ı. Exploring the sources of default

ernational convergence of capital
work. Bank for International Settle-

/sis of high dimensional VARs fea-
an Central Bank.
of the vector autoregressive model
AT *2008: Proceedings in Computa-*
9–501. Physica, Heidelberg.
Frailty correlated default. *Journal of*

d corporate default prediction with
*:s* **83**, 635–665.
nigration models with application to
188–226.
Education, Upper Saddle River, NJ.
dit risk models. *The Journal of Credit*

sk factor modeling and the Basel II
nd Financial Supervision, Deutsche

sting credit event frequency: empir-
*f Risk* **9**(1), 75–98.
ı University Press.
king Paper 14, Bank for International

Modeling frailty-correlated defaults
*Econometrics* **162**(2), 312–325.
A. (2009). Credit cycles and macro
-54.
*ne Series Analysis.* Springer.
*eries Econometrics.* Cambridge Uni-

Weiner, S. (2006). Macroeconomic
*urnal of Money, Credit and Banking*



The Journal of Risk Model Validation (25–44) Volume 5/Number 2, Summer 2011

# Empirical performance of loss given default prediction models

**Benjamin Bade**
Institute of Banking and Finance, Leibniz University of Hannover,
Königsworther Platz 1, 30167 Hannover, Germany;
email: benjamin.bade@finance.uni-hannover.de

**Daniel Rösch**
Institute of Banking and Finance, Leibniz University of Hannover,
Königsworther Platz 1, 30167 Hannover, Germany;
email: daniel.roesch@finance.uni-hannover.de

**Harald Scheule**
Department of Finance, Faculty of Business and Economics,
University of Melbourne, Victoria 3010, Australia;
email: hscheule@unimelb.edu.au

*The global financial crisis highlighted the fact that default and recovery rates of multiple borrowers generally deteriorate jointly during economic downturns. The vast majority of the literature, as well as many industry credit-portfolio risk models, ignore this and analyze default probabilities and recoveries in the event of default separately. As a result, the models project losses that are too low in economic downturns such as the recent financial crisis. Nevertheless, alternatives that incorporate the dependence between probabilities of default and recovery rates have been proposed. This paper is the first of its kind to assess the performance of these structurally different approaches. Four banks using different estimation procedures are compared. We use root mean square errors and relative absolute errors to measure the predictive accuracy of each procedure. The results show that models accounting for the correlation of default and recovery do indeed perform better than models ignoring it.*

## 1 INTRODUCTION

Calculating an accurate measurement of the credit risk underlying defaultable obligations such as loans or bonds is probably one of the most challenging tasks involved in

The authors gratefully acknowledge the support of the Australian Centre for Financial Studies, the Frankfurt Institute for Risk Management and Regulation, the Hong Kong Institute for Monetary Research and the Thyssen Krupp Foundation.

the risk management of a financial institution. The trade-off between complying with the Basel capital requirements and the opportunity costs of tying up too much capital makes this task even more challenging. Appropriate models for the probability of a default event (PD), the exposure at the time of default (EAD) and the loss given a default event (LGD) have to be defined and calibrated by empirical data. In particular, the test of modeling PD and LGD deals with a high level of uncertainty.

Looking at the theoretical and empirical realization of this task in theory as well as in practice, several gaps are identifiable. First of all, there is a wide range of literature on analyzing the drivers of either PD (see, for example, Leland (1994); Jarrow and Turnbull (1995); Longstaff and Schwartz (1995); Madan and Unal (1995); Leland and Toft (1996); Jarrow *et al* (1997); Duffie and Singleton (1999); Shumway (2001); McNeil and Wendin (2007); and Duffie *et al* (2007)) or LGD (see, for example, Carey (1998); Citron *et al* (2003); Dermine and de Carvalho (2006); Acharya *et al* (2007); Altman (2009); Qi and Yang (2009); Grunert and Weber (2009); and Calabrese and Zenga (2010)). Many industry credit-portfolio risk models are also based on isolated modules for default probabilities and recoveries in the event of default. In contrast, approaches to joint modeling and estimation are scarce (exceptions are, for example, Pykhtin (2003); Rösch and Scheule (2005); Kupiec (2008); Bruche and González-Aguado (2010); and Rösch and Scheule (2010)), although empirical data shows that default and recovery rates jointly deteriorate during economic downturns. Figure 1 on the facing page highlights this stylized fact for the recession years 1990 and 1991 (the time of the Persian Gulf War), 2001 and 2002 (the period following the September 11, 2001 terrorist attacks and the general downturn in the US technology industry) as well as 2008 and 2009 (the global financial crisis).

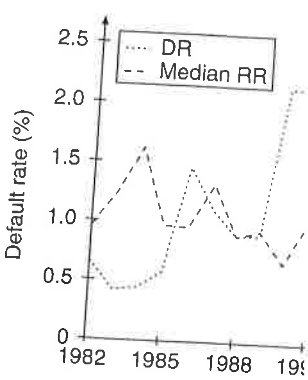
Bade *et al* (2011) provide empirical evidence that default process and recovery process are indeed highly correlated by applying US nonfinancial corporate bond data to an econometric extension of the economic model introduced by Pykhtin (2003).

The second gap in the literature is performance comparisons among the several different approaches to PD and LGD forecasting. Besides the most recent contribution of Qi and Zhao (2011), one exception is Bastos (2010), who compares simple ordinary least squares (OLS) estimation procedures of LGD with a nonparametric regression tree approach on the basis of root mean squared errors (RMSEs) and relative absolute errors (RAEs). Nevertheless, the authors of both papers use data solely from defaulted obligations, as do their predecessors from this strand of literature (see, for example, Bellotti and Crook (2007) and Caselli *et al* (2008)).

This paper addresses these weaknesses by comparing predictions derived from the model by Bade *et al* (2011) with a quick and dirty mean prediction, a simple OLS model and a model incorporating a perfect correlation between default and recovery process as proposed by Rösch and Scheule (2009). Following Bastos (2010) we do this by calculating RMSEs and RAEs for the recovery rate estimates of defaulted bonds.



**FIGURE 1** Default rates and




This figure shows that default and recovery ratio of defaulted bond issues to total bond debt obligations thirty days after the occurr detailed description of the data, see Sectior

In addition, we apply these mea between portfolio default rate a expected loss (EL).

The paper proceeds as follows. ing their estimation and the calcul on the derived parameter estimates framework of our analysis. The res the paper.

## 2 THEORETICAL FRAMEW

### 2.1 The general default and

Generally, we assume that the defa time period $t$ ($i = 1, \dots, N_t, t = 1$ value return $V_{it}$ as introduced by M value return, specified by:

$$V_{it} =$$

crosses a threshold, generally assu observable and deterministic firm-sp factors that influence the asset value

trade-off between complying with
costs of tying up too much capital
ite models for the probability of a
:fault (EAD) and the loss given a
ed by empirical data. In particular,
ı level of uncertainty.
on of this task in theory as well as
. there is a wide range of literature
mple, Leland (1994); Jarrow and
Madan and Unal (1995); Leland
:gleton (1999); Shumway (2001);
or LGD (see, for example, Carey
ho (2006); Acharya *et al* (2007);
Veber (2009); and Calabrese and
nodels are also based on isolated
the event of default. In contrast,
rce (exceptions are, for example,
: (2008); Bruche and González-
hough empirical data shows that
conomic downturns. Figure 1 on
ession years 1990 and 1991 (the
period following the September
ı the US technology industry) as

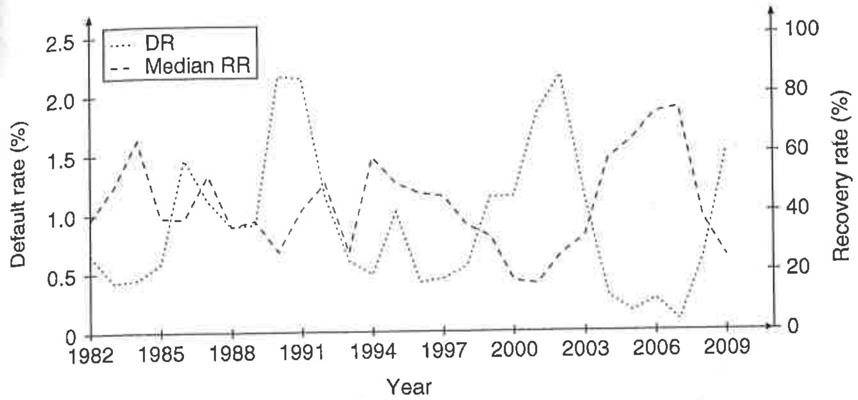
ıt default process and recovery
onfinancial corporate bond data
ntroduced by Pykhtin (2003).
omparisons among the several
des the most recent contribution
who compares simple ordinary
ith a nonparametric regression
(RMSEs) and relative absolute
; use data solely from defaulted
of literature (see, for example,

g predictions derived from the
ean prediction, a simple OLS
between default and recovery
wing Bastos (2010) we do this
estimates of defaulted bonds.





**FIGURE 1** Default rates and recovery rates of nonfinancial bond issues 1982–2009.




This figure shows that default and recovery rates vary over time and are negatively correlated. The default rate is the ratio of defaulted bond issues to total bond issues per year. The recovery rate is the ratio of the price of defaulted debt obligations thirty days after the occurrence of a default event to the par value. Source: Moody's. For a more detailed description of the data, see Section 3.

In addition, we apply these measures to the portfolio level: namely, the difference between portfolio default rate and PD as well as between portfolio loss rate and expected loss (EL).

The paper proceeds as follows. Section 2 briefly introduces the models used, including their estimation and the calculation procedures of the required risk measures based on the derived parameter estimates. In Section 3 we describe the empirical data and the framework of our analysis. The results are presented in Section 4. Section 5 concludes the paper.

## 2 THEORETICAL FRAMEWORK

### 2.1 The general default and recovery process specification

Generally, we assume that the default process of a single borrower or bond issuer $i$ in time period $t$ ($i = 1, \dots, N_t$, $t = 1, \dots, T$) is driven by a normally distributed asset value return $V_{it}$ as introduced by Merton (1974). A default event occurs if the asset value return, specified by:

$$V_{it} = \beta_0 + \beta' x_{it}^V + Z_{it}^V \tag{2.1}$$

crosses a threshold, generally assumed to be zero. $x_{it}^V = (x_{it1}^V, \dots, x_{itK}^V)'$ are $K$ observable and deterministic firm-specific, industry-specific or macroeconomic risk factors that influence the asset value return. $\beta = (\beta_1, \dots, \beta_K)'$ are the sensitivities with respect to these factors and $\beta_0$ is a constant. $Z_{it}^V$ is an idiosyncratic independent and identically distributed $N \sim (0, 1)$ random variable driving the return of borrower $i$'s assets in time period $t$.

Following Bade *et al* (2011) we specify the recovery process by:

$$Y_{it} = \gamma_0 + \gamma' x_{it}^Y + \sigma \rho^U Z_{it}^V + \sigma \sqrt{1 - (\rho^U)^2} Z_{it}^Y \tag{2.2}$$

where $Y_{it}$ is the logarithm of the recovery rate and is thus interpretable as (potential) return on the debt amount outstanding. $x_{it}^Y = (x_{it1}^Y, \dots, x_{itL}^Y)'$ are $L$ deterministic observable risk factors driving the recovery, $\gamma = (\gamma_1, \dots, \gamma_L)'$ represent the loadings of these factors, and $\gamma_0$ is a constant. $Z_{it}^Y$ is independent and identically distributed $N \sim (0, 1)$ and $\sigma$ is a constant parameter. Yet, since $Z_{it}^V$ is part of (2.1) and (2.2), the parameter $\rho^U$ is the correlation between both firm-specific errors as well as the conditional correlation between the asset return and the log-recovery process given the observable covariates.

Besides the possible correlation of the default process and the recovery process introduced in the model presented above, the second feature we would like to introduce is that, in general, the recovery rate of a debt obligation is only observable in the case of default. In order to account for this fact, Bierens (2007) derives a maximum likelihood procedure to simultaneously estimate the parameters for such a statistical model firstly introduced by Heckman (1979). The log-likelihood for a single observation $i$ in period $t$ takes the following form:

$$\begin{aligned} \mathcal{L}_{it} = {} & (1 - d_{it}) \ln \Phi(\beta_0 + \beta' x_{it}^V) + d_{it} \ln(1 - \Phi(\beta_0 + \beta' x_{it}^V)) \\ & + d_{it} \ln \frac{\phi((y_{it} - (\gamma_0 + \gamma' x_{it}^Y))/\sigma)}{\sigma(1 - \Phi(\beta_0 + \beta' x_{it}^V))} \\ & + d_{it} \ln \left(1 - \Phi\left[\frac{(\rho^U/\sigma)(y_{it} - (\gamma_0 + \gamma' x_{it}^Y)) + (\beta_0 + \beta' x_{it}^V)}{\sqrt{1 - (\rho^U)^2}}\right]\right) \end{aligned} \tag{2.3}$$

$\phi(\cdot)$ specifies the density function and $\Phi(\cdot)$ the cumulative distribution function of the standard normal distribution. $d_{it}$ indicates whether the observed obligation defaults ($d_{it} = 1$) or not ($d_{it} = 0$). Thus, all parameters may be estimated without the knowledge of values for $V_{it}$. Equation (2.3) is then maximized over $n_t$ observations per period and $T$ periods:

$$\ell = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathcal{L}_{it} \tag{2.4}$$

## 2.2 Model assumptions and consequences

For the general framework presented above two restrictive assumptions are of particular interest. The first one is the assumption that, conditional on given realizations of the observable risk factors, both processes are uncorrelated, ie, $\rho^U = 0$. In this case the observed log-recoveries are normally distributed (see the dark bars in Figure 2 on the next page). The assumption of uncorrelated error terms allows a separate estimation of the parameters underlying both processes in the model, since $\mathcal{L}_{it}$ simplifies to:

$$\mathcal{L}_{it}^{\text{uncorr}} = \underbrace{(1 - d_{it}) \ln \Phi(\beta_0 + \beta' x_{it}^V) + d_{it} \ln(1 - \Phi(\beta_0 + \beta' x_{it}^V))}_{\mathcal{L}_{it}^{\text{probit}}} + \underbrace{d_{it} \ln \frac{\phi((y_{it} - (\gamma_0 + \gamma' x_{it}^Y))/\sigma)}{\sigma}}_{\mathcal{L}_{it}^{\text{recovery}}} \quad (2.5)$$

The parameters of $\mathcal{L}_{it}^{\text{probit}}$ are estimated by a standard probit procedure via maximum likelihood (see, for example, Gordy and Heitfield (2000); Gordy and Heitfield (2002); and Hamerle *et al* (2003)):

$$\ell^{\text{probit}} = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathcal{L}_{it}^{\text{probit}} \quad (2.6)$$

Due to the independence of the recovery process from the default process, the parameters of $\mathcal{L}_{it}^{\text{recovery}}$ need not necessarily be estimated via maximum likelihood. For convenience, a simple OLS regression of the observed log-recoveries may be run.[1]

The second possible restrictive assumption to the model is that default and recovery processes are perfectly positively correlated, ie, $\rho^U = 1$, and that $\beta_0 = \gamma_0/\sigma$ as well as $\beta = \gamma/\sigma$. In other words, both processes are driven by the same explanatory variables and each variable has the same standardized exposure in both processes. Thus, the default barrier translates into a cutoff point for the observed log-recoveries. Their distribution equals a truncated normal distribution (see the lighter bars in Figure 2 on the next page).

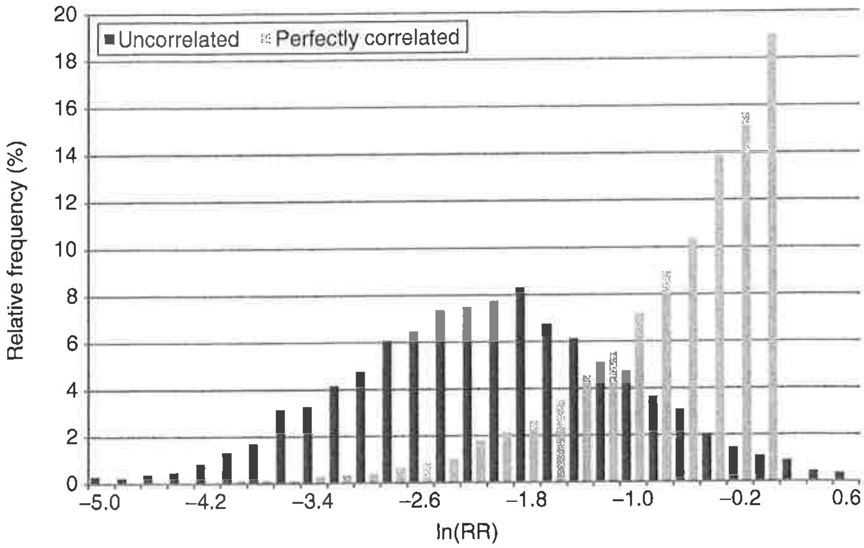
The log-likelihood for a single observation under this restriction simplifies to the log-likelihood of a Tobit model:[2]

$$\mathcal{L}_{it}^{\text{Tobit}} = (1 - d_{it}^{\text{Tobit}}) \ln \Phi\left(\frac{\gamma_0 + \gamma' x_{it}^Y}{\sigma}\right) + d_{it}^{\text{Tobit}} \ln \frac{\phi((y_{it} - (\gamma_0 + \gamma' x_{it}^Y))/\sigma)}{\sigma} \quad (2.7)$$

[1] Please note that many other transformations of the recovery rates, such as logit or probit, are possible (see, for example, Dermine and de Carvalho (2006) or Bastos (2010)), but in order to ensure that results are comparable to the unrestricted model we focus on the logarithmic transformation.

[2] For the derivation of such a likelihood, see Bierens (2004). For an empirical application for bond defaults and recoveries, see Rösch and Scheule (2009).

**FIGURE 2** Distributions of observable log-recoveries for a sample portfolio of 100 000 obligors and differently correlated error terms.




This figure presents distributions of log-recoveries for defaulted obligors in a sample portfolio of 100 000 obligors under different assumptions concerning the correlation between default and recovery process. The underlying parameters of the simulation for uncorrelated error terms (dark bars), ie, $\rho^U = 0$, are $\beta_0 = 1.6449$ (which corresponds to a PD of 5%), $\gamma_0 = -2.3551$ and $\sigma = 1$. The underlying parameters of the simulation for perfectly correlated error terms (light bars), ie, $\rho^U = 1$, are $\beta_0 = 1.6449$ (which corresponds to a PD of 5%), $\gamma_0 = 2.46735$ and $\sigma = 1.5$.

Since the default barrier is generally assumed to be zero, the truncation of the log-recoveries is made at zero too. Nevertheless, real data may contain recovery rates greater than 1, ie, log-recoveries greater than 0. These observations should be treated as nondefaults, such that $d_{it}^{\text{Tobit}} \neq d_{it}$ in these cases. The maximum likelihood function is:

$$\ell^{\text{Tobit}} = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathcal{L}_{it}^{\text{Tobit}} \tag{2.8}$$

## 2.3 Calculation of risk measures

In order to predict the risk of a debt obligation, the parameters derived by the methods presented above are only of secondary interest. The primary risk measures of importance are the PD, the EL and the recovery rate in the case where such an obligation defaults (expected recovery given default (ERGD)). Generally, these three ratios are linked by:

$$\text{ERGD}_{it} = 1 - \frac{\text{EL}_{it}}{\text{PD}_{it}} \tag{2.9}$$

Since we assume an asset value process for the default event, the PD is given as the probability that $V_{it}$ falls below zero (given the observable covariates). Under the normality assumption, we obtain:

$$\mathrm{PD}_{it} = 1 - \Phi(\beta_0 + \beta' x_{it}^V) \tag{2.10}$$

For EL and ERGD, respectively, the assumptions concerning the link between default and recovery process have to be considered. In the general case the parameter estimates of (2.4) are used to calculate the expected loss by:

$$\begin{aligned}\mathrm{EL}_{it}^{\mathrm{general}} &= \Phi_2\biggl[-(\beta_0 + \beta' x_{it}^V), -\frac{\gamma_0 + \gamma' x_{it}^Y}{\sigma}, \rho^U\biggr] \\ &\quad - \exp(\gamma_0 + \gamma' x_{it}^Y + \tfrac{1}{2}\sigma^2) \\ &\qquad \times \Phi_2\biggl[-(\beta_0 + \beta' x_{it}^V) - \sigma\rho^U, -\frac{\gamma_0 + \gamma' x_{it}^Y}{\sigma} - \sigma, \rho^U\biggr]\end{aligned} \tag{2.11}$$

$\Phi_2[\cdot,\cdot,\cdot]$ represents the distribution function of the bivariate normal distribution. For the more restrictive case of uncorrelated error terms it is most convenient to calculate the expected recovery given default first and the expected loss afterward by applying the parameter estimates of (2.6) to the PD and rearranging (2.9). If the parameters of the recovery process with log-recoveries as a dependent variable are estimated by simple OLS, ERGD is calculated by:

$$\mathrm{ERGD}_{it}^{\mathrm{OLS}} = \exp(\gamma_0 + \gamma' x_{it}^Y + 0.5) \tag{2.12}$$

With the parameters derived under the assumptions of the Tobit approach in (2.7) we obtain EL by:

$$\mathrm{EL}_{it}^{\mathrm{Tobit}} = \Phi\biggl(-\frac{\gamma_0 + \gamma' x_{it}^Y}{\sigma}\biggr) - \exp(\gamma_0 + \gamma' x_{it}^Y + 0.5\sigma^2)\Phi\biggl(-\frac{\gamma_0 + \gamma' x_{it}^Y + \sigma^2}{\sigma}\biggr) \tag{2.13}$$

Please note that in the Tobit case the PD is computed as:

$$\mathrm{PD}_{it}^{\mathrm{Tobit}} = \Phi\biggl(-\frac{\gamma_0 + \gamma' x_{it}^Y}{\sigma}\biggr) \tag{2.14}$$

## 3 DATA AND METHODOLOGY OF THE PERFORMANCE COMPARISON

### 3.1 Default and recovery data

The data sample underlying the empirical analysis is provided by Moody's credit rating agency and is the same as the one used by Bade *et al* (2011). The data set

**TABLE 1** Number of observations, default rate and mean recovery.

| Rating | $N_{obs}$ | Percentage of all observations | $N_{def}$ | Percentage of all default observations | DR | RRGD$_{\emptyset}$ |
|---|---|---|---|---|---|---|
| IG | 146 582 | 78.120 | 51 | 3.074 | 0.035 | 46.823 |
| Ba | 15 262 | 8.134 | 87 | 5.244 | 0.570 | 48.607 |
| B | 20 132 | 10.729 | 530 | 31.947 | 2.633 | 39.890 |
| C | 5 662 | 3.018 | 991 | 59.735 | 17.503 | 34.836 |
| Total | 187 638 | 100.000 | 1659 | 100.000 | 0.884 | 37.541 |

This table reports descriptive statistics on defaults and recoveries of nonfinancial bonds from 1982 to 2009. The data set provided by Moody's is split up into four rating categories: investment grade (IG), which contains all observations with a Moody's rating higher than Ba; those with a rating of Ba; a rating of B; and a rating of C, which contains all observations with a Moody's rating lower than B. $N_{obs}$ is the number of observations. $N_{def}$ is the number of defaults. DR (default rate) is the ratio of the number of defaults to the number of observations in each rating grade. RRGD$_{\emptyset}$ is the mean recovery rate of the defaulted bonds in each rating grade. The recovery rate is the ratio of the price of defaulted debt obligations thirty days after the occurrence of a default event to par value.

contains the annual ratings of regular US bond issues, as well as default dates and recovery rates given default. Moody's records a default event if interest or principal payments are missed or delayed, Chapter 11 or Chapter 7 bankruptcy is filed or a distressed exchange, such as a reduction in a financial obligation, occurs. The recovery rate is equal to the price of a defaulted bond measured thirty days after a default event in relation to the face value of the bond.

Table 1 summarizes important descriptive statistics for the data set, which consists of 187 638 observations for regular US bond issues of nonfinancial institutions from 1982 to 2009. Coincident with a change in Moody's rating methodology in 1982 and the role of ratings in the subsequent analysis, earlier observations are excluded from this empirical study.

During the observation period, a total of 1659 defaults occurred, which yields a default rate (DR) of 0.884%. The mean recovery rate for all defaulted bonds is 37.541%; the median recovery rate is 32%.

Table 1 also shows the descriptive statistics per rating category: all bond issues with a rating higher than Ba are aggregated to an investment grade (IG) rating, and all bond issues with a rating lower than B are aggregated to rating C. This categorization addresses the limited number of default events in the subcategories. The table shows that, as one may expect, the default rate increases from rating IG to C. The mean recovery rate decreases from rating IG to C, except for grades Ba (48.607%) and IG (46.823%), which may be due to the small number of defaults, and hence the small number of recovery events in both grades.

Since the rating grade as well as the rating shift in the year prior to the observed rating status ($\text{rating}_{it} - \text{rating}_{it-1}$) are statistically and economically significant for



the data set, we include rat
shift as explanatory variabl
In order to account for th
in Figure 1 on page 27, we ir
variable in the study.
Since all explanatory vari
quantities when predicting I

## 3.2 Model validation fr

In the empirical study we c
projection of future defaults

- Bank 1 simply estimat
probably the most conv
predicting future defau
- Bank 2 follows the res
the probit approach of
industry-specific and n
marginal effect of each
regard to LGD forecasts
arithm of the recovery r
- Bank 3 uses the Tobit ap
the historical data.
- Bank 4 uses the general
simultaneously.

In detail, the model validatior
of five steps, which are repeatec

- Step 1: we select 90% of
10% of the data as out-of-

[3] Since, for bonds originated in the yea
(MV), we include a dummy variable
Through this, we are able to keep th
observations with $\text{rating}_{it} - \text{rating}_{it-}$

recovery.

| ige | | |
|---|---|---|
| t ons | DR | RRGD$_\emptyset$ |
| 4 | 0.035 | 46.823 |
| 4 | 0.570 | 48.607 |
| 7 | 2.633 | 39.890 |
| 5 | 17.503 | 34.836 |
| 0 | 0.884 | 37.541 |

ıcial bonds from 1982 to 2009. The data ɔe (IG), which contains all observations B; and a rating of C, which contains all rvations. $N_{\text{def}}$ is the number of defaults. ervations in each rating grade. RRGD$_\emptyset$ recovery rate is the ratio of the price of to par value.

as well as default dates and event if interest or principal er 7 bankruptcy is filed or a igation, occurs. The recovery irty days after a default event

ır the data set, which consists ıonfinancial institutions from ng methodology in 1982 and servations are excluded from

ıults occurred, which yields te for all defaulted bonds is

ng category: all bond issues ıent grade (IG) rating, and all rating C. This categorization bcategories. The table shows m rating IG to C. The mean grades Ba (48.607%) and IG ɟefaults, and hence the small

he year prior to the observed economically significant for





the data set, we include rating dummies as well as an ordinal variable for the rating shift as explanatory variables in the empirical study.[3]

In order to account for the time series variation of default and recovery rates shown in Figure 1 on page 27, we include the lagged change of GPDI as a further explanatory variable in the study.

Since all explanatory variables are lagged by one year, they can be treated as known quantities when predicting PD, EL and ERGD.

## 3.2 Model validation framework

In the empirical study we compare four banks with competing approaches to the projection of future defaults and losses:

- Bank 1 simply estimates PD, EL and ERGD by historical averages, which is probably the most convenient but most likely also the least accurate method for predicting future default or recovery rates.
- Bank 2 follows the restrictive approach of (2.5), ie, it estimates the PD with the probit approach of (2.6), which allows an incorporation of firm-specific, industry-specific and macroeconomic covariates and an explanation of the marginal effect of each considered variable on the likelihood of a default. With regard to LGD forecasts, Bank 2 uses an OLS regression with the natural logarithm of the recovery rate of defaulted bonds as dependent variable.
- Bank 3 uses the Tobit approach of (2.7) to obtain the relevant parameters from the historical data.
- Bank 4 uses the general Heckman approach of (2.3) to forecast PD and LGD simultaneously.

In detail, the model validation framework for our performance comparison consists of five steps, which are repeated 10 000 times in order to exclude sample effects:

- Step 1: we select 90% of the data as a random sample and treat the remaining 10% of the data as out-of-sample.

[3] Since, for bonds originated in the year of observation, $\text{rating}_{it} - \text{rating}_{it-1}$ yields a missing value (MV), we include a dummy variable for these observations and set $\text{rating}_{it} - \text{rating}_{it-1} = 0$. Through this, we are able to keep these observations in the data set and differentiate between observations with $\text{rating}_{it} - \text{rating}_{it-1} = 0$ and $\text{rating}_{it} - \text{rating}_{it-1} = \text{MV}$.

**TABLE 2** Results for RMSE on the recovery rate level.

(a) $\text{RMSE}_{\text{RR}}^{\text{in-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.28623<br>(0.00060) | 0.27280<br>(0.00130) | 0.26855<br>(0.00126) | 0.26632<br>(0.00134) |
| 2 | 0.28623<br>(0.00059) | 0.27281<br>(0.00130) | 0.26821<br>(0.00123) | 0.26592<br>(0.00133) |
| 3 | 0.28623<br>(0.00060) | 0.27231<br>(0.00133) | 0.26864<br>(0.00127) | 0.26553<br>(0.00137) |
| 4 | 0.28623<br>(0.00060) | 0.27224<br>(0.00134) | 0.26833<br>(0.00126) | 0.26520<br>(0.00138) |

(b) $\text{RMSE}_{\text{RR}}^{\text{out-of-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.28633<br>(0.01386) | 0.27380<br>(0.01195) | 0.26842<br>(0.01167) | 0.26665<br>(0.01258) |
| 2 | 0.28631<br>(0.01394) | 0.27402<br>(0.01152) | 0.26808<br>(0.01141) | 0.26636<br>(0.01241) |
| 3 | 0.28629<br>(0.01406) | 0.27373<br>(0.01202) | 0.26851<br>(0.01171) | 0.26617<br>(0.01281) |
| 4 | 0.28630<br>(0.01413) | 0.27388<br>(0.01218) | 0.26823<br>(0.01167) | 0.26598<br>(0.01279) |

RMSE is calculated by (3.1). Standard deviations are reported in parentheses.

- Step 2: with the in-sample data we estimate the relevant parameters of the models underlying the banks' prediction techniques. For each model, we investigate four different specifications containing the following explanatory variables:
  - Specification 1: ratings.
  - Specification 2: ratings and lagged GPDI change.
  - Specification 3: ratings and rating shift.
  - Specification 4: ratings, rating shift and lagged GPDI change.
- Step 3: these parameters are incorporated to estimate PD, ERGD and EL for each observation of the in-sample data set as well as for the out-of-sample data set.

| ank 3 | Bank 4 |
|---|---|
| 26855 | 0.26632 |
| 0126) | (0.00134) |
| 26821 | 0.26592 |
| 0123) | (0.00133) |
| 26864 | 0.26553 |
| 0127) | (0.00137) |
| 26833 | 0.26520 |
| 0126) | (0.00138) |

| ank 3 | Bank 4 |
|---|---|
| 26842 | 0.26665 |
| 01167) | (0.01258) |
| 26808 | 0.26636 |
| 01141) | (0.01241) |
| 26851 | 0.26617 |
| 01171) | (0.01281) |
| 26823 | 0.26598 |
| 01167) | (0.01279) |

eses.

elevant parameters of the mod- For each model, we investigate wing explanatory variables.

change.

gged GPDI change.

stimate PD, ERGD and EL for well as for the out-of-sample





- Step 4: on the single borrower level we follow the approach by Bastos (2010) and compare the realized recovery rates of the defaulted bonds $\mathrm{RR}_{jt}$, where $j = 1, \dots, n_t^{\mathrm{def}}$, in each data subset with their estimates via RMSE:

$$\mathrm{RMSE}_{\mathrm{RR}} = \sqrt{\left(\sum_{t=1}^{T} n_t^{\mathrm{def}}\right)^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n_t^{\mathrm{def}}} (\mathrm{RR}_{jt} - \mathrm{ERGD}_{jt}^{\mathrm{model}})^2} \tag{3.1}$$

and RAE:

$$\mathrm{RAE}_{\mathrm{RR}} = \frac{\sum_{t=1}^{T} \sum_{j=1}^{n_t^{\mathrm{def}}} |\mathrm{RR}_{jt} - \mathrm{ERGD}_{jt}^{\mathrm{model}}|}{\sum_{t=1}^{T} \sum_{j=1}^{n_t^{\mathrm{def}}} |\mathrm{RR}_{jt} - \mathrm{ERGD}_{jt}^{\mathrm{simple}}|} \cdot 100 \tag{3.2}$$

RMSE measures the accuracy of the estimates in absolute terms while RAE measures the accuracy relative to a benchmark estimator. For convenience we use the arithmetic mean of the realized recovery rates calculated by Bank 1 for the corresponding rating grade of each observation as a simple predictor.

- Step 5: on the portfolio level we aggregate the PDs and ELs of the borrowers in both subsamples to portfolio PDs and ELs by:

$$\mathrm{PD}^{\mathrm{PF}} = \left(\sum_{t=1}^{T} n_t\right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathrm{PD}_{it} \tag{3.3}$$

and:

$$\mathrm{EL}^{\mathrm{PF}} = \left(\sum_{t=1}^{T} n_t\right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n_t} \mathrm{EL}_{it} \tag{3.4}$$

Since we only get one value per risk measure and portfolio that is compared with the realized portfolio default rate and portfolio loss rate, respectively, we have to calculate RMSE and RAE over the 10 000 iterations of this random sampling procedure. We do this for the out-of-sample portfolio.

## 4 RESULTS

### 4.1 Single borrower level

Table 2 on the facing page shows the RMSEs in-sample and out-of-sample by bank and specification. On average, the least accurate predictive power is reached by using the arithmetic mean of observed recovery rates (Bank 1) as a forecast for ERGD in-sample as well as out-of-sample. Despite the highest average RMSE, the standard

**TABLE 3** Results for RAE on the recovery rate level.

(a) $\text{RAE}_{\text{RR}}^{\text{in-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 99.067 | 97.275 | 95.026 |
| | — | (0.417) | (0.433) | (0.590) |
| 2 | 100 | 99.121 | 97.121 | 94.897 |
| | — | (0.413) | (0.424) | (0.571) |
| 3 | 100 | 98.367 | 96.856 | 94.271 |
| | — | (0.418) | (0.437) | (0.605) |
| 4 | 100 | 98.376 | 96.841 | 94.240 |
| | — | (0.420) | (0.437) | (0.621) |

(b) $\text{RAE}_{\text{RR}}^{\text{out-of-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 99.657 | 97.477 | 95.323 |
| | — | (7.272) | (6.525) | (5.712) |
| 2 | 100 | 99.785 | 97.322 | 95.217 |
| | — | (7.142) | (6.399) | (5.600) |
| 3 | 100 | 99.114 | 97.070 | 94.654 |
| | — | (7.236) | (6.451) | (5.537) |
| 4 | 100 | 99.214 | 97.065 | 94.664 |
| | — | (7.385) | (6.523) | (5.611) |

RAE is calculated by (3.2). Standard deviations are reported in parentheses.

deviation for the in-sample RMSEs of Bank 1 (0.0006) is the lowest of all four banks in each specification. In contrast, the out-of-sample standard deviation of Bank 1's RMSEs is the highest. Thus, Bank 1 not only has the least accurate method to predict future (ie, out-of-sample) recovery rates for defaulted bonds on average, but also the most insecure method.

Using a simple OLS regression and calculating ERGDs on the basis of the regression results yields improved results compared with Bank 1's approach. The RMSEs are reduced on average and for the out-of-sample data in standard deviation, too. The more elaborate the model specification, the lower the average $\text{RMSE}_{\text{RR}}^{\text{in-sample}}$. Out-of-sample, adding GPDI to the regression model, ie, switching from Specification 1 to 2 or from 3 to 4, reduces the predictive accuracy. With the exception of switching from Specification 1 to 2 for the out-of-sample data, the standard deviation of the RMSEs increases with the number of variables taken into account in both subsamples.

| nk 3 | Bank 4 |
|---|---|
| 275 | 95.026 |
| 433) | (0.590) |
| 121 | 94.897 |
| 424) | (0.571) |
| 856 | 94.271 |
| .437) | (0.605) |
| .841 | 94.240 |
| .437) | (0.621) |

| nk 3 | Bank 4 |
|---|---|
| .477 | 95.323 |
| .525) | (5.712) |
| .322 | 95.217 |
| .399) | (5.600) |
| .070 | 94.654 |
| .451) | (5.537) |
| .065 | 94.664 |
| .523) | (5.611) |

es.

ı is the lowest of all four banks tandard deviation of Bank 1's ast accurate method to predict bonds on average, but also the

GDs on the basis of the regres- ınk 1's approach. The RMSEs in standard deviation, too. The verage $RMSE_{RR}^{in\text{-}sample}$. Out-of- ching from Specification 1 to 2 ıe exception of switching from ndard deviation of the RMSEs ıunt in both subsamples.





**TABLE 4** Robustness check results for RMSE on the recovery rate level.

(a) $RMSE_{RR}^{in\text{-}sample}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.28625<br>(0.00066) | 0.27267<br>(0.00145) | 0.26810<br>(0.00143) | 0.26618<br>(0.00151) |
| 2 | 0.28625<br>(0.00065) | 0.27251<br>(0.00144) | 0.26761<br>(0.00142) | 0.26596<br>(0.00151) |
| 3 | 0.28625<br>(0.00066) | 0.27167<br>(0.00145) | 0.26700<br>(0.00145) | 0.26482<br>(0.00152) |
| 4 | 0.28625<br>(0.00066) | 0.27152<br>(0.00145) | 0.26674<br>(0.00143) | 0.26487<br>(0.00153) |

(b) $RMSE_{RR}^{out\text{-}of\text{-}sample}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.28638<br>(0.01579) | 0.27407<br>(0.01324) | 0.26793<br>(0.01318) | 0.26649<br>(0.01400) |
| 2 | 0.28638<br>(0.01566) | 0.27421<br>(0.01329) | 0.26746<br>(0.01315) | 0.26639<br>(0.01390) |
| 3 | 0.28638<br>(0.01582) | 0.27335<br>(0.01359) | 0.26685<br>(0.01339) | 0.26527<br>(0.01416) |
| 4 | 0.28637<br>(0.01553) | 0.27351<br>(0.01355) | 0.26658<br>(0.01328) | 0.26543<br>(0.01396) |

RMSE is calculated by (3.1). Standard deviations are reported in parentheses.

The Tobit procedure used by Bank 3 yields a further improvement of the results compared with Banks 1 and 2. Nevertheless, some qualitative differences are apparent. In contrast to Bank 2, the incorporation of GPDI yields more accurate recovery rate forecasts; RMSE decreases on average as well as its standard deviation. On the other hand, incorporating the rating shift increases RMSE as well as its standard deviation and thus lowers the predictive power of the Tobit model for recovery rates. Unusually, the average $RMSE_{RR}^{out\text{-}of\text{-}sample}$ is lower than the average $RMSE_{RR}^{in\text{-}sample}$ for each model specification.

The Heckman model implemented by Bank 4 yields the best results for the average RMSE. The predictive power rises by adding more explanatory variables to the model specification. The standard deviation of RMSE, though, is higher than for the other banks' predictions in-sample and higher than Bank 2's predictions out-of-sample. Due

**TABLE 5** Robustness check results for RAE on the recovery rate level.

(a) $\mathrm{RAE}_{\mathrm{RR}}^{\text{in-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 98.641 | 96.450 | 94.778 |
| | — | (0.470) | (0.490) | (0.617) |
| 2 | 100 | 98.575 | 96.225 | 94.611 |
| | — | (0.469) | (0.494) | (0.612) |
| 3 | 100 | 97.864 | 95.648 | 93.879 |
| | — | (0.474) | (0.494) | (0.606) |
| 4 | 100 | 97.826 | 95.576 | 93.984 |
| | — | (0.473) | (0.500) | (0.618) |

(b) $\mathrm{RAE}_{\mathrm{RR}}^{\text{out-of-sample}}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 99.440 | 96.700 | 95.121 |
| | — | (8.342) | (7.253) | (6.517) |
| 2 | 100 | 99.483 | 96.485 | 94.989 |
| | — | (8.381) | (7.310) | (6.615) |
| 3 | 100 | 98.778 | 95.916 | 94.264 |
| | — | (8.516) | (7.365) | (6.573) |
| 4 | 100 | 98.830 | 95.838 | 94.389 |
| | — | (8.407) | (7.290) | (6.577) |

RAE is calculated by (3.2). Standard deviations are reported in parentheses.

to the computational complexity of the likelihood function, too little recovery data in the sample might be an explanation for a higher number of outliers for the recovery rate estimates compared with the other models. Since such outliers have a higher loading in a quadratic measure like RMSE than for a measure based on the absolute value like RAE, the distribution of RMSE itself is more sensitive to these. Thus, a higher standard deviation of RMSEs, which itself is the square root of a quadratic measure, is likely to be caused by this connection.

The results for $\mathrm{RAE}_{\mathrm{RR}}^{\text{in-sample}}$ and $\mathrm{RAE}_{\mathrm{RR}}^{\text{out-of-sample}}$ presented in Table 3 on page 36 broadly confirm the results above. Relative to the results of Bank 1, the Heckman model performs best, followed by the Tobit approach and the OLS approach, which only performs a little better than the historical average. It is notable that in-sample the standard deviation of the RAEs increases with decreasing average RAE, while out-of-sample the result is the opposite.



**TABLE 6** Results for RMSE

| Specification |
|---|
| 1 |
| 2 |
| 3 |
| 4 |

| Specification |
|---|
| 1 |
| 2 |
| 3 |
| 4 |

In order to check whether t
that have a missing value (62
the study. Table 4 on page 3
this robustness check for RM
unchanged. In absolute term
for Bank 1 rises on average,
three banks. Relative to Banl
for the whole data set, as Tab

## 4.2 Portfolio level

The results of the performan
in Table 6.[4] The probit appro
yield almost identical RMSEs
Tobit approach yields the wor
specifications. Specification 4
between Bank 2 and Bank 4

[4] Note that the portfolio loss rate m
considered in this paper. This woul
to capture the comovement of defa
models.

very rate level.

| ank 3 | Bank 4 |
|---|---|
| .450 | 94.778 |
| .490) | (0.617) |
| .225 | 94.611 |
| .494) | (0.612) |
| .648 | 93.879 |
| .494) | (0.606) |
| .576 | 93.984 |
| .500) | (0.618) |

| ank 3 | Bank 4 |
|---|---|
| .700 | 95.121 |
| .253) | (6.517) |
| .485 | 94.989 |
| .310) | (6.615) |
| .916 | 94.264 |
| .365) | (6.573) |
| .838 | 94.389 |
| .290) | (6.577) |

es.

tion, too little recovery data in
er of outliers for the recovery
e such outliers have a higher
measure based on the absolute
ore sensitive to these. Thus, a
the square root of a quadratic

sented in Table 3 on page 36
ults of Bank 1, the Heckman
and the OLS approach, which
e. It is notable that in-sample
creasing average RAE, while





**TABLE 6** Results for RMSE and RAE of the portfolio default rate.

(a) $RMSE_{DR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.000668 | 0.000668 | 0.000674 | 0.000669 |
| 2 | 0.000679 | 0.000675 | 0.000680 | 0.000675 |
| 3 | 0.000681 | 0.000676 | 0.000684 | 0.000677 |
| 4 | 0.000684 | 0.000675 | 0.000683 | 0.000676 |

(b) $RAE_{DR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 100.000 | 100.591 | 99.941 |
| 2 | 100 | 99.256 | 99.837 | 99.192 |
| 3 | 100 | 99.099 | 100.207 | 99.227 |
| 4 | 100 | 98.668 | 99.652 | 98.746 |

In order to check whether the results are data specific, we excluded all observations that have a missing value (62 990 observations) for $\text{rating}_{it} - \text{rating}_{it-1}$ and repeated the study. Table 4 on page 37 and Table 5 on the facing page present the results of this robustness check for RMSE and RAE. Qualitatively, the previous results remain unchanged. In absolute terms the data reduction has contrary effects. While RMSE for Bank 1 rises on average, it decreases for almost every specification of the other three banks. Relative to Bank 1, each of the other three banks performs better than for the whole data set, as Table 5 on the facing page shows.

## 4.2 Portfolio level

The results of the performance analysis for the portfolio default rate are presented in Table 6.[4] The probit approach of Bank 2 and the Heckman approach of Bank 4 yield almost identical RMSEs, which are lower than for Bank 1 and Bank 3. Bank 3's Tobit approach yields the worst predictions of all four banks for the first three model specifications. Specification 4 shows a slightly higher $RMSE_{DR}$ for Bank 1. The draw between Bank 2 and Bank 4 is confirmed by $RAE_{DR}$. For Specifications 1 and 2

[4] Note that the portfolio loss rate may also be compared with the value-at-risk of the LGD models considered in this paper. This would require accounting for an unobservable systematic risk factor to capture the comovement of default and recovery processes. Bade *et al* (2011) introduce such models.

**TABLE 7** Results for RMSE and RAE of the portfolio loss rate.

(a) $RMSE_{LR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.000462 | 0.000616 | 0.000550 | 0.000463 |
| 2 | 0.000466 | 0.000618 | 0.000554 | 0.000466 |
| 3 | 0.000463 | 0.000626 | 0.000557 | 0.000465 |
| 4 | 0.000470 | 0.000629 | 0.000560 | 0.000469 |

(b) $RAE_{LR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 136.607 | 120.479 | 100.912 |
| 2 | 100 | 135.682 | 120.010 | 99.967 |
| 3 | 100 | 137.819 | 121.082 | 100.214 |
| 4 | 100 | 137.341 | 120.604 | 100.197 |

the Heckman approach yields a lower value for $RAE_{DR}$ and for the remaining two specifications the probit approach is advantageous.

The results for the portfolio loss rate shown in Table 7 are much more widespread. Here, the simple prediction by historical average is the best predictor for future portfolio loss rates, followed closely by the Heckman approach of Bank 4. The Tobit approach performs rather poorly with a 20% worse loss estimation against the historical average, indicating that default and recovery process are not perfectly correlated. The worst performance is reached by Bank 2, with more than 35% fewer accurate loss rate predictions. Thus, an estimation of two separate models for PD and LGD followed by a calculation of the expected loss based on the parameters derived from both models is not suitable. It results in a high degree of misspecification, since the possible correlation between the processes is ignored.

We provide the same robustness check on the portfolio level as on the single borrower level. Table 8 on the facing page shows the results. Due to the data reduction, $RMSE_{DR}$ and $RMSE_{LR}$ deteriorate for all four banks. The draw between Bank 2 and Bank 4 concerning the default rate forecast switches to a marginal advantage for the probit approach. The portfolio loss rate predictions of Banks 2–4 relative to Bank 1 improve compared with the primary results. Yet the Heckman approach yields the best predictions if more explanatory variables than the rating grade are taken into consideration.

**TABLE 8** Robustness check results for RMSE and RAE on the portfolio level.

(a) $RMSE_{DR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.000874 | 0.000874 | 0.000878 | 0.000874 |
| 2 | 0.000862 | 0.000859 | 0.000863 | 0.000859 |
| 3 | 0.000875 | 0.000868 | 0.000874 | 0.000870 |
| 4 | 0.000876 | 0.000867 | 0.000872 | 0.000868 |

(b) $RAE_{DR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 100 000 | 100 306 | 99 966 |
| 2 | 100 | 99 628 | 99 988 | 99 631 |
| 3 | 100 | 99 263 | 99 804 | 99 374 |
| 4 | 100 | 98 882 | 99 366 | 98 975 |

(c) $RMSE_{LR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 0.000595 | 0.000777 | 0.000664 | 0.000599 |
| 2 | 0.000587 | 0.000766 | 0.000652 | 0.000586 |
| 3 | 0.000600 | 0.000782 | 0.000666 | 0.000595 |
| 4 | 0.000591 | 0.000773 | 0.000658 | 0.000587 |

(d) $RAE_{LR}$

| Specification | Bank 1 | Bank 2 | Bank 3 | Bank 4 |
|---|---|---|---|---|
| 1 | 100 | 131 485 | 110 730 | 100 292 |
| 2 | 100 | 131 982 | 110 707 | 99 698 |
| 3 | 100 | 132 489 | 111 229 | 99 334 |
| 4 | 100 | 133 427 | 111 760 | 99 252 |

## 5 CONCLUSION

Various work in the literature on default rates and recovery rates, as well as recent contributions suggesting a joint modeling of both variables, shows the high complexity of these quantities and the challenge involved in obtaining an accurate measurement. While many previous contributions focused on the qualitative and quantitative drivers of both variables, this paper compares the predictive performance of several modeling approaches. RMSEs and RAEs are calculated for four banks, with each bank using a different approach to forecast future defaults and losses. In order to check their contribution to the predictive power of each bank's approach, four different combinations of explanatory variables are investigated.

The results show that a disjunct consideration of default and recovery ignoring the high correlation between both quantities yields not only biased parameter estimates, but also a worse predictive power for future losses than the general approach applied by Bade *et al* (2011). Especially on the portfolio loss level, the relative inaccuracy is severe.

While the portfolio default rate estimates may not be considered as significantly differing among the four banks, the portfolio loss rate and the recovery rate of a single borrower are predicted best with the general model allowing default and recovery to be correlated. Nevertheless, the quick and dirty solution also yields a relatively accurate measure of the future portfolio loss rate.

Thus, accounting for the high correlation of default and recovery rates highlighted during past economic downturns – most recently by the global financial crisis – is a necessary condition for a suitable credit risk model. This paper provides further evidence that the model suggested by Pykhtin (2003) and adopted by Bade *et al* (2011) is a suitable model fulfilling this requirement.

## REFERENCES


Acharya, V. V., Bharath, S. T., and Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics* **85**(3), 787–821.

Altman, E. (2009). Default recovery rates and LGD in credit risk modeling and practice: an updated review of the literature and empirical evidence. Working Paper, Stern School of Business, New York University.

Bade, B., Rösch, D., and Scheule, H. (2011). Default and recovery risk dependencies in a simple credit risk model. *European Financial Management* **17**(1), 120–144.

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

Bellotti, T., and Crook, J. (2007). Modelling and predicting loss given default for credit cards. Working Paper, University of Edinburgh.

Bierens, H. J. (2004). The Tobit model. Working Paper, Pennsylvania State University.

Bierens, H. J. (2007). Maximum likelihood estimation of Heckman's sample selection model. Working Paper, Pennsylvania State University.

Bruche, M., and González-Aguado, C. (2010). Recovery rates, default probabilities, and the credit cycle. *Journal of Banking and Finance* **34**(4), 754–764.

Calabrese, R., and Zenga, M. (2010). Bank loan recovery rates: measuring and nonparametric density estimation. *Journal of Banking and Finance* **34**(5), 903–911.

Carey, M. (1998). Credit risk in private debt portfolios. *Journal of Finance* **53**(4), 1363–1387.

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

Citron, D., Wright, M., Ball, R., and Rippington, F. (2003). Secured creditor recovery rates from management buy-outs in distress. *European Financial Management* **9**(2), 141–161.

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

Duffie, D., and Singleton, K. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies* **12**(4), 687–720.

Duffie, D., Saita, L., and Wang, K. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics* **83**(3), 635–665.

Gordy, M., and Heitfield, E. (2000). Estimating factor loadings when ratings performance data are scarce. Working paper, Board of Governors of the Federal Reserve System, Division of Research and Statistics.

Gordy, M., and Heitfield, E. (2002). Estimating default correlations from short panels of credit rating performance data. Working Paper, Federal Reserve Board.

Grunert, J., and Weber, M. (2009). Recovery rates of commercial lending: empirical evidence for German companies. *Journal of Banking and Finance* **33**(3), 505–513.

Hamerle, A., Liebig, T., and Rosch, D. (2003). Benchmarking asset correlations. *Risk* **16**(11), 77–82.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**(1), 153–161.

Jarrow, R., and Turnbull, S. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance* **50**(1), 53–85.

Jarrow, R., Lando, D., and Turnbull, S. (1997). A Markov model for the term structure of credit risk spreads. *Review of Financial Studies* **10**(2), 481–523.

Kupiec, P. H. (2008). A generalized single common factor model of portfolio credit risk. *Journal of Derivatives* **15**(3), 25–40.

Leland, H. (1994). Corporate debt value, bond covenants and optimal capital structure. *Journal of Finance* **49**(4), 1213–1252.

Leland, H., and Toft, K. (1996). Optimal capital structure, endogeneous bankruptcy, and the term structure of credit spreads. *Journal of Finance* **51**(3), 987–1019.

Longstaff, F., and Schwartz, E. (1995). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance* **50**(3), 789–819.

Madan, D., and Unal, H. (1995). Pricing the risk of recovery in default with APR violation. *Journal of Banking and Finance* **27**(6), 1001–1218.

McNeil, A., and Wendin, J. (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* **14**(2), 131–149.

Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **29**(2), 449–470.

Pykhtin, M. (2003). Unexpected recovery risk. *Risk* **16**(8), 74–78.

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

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

Rösch, D., and Scheule, H. (2005). A multi-factor approach for systematic default and recovery risk. *Journal of Fixed Income* **15**(2), 63–75.

Rösch, D., and Scheule, H. (2009). An unbiased approach for PD and recovery estimation. Working Paper, University of Hannover and The University of Melbourne.

Rösch, D., and Scheule, H. (2010). Downturn credit portfolio risk, regulatory capital and prudential incentives. *International Review of Finance* **10**(2), 185–207.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard-rate model. *Journal of Business* **74**(1), 101–124.





# A practical anat
# risk charge moc

**Marcus R. W. Martin**
University of Applied Sciences, I
email: marcus.martin@h-da.de

**Helmut Lutz**
DekaBank, Hahnstrasse 55, 605
email: helmut.lutz@deka.de

**Carsten S. Wehn**
DekaBank, Hahnstrasse 55, 605
email: wehn@gmx.de

*This paper considers the d*
*tal risk charge. We show tl*
*reasoning behind this. We s*
*multistate credit risk mode*
*credit risk modeling with t*
*charge and successively rea*
*tory requirements. We analy*
*factor reduction, migration*
*efficient semianalytical incr*
*the CRM. This analysis gives*
*and thus into the models' ana*
*plexity is highly relevant to p*
*applying the different approa*

Marcus R. W. Martin is Professor of
Applied Sciences, Darmstadt. Helmut
Frankfurt. Carsten S. Wehn is head o
modeling team is responsible for develo
risks, credit risks, liquidity risks and o
responsible for validating the adequacy
testing and economic capital models in
own and should not be cited as being
described herein is claimed to be in act