Loss functions for Loss Given Default model comparison[*]

Christophe Hurlin[†], Jérémy Leymarie[‡], Antoine Patin[§]

May 27, 2017

Abstract

We propose a new approach for comparing Loss Given Default (LGD) models which is based on loss functions defined in terms of regulatory capital charge. Our comparison method improves the banks' ability to absorb their unexpected credit losses, by penalizing more heavily LGD forecast errors made on credits associated with high exposure and long maturity. We also introduce asymmetric loss functions that only penalize the LGD forecast errors that lead to underestimate the regulatory capital. We show theoretically that our approach ranks models differently compared to the traditional approach which only focuses on LGD forecast errors. We apply our methodology to six competing LGD models using a unique sample of almost 10,000 defaulted credit and leasing contracts provided by an international bank. Our empirical findings clearly show that model rankings based on capital charge losses differ drastically from those based on the LGD loss functions currently used by regulators, banks, and academics.

Keywords: Risk management, Loss Given Default (LGD), Credit Risk Capital Requirement, Loss Function, Forecasts Comparison

JEL classification: G21, G28 .

1 Introduction

Since the Basel II agreements, banks have the possibility to develop internal rating models to compute their regulatory capital charge for credit risk, through the internal rating-based approach (IRB). The IRB approach can be viewed as an external risk model based on the asymptotic single risk factor (ASRF) model. This risk model relies on four key risk parameters: the exposure at default (EAD), the probability of default (PD), the loss given default (LGD), and the effective maturity (M). The Basel Committee on Banking Supervision (BCBS) allows financial institutions to use one of the following two methods: (1) the Foundation IRB (FIRB), in which banks only estimate the PD, the other parameters being arbitrarily set; (2) the Advanced IRB (AIRB), in which banks estimate both the PD and the LGD using their own internal risk models.[1]

In this paper, we propose a new approach for comparing LGD models which is based on loss functions defined in terms of regulatory capital charge. Given the importance of the LGD parameter in the Basel risk weight function and the regulatory capital for credit risk, the LGD model comparison is a crucial problematic for banks and regulators. Unlike PD, the LGD estimates enter the capital requirement formula in a linear way and, as a consequence, the estimation errors have a strong impact on required capital. Furthermore, there is no benchmark model emerging from the "zoo" of LGD models currently used by regulators, banks, and academics.[2] Indeed, the academic literature on LGD definition, measurement, and modelling is surprisingly underdeveloped (see Schuermann (2004) for a survey) and is particularly dwarfed by the one on PD models. The LGD can be broadly defined as the ratio (expressed as percentage of the EAD) of the loss that will never be recovered by the bank, or equivalently by one minus the recovery rate. While this definition is clear, the measurement and the modelling of the LGD raise numerous issues in practice. Regarding the measurement, both the BCBS and the European Banking Authority (see, for instance, EBA (2016)) made tremendous efforts to clarify the notion of default and the scope of losses that should be considered by the banks to measure the *workout* LGD. On the contrary, no particular guidelines have been provided

---

[1]In the FIRB approach, the LGD is fixed at 45% for senior claims on corporate, sovereigns, and banks not secured by recognized collateral, 75% for all subordinated claims on corporate, sovereigns, and banks. The effective maturity M is fixed at 2.5 years for corporate exposures except for repo-style transactions where the maturity is fixed at 6 months (Roncalli (2009)).

[2]By analogy with the "factor zoo" evoked by Cochrane (2011).

for LGD models. This may explain why there is such a large heterogeneity in the modelling approaches used by AIRB banks and academics. Commonly used models include (1) simple look-up (contingency) tables, (2) sophisticated machine learning regression methods (regression tree, bagging, random forests, gradient boosting, artificial neural network, support vector machine etc.), (3) parametric approaches based on beta, exponential-gamma, inflated beta distributions, fractional response regression, and Tobit models, etc. (4) non(semi)-parametric approaches such as kernel density estimators, quantile regressions, multivariate adaptive regression splines, and mixture models, etc. Thus, Loterman et al. (2012) evaluate 24 regression techniques for the LGD of six major international banking institutions, whereas Qi and Zhao (2011) compare six models that provide very different results.

How should LGD models be compared? The benchmarking method currently adopted by banks and academics simply consists in (1) considering a sample of defaulted credits split in a training set and a test set, (2) estimating the competing models on the training set and then, (3) evaluating the LGD forecasts on the test set with standard statistical criteria such as the mean square error (MSE) or the mean absolute error (MAE). Thus, LGD model comparison is made independently from the other Basel risk parameters (EAD, PD, M). The first shortcoming of this approach is its lack of economic interpretation of the loss function applied to the LGD estimates. What do a MSE of 10% or a MAE of 27% exactly imply in terms of financial stability? These figures give no information whatsoever about the estimation error made on capital charge and bank's ability to face an unexpected credit loss. The second shortcoming is related to the two-step structure of the AIRB approach. The LGD forecasts produced by the bank's internal models are, in a second step, introduced in the regulatory formula to compute the capital charge. If LGD models are compared independently from this second step, the same weight is given to a LGD estimation error of 10% made on two contracts with an EAD of 1,000€ and 1,000,000€, respectively. Similarly, it gives the same weight to a LGD estimation error of 10% made on two contracts, one with a PD of 5% and another with a PD of 15%.

In our approach, the LGD forecast errors are assessed in terms of regulatory capital and ultimately, in terms of bank's capacity to face unexpected losses on its credit portfolio. To do so, we define a set of expected loss functions for the LGD forecasts, which are expressed

3

in terms of *capital charge* induced by these forecasts. Hence, these loss functions take into account the EAD, PD, and maturity of the loans. For instance, they penalize more heavily the LGD forecast errors made on credits associated to high exposure and long maturity. Furthermore, we propose asymmetric loss functions that only penalize the LGD forecast errors that lead to underestimating the regulatory capital. Such asymmetric functions may be preferred by the banking regulators in order to neutralize the impact of the LGD forecast errors on the required capital and ultimately, to enhance the soundness and stability of the banking system. We show theoretically that the model ranking determined by a LGD-based loss function may differ from the ranking based on the capital charge loss function. In particular, we demonstrate the conditions under which both rankings are consistent. These conditions have no particular reasons to be valid in practice. This theoretical analysis may be related to the notion of model ranking consistency introduced by Hansen and Lunde (2006), Patton (2011), and Laurent, Rombouts and Violante (2013).

We apply our methodology to six competing LGD models using a unique sample of almost 10,000 defaulted credit and leasing contracts provided by the bank of a worldwide leader automotive company.[3] To the best of our knowledge, this is the first time that such a dataset is used in the academic literature. While most of the empirical literature on LGD is based on corporate bonds (given the public availability of data) and market LGDs, our dataset consists of an actual portfolio of personal loans and leasing for which we observe the workout LGDs. We find that the model ranking based on the LGD loss function is generally different from the model ranking obtained with the capital charge loss functions. Such a difference clearly illustrates that the consistency conditions previously mentioned are not fulfilled, at least in our sample. Our findings are robust to (1) the choice of the explanatory variables considered in the LGD models, (2) the inclusion (or not) of the EAD as a covariate, and (3) the use of the Basel PDs (collected one year before the default) in the capital charge loss function. We also find that the LGD forecast errors are generally right-skewed. In this context, the use of asymmetric loss functions provides a model ranking which is very different from the ranking obtained with symmetric loss functions.

---

[3]The six competing LGD models include (1) the fractional response regression model, (2) the regression tree, (3) the random forest, (4) the gradient boosting, (5) the artificial neural network, and (6) the support vector machine.

The main contribution of this paper is to propose a comparison method for LGD models which improves the banks' solvability. Within the BCBS framework, the level of regulatory capital is determined such as to cover unexpected credit losses. This level depends on estimated risk parameters, and in particular on the LGD. As a consequence, any under-estimation of these risk parameters induces an underestimation of the regulatory capital and in fine, a lowest bank's solvency. In this context, when considering a set of competing LGD models that produce different LGD forecasts, an appropriate comparison methods should select the model associated with the lowest estimation errors on the regulatory capital. This is not the case with the comparison method currently used by banks and academics which is only based on the LGD estimation errors. Conversely, our approach allows us to select the LGD model which induces the lowest estimation errors on the regulatory capital. Hence, we believe that adopting this new model comparison approach should be of general interest.

The rest of this paper is structured as follows. We discuss in Section 2 the main features of the AIRB approach and the regulatory capital for credit risk portfolios. In this presentation, we focus on LGD models and the current methods used to compare them. In Section 3, we present the capital charge loss function that is at the heart of our comparison methodology. In Section 4, we describe the dataset as well the six competing LGD models. In section 5, we conduct our empirical analysis and display our main takeaways. We summarize and conclude our paper in Section 6.

2   Capital charge for credit risk portfolios

In this section, we propose a brief overview of the role of the LGD in the computation of the regulatory capital under the AIRB approach. Then, we present the main issues related to LGD measurement and modelling. We then discuss the method which is currently used to compare the LGD models.

2.1   Capital requirement, individual risk contributions, and LGD

Let us consider a portfolio of n credits indexed by $i = 1, ..., n$. Each credit is characterized by (1) an EAD defined as the outstanding debt at the time of default, (2) a LGD defined as the percentage of exposure at default that is lost if the debtor defaults, (3) a PD that measures

the likelihood of the default risk of the debtor over a horizon of one year, and (4) an effective maturity M, expressed in years. The credit portfolio loss is then equal to

$$L = \sum_{i=1}^{n} \text{EAD}_i \times \text{LGD}_i \times D_i \tag{1}$$

where D_i is a binary random variable that takes a value 1 if there is a default before the residual maturity M_i and 0 otherwise.

In the AIRB approach, the regulatory capital (RC) charge is designed to cover the unexpected bank's credit loss. The unexpected loss is defined as the difference between the 99.9% Value-at-Risk (VaR) of the portfolio loss and the expected loss $\mathbb{E}(L)$. In order to compute the unexpected credit loss, the Basel Committee considers the ASRF model. This model is based on the seminal Merton-Vasicek "model of the firm" (Merton (1974), Vasicek (2002)) with additional assumptions such as the infinite granularity of considered portfolios, the normal distribution of the risk factor, and a time horizon of one year (BCBS (2004, 2005)). Under these assumptions (cf. appendix A), the unexpected loss, and hence the regulatory capital, can be decomposed as a sum of independent risk contributions (RC_i) which only depend on the characteristics of the ith credit (Genest and Brie (2013), Roncalli (2009)). The regulatory capital is then equal to

$$\text{RC} = \sum_{i=1}^{n} \text{RC}_i \tag{2}$$

The supervisory formula for the risk contribution RC_i is given by

$$\text{RC}_i \equiv \text{RC}_i(\text{EAD}_i, \text{PD}_i, \text{LGD}_i, \text{M}_i) = \text{EAD}_i \times \text{LGD}_i \times \delta(\text{PD}_i) \times \gamma(\text{M}_i) \tag{3}$$

with

$$\delta(\text{PD}_i) = \Phi\left(\frac{\Phi^{-1}(\text{PD}_i) + \sqrt{\rho(\text{PD}_i)}\Phi^{-1}(99.9\%)}{\sqrt{1 - \rho(\text{PD}_i)}}\right) - \text{PD}_i \tag{4}$$

where $\Phi(.)$ denotes the cdf of a standard normal distribution, $\rho(\text{PD})$ a parametric decreasing function for the default correlation, and $\gamma(\text{M})$ a parametric function for the maturity adjustment. The maturity adjustment and the correlation functions suggested by the BCBS depend on the type of exposure: corporate, sovereign or bank exposures, versus residential mortgage, revolving, or other retail exposures. The functions $\rho(\text{PD})$ and $\gamma(\text{M})$ suggested by the BCBS are reported in appendix B.

The Basel II formula (Equations 2, 3, and 4) highlights the key role of the LGD. Since LGD enters the capital requirement formula in a linear way, the LGD forecast errors have necessarily a strong impact. Thus, the choice of the LGD model is crucial for banking regulation.

2.2 LGD data and models

As previously mentioned, the AIRB approach allows banks to develop internal models for estimating PD, LGD, and EAD.[4] However, the LGD definition, measurement, and modelling raise numerous practical issues.

The LGD is defined as the ratio of losses (expressed as percentage of the EAD) that will never be recovered by the lender, or equivalently by one minus the recovery rate. The Basel II Accord requires that all relevant factors that may reduce the final economic value of recovered portion of an exposure must be taken into account into the LGD calculation. These factors correspond to (1) the direct (external) costs associated to the loss of principal and the foregone interest income, (2) the indirect (internal) costs incurred by the bank for recovery in the form of workout costs (administrative costs associated with collecting information on the exposure, legal costs, etc.), and (3) the funding costs reflected by an appropriate discount rate tied to the time span between the emergence of default and the actual recovery.

Schuermann (2004) identifies three main ways of measuring LGD. The *market* LGD is calculated as one minus the ratio of the trading price of the asset some time after default to the trading price at the time of default. The *implied market* LGD is derived from risky (but not defaulted) bond prices using a theoretical asset pricing model. As they are based on trading prices, the market and implied market LGDs are generally available only for bonds and loans issued by large firms. On the contrary, the *workout* LGD can be measured for any type of instrument. The workout LGD estimation is based on an economic notion of loss including all the relevant costs tied to the collection process, but also the effect deriving from the discount of cash flows. The scope of necessary data for proper LGD estimation is very broad and includes not only the date of default, all cash flows and events after default, but also all relevant information about the obligors and transactions that could be used as risk drivers (collateral, etc.). The workout approach is clearly preferred by the regulators.

---

[4]For standard credits and loans, the EAD are observed. However, for off-balance sheet EAD, the bank has to estimate a credit conversion factor (CCF). See for instance, Gürtler, Hibbeln and Usselman (2017).

For instance, in its guidelines on LGD estimation, the EBA (November 2016) states that "*the workout LGD is considered to be the main, superior methodology that should be used by institutions. It is essential that LGD estimates are based on the institutions' own loss and recovery experience in order to make sure that the estimates are adequate for the institutions portfolios and policies, and in particular that they are consistent with the recovery processes*" (EBA (2016), page 11). Notice that most empirical academic studies neglect workout costs because of data limitations, even if Khieu et al. (2012) found evidence that market LGDs are biased and inefficient estimates of the workout LGD.[5]

The general purpose of the LGD (internal) models consists in providing an estimate (or a forecast) of the LGD for the credits which are currently in the bank's portfolio and for which the bank does not observe the potential losses induced by a default of the borrower. These models are generally estimated on a sample of defaulted credits for which the ex-post workout LGD is observed. By identifying the main characteristics of these contracts and the key factors of the recovery rates, it is then possible to forecast the LGD for all the similar non-defaulted credits which are in the bank's portfolio.

Töws (2016) identifies two major challenges in estimating recovery rates with respect to defaulted bank loans or bonds. First, the LGD theoretically ranges between 0 and 100% of the EAD, meaning that the bank cannot recover more than the outstanding amount and that the lender cannot lose more than the outstanding amount. However, several studies (Schmit and Stuyck (2002), Schmit (2004), Schuermann (2004), Töws (2016)) show that when workout costs are incorporated, the LGD is sometimes larger than 100%. The second challenge in estimating recovery rates is the bimodal nature of the LGD distribution. Indeed, recovery as a percentage of exposure is generally either relatively high (around 70-80%) or low (around 20-30%). Hence, thinking about an "average" LGD can be misleading.

Because of the specific nature of the LGD distribution, a large variety of LGD models are currently used by academic and practitioners. Furthermore, there is no benchmark model emerging from this "zoo" of LGD models. This explains why LGD model comparison is so important.[6] Four categories of LGD models can be identified. The first category corresponds

---

[5]Conversely, Gürtler and Hibbeln (2013) identify several problems in modeling workout LGDs that can lead to inaccurate LGD forecasts. In particular, the LGDs within the modeling data can be significantly biased downwards if all available defaults with completed workout processes are considered.

[6]This variety explains the number of benchmarking studies proposed in the academic literature these last

to nonparametric approaches. A popular method corresponds to the contingency or "look-up" tables which contain LGD averages grouped by credit characteristics (segmentation). For instance, a cell of this table might include senior unsecured loans for the automotive industry during a recession. An average LGD is computed from the observations of the training set that belong to this cell and then, is applied to the similar credits of the bank's portfolio. These tables have the advantage of being easy to build and easy to use. However, with enough cuts one quickly runs out of data: many cells in this contingency table will likely go unfilled or have only very few observations to compute average. The second category corresponds to parametric approaches. Since LGD is theoretically defined over $[0, 1]$, the parametric models are generally based on beta (Credit Portfolio View of Mc Kinsey), exponential-gamma (Gouriéroux, Monfort, and Polimenis (2006)), inflated beta (Ospina and Ferrari (2010)), or logistic-Gaussian distributions. In a similar way, the fractional response regression model which keeps the predicted values in the unit interval, has also been used for LGD estimates by Dermine and Carvalho (2006), Bastos (2010), Qi and Zhao (2011), or Bellotti and Crook (2012). Using a different approach, Tanoue, Kawada and Yamashita (2017) propose a parametric multi-step approach for the LGD of bank loans in Japan. The third category encompasses the kernel density estimators, quantile regressions and mixture models. Calabrese and Zenga (2010) consider a mixture of a Bernoulli random variable and a continuous random variable on the unit interval to model the LGD of a large dataset of defaulted Italian loans. Similarly, Calabrese (2014) suggests a mixture distribution approach for the downturn LGD. Renault and Scaillet (2004) or Hagmann, Renault, and Scaillet (2005) consider various kernel density estimator of the LGD distribution, whereas Krüger and Rösh (2017) consider quantile regressions for modelling downturn LGD. We can also mention the use of multivariate adaptive regression splines (Loterman et al. (2012)), support vector machine (Yao, Crook, and Andreeva (2015)) and least squares support vector machine (Loterman et al. (2012)). These approaches have the common advantage to reveal a number of bumps which can be larger than those obtained with parametric distributions. Finally, the last category corresponds to machine learning approaches such as regression tree algorithms (Hartmann-Wendels, Miller, and Töws (2014)), artificial neural networks, random forest, gradient boosting, and many others. Qi and Zhao (2011), and Bastos (2010) compare fractional response regression models to other parametric

years (Bastos (2010), Qi and Zhao (2011), Loterman et al. (2012), Töws (2016) among others).

and nonparametric modeling methods for LGD. They conclude that machine learning methods, such as regression trees and neural networks, perform better than parametric methods when overfitting is properly controlled for. A similar conclusion is get by Loterman et al. (2012) who show that non-linear techniques, and in particular support vector machine and neural networks, perform significantly better than more traditional linear techniques.

2.3 LGD model comparison

Choosing the best methodology for fitting the recovery rates curve among the set of potential LGD models, implies to compare the predictive performances of the models. In the sequel, we briefly present the comparison method currently used both by academics and banks.

Consider a set of $\mathcal{M}$ LGD models indexed by $m = 1, ..., \mathcal{M}$ and a sample of n_d defaulted credits which is randomly split into a training set including n_t credits and a test set including n_v credits, with $n_t + n_v = n_d$. In a first step, the models are estimated (for parametric models) or calibrated on the training set.[7] In a second step, the models are used to produce pseudo out-of-sample forecasts of the LGD for the credits of the test set. The test set is then used solely to assess the prediction performances of the models. Denote by LGD_i the true LGD observed for the ith credit, for $i = 1, ..., n_v$ and by $\widehat{\mathrm{LGD}}_{i,m}$ the corresponding forecast issued from model m.

The assessment of the prediction performances of the LGD models is generally based on an expected loss $\mathcal{L}$ defined as

$$\mathcal{L}_m \equiv \mathcal{L}\left(\mathrm{LGD}_i, \widehat{\mathrm{LGD}}_{i,m}\right) = \mathbb{E}\left(L\left(\mathrm{LGD}_i, \widehat{\mathrm{LGD}}_{i,m}\right)\right) \tag{5}$$

where $L(.,.)$ is an integrable loss function, with $L : \Omega^2 \to \mathbb{R}^+$, that satisfies the main standard properties discussed in Granger (1999).[8] Since the LGD is a continuous variable defined over a subspace Ω of $\mathbb{R}^+$ (typically $[0,1]$ or $[0,\delta]$ with $\delta > 1$), the loss function is similar to those generally used for any standard regression models. The loss functions generally considered in the academic literature (Gupton and Stein (2002), Caselli, Gatti, and Querci

---

[7]For the machine learning methods (regression trees, neural networks, etc.), the training set is further split into training and validation subsets. The validation set is used to select the criterion for evaluating the candidate splitting rules, the depth of the tree or any other parameter required by these methods.

[8]If we denote by $e = x - \widehat{x}$ the error and rewrite the loss function as a function of e, these required properties can be summarized as follows: (i) $L(0) = 0$, (ii) $\min L(e) = 0$ so that $L(e) \geq 0$, (iii) $L(e)$ is monotonically non-decreasing as e moves away from zero so that $L(e_1) \geq L(e_2)$ if $e_1 > e_2 > 0$, and if $e_1 < e_2 < 0$.

(2008), Matuszyk, Mues, and Thomas (2010), Loterman et al. (2012), etc.) for the LGD estimates are the quadratic loss function $L(x, \widehat{x}) = (x - \widehat{x})^2$ and the absolute loss function $L(x, \widehat{x}) = |x - \widehat{x}|$. Thus, the prediction performances of the LGD models are compared through the empirical mean of their losses computed on the test set, defined as

$$\widehat{\mathcal{L}}_m = \frac{1}{n_v} \sum_{i=1}^{n_v} L\left(\text{LGD}_i, \widehat{\text{LGD}}_{i,m}\right) \tag{6}$$

Given the functional form of the loss function, the empirical mean $\widehat{\mathcal{L}}_m$ corresponds to a common measure of predictive accuracy such as the MSE, MAE, or RAE, with

$$\text{MSE: } \widehat{\mathcal{L}}_m = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(\text{LGD}_i - \widehat{\text{LGD}}_{i,m}\right)^2$$

$$\text{MAE: } \widehat{\mathcal{L}}_m = \frac{1}{n_v} \sum_{i=1}^{n_v} \left|\text{LGD}_i - \widehat{\text{LGD}}_{i,m}\right|$$

$$\text{RAE: } \widehat{\mathcal{L}}_m = \sum_{i=1}^{n_v} \left|\text{LGD}_i - \widehat{\text{LGD}}_{i,m}\right| / \sum_{i=1}^{n_v} \left|\text{LGD}_i - \overline{\text{LGD}}_i\right|$$

The LGD models are compared and ranked according to the realization of the statistic $\widehat{\mathcal{L}}_m$ on the test set. A model m is preferred to a model m' as soon as $\widehat{\mathcal{L}}_m < \widehat{\mathcal{L}}_{m'}$. Denote by $\widehat{m}^*$ the model associated to the minimum realization $\widehat{\mathcal{L}}_m$ for $m = 1, ..., \mathcal{M}$. Under some regularity conditions, $\widehat{\mathcal{L}}_m$ converges to $\mathcal{L}_m$, and the model $\widehat{m}^*$ corresponds to the optimal model m^* defined as

$$m^* = \underset{m=1,...,\mathcal{M}}{\arg \min} \mathbb{E}\left(L\left(\text{LGD}_i, \widehat{\text{LGD}}_{i,m}\right)\right) \tag{7}$$

This general approach has three main shortcomings. The first one, which is not specific to the context of LGD models, is that the models are compared on the basis of the empirical means. Nothing guarantees that the observed differences are statistically significant. A simple solution consists in testing the null hypothesis of no difference in the accuracy of two competing forecasts with a DM-type test (Diebold and Mariano (1995)) or to identify a model confidence set (Hansen, Lunde, and Nason (2011)) which contains the "best" models for a given level of confidence. The second drawback of the current approach is the lack of economic interpretation for the loss function. What do a MSE of 10% or a MAE of 27% exactly imply in terms of regulatory capital? These figures give no information about the estimation error made on the capital charge, and ultimately on the ability of the bank to absorb unexpected

losses. The last pitfall is related to the two-step structure of the AIRB approach. The output of the bank's internal models, including the LGD models, are the Basel risk parameter estimates. These estimates are, in a second step, introduced in the ASRF model to compute the capital charge for each credit. As shown in the top panel of Figure 1, the LGD model comparison is currently done independently of this second step and, as a consequence, of the ASRF model and the other risk parameters (EAD, PD, etc.).

Figure 1: Comparison of LGD models in the regulatory framework
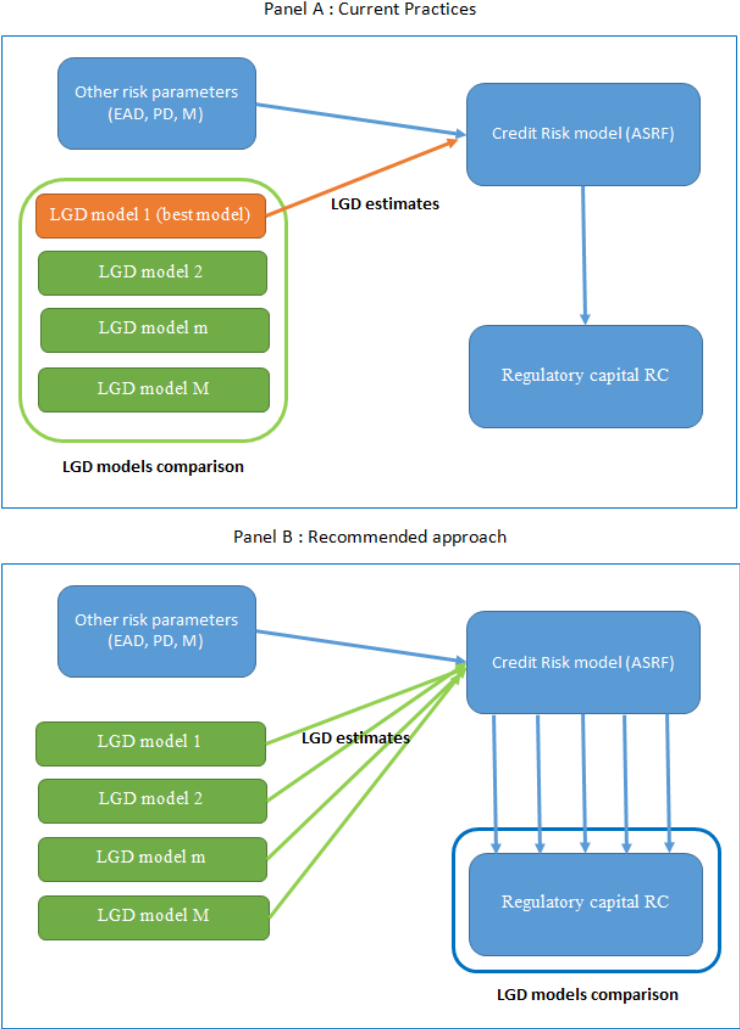
3 Capital charge loss functions for LGD models

"*Of great importance, and almost always ignored, is the fact that the economic loss associated with a forecast may be poorly assessed by the usual statistical metrics. That is, forecasts are used to guide decisions, and the loss associated with a forecast error of a particular sign and size is induced directly by the nature of the decision problem at hand.*". Diebold and Mariano (1995), page 2.

This quotation issued from the seminal paper of Diebold and Mariano (1995), perfectly illustrates the drawbacks of the current practices of LGD model comparison. In the BCBS framework, the LGD estimates are only *inputs* of the ASRF model which produces the key estimate, namely the capital charge for credit risk. Consequently, the economic loss associated to the LGD models has to be assessed in terms of regulatory capital. The bottom panel of Figure 1 summarizes the alternative approach that we recommend for LGD model comparison. The LGD forecasts issued from the competing models and the other risk parameters (EAD, PD, etc.) are jointly used to compute the capital charges. Then, our approach consists in comparing the LGD models not in terms of forecasting abilities for the LGD itself, but in terms of forecasting abilities for the regulatory capital charges. The main advantage of this approach is that it favors the LGD model that leads to the lowest estimation errors associated to the loans with the highest EAD and PD.

3.1 Capital charge expected loss

The capital charge expected loss $\mathcal{L}_{CC,m}$ is simply defined as the expected loss defined in terms of regulatory capital charge, which is associated to a LGD model m. Formally, we have

$$\mathcal{L}_{CC,m} \equiv \mathcal{L}\left(\text{RC}_i, \widehat{\text{RC}}_{i,m}\right) = \mathbb{E}\left(L\left(\text{RC}_i, \widehat{\text{RC}}_{i,m}\right)\right) \tag{8}$$

where $L\left(.,.\right)$ is an integrable *capital charge* loss function with $L : \mathbb{R}^{+2} \to \mathbb{R}^+$, and

$$\text{RC}_i = \text{EAD}_i \times \text{LGD}_i \times \delta\left(\text{PD}\right) \times \gamma\left(\text{M}_i\right)$$

$$\widehat{\text{RC}}_{i,m} = \text{EAD}_i \times \widehat{\text{LGD}}_{i,m} \times \delta\left(\text{PD}\right) \times \gamma\left(\text{M}_i\right)$$

The variable RC_i denotes the risk contribution of the ith credit, defined by the regulatory formula (Equation 3). This risk contribution depends on the risk parameters associated to

the credit i, namely EAD_i, LGD_i, and M_i.[9] Notice that PD is not indexed by i, meaning that we consider the same default probability for all the credits. As we only consider defaulted credits in the test set, PD is fixed to an arbitrary value, typically close to 1. Similarly, $\widehat{\text{RC}}_{i,m}$ denotes the estimated risk contribution for credit i, which is based on the individual risk parameters (EAD_i and M_i), the common value for the PD and the LGD forecast issued from model m.

Given the functional form of $L(.,.)$, the empirical counterpart $\widehat{\mathcal{L}}_{CC,m}$ can be defined in terms of MSE, MAE, RAE or any usual model comparison criteria, with for instance

$$\text{Capital Charge MSE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(\text{RC}_i - \widehat{\text{RC}}_{i,m} \right)^2$$

$$\text{Capital Charge MAE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v} \sum_{i=1}^{n_v} \left| \text{RC}_i - \widehat{\text{RC}}_{i,m} \right|$$

$$\text{Capital Charge RAE: } \widehat{\mathcal{L}}_{CC,m} = \sum_{i=1}^{n_v} \left| \text{RC}_i - \widehat{\text{RC}}_{i,m} \right| / \sum_{i=1}^{n_v} \left| \text{RC}_i - \overline{\text{RC}}_i \right|$$

where n_v denotes the size of the test set of defaulted credits. Beyond these traditional statistical criteria, we also introduce asymmetric criteria especially designed to improve financial stability. These loss functions only penalize the capital charge underestimates and they do not take into account the overestimations. As the regulatory capital is designed to absorb the unexpected credit losses, any underestimate of this charge can threaten the bank's solvability. Thus, we propose asymmetric loss functions defined as

$$\text{Asymmetric MSE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v^+} \sum_{i=1}^{n_v^+} \left(\text{RC}_i - \widehat{\text{RC}}_{i,m} \right)^2 \times \mathbb{I}_{\left(\text{RC}_i > \widehat{\text{RC}}_{i,m} \right)}$$

$$\text{Asymmetric MAE: } \widehat{\mathcal{L}}_{CC,m} = \frac{1}{n_v^+} \sum_{i=1}^{n_v^+} \left| \text{RC}_i - \widehat{\text{RC}}_{i,m} \right| \times \mathbb{I}_{\left(\text{RC}_i > \widehat{\text{RC}}_{i,m} \right)}$$

where $\mathbb{I}_{(.)}$ denotes the indicator function that takes a value 1 when the event occurs and 0 otherwise, and n_v^+ is the number of defaulted credits for which we observe $\text{RC}_i > \widehat{\text{RC}}_{i,m}$. These loss functions are particularly suitable to compare LGD models which produce skewed LGD estimation errors (cf. Section 5).

---

[9]For simplicity, we assume that there is no off-balance sheet exposure and that the exposure at default is not estimated, but observed ex-post. The maturity is also observed.

The expected loss $\widehat{\mathcal{L}}_{CC,m}$ has a direct economic interpretation. For instance, the capital charge MAE represents the average absolute estimation error observed between the capital charge estimates (associated to the LGD estimates issued from a given model) and the true ones (based on the observed LGD for the defaulted credit). Similarly, the asymmetric MSE corresponds to the variance of the capital charge underestimates produced by a given LGD model. Furthermore, these comparison criteria take into account the exposure and the maturity of the credits. Finally, the comparison rule for the LGD models is the same as before. A model m is preferred to a model m' as soon as $\widehat{\mathcal{L}}_{CC,m} < \widehat{\mathcal{L}}_{CC,m'}$. Denote by $\widehat{m}^*_{CC}$ the model associated to the minimum empirical mean $\widehat{\mathcal{L}}_{CC,m_{CC}}$ among the set of $\mathcal{M}$ models. Under some regularity conditions, $\widehat{\mathcal{L}}_{CC,m_{CC}}$ converges to $\mathcal{L}_{CC,m_{CC}}$, and allows to identify the optimal model in terms of capital charge expected loss.

As previously mentioned, the expected loss expressed in terms of capital charge depends on the value of PD chosen for the defaulted credits that belong to the test set.

Proposition 1 *The ranking of the LGD models based on the capital charge expected loss, does not depend on the choice of the PD value.*

The proof of proposition 1 is straightforward. Since $\delta\,(\mathrm{PD})$ is a constant term that does not depend on the contract i or the model m, the choice of PD does not affect the *relative* values of the expected losses observed for two alternative models, m and m'. This choice only affects the *absolute* value of the expected losses $\mathcal{L}_{CC,m}$ and $\mathcal{L}_{CC,m'}$.

Equation 4 implies that $\delta\,(1) = 0$ and $\delta\,(0) = 0$. As a consequence, the PD value has to be chosen on the interval $]0,1[$. Here, we recommend to use the value PD* that maximizes the value of $\delta\,(\mathrm{PD})$ and hence, the regulatory capital since RC_i is an increasing function of $\delta\,(\mathrm{PD})$. The profile of the capital charge coefficient $\delta\,(\mathrm{PD})$ depends on the type of exposure (cf. appendix B) and is displayed on Figure 2. The capital charge coefficient increases with PD until an inflexion point, and then decreases to 0 when the PD tends to 1. This profile is explained by the fact that once this inflexion point is reached, losses are no longer absorbed by the regulatory capital (which covers the unexpected bank's credit loss), but by the provisions done for the expected credit losses $\mathbb{E}\,(L)$ (Genest and Brie (2013)). The maximum of the $\delta\,(.)$ function is reached for a PD value of 28.76% in the case of residential mortgage, 38.98% for revolving retail, and 40.45% for other exposures.

15

Figure 2: δ function for different types of retail exposure
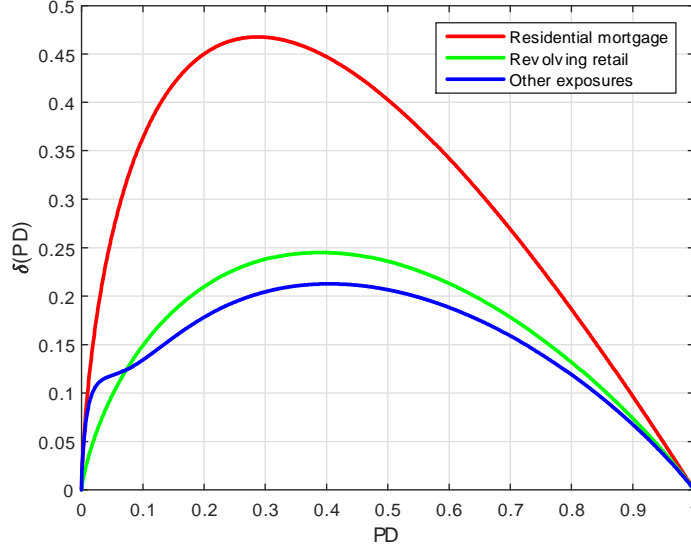
3.2 Ranking consistency

The aim of this section is to determine the conditions under which the model ranking induced by a LGD-based loss function differs (or not) from the ranking based on a capital charge-based loss function. This analysis may be related to the notion of ranking consistency introduced by Hansen and Lunde (2006), Patton (2011, 2016), and Laurent, Rombouts, and Violante (2013).[10] In the sequel, we state that two LGD model rankings are consistent when they are exactly the same whether they are based on LGD-based or capital-based estimation errors. Consider the following assumptions on the LGD loss functions.

Assumption A1: $L(x, \widehat{x}) = g(x - \widehat{x})$ with $g : \mathbb{R} \rightarrow \mathbb{R}^{+}$, a continuous and integrable function.

Assumption A2: The function $g(.)$ is multiplicative: $\forall k \in \mathbb{R}$, $g(k(x - \widehat{x})) = g(k) g(x - \widehat{x})$. Notice that assumptions A1 and A2 are satisfied by the usual loss functions considered in the LGD literature.[11]

---

[10] In the context of Patton (2011) and Laurent, Rombouts, and Violante (2013), a volatility model ranking is said consistent when it is the same whether it is based on the true conditional variance or a conditionally unbiased proxy.

[11] For instance, the quadratic loss function $L(x, \widehat{x}) = (x - \widehat{x})^2$ with $g(y) = y^2$ implies that $L(kx, k\widehat{x}) = L(g(k(x - \widehat{x}))) = k^2 (x - \widehat{x})^2 = g(k) g(x - \widehat{x})$. Regarding the absolute loss function $L(x, \widehat{x}) = |x - \widehat{x}|$ with $g(y) = |y|$, we have $L(kx, k\widehat{x}) = L(g(k(x - \widehat{x}))) = |k| |x - \widehat{x}| = g(k) g(x - \widehat{x})$.

Consider a set of $\mathcal{M}$ LGD models, indexed by $m = 1, \cdots, \mathcal{M}$. We refer to the ordering based on the expected loss as the true ranking and we assume that LGD-based expected losses are ranked as follows

$$\mathcal{L}_1 < \mathcal{L}_2 < ... < \mathcal{L}_\mathcal{M} \tag{9}$$

with $\mathcal{L}_m = \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right)$ and $\varepsilon_{i,m} = \text{LGD}_i - \widehat{\text{LGD}}_{i,m}$, $\forall m = 1, ..., \mathcal{M}$. Now, define the corresponding capital charge expected loss, $\mathcal{L}_{CC,m}$, for the model m as

$$\mathcal{L}_{CC,m} = \mathbb{E}\left(g\left(\eta_{i,m}\right)\right) \tag{10}$$

with $\eta_{i,m} = \text{RC}_i - \widehat{\text{RC}}_{i,m}$. By definition of the regulatory capital charge, we have[12]

$$\eta_{i,m} = \text{EAD}_i \times \delta\left(\text{PD}\right) \times \gamma\left(\text{M}\right) \times \varepsilon_{i,m} \tag{11}$$

Proposition 2 *The model rankings produced by LGD-based and capital charge-based expected losses are consistent, i.e. $\mathcal{L}_1 < \mathcal{L}_2 < ... < \mathcal{L}_\mathcal{M}$ and $\mathcal{L}_{CC,1} < \mathcal{L}_{CC,2} < ... < \mathcal{L}_{CC,\mathcal{M}}$, as soon as, $\forall m = 1, ..., \mathcal{M} - 1$*

$$cov\left(g\left(EAD_i\right), g\left(\varepsilon_{i,m}\right)\right) - cov\left(g\left(EAD_i\right), g\left(\varepsilon_{i,m+1}\right)\right) < \mathbb{E}\left(g\left(EAD_i\right)\right)\left(\mathcal{L}_{m+1} - \mathcal{L}_m\right) \tag{12}$$

The proof of proposition 2 is reported in appendix C. Since $\mathcal{L}_m < \mathcal{L}_{m+1}$, the consistency condition of proposition 2 is satisfied as soon as $cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) < cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m+1}\right)\right)$. Thus, both rankings are consistent as soon as the covariances of the LGD forecast errors with the exposures are ranked in the same manner as the LGD models themselves. Consider a simple case with two LGD models A and B, where model A has a smaller LGD-based MSE than model B. Model A will have also a smaller MSE in terms of capital charge, if its squared LGD estimation errors are less correlated to the squared EAD than the errors of model B. For instance, if model B produces large LGD estimation errors for high exposures and low LGD errors for low exposures, whereas it is not the case for model A, both model comparison approaches will provide the same rankings. Obviously, this condition is very particular and in the general case, the two comparison approaches are likely to provide inconsistent LGD model rankings.

Proposition 2 has a direct interpretation in the special case where the exposures are independent from the estimation errors of the LGD models.

---

[12] For simplicity, we assume that the credits have the same maturity M. In the general case, the consistency condition of the model rankings can be easily deduced from the formula given in this benchmark case.

Corollary 3 *As soon as the EAD_i and the LGD estimation errors $\varepsilon_{i,m}$ are independent, the model rankings based on the LGD and capital charge expected losses are consistent.*

The proof is provided in the appendix D. This corollary implies that when credit exposures and LGD estimation errors $\varepsilon_{i,m}$ are independent, the current model comparison approach that consists to compare the MSE, MAE or RAE in terms of LGD estimation errors is sufficient. However, the independence assumption is likely to be violated. First, Schuermann (2004) states that the size of exposure has no strong effect on losses. But, even if the variables EAD_i and LGD_i are independent, it does not necessarily imply that EAD_i and $LGD_i - \widehat{LGD}_{i,m}$ are independent. Second, it is important to notice that the introduction of the EAD as an explanatory variable in the LGD model, does not necessarily guarantee that the EAD and estimation errors are independent. It depends on the model (linear or not) and the estimation method used. For instance, the independence assumption is satisfied for linear regression model estimated by OLS. Conversely, for nonlinear models or machine learning methods, such as regression tree, support vector machine, or random forest, the forecast errors may be correlated with the explanatory variables.

4 LGD data and models

In the sequel, we propose an empirical application of our comparison approach for LGD models. The objective consists in comparing the model ranking obtained with our approach to the traditional ranking produced by LGD-based expected losses. In this section, we describe our dataset as well the six competing LGD models.

4.1 Data description

Our dataset, one of the first of its kind to be used in an academic study, was provided by an international bank specialized in financing, insurance, and related activities for a worldwide leader automotive company. Contrary to the existing literature on LGD, which is for the most part related to corporate bonds (given the public availability of data) and market LGD, our dataset consists in a portfolio of retail loans (personal loans and leasing) for which we observe the workout LGD.[13]

---

[13]Workout recoveries are also used by Khieu et al. (2012), Dermine and Neto de Carvalho (2006) or Töws (2016). Miller and Töws (2017) also consider workout LGDs for leasing.

The initial sample includes 23,933 loans. We limit our analysis to the 9,738 closed recovery processes for which we observe the final losses. The corresponding sample covers 6,946 credit and 2,792 leasing contracts granted to individual (6,521 contracts) and professional (3,217 contracts) Brazilian customers that defaulted between January 2011 and November 2014. For each contract, we observe the characteristics of the loan (e.g. type of contract, interest rate, duration, etc.) and the borrower (professional, individual, etc.), as well as the LGD and EAD. All the contracts are in default, so by definition their PD is equal to 1 (certain event). However, we collect for each contract the PD calculated by the internal bank's risk model one year before the default occurs. For the contracts that entered in default in less than one year, the PD is set to the value determined by the internal bank's risk model at the granting date. Finally, we complete the database with the Brazilian GDP growth rate, unemployment rate and interest rate. These three macroeconomic variables will be introduced in the LGD models in order to take into account the influence of the economic cycles on the recovery rate, as in Bellotti and Crook (2012). The name and the description of the dataset variables are reported in Table 6 (appendix E).

Table 1: Descriptive statistics on LGD, PD, and EAD

	Nb of obs	LGD(%)		PD (%)		EAD (BRL)	
Panel A. By year							
	–	min	max	min	max	min	max
2011	1573	0.00	116.14	0.06	71.03	0.01	104,959
2012	2430	0.00	114.65	0.09	61.50	383	123,551
2013	2946	0.00	102.97	0.10	61.26	0.18	114,858
2014	2789	0.00	101.59	0.06	70.53	350	92,595
Panel B. By exposure							
	–	min	max	min	max	min	max
Credit	6946	0.00	116.14	0.06	71.03	0.01	118,284
Leasing	2792	0.00	114.33	0.06	71.03	411	123,551
Panel C. By customer type							
	–	min	max	min	max	min	max
Individuals	6521	0.00	115.35	0.06	53.61	0.18	114,858
Professionals	3217	0.00	116.14	0.11	71.03	0.01	123,551

Table 1 displays some descriptive statistics about the LGD, PD, and EAD by year, by exposure and customer type. For confidentiality reasons, we do not report the average values. The number of defaulted contracts per year ranges between 1,573 and 2,946. The maximum and the average (not reported) losses tend to decrease between 2011 and 2014. Credit and leasing have approximately the same average and maximum loss rate. The EAD ranges from less than 1 BRL to 123,550 BRL. The PD ranges from less than 1% to 71%, but almost 3/4 of the PD values are below 10%.

Figure 3: Empirical distribution of the LGDs
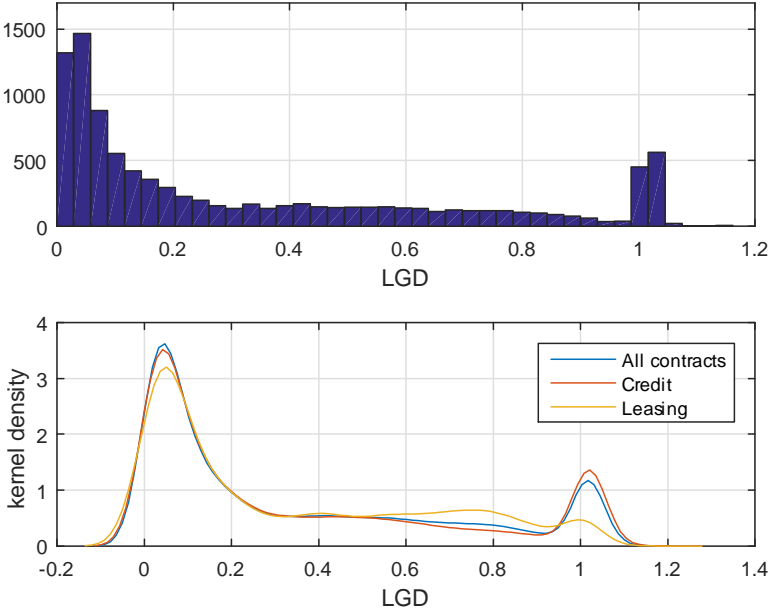
These figures hide a great heterogeneity of the recovery rates. The empirical distribution of the 9,738 workout LGDs is displayed on the top panel of Figure 3. Three remarks should be made here. First, 10.58% of the defaulted contracts have a recovery rate that exceeds 100%, with a maximum value of 116.14%, due to the workout costs.[14] Second, the kernel density estimate of the LGD distribution is bimodal (bottom panel of Figure 3), meaning that the percentage of exposure is either relatively high or low. This finding confirms the property

---

[14]Notice that this percentage is smaller than those generally observed in the literature. For instance, in the leasing industry, Schmit and Stuyck (2002) or Schmit (2004) report that up to 59% of all defaulted contracts in their sample have a recovery rate that exceeds 100%.

largely documented in the literature (Schuermann (2004)). Finally, the LGD distributions for the credit and leasing contracts are relatively close, except for the right part of the distribution. The probability to observe a high loss is less important for leasing than for credit. This difference illustrates the role of the collateral in the recovery processes (in the case of leasing, the vehicle belongs to the bank and plays the same role as a collateral).

Figure 4: Scatter plot of LGD versus log(EAD)
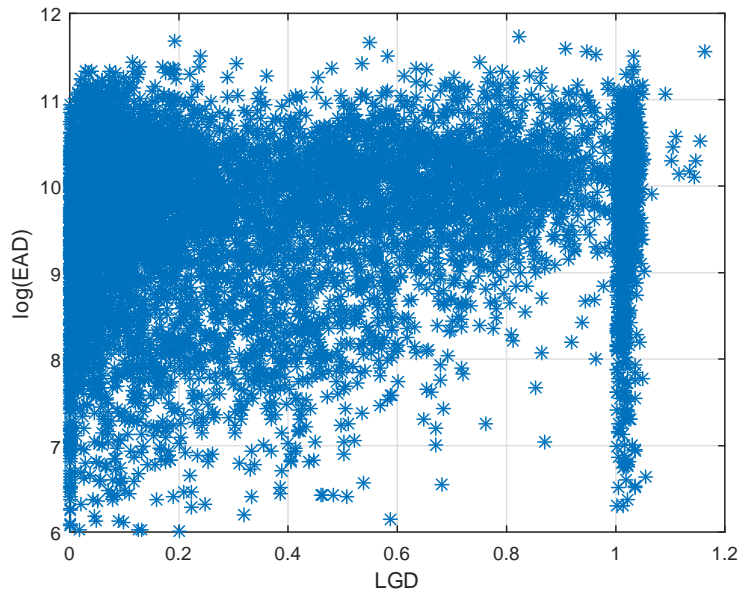
Finally, Figure 4 displays the scatter plot of LGD versus the logarithm of EAD. We find a positive correlation between the LGD and the EAD. The correlation is relatively small (0.11), but significant. This observation justifies the introduction of the exposure as explanatory variable in our LGD models.

4.2 Competing LGD Models

For our comparison, we consider six competing LGD models which are commonly used in academic and practitioner literature (see for instance Bastos (2010), Qi and Zhao (2011), Loterman et al. (2012), etc.), namely (1) the fractional response regression model, (2) the regression tree, (3) the random forest, (4) the gradient boosting, (5) the artificial neural network, and (6) the support vector machine. In the sequel, we briefly present these competing

models and mention the main references for further details.

4.2.1 Fractional response regression

The fractional response regression (FRR) model, proposed by Papke and Wooldridge (1996), allows to model the conditional mean of continuous variable defined over $[0, 1]$. The FRR specification is defined as

$$\mathbb{E}\left(\text{LGD}_i | X_i\right) = G\left(X_i'\beta\right) \tag{13}$$

where X_i is a k-vector of explanatory variables for the ith loan, β a k-vector of parameters and $G\left(.\right)$ a link function, with $G : \mathbb{R} \to [0, 1]$. A natural choice for the link function is the logistic function with

$$G\left(X_i'\beta\right) = \frac{1}{1 + \exp\left(-X_i'\beta\right)} \tag{14}$$

The model parameters are estimated by quasi-maximum likelihood (QML), where the quasi likelihood is defined as a modified Bernouilli likelihood. If we denote by $\widehat{\beta}$ the QML estimator of β, the LGD estimator is then given by $\widehat{\text{LGD}}_i = G(X_i'\widehat{\beta})$.

4.2.2 Regression tree

The regression tree (TREE), initially introduced by Breiman et al. (1984), is a machine-learning forecasting method. For a continuous variable, the tree is obtained by recursively partitioning the covariates space according to a prediction error (defined as the squared difference between the observed and predicted values) and then, by fitting a simple mean prediction within each partition.

The sketch of a regression tree algorithm is the following. The algorithm starts with a root node gathering all observations. For each covariate X, find the set R that minimizes the sum of the node impurities in the two child nodes and choose the split that gives the minimum overall X and R. The splitting procedure continues until no significant further reduction of the sum of squared deviations is possible. At the end of the procedure, we get a partition into K regions $R_1, \ldots, R_K$, also called terminal nodes or leaves. For each terminal node k, the LGD forecast is then given by the average LGD, denoted $\overline{\text{LGD}}_k$, estimated from all the contracts that belong to the region R_k, with

$$\widehat{\text{LGD}}_i = \sum_{k=1}^{K} \overline{\text{LGD}}_k \times \mathbb{I}_{(X_i \in R_k)} \tag{15}$$

There exist many algorithms for regression trees. Here, we consider the CART algorithm (Breiman et al. (1984)).

4.2.3 Random forest

Random forest (RF), introduced by Breiman (2001), is a Boostrap aggregation method of regression trees, trained on different parts of the same training set, with the goal of reducing overfitting (or, equivalently estimator variance). Random forest generally induces a small increase in the bias compared to regression trees and a loss of interpretability, but generally greatly boosts the performance of the model. In addition to constructing each tree using a different bootstrap sample of the data as in bagging approaches, random forests change how the regression trees are constructed. Indeed, each node is split using the best among a subset of covariates randomly chosen at that node. Assume that B regression trees are combined and denote by $\widehat{\text{LGD}}_{i,b}$ the prediction of the bth tree, then the random forest prediction is defined as

$$\widehat{\text{LGD}}_i = \frac{1}{B} \sum_{b=1}^{B} \widehat{\text{LGD}}_{i,b} \tag{16}$$

4.2.4 Gradient boosting

Gradient boosting (GB) is an iterative aggregation procedure that consecutively fits new models (typically regression trees) to provide a more accurate estimate of the dependent variable (Friedman (2001)). The general feature of this algorithm consists in constructing for each iteration, a new base-learner which is maximally correlated with the negative gradient of a loss function, evaluated at the previous iteration over the whole sample. In general, the choice of the loss function is up to the researcher, but most of the studies consider the quadratic loss function.[15]

The gradient boosting algorithm can be summarized as follows. A first regression tree is built on the LGD training set. Denote by $f_0(X_i)$ the prediction for the ith loan and define the corresponding residuals $r_{i0} = \text{LGD}_i - f_0(X_i)$ for $i =, .., n_t$. At the first iteration, a new regression tree is applied to the residuals r_{i0}. The LGD predictions are then updated using the iterative formula $f_1(X_i) = f_0(X_i) + r_{i1}$, where r_{i1} denotes the adjusted residuals issued

---

[15] Another possibility would consist to use our capital charge loss function for the gradient boosting algorithm. But, this new estimation method for LGD is beyond the scope of this paper.

from the regression tree. After M iterations, algorithm stops and the final LGD predictions are given by

$$\widehat{\text{LGD}}_i = f_0(X_i) + \sum_{m=1}^{M} r_{im} \tag{17}$$

4.2.5 Artificial neural network

Artificial neural networks (ANN) are a class of flexible non-linear models, initially introduced by Bishop (1995). It produces an output value by feeding inputs through a network whose subsequent nodes apply some chosen activation function to a weighted sum of incoming values. The type of ANN considered in this study is a multilayer perceptron similar to that used by Qi et Zhao (2011) for the LGD forecasts. It consists in a three-layer network based on input-layer units, hidden-layer units, and output-layer units. The central idea of the algorithm is (1) to extract linear combinations of the covariates from the input-layer units to the hidden-layer units, and (2) to apply nonlinear function on these derived features in the output-layer units to predict the dependent variable.

Let f be the unknown underlying function, through which a vector of input variables X explains LGD, i.e. $LGD_i = f(X_i)$. Derived features Z_m are created using linear combinations of the covariates such as

$$Z_{im} = G(\alpha'_m X_i), \quad \forall m = 1, \dots, M \tag{18}$$

where M is the number of hidden layer units, α_m a vector of coefficients (including a constant term) from the input-layer units to the hidden-layer units and $G\left(.\right)$ the logistic function, which is the common activation function used in neural network. The LGD are then modeled as a function of these linear combinations such that

$$f\left(X_i\right) = \beta_0 + \sum_{m=1}^{M} \beta_m Z_{im} + \varepsilon_i \tag{19}$$

where β_m are coefficients from the hidden-layer units to the output-layer units. The LGD forecasts are then given by $\widehat{\text{LGD}}_i = f\left(X_i\right).$

4.2.6 Support vector machine

Initially introduced by Vapnik (1995), support vector machine (SVM) is a machine learning tool for classification and regression. SVM has become popular for its ability to deal with

large data, its small number of meta-parameters, and its good results in practice. Given the continuous nature of the LGD variable, we consider SVM regression method as in Yao, Crook and Andreeva (2015). The objective in a SVM regression consists in approximating the response variable y_i by a function $f(.)$ that do not deviate from y_i more than a margin value ε for each observation of the sample, while simultaneously controlling for the model complexity. The details of the method are explained in appendix F.

5   Empirical application

The competing models are estimated on a training set of 7,791 credits (80% of the sample) and the pseudo out-of-sample forecasts are evaluated on a test set of 1,947 credits. For each model, we consider the same set of explanatory variables including the exposure at default, the contract duration, the interest or renting rate, the type of exposure (credit versus leasing), the customer type (individual or professional), the brand of the car, and the state of the car (new or second-hand).

Table 2: Descriptive statistics on the LGD and regulatory capital forecast errors

	FRR	ANN	TREE	SVM	RF	GB
			LGD errors			
minimum	-0.662	-0.681	-0.475	-0.511	-0.626	-0.505
maximum	0.888	1,001	0.874	1.058	0.990	0.890
mean	0.011	0.011	0.010	0.156	0.009	0.010
median	-0.136	-0.122	-0.142	0.005	-0.114	-0.138
variance	0.117	0.116	0.118	0.119	0.117	0.116
skewness	0.824	0.804	0.817	0.843	0.791	0.830
excess kurtosis	-0.618	-0.525	-0.605	-0.606	-0.473	-0.626
			Regulatory capital errors			
minimum	-11,689	-12,027	-8,279	-9,018	-10,135	-8,910
maximum	8,588	8,786	8,973	10,213	10,054	8,367
mean	60	44	66	732	38	66
median	-257	-231	-256	13	-232	-257
variance	3,813,596	3,799,691	3,730,561	4,071,567	3,737,503	3,717,518
skewness	0.55	0.41	0.80	1.36	0.61	0.79
excess kurtosis	2.52	2.83	2.01	2.84	2.73	1.91

Table 2 displays some figures about LGD and regulatory capital forecast errors, respectively defined by $\mathrm{LGD}_i - \widehat{\mathrm{LGD}}_{i,m}$ and $\mathrm{RC}_i - \widehat{\mathrm{RC}}_{i,m}$. Notice that, given this notation, a positive error implies an underestimation of the true value. The regulatory capital charges are computed with a PD of 40.45%, which corresponds to the maximal charge for the retail exposures. We observe that the empirical means of the LGD and RC forecast errors are slightly positive, whereas the medians are generally negative. This feature is due to the positive skewness observed for the errors of all models, and in particular for the SVM model. The kernel density estimates of the forecast errors distributions displayed in Figure 5, show that one can frequently observe capital requirement underestimates larger than 4,000 BRL, whereas similar overestimates are much rarer. Such a feature is problematic within a regulatory perspective, and justifies the use of asymmetric loss functions for comparing LGD models. We also observe in Table 2 that the excess kurtosis for the RC are positive, indicating fat tails for the error distributions. As concerned the LGD errors, the artificial neural network and the gradient boosting have the smallest variance. However, it is no longer the case for the artificial neural network when one considers the RC forecast errors. This result clearly illustrates the usefulness of our comparison approach.

Figure 6 displays the scatter plot of the LGD forecast errors (x-axis) and the RC forecast errors (y-axis), obtained with the support vector machine. Each point represents a contract (credit or leasing). This plot shows the great heterogeneity that exists between both type of errors. Due to the differences in EAD across borrowers, the magnitudes of the RC errors can drastically differ for the same level of LGD forecast error. Consider the two credits represented by the symbols A and B, with an EAD equal to 51,983 BRL and 6,024 BRL, respectively. For the same level of LGD forecast error (74.4%), the support vector machine slightly underestimates the capital requirement (953 BRL) in the case of the credit B, whereas the underestimation reaches 8,231 BRL in the case of credit A. Obviously, from a regulatory perspective, the second LGD error should be more penalized than the first one, as its consequence on the RC estimates are more drastic. The dispersion of the observations within the y-axis fully justifies our comparison approach for LGD models, based on expected loss functions expressed in terms of capital charge. Furthermore, the scatter plot confirms the asymmetric pattern of the errors distribution associated to the support vector machine model. This model

Figure 5: Kernel density estimate of the estimation error
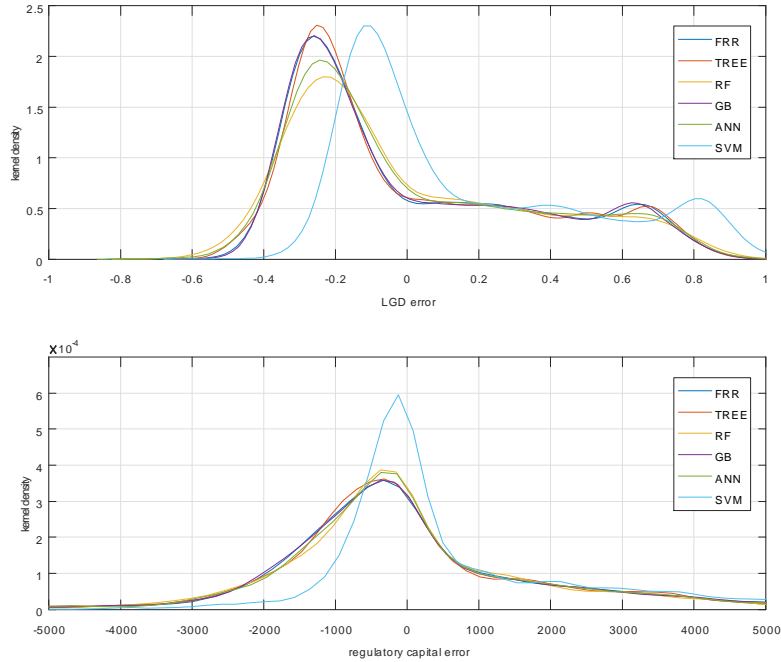
leads to relatively few overestimates (negative errors), both for LGD and RC, while it leads to large underestimates (positive errors). Thus, any competing LGD model that leads to less underestimates than the SVM should be preferred from a regulatory perspective. For this reason, we recommend the use of asymmetric loss functions.

These features (heterogeneity and asymmetry) are not specific to the SVM model, even if the skewness of the errors is more pronounced for this model compared to the other ones. Figure 7 shows that the profile of the scatter plots of the LGD and RC errors are quite similar for the 6 competing models. This similarity is due to the fact that we use the same set of covariates for all the models.

Table 3 displays the model rankings issued from two usual loss functions, namely the MSE and the MAE, associated to the LGD and regulatory capital forecast errors.[16] We also report the rankings issued from the corresponding asymmetric expected losses. The corresponding values of the losses are displayed in Table 7 (appendix G). Regarding the MSE, the gradient

---

[16]The values of the losses are displayed in Table 7 in appendix G.

Figure 6: Scatter plot of LGD versus regulatory capital forecast errors for the support vector machine (SVM) model
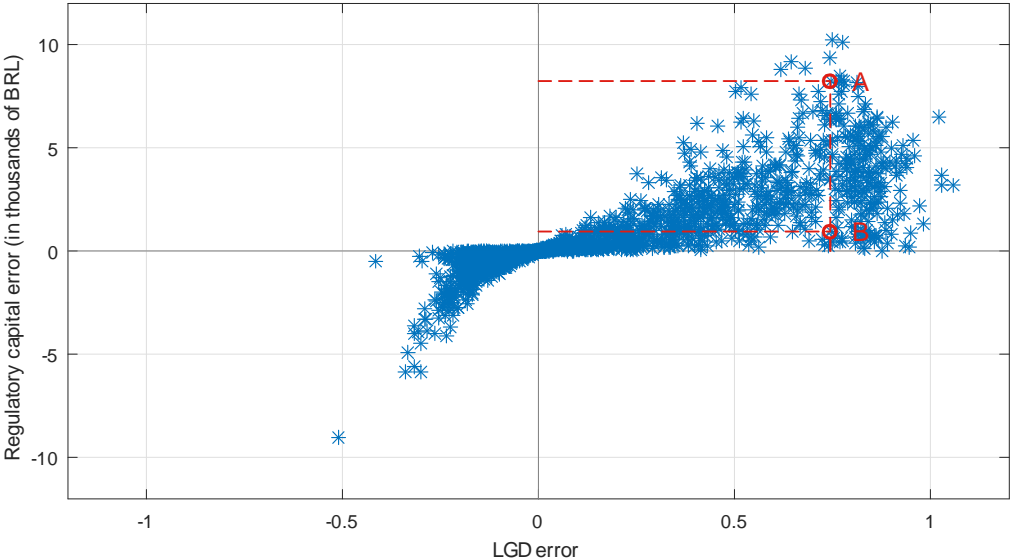
boosting is ranked as the best model, either for LGD or RC symmetric losses. But, when one considers asymmetric losses, it collapses to the penultimate rank and the random forest exhibits the best forecasting abilities. Two comments should be made here. First, in this empirical application, the best models identified by the LGD and RC-based approaches are the same. Nevertheless, this result should not be generalized. As we can observe, the rest of the LGD model rankings are not consistent. For instance, artificial neural network is identified as the second-best model with the LGD loss while it holds the fourth rank with the capital charge loss. Conversely, regression tree is ranked second with the capital charge loss while it is ranked at the penultimate position with the LGD-based loss. Similar results are obtained when one compares the rankings associated to the asymmetric LGD and RC-based loss functions. These inversions prove that the ranking consistency condition of proposition 2 is not valid, at least in our sample, for some couples of models. Second, our results highlight the usefulness of asymmetric loss functions. These functions penalize the models with the largest positive errors (underestimates) as the gradient boosting for instance. Notice that the support vector machine is the worst model, no matter if the loss is symmetric or not. As concerned the MAE criteria, we get similar conclusions, except for the support vector machine, which is ranked as

28

Figure 7: Scatter plot of the LGD and regulatory capital forecast errors (all models)
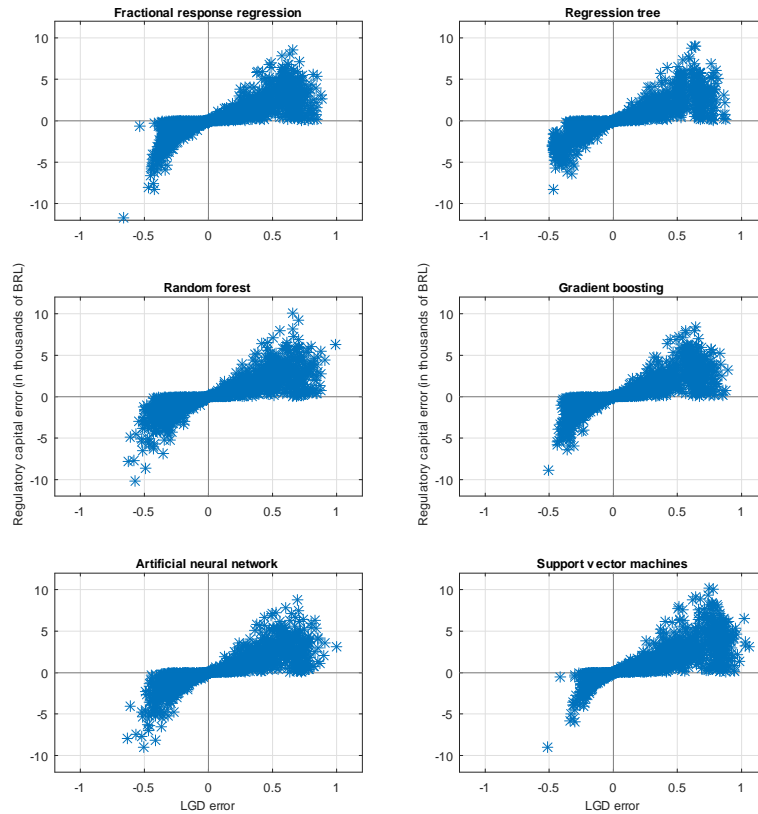
the best model when one considers symmetric LGD-based loss. Indeed, this model generates relatively few, but large errors. As a consequence, it is less penalized by the MAE criteria than by the MSE, which is more sensitive to extreme values. However, the support vector machine remains the worst model when one considers asymmetric loss functions, due to the large skewness of its forecast errors. Finally, we also observe some differences between the LGD-based and the RC-based rankings, even if these changes are less frequent than with the MSE criteria, confirming the inconsistency of both rankings.

6   Robustness checks

Our empirical results are robust to a variety of robustness checks. Firstly, instead of considering a common PD value for all the credits in the computation of the capital charges, we use

Table 3: Model rankings based on LGD and capital charge expected loss functions

Ranking	LGD Loss	CC Loss	Asym. LGD Loss	Asym. CC Loss
		Mean squared error		
1.	GB	GB	RF	RF
2.	ANN	TREE	FRR	ANN
3.	FRR	RF	ANN	FRR
4.	RF	ANN	TREE	TREE
5.	TREE	FRR	GB	GB
6.	SVM	SVM	SVM	SVM
		Mean absolute error		
1.	SVM	SVM	RF	RF
2.	RF	RF	FRR	ANN
3.	ANN	ANN	ANN	FRR
4.	GB	TREE	TREE	TREE
5.	FRR	GB	GB	GB
6.	TREE	FRR	SVM	SVM

the individual PD calculated by the internal bank's risk model one year before the default occurs. The corresponding LGD model rankings are reported in Table 4. The corresponding values of the losses are displayed in the bottom part of Table 7 in appendix G. The rankings based on the MSE are similar to those obtained with a common PD (cf. Table 3). The only change concerns the gradient boosting and the regression tree models in the case of asymmetric capital charge loss function. For the symmetric capital charge loss function, the only change concerns the random forest and the regression tree. As a consequence, we still observe model ranking inversions compared to the ranking based on the LGD loss functions.

Secondly, we extend the set of explanatory variables considered for the six competing LGD models. As several studies show that recoveries in recessions are lower than during expansions (Schuermann (2004), Bellotti and Crook (2012)), we introduce three additional macroeconomic variables in order to take into account the influence of the economic cycles on the recovery rate, namely the Brazilian GDP growth, unemployment and interest rates. Table 5 displays the corresponding LGD model rankings obtained for a common PD value. Similar (not reported) results are obtained when one considers the Basel PD estimates. With the MSE criterion, random forest outperforms all competing models whatever the loss function

Table 4: Model rankings based on LGD and capital charge (with Basel PD) expected loss functions

Ranking	LGD loss	CC loss	Asym. LGD loss	Asym. CC loss
Mean squared error				
1.	GB	GB	RF	RF
2.	ANN	RF	FRR	ANN
3.	FRR	TREE	ANN	FRR
4.	RF	ANN	TREE	GB
5.	TREE	FRR	GB	TREE
6.	SVM	SVM	SVM	SVM
Mean absolute error				
1.	SVM	SVM	RF	RF
2.	RF	RF	FRR	ANN
3.	ANN	ANN	ANN	FRR
4.	GB	TREE	TREE	TREE
5.	FRR	GB	GB	GB
6.	TREE	FRR	SVM	SVM

considered. It is also the case for the asymmetric MAE criterion. As in the previous cases, we observe a ranking inconsistency for other models, meaning that the condition of proposition 2 is not valid for these couples of models. The loss values reported in Table 8 (appendix G) are generally smaller than those obtained without macroeconomic variables, confirming the influence of the business cycle on the recovery rates.

Finally, we also consider the same type of regressions by excluding the exposure at default from the set of explanatory variables. The qualitative results (not reported) remain the same: we observe a global inconsistency of the LGD model rankings based on the LGD estimates or on the capital charge estimates. So, include (or exclude) the EAD as explanatory variable in the LGD models, has no consequence on the validity of the condition of proposition 2, since we only consider non-linear LGD models in our application.

Table 5: Model rankings based on LGD and capital charge expected loss functions: LGD models with macroeconomic variables and common PD

Ranking	LGD Loss	CC Loss	Asym. LGD Loss	Asym. CC Loss
	Mean squared error			
1.	RF	RF	RF	RF
2.	GB	TREE	ANN	ANN
3.	ANN	GB	TREE	TREE
4.	TREE	ANN	GB	GB
5.	FRR	FRR	FRR	FRR
6.	SVM	SVM	SVM	SVM
	Mean absolute error			
1.	SVM	SVM	RF	RF
2.	RF	RF	ANN	ANN
3.	ANN	ANN	TREE	TREE
4.	GB	TREE	GB	GB
5.	TREE	GB	FRR	FRR
6.	FRR	FRR	SVM	SVM

7 Conclusion

LGD is one of the key modeling components of the credit risk capital requirements. According to the AIRB approach adopted by most major international banks, the LGD forecasts are issued from internal risk models. While the practices seem to be well established for the PD modelling, no particular guideline has been proposed concerning how LGD models should be compared, selected, and evaluated. As a consequence, the model benchmarking method generally adopted by banks and academics simply consists in evaluating the LGD forecasts on a test set, with standard statistical criteria such as MSE, MAE, etc., as for any continuous variable. Thus, the LGD model comparison is done regardless of the other Basel risk parameters (EAD, PD, M) and by neglecting the impact of the LGD forecast errors on the regulatory capital. This approach may lead to select a LGD model that has the smallest MSE among all the competing models, but that induces small errors on small exposures, but large errors on large exposures.

We propose an alternative comparison methodology for the LGD models which is based on expected loss functions expressed in terms of regulatory capital charge. These loss func-

tions penalize more heavily the LGD forecast errors associated to large exposure or to long credit maturity. We also define asymmetric loss functions that only penalize the LGD models which lead to underestimating the regulatory capital, since these underestimations weaken the bank's ability to absorb unexpected credit losses. Using a sample of credits provided by an international bank, we illustrate the interest of our method by comparing the rankings of six competing LGD models. Our approach allows to identify the best LGD models associated with the lowest estimation errors on the regulatory capital. Besides, the empirical results confirm that the ranking based on a naive LGD loss function are drastically different from the models ranking obtained with the capital charge symmetric (or asymmetric) loss.

A natural extension of this work will consist to propose statistical tests designed to compare the expected capital charge losses for a pair of LGD models (DM-type test) or to identify a model confidence set (Hansen, Lunde, and Nason (2011)) that contains the "best" LGD models, for a given level of confidence.

A Asymptotic Single Risk Factor model

Here, we detail the sketch of the proof of the regulatory formula for the credit capital charge (for more details, see Roncalli (2009) or Gouriéroux and Tiomo (2007)). Let us consider a portfolio of n credits indexed by $i = 1, ..., n$. The portfolio loss is equal to

$$L = \sum_{i=1}^{n} \text{EAD}_i \times \text{LGD}_i \times D_i$$

where EAD_i is the exposure at default for the ith credit (assumed to be constant), LGD_i is the loss given default (random variable) and D_i is a binary random variable that takes a value 1 if there is a default before the residual maturity M_i and 0 otherwise. Formally, $D_i = 1_{(\tau_i \leq M_i)}$ where τ_i is the default time (random variable).

Assumption A1: *The default depends on a set of factors X and we denote by x the realization of X.*

Assumption A2: *The loss given default LGD_i is independent from the default time τ_i.*

Assumption A3: *The default times τ_i, $i = 1, ..., n$ are independent conditionally to the X factors.*

Assumption A4: *The portfolio is infinitely fine-grained, which means that there is no concentration, with*

$$\lim_{n \to \infty} \max \frac{\text{EAD}_j}{\sum_{i=1}^{n} \text{EAD}_i} = 0 \quad \forall j$$

Under assumptions A1-A4, it is possible to show that the conditional distribution of L given X degenerates to the conditional expectation $\mathbb{E}_X (L) = \mathbb{E} (L | X = x)$ and we get

$$L | X \xrightarrow{p} \mathbb{E}_X (L) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E} (\text{LGD}_i) \times p_i (x)$$

where $p_i (x) = \mathbb{E}_X (D_i) = \mathbb{E} (D_i = 1 | X = x)$ is the conditional default probability. Notice that under assumption A2, $\mathbb{E}_X (\text{LGD}_i) = \mathbb{E} (\text{LGD}_i)$. As a consequence, the portfolio loss has a marginal distribution given by

$$L \xrightarrow{d} g (X) = \sum_{i=1}^{n} \underbrace{\text{EAD}_i}_{\text{constant term}} \times \underbrace{\mathbb{E} (\text{LGD}_i)}_{\text{constant term}} \times \underbrace{p_i (X)}_{\text{random var.}}$$

Denote by F_L the cdf of L such that $F_L (l) \equiv \Pr (L \leq l) = \Pr (g (X) \leq l)$.

Assumption A5: There is only one factor X, with a cdf $F_X (.)$ and $p_i (X)$ is a decreasing function of X.

Under assumption A5, the α-VaR of the portfolio loss L is defined as $VaR_L (\alpha) = F_L^{-1} (\alpha) = g \left(F_X^{-1} (1 - \alpha) \right)$ or equivalently by

$$VaR_L (\alpha) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E} (\text{LGD}_i) \times p_i \left(F_X^{-1} (1 - \alpha) \right) = \sum_{i=1}^{n} RC_i$$

where RC_i denotes the risk contribution of the credit i. The VaR of an infinitely fine-grained portfolio can be decomposed as a sum of independent risk contributions, since RC_i only depends on the characteristics of the ith credit (exposure at default, loss given default and probability of default). Similarly, the marginal loss expectation is defined as

$$\mathbb{E}(L) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times p_i$$

where $p_i = \Pr(D_i = 1)$ corresponds to the unconditional probability of failure.

Assumption 6: *Let Z_i be the normalized asset value of the entity i. The default occurs when Z_i is below a given barrier B_i (level of debt), with*

$$D_i = 1 \quad \text{if} \quad Z_i \leq B_i$$

Assumption 7: *The asset value Z_i depends on a common risk factor X and an idiosyncratic risk factor ε_i, with*

$$Z_i = \sqrt{\rho}X + \sqrt{1-\rho}\varepsilon_i$$

where X and ε_i are two independent standard normal random variables, and ρ is the asset's correlation (or with the factor).

Under assumptions A6-A7, the conditional probability of default is equal to

$$p_i(x) = \Phi\left(\frac{B_i - \sqrt{\rho}x}{\sqrt{1-\rho}}\right)$$

where $\Phi(.)$ is the cdf of the standard normal distribution and the barrier B_i corresponds to the quantile associated to the unconditional probability of default, $B_i = \Phi^{-1}(p_i)$. Since $\Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$, we get

$$VaR_L(\alpha) = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times \Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$$

In order to determine the regulatory capital (RC), the BCBS considers the unexpected loss as the credit risk measure

$$\text{RC} = UL(\alpha) = VaR_L(\alpha) - \mathbb{E}(L)$$

Then, we get

$$\text{RC} = \sum_{i=1}^{n} \text{EAD}_i \times \mathbb{E}(\text{LGD}_i) \times \left(\Phi\left(\frac{\Phi^{-1}(p_i) + \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right) - p_i\right)$$

By considering a risk level $\alpha = 99.9\%$ and by denoting PD the unconditional probability of default, we get the IRB formula (without maturity adjustment).

B   Maturity adjustment and correlation functions

The maturity adjustment suggested by the BCBS depends on the type of exposure. For the corporate, sovereign, and bank exposures, it is defined as

$$\gamma\left(\mathrm{M}\right) = \frac{1 + \left(\mathrm{M} - 2.5\right) \times b\left(\mathrm{PD}\right)}{1 - 1.5 \times b\left(\mathrm{PD}\right)}$$

with the smoothed maturity adjustment equal to

$$b\left(\mathrm{PD}\right) = \left(0.11852 - 0.05478\log\left(\mathrm{PD}\right)\right)^2$$

For the retail exposures, there is no maturity adjustment, i.e. $\gamma\left(\mathrm{M}\right) = 1$. The correlation function $\rho\left(\mathrm{PD}\right)$ describes the dependence of the asset value of a borrower on the general state of the economy. Different asset classes show different degrees of dependency on the overall economy, so it's necessary to adapt the correlation coefficient to these classes. The correlation function $\rho\left(\mathrm{PD}\right)$ for corporate, sovereign, and bank exposures is defined as

$$\rho\left(\mathrm{PD}\right) = 0.12 \times \left(\frac{1 - e^{-50\ \mathrm{PD}}}{1 - e^{-50}}\right) + 0.24 \times \left(1 - \left(\frac{1 - e^{-50\ \mathrm{PD}}}{1 - e^{-50}}\right)\right)$$

For small and medium-sized enterprises (SME), a firm-size adjustment is introduced that depends on the sales. In the sequel, we neglect this adjustment for simplicity. For retail exposures, the correlation function $\rho\left(\mathrm{PD}\right)$ depends on the exposures. For the residential mortgage exposures, the BCBS recommends to fix the correlation at 0.15, for revolving retail exposures at 0.04 and for other retail exposures, to use the following formula

$$\rho\left(\mathrm{PD}\right) = 0.03 \times \left(\frac{1 - e^{-35\ \mathrm{PD}}}{1 - e^{-35}}\right) + 0.16 \times \left(1 - \left(\frac{1 - e^{-35\ \mathrm{PD}}}{1 - e^{-35}}\right)\right)$$

C   Proof of proposition 2

Proof. Under assumptions A1-A2, the capital charge expected loss can be expressed as

$$
\begin{aligned}
\mathcal{L}_{CC,m} &= \mathbb{E}\left(g\left(\eta_{i,m}\right)\right) \\
&= \mathbb{E}\left(g\left(\mathrm{EAD}_i \times \delta\left(\mathrm{PD}\right) \times \gamma\left(\mathrm{M}\right) \times \varepsilon_{i,m}\right)\right) \\
&= g\left(\delta\left(\mathrm{PD}\right)\right) \times g\left(\gamma\left(\mathrm{M}\right)\right) \times \mathbb{E}\left(g\left(\mathrm{EAD}_i \times \varepsilon_{i,m}\right)\right)
\end{aligned}
$$

since $\delta\left(\mathrm{PD}\right)$ and $\gamma\left(\mathrm{M}\right)$ are positive constant terms. Rewrite $\mathcal{L}_{CC,m}$ as

$$\mathcal{L}_{CC,m} = \Delta \times cov\left(g\left(\mathrm{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) + \Delta \times \mathbb{E}\left(g\left(\mathrm{EAD}_i\right)\right) \times \mathcal{L}_m$$

with $\Delta = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right)$ and $\mathcal{L}_m = \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right)$. Consider two LGD models m and $m+1$. The rankings of the two models are consistent as soon as $\mathcal{L}_m < \mathcal{L}_{m+1}$ and $\mathcal{L}_{CC,m} < \mathcal{L}_{CC,m+1}$. Since $\Delta > 0$, these conditions can be expressed as

$$cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) + \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathcal{L}_m < cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m+1}\right)\right) + \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathcal{L}_{m+1}$$

Or equivalently as

$$cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m}\right)\right) - cov\left(g\left(\text{EAD}_i\right), g\left(\varepsilon_{i,m+1}\right)\right) < \mathbb{E}\left(g\left(\text{EAD}_i\right)\right)\left(\mathcal{L}_{m+1} - \mathcal{L}_m\right)$$

with $\mathcal{L}_{m+1} - \mathcal{L}_m < 0$ and $\mathbb{E}\left(g\left(\text{EAD}_i\right)\right) > 0$. ∎

D   Proof of corollary 3

Proof. If the variables EAD_i and $\varepsilon_{i,m}$ are independent, the variables $g\left(\text{EAD}_i\right)$ and $g\left(\varepsilon_{i,m}\right)$ are also independent. Then, the capital charge expected loss becomes

$$\mathcal{L}_{CC,m} = \mathbb{E}\left(g\left(\eta_{i,m}\right)\right) = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right) \times \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) \times \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right)$$

Consider two LGD models m and $m+1$, $\forall m = 1, ..., \mathcal{M} - 1$, for which $\mathcal{L}_m < \mathcal{L}_{m+1}$, then we have

$$\Delta_i \times \mathbb{E}\left(g\left(\varepsilon_{i,m}\right)\right) < \Delta_i \times \mathbb{E}\left(g\left(\varepsilon_{i,m+1}\right)\right)$$

with $\Delta_i = g\left(\delta\left(\text{PD}\right)\right) \times g\left(\gamma\left(\text{M}\right)\right) \times \mathbb{E}\left(g\left(\text{EAD}_i\right)\right) > 0$. The ranking of LGD models are necessarily consistent, i.e. $\mathcal{L}_{CC,m} < \mathcal{L}_{CC,m+1}$. ∎

E Dataset description

Table 6: List of the variables

Variables type	Variables name	Description
Contract	Duration	Duration of the contract
	Time to default	Number of months before default
	Relative duration	Time to default divided by duration
	Interest rate	Interest (or renting) rate
	Exposition type	Credit or leasing
	Customer type	Individual, professional (natural or legal)
	Brand of the car	Brand name of the car
	State of the car	New or second-hand
Macroeconomic	GDP Growth rate	Brazil, quaterly
	Unemployment rate	Brazil, monthly
	Interbank market interest rate	Brazil, monthly
Basel parameters	EAD	Exposure at default
	PD	Basel default probability estimated by the bank
	LGD	Loss Given Default

F   Support vector machine

Suppose a set of training data $\{y_i, X_i\}_{i=1}^N$ in which y_i is the observed response value (i.e. LGD_i in our case) and X_i the associated k-vector of explanatory variables for the ith individual. Let us assume that y_i can be approximated by a linear function such that

$$f(X_i) = X_i'\beta$$

where β is a k-vector of unknown parameters. In a SVM regression, β is determined by solving a risk minimization problem with respect to an ϵ-insensitive loss function ($\epsilon \geq 0$). This ϵ-insensitive loss function belongs to the so-called *robust regression* family and is known to provide reliable forecasts for many distributional hypothesis made on the regression noise (see Vapnik (1995) for more details). In the following, we consider the well-known *linear* ϵ-insensitive loss function defined as

$$L_\epsilon(y, f(X)) = \begin{cases} 0 & \text{if } |y - f(X)| \leq \epsilon \\ |y - f(X)| - \epsilon & \text{otherwise} \end{cases}$$

SVM regression hence consists in finding the value of β that solves a convex minimization problem subject to a L_ϵ-based constraint. One has to minimize

$$J(\beta) = \frac{1}{2}\beta'\beta$$

subject to

$$|y_i - Xi'\beta| \leq \epsilon, \qquad \forall i = 1, \ldots, N$$

The minimization of the objective $J(\beta)$ allows to control appropriately for overfitting, while the constraints impose $f(X_i)$ to deviate from y_i by a value no greater than ϵ for all the observations. As the constraints cannot be satisfied for some observations, slack variables $\{\xi_i, \xi_i^*\}_{i=1}^N$ are introduced in order to get a feasible problem. With these slack variables, the primal formula becomes

$$\Phi(\beta, \xi^*, \xi) = \frac{1}{2}\beta'\beta + C\left(\sum_{i=1}^N \xi_i^* + \sum_{i=1}^N \xi_i\right)$$

under the constraints

$$y_i - X_i'\beta \leq \epsilon + \xi_i^*, \quad i = 1, \ldots, N$$
$$X_i'\beta - y_i \leq \epsilon + \xi_i, \quad i = 1, \ldots, N$$
$$\xi_i^* \geq 0, \qquad i = 1, \ldots, N$$
$$\xi_i \geq 0, \qquad i = 1, \ldots, N$$

The constant C is the box constraint, a positive regularization parameter that controls the penalty imposed on observations that lie outside the ϵ margin. Therefore, this parameter determines the trade-off between the model complexity (flatness) and the degree to which deviations larger than ϵ are tolerated. This optimization problem can be solved in a simpler way using its Lagrange dual formulation counterpart. The dual formula requires the introduction of nonnegative multipliers denoted $\{\alpha_i, \alpha_i^*\}_{i=1}^N$ in the optimization problem leading to the maximization of

$$W\left(\alpha, \alpha^*\right) = -\epsilon \sum_{i=1}^N \left(\alpha_i^* + \alpha_i\right) + \sum_{i=1}^N y_i \left(\alpha_i^* - \alpha_i\right) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right) X_i' X_j$$

subject to constraints

$$\sum_{i=1}^N \alpha_i^* = \sum_{i=1}^N \alpha_i,$$

$$0 \leq \alpha_i^* \leq C, \qquad i = 1, \ldots, N$$

$$0 \leq \alpha_i \leq C, \qquad i = 1, \ldots, N$$

Using optimized $\{\alpha_i, \alpha_i^*\}_{i=1}^N$ multipliers and $\{X_i\}_{i=1}^N$ covariates allow to compute

$$\beta = \sum_{i=1}^N \left(\alpha_i^* - \alpha_i\right) X_i$$

We finally get

$$f\left(X_i\right) = X_i' \sum_{i=1}^N \left(\alpha_i^* - \alpha_i\right) X_i$$

G Empirical means of the losses

Table 7: Empirical means of the LGD and capital charge losses

			FRR	ANN	TREE	SVM	RF	GB
			Common PD (0.40451)					
MSE	Standard	LGD	0.1169	0.1160	0.1176	0,1428	0.1170	0.1160
		CC	3,815,235	3,799,711	3,732,987	4,604,789	3,737,030	3,720,010
	Asymmetric	LGD	0.2010	0.2014	0.2018	0.2646	0.1982	0.2019
		CC	6,036,659	5,898,476	6,131,352	8,212,090	5,835,150	6,179,716
MAE	Standard	LGD	0.2906	0.2856	0.2908	0.2729	0.2856	0.2896
		CC	1,373.96	1,353.61	1,366.04	1,306.95	1,342.20	1,367.34
	Asymmetric	LGD	0.3815	0.3817	0.3817	0.4243	0.3752	0.3843
		CC	1,815.28	1,800.22	1,819.83	2,016.85	1,758.73	1,836.43
			Basel PD					
MSE	Standard	LGD	0.1169	0.1160	0.1176	0.1428	0.1170	0.1160
		CC	1,041,302	1,034,879	1,009,601	1,173,002	1,008,480	1,005,158
	Asymmetric	LGD	0.2010	0.2014	0.2018	0.2646	0.1982	0.2019
		CC	1,450,604	1,435,094	1,504,186	2,008,958	1,414,640	1,502,921
MAE	Standard	LGD	0.2906	0.2856	0.2908	0.2729	0.2856	0.2896
		CC	693.70	682.04	686.89	651.49	677.57	688.98
	Asymmetric	LGD	0.3815	0.3817	0.3817	0.4243	0.3752	0.3843
		CC	869.97	861.95	878.73	975.01	844.15	884.70

Table 8: Empirical means of the LGD and capital charge losses (LGD models with macroeconomic variables)

			FRR	ANN	TREE	SVM	RF	GB
			Common PD (0,40451)					
MSE	Standard	LGD	0.1125	0.1108	0.1117	0.1307	0.1092	0.1101
		CC	3,589,815	3,530,643	3,470,465	4,107,855	3,441,599	3,476,828
	Asymmetric	LGD	0.1994	0.1870	0.1904	0.2425	0.1843	0.1921
		CC	5,792,513	5,480,483	5,678,123	7,376,466	5,443,856	5,716,877
MAE	Standard	LGD	0.2805	0.2731	0.2768	0.2626	0.2705	0.2766
		CC	1,316.02	1,279.29	1,296.38	1,243.49	1,270.13	1,304.60
	Asymmetric	LGD	0.3807	0.3613	0.3683	0.4030	0.3583	0.3721
		CC	1,784.87	1,691.39	1,741.11	1,897.82	1,686.07	1,758.48
			Basel PD					
MSE	Standard	LGD	0.1125	0.1108	0.1117	0.1307	0.1092	0.1101
		CC	991,435	971,479	952,024	1,066,160	942,159	945,860
	Asymmetric	LGD	0.1994	0.1870	0.1904	0.2425	0.1843	0.1921
		CC	1,442,541	1,389,007	1,462,395	1,851,081	1,367,273	1,412,801
MAE	Standard	LGD	0.2805	0.2731	0.2768	0.2626	0.2705	0.2766
		CC	666.01	646.16	651.15	620.44	641.51	658.36
	Asymmetric	LGD	0.3807	0.3613	0.3683	0.4030	0.3583	0.3721
		CC	861.93	814.91	850.51	921.58	813.93	849.72

References

[1] Bastos J.A. (2010). Forecasting Bank Loans Loss-Given-Default. *Journal of Banking and Finance*, 34(10), 2510–2517.

[2] Basel Committee on Banking Supervision (2004). An Explanatory Note on the Basel II IRB Risk Weight Functions. Consultation Paper, October.

[3] Basel Committee on Banking Supervision (2005). An Explanatory Note on the Basel II IRB Risk Weight Functions. Consultation Paper, July.

[4] Bellotti T. & Crook J. (2012). Loss Given Default Models Incorporating Macroeconomic Variables for Credit Cards. *International Journal of Forecasting*, 28, 171-182.

[5] Bishop C.M. (1995). Neural Networks for Pattern Recognition. *Oxford University Press.*

[6] Breiman L., Friedman J., Stone C.J. & Olshen R.A. (1984). Classification and Regression Trees. *Chapman and Hall/CRC.*

[7] Breiman L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

[8] Calabrese R. (2014). Downturn Loss Given Default: Mixture Distribution Estimation. *European Journal of Operational Research*, 237, 271-277.

[9] Calabrese R. & Zenga M. (2010). Bank Loan Recovery Rates: Measuring and Nonparametric Density Estimation. *Journal of Banking and Finance*, 34(5), 903–911.

[10] Caselli S., Gatti S. & 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.

[11] Cochrane J.H. (2011). Presidential Address: Discount Rates. *Journal of Finance*, 66(4), 1047–1108.

[12] Dermine J. & Neto de Carvalho C. (2006). Bank Loan Losses-Given-Default: A Case Study. *Journal of Banking and Finance*, 30(4), 1219–1243.

[13] Diebold F.X. & Mariano R.S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13(3), 253-263.

[14] European Banking Authority (2016). Guidelines on PD Estimation, LGD Estimation and the Treatment of Defaulted Exposures. Consultation Paper, November.

[15] Friedman J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232.

[16] Gouriéroux C. & Tiomo A. (2007). Risque de crédit : une approche avancée. *Economica.*

[17] Gouriéroux C., Monfort A. & Polimenis V. (2006). Affine Models for Credit Risk Analysis. *Journal of Financial Econometrics*, 4(3), 494–530.

[18] Genest B. & Brie L. (2013). Basel II IRB Risk Weight Functions: Demonstration and Analysis. Working paper.

[19] Granger C.W.J. (1999). Outline of Forecast Theory Using Generalized Cost Functions. *Spanish Economic Review*, 1(2), 161–173.

[20] Gupton G.M. & Stein R.M. (2002). LOSSCALC: Model for Predicting Loss Given Default. Moody's technical report.

[21] Gürtler M. & Hibbeln M.T. (2013). Improvements in Loss Given Default Forecasts for Bank Loans. *Journal of Banking and Finance*, 37, 2354-2366.

[22] Gürtler M., Hibbeln M.T. & Usselman P. (2017). Exposure at Default Modeling: A Theoretical and Empirical Assessment of Estimation Approaches and Parameter Choice. Forthcoming in *Journal of Banking and Finance.*

[23] Hagmann M., Renault O. & Scaillet O. (2005). Estimation of Recovery Rate Densities: Nonparametric and Semi-parametric Approaches versus Industry Practice. *Recovery Risk: The Next Challenge in Credit Risk Management*, Altman, Resti, Sironi (ed.), 323–346.

[24] Hansen P.R. & Lunde A. (2006). Consistent Ranking of Volatility Models. *Journal of Econometrics*, 131(1–2), 97–121.

[25] Hansen P.R., Lunde A. & Nason J.M. (2011). The Model Confidence Set. *Econometrica*, 79(2), 453–497.

[26] Hartmann-Wendels T., Miller P. & Töws E. (2014). Loss Given Default for Leasing: Parametric and Nonparametric Estimations. *Journal of Banking and Finance*, 40, 364-375.

[27] Khieu H.D., Mullineaux D.J. & Yi H. (2012). The Determinants of Bank Loan Recovery Rates. *Journal of Banking and Finance*, 36(4), 923–933.

[28] Krüger S. & Rösh D. (2017). Downturn LGD Modeling using Quantile Regression. *Journal of Banking and Finance*, 79, 42-56.

[29] Laurent S., Rombouts J.V.K. & Violante F. (2013). On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models. *Journal of Econometrics*, 173(1), 1–10.

[30] Loterman G., Brown I., Martens D., Mues C. & Baesens B. (2012). Benchmarking Regression Algorithms for Loss Given Default Modeling. *International Journal of Forecasting*, 28(1), 161–170.

[31] Matuszyk A., Mues C. & Thomas L.C. (2010). Modelling LGD for Unsecured Personal Loans: Decision Tree Approach. *Journal of the Operational Research Society*, 61(3), 393–398.

[32] Merton R.C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29(2), 449–470.

[33] Miller P. & Töws E. (2017). Loss Given Default Adjusted Workout Processes for Leases. forthcoming in the *Journal of Banking and Finance*.

[34] Ospina R. & Ferrari S.L.P. (2010). Inflated Beta Distributions. *Statistical Papers*, 51(1), 111–126.

[35] Papke L.E. & Wooldridge J.M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, 11(6), 619–632.

[36] Patton A.J. (2011). Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics*, 160(1), 246–256.

[37] Patton A.J. (2016). Comparing Possibly Misspecified Forecasts. Working Paper.

[38] Qi M. & Zhao X. (2011). Comparison of Modeling Methods for Loss Given Default. *Journal of Banking and Finance*, 35(11), 2842–2855.

[39] Renault O. & Scaillet O. (2004). On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities. *Journal of Banking and Finance*, 28(12), 2915–2931.

[40] Roncalli T. (2009). La Gestion des Risques Financiers. *Economica*, 2ème édition.

[41] Schmit M. (2004). Credit Risk in the Leasing Industry. *Journal of Banking and Finance*, 28(4), 811–833.

[42] Schmit M. & Stuyck J. (2002). Recovery Rates in the Leasing Industry. Working Paper.

[43] Schuermann T. (2004). What Do We Know About Loss Given Default? *Credit Risk Models and Management*, Shimko (ed.).

[44] Tanoue Y., Kawada A. & Yamashita S. (2017). Forecasting Loss Given Default of Bank Loans with Multi-Stage Model. *International Journal of Forecasting*, 33, 513-522.

[45] Töws E. (2016). Advanced Methods for Loss Given Default Estimation. PhD thesis, Universität zu Köln

[46] Vapnik V.N. (1995). The Nature of Statistical Learning Theory. *Springer*, First Edition.

[47] Vasicek O.A. (2002). The Distribution of Loan Portfolio Value. *Risk*, 15(12), 160–162.

[48] Yao X., Crook J. & Andreeva G. (2015). Support Vector Regression for Loss Given Default Modelling. *European Journal of Operational Research,* 240, 528-538.