>
<td>119,211</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2011</td>
</tr>
<tr>
<td>Loterman et al.</td>
<td>Mortgage loan</td>
<td>3,351</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2011</td>
</tr>
<tr>
<td>Qi and Yang</td>
<td>Mortgage insurance</td>
<td>241,293</td>
<td>1990-2003</td>
<td>Max. 0.568</td>
<td>-</td>
<td>US and other</td>
<td>2009</td>
</tr>
<tr>
<td>Schuemann</td>
<td>Bank loans</td>
<td>151</td>
<td>1970-2003</td>
<td>0.631</td>
<td>0.655</td>
<td>-</td>
<td>2006</td>
</tr>
<tr>
<td>Thomas et al.</td>
<td>Personal loan</td>
<td>50,000</td>
<td>1989-2004</td>
<td>-</td>
<td>-</td>
<td>UK</td>
<td>2010</td>
</tr>
<tr>
<td>Zhang and Thomas</td>
<td>Personal loan</td>
<td>27,278</td>
<td>1987-1999</td>
<td>0.420</td>
<td>-</td>
<td>UK</td>
<td>2010</td>
</tr>
</tbody>
</table>

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 difficult 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