Journal of Banking & Finance 37 (2013) 2354–2366

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Improvements in loss given default forecasts for bank loans

Marc Gürtler [1], Martin Hibbeln *

Braunschweig Institute of Technology, Department of Finance, Abt-Jerusalem-Str. 7, 38106 Braunschweig, Germany

ARTICLE INFO

Article history:
Received 28 June 2012
Accepted 20 January 2013
Available online 6 February 2013

JEL classification:
G21
G28

Keywords:
Bank loans
Credit risk
Forecasting
Loss given default
Workout process

ABSTRACT

An accurate forecast of the parameter loss given default (LGD) of loans plays a crucial role for risk-based decision making by banks. We theoretically analyze problems arising when forecasting LGDs of bank loans that lead to inconsistent estimates and a low predictive power. We present several improvements for LGD estimates, considering length-biased sampling, different loan characteristics depending on the type of default end, and different information sets according to the default status. We empirically demonstrate the capability of our proposals based on a data set of 69,985 defaulted bank loans. Our results are not only important for banks, but also for regulators, because neglecting these issues leads to a significant underestimation of capital requirements.

1. Introduction

The most central risk parameters of a loan are the probability of default (PD) and the loss given default (LGD). A decade ago, the focus of academic research and banking practice was mainly on the prediction of PDs, but more recently, substantial effort has been put into modeling LGDs. One reason for this is the requirement of the Basel II/III framework for banks to provide their own estimates of the LGD when using the advanced internal ratings-based (A-IRB) approach for corporates or the IRB approach for retail exposures. In addition to the regulatory requirement, accurate predictions of LGDs are important for risk-based decision making, e.g. the risk-adjusted pricing of loans, economic capital calculations, and the pricing of asset-backed securities or credit derivatives (cf. Jankowitsch et al., 2008). Consequently, banks using LGD models with high predictive power can generate competitive advantages, whereas weak predictions can lead to adverse selection.

There are different streams of LGD-related literature. Some studies seek to estimate the distribution of LGDs for credit portfolio modeling (cf. Renault and Scaillet, 2004; Calabrese and Zenga, 2010), whereas others analyze the factors influencing individual LGDs. Furthermore, some studies deal with the relation between PDs and LGDs (cf. Frye, 2000; Altman et al., 2005; Acharya et al., 2007; Bade et al., 2011). Although most of the literature consists of empirical studies for corporate bonds, a smaller fraction focuses on bank loans, whether retail or corporate, mainly due to limited data availability. A survey of empirical studies of LGDs with a classification into bank and capital market data can be found in Grunert and Weber (2009).

For bank loans, the estimation of LGDs is usually based on discounted recovery cash flows, leading to workout LGDs. A first step has been taken towards forecasting individual LGDs for bank loans by empirical studies reporting LGDs for different categories of influence factors (cf. Asarnow and Edwards, 1995; Felsovalyi and Hurt, 1998; Eales and Bosworth, 1998; Araten et al., 2004; Franks et al., 2004). More recent studies analyze factors influencing LGDs via linear regressions (cf. Citron et al., 2003; Caselli et al., 2008; Grunert and Weber, 2009), log regressions (cf. Caselli et al., 2008), or log–log regressions (cf. Dermine and Neto de Carvalho, 2006; Bastos, 2010). Bellotti and Crook (2012) and Loterman et al. (2012) compare the performance of different models constructed as combinations of different modeling algorithms and different transformations of the recovery rate, e.g. OLS regressions or decision trees, on the one hand, and log or probit transformations

☆ We thank Arnie Cowan, Jens Grunert, Hannes Klein, Philippe Louis, seminar participants at the 2011 Banking Workshop (Münster), the 2011 Australasian Finance & Banking Conference (Sydney), the University of Düsseldorf, the 2012 Midwest Finance Association meeting (New Orleans), and an anonymous referee for their helpful suggestions and comments.

* Corresponding author. Tel.: +49 531 391 2898; fax: +49 531 391 2899.
E-mail addresses: marc.guertler@tu-bs.de (M. Gürtler), martin.hibbeln@tu-bs.de (M. Hibbeln).
[1] Tel.: +49 531 391 2895; fax: +49 531 391 2899.

0378-4266/$ - see front matter
http://dx.doi.org/10.1016/j.jbankfin.2013.01.031

on the other hand. Bastos (2010) proposes to model LGDs with nonparametric and nonlinear regression trees.

The main motivation of this paper is to improve forecasts of LGDs for bank loans. We theoretically analyze several problems that arise when forecasting LGDs and derive recommendations for action in order to get consistent estimates with high predictive power. We apply the proposed methods to a bank internal data set consisting of 69,985 defaulted loans of a large German bank and analyze the improvements that can be achieved. We discuss our improvements within the typical steps of the modeling process. After all payments during the workout processes have been collected for the modeling data set, which consists of historical data of defaulted loans, the realized workout LGDs have to be calculated.[2] Within the calculation of LGDs, we observe the effect that samples of historical LGDs are usually biased, due to differences in the duration of the workout process. As the modeling data set usually consists of defaults with completed workout process, defaults with a long workout process are underrepresented because it is more likely that the workout process started before or will end after the available observation period. However, defaults with a long workout process, on average, have higher LGDs. Consequently, this underrepresentation of defaults with high LGDs leads to an underestimation of LGDs. To avoid the resulting underestimation of LGDs, we propose a procedure for restricting the modeling data to get consistent LGD estimates (Improvement 1).

Using calculated LGDs for the modeling data set, prediction models for LGDs can be developed to apply them to the scoring data set, which consists of new loans, non-defaulted existing loans, and defaulted existing loans. For new loans, LGD estimates are not only required to determine the required capital backing, but also a high accuracy of individual LGD estimates is essential to avoid adverse selection. In the literature, the prediction of individual LGDs is mostly based on a direct regression on LGDs. However, the estimation of LGDs with a single model often performs poorly. We discover that it is important to distinguish between recovered loans and write-offs in the model design because the characteristics of both types of default end can be very different. Against this background, we propose a two-step estimation of LGDs that strongly outperforms the direct regression approach. In the first step, the probability of a recovery/write-off is estimated. In the second step, the LGDs of recovered loans, as well as the LGDs of write-offs, are predicted separately. These predictions are combined into the total LGD forecast (Improvement 2).

Furthermore, the existing literature on LGD modeling does not explicitly deal with LGDs for defaulted loans, although for defaulted loans with an active default status, estimates of LGDs are required, e.g. for regulatory and economic capital calculations. In this case, only the portfolio LGD, and not the individual LGD, is of interest. However, if the average LGD of the modeling data is assigned to the portfolio, the LGD is significantly underestimated. The reason is that the information set of defaulted loans differs from the information set of non-defaulted loans. For defaulted loans an estimator conditional on the specific default status of a loan is required, whereas the average LGD is an unconditional estimator and leads to inconsistent LGD estimates. Against this background, we propose a consistent estimator for defaulted loans (Improvement 3).

The proposed three improvements have a significant impact on LGD forecasts and should be considered when modeling LGDs because neglecting these issues leads to a significant underestimation and low accuracy of LGDs. However, to the best of our knowledge, no research has addressed these issues as yet. The remainder of this paper is structured as follows. Section 2 contains a theoretical derivation of the three Improvements. In Section 3, we present an empirical study in which we analyze the extent of each improvement on the basis of real data. Our conclusions are presented in Section 4.

2. Theoretical analysis of LGD forecasts

2.1. Calculation of workout LGDs

There are some relevant differences between LGDs of corporate bonds and bank loans. First, LGDs of bank loans are typically lower than LGDs of corporate bonds. According to Schuermann (2006), this empirical finding is mainly a result of the higher seniority of loans (on average), and better monitoring. Second, LGDs of corporate bonds are typically determined on the basis of market values, resulting in "market LGDs", whereas the LGDs of bank loans are usually "workout LGDs". If the market value of a bond directly after default is divided by the exposure at default (EAD), which is the face value at the default event, we obtain the market recovery rate (RR). Application of the equation $LGD = 1 - RR$ results in the market LGD. Conversely, the workout LGD is based on actual cash flows that are connected with the defaulted debt position. These are mainly discounted recovery cash flows, but they are also discounted costs of the workout process. If these cash flows are divided by the EAD, we obtain the workout LGD. Even though the calculation of workout LGDs is more complex, the advantage is that the results are more accurate and that this approach is applicable for all types of debt (cf. Calabrese and Zenga, 2010).

For the forecasting of LGDs, we have to calculate historical workout LGDs for our modeling data. Let S be a set of loans and $i \in S$ an individual loan. The workout LGD of loan i is typically expressed as follows:[3]

$$LGD_i = 1 - \frac{RCF_i - C_i}{EAD_i}, \tag{1}$$

where RCF_i stands for the sum of discounted recovery cash flows of loan i, C_i represents the sum of discounted direct and indirect costs of loan i, and EAD_i is the exposure at default of loan i.[4] Eq. (1) leads to LGD = 0 if the recovery cash flow equals the exposure at default plus the costs of the workout process. In this context, it is important to notice that usually only direct costs can be charged to the obligor whereas indirect costs have to be borne by the bank.[5] If the loan defaults completely, the LGD can even be higher than 1 if there are additional costs that arise during the workout process. However, a defaulted loan can have two different types of default ends, which directly influence the calculation of LGDs: some contracts can be recovered, whereas other contracts have to be written off.

- *Recoveries (RCs):* In the case of a recovery, the default reason no longer exists, e.g. the obligor paid the amount that was in arrears, or a new payment plan has been arranged. Thus, the contract is henceforth handled as a normal non-defaulted loan.
- *Write-offs (WOs):* If the chance of recovering additional money from the obligor or the realization of collateral is considered to be small, the contract will be written off. Thus, there are generally no further payments for this contract.

[2] For retail loans, a default is usually assigned at a contract level. Conversely, for corporate loans, a default is generally determined at a firm level, so that several contracts of a firm default simultaneously. This is in line with the regulatory requirements, see Basel Committee on Banking Supervision (2005b), §455, and has to be considered in the calculation of LGDs.

[3] Cf. Franks et al. (2004) or Calabrese and Zenga (2010).

[4] We used the effective interest rate to discount the cash flows as this method has been favored by the national banking supervisor. For details regarding appropriate discount rates, see Basel Committee on Banking Supervision (2005a) and Maclachlan (2005).

[5] A description of direct and indirect costs in context of calculating LGDs can be found in Franks et al. (2004).