Jérémy LEYMARIE

Address :
Université d'Orléans
Laboratoire d'Economie d'Orléans
Rue de Blois
45100 Orléans, France

Born July 21, 1989
French citizen
Phone : +33 6 73 63 08 82
Email : jeremy.leymarie@univ-orleans.fr
http://sites.google.com/site/researchjeremyleymarie

Education

2016 – present **Ph.D. in Finance and Econometrics, University of Orléans**
Thesis title : « Three Essays in Financial Econometrics »
Supervisors : Prof. Christophe Hurlin, and Prof. Denisa Banulescu-Radu

oct. nov. 2018 **Visiting Scholar, Maastricht University, the Netherlands**
Sponsor : Prof. Alain Hecq

2014 – 2015 **Master in Econometrics and Applied Statistics, University of Orléans**
high honours, ranked 1st

2009 – 2013 **Bachelor in Economic Policy Analysis, University of Auvergne**
high honours, ranked 2nd

References

Christophe Hurlin
Full professor
University of Orléans, France
christophe.hurlin@univ-orleans.fr

Denisa Banulescu-Radu
Associate professor
University of Orléans, France
georgiana.banulescu@univ-orleans.fr

Research interests

Primary : Capital requirement, Financial econometrics, Financial regulation, Financial stability, Forecasting, Hypothesis testing, Model comparison, Inference, risk measure, risk management.

Secondary : Methods for Big Data, Machine Learning Algorithms.

Publications

- "**Loss Functions for Loss Given Default Model Comparison**" with C. Hurlin, and A. Patin, 2018. *European Journal of Operational Research, Vol 268, Issue 1, 348-360*

  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.

Working Papers

— "**Backtesting Marginal Expected Shorfall and Related Systemic Risk Measures**" with D. Banulescu, C. Hurlin, O. Scaillet, 2017 (Research grant, Fondation Banque de France)

This paper proposes two backtesting tests to assess the validity of the systemic risk measure forecasts. This new tool meets the need of financial regulators of evaluating the quality of systemic risk measures generally used to identify the financial institutions contributing the most to the total risk of the financial system (SIFIs). The tests are based on the concept of cumulative violations and it is built up in analogy with the recent backtesting procedure proposed for ES (Expected Shortfall). First, we introduce two backtests that apply for the case of the MES (Marginal Expected Shortfall) forecasts. The backtesting methodology is then generalised to MES-based systemic risk measures (SES, SRISK) and to the Delta CoVaR. Second, we study the asymptotic properties of the tests in presence of estimation risk and we investigate their finite sample performances via Monte Carlo simulations. Finally, we use our backtests to assess the validity of the MES, SRISK and Δ CoVaR forecasts on a panel of EU financial institutions.

— "**Backtesting Expected Shortfall via Multi-Quantile Regression**", with O. Couperier, 2018 (R&R in *Journal of Business and Economic Statistics*)

In this article, we propose a new approach to backtest Expected Shortfall (ES) exploiting the definition of ES as a function of Value-at-Risk (VaR). Our methodology jointly examines the validity of the VaR forecasts along the tail distribution of the risk model, and hence encompasses the Basel Committee recommendation of verifying quantiles at risk levels 97.5%, and 99%. We introduce four easy-to-use backtests in which we regress the ex-post losses on the VaR forecasts in a multi-quantile regression model, and test the resulting parameter estimates. Monte-Carlo simulations show that our tests are powerful to detect various model misspecifications. We apply our backtests on S&P500 returns over the period 2007-2012. Our tests clearly identify misleading ES forecasts in this period of financial turmoil. Empirical results also show that the detection abilities are higher when the evaluation procedure involves more than two quantiles, which should accordingly be taken into account in the current regulatory guidelines.

— "**Big Data : A new trick for credit scoring ?**", 2017

FinTechs and online banks argue that Big Data will disrupt the creditworthiness assessment in general, and the credit scoring practices in particular. Big Data credit scoring consists in building new credit scores, especially for online credits, or refining the current scores, with thousands of pieces of data about the loan applicants, collected from a variety of online sources such as social and professional media, Google search terms, IP address, device used, etc. In this paper, we propose an original and unifying methodological approach that allows evaluating the Big Data credit scoring models. Our goal is threefold : (1) evaluate the classification gains associated to Big Data scoring methodologies, (2) identify the effects of the new modelling and variable selection methods required by Big Data, (3) study the characteristics of individuals for which Big Data scoring has changed the creditworthiness assessment, compared to that obtained with traditional scoring models.

Conferences and workshops

1. **CFE**, 12th International Conference on Computational and Financial Econometrics, University of Pisa, Italy, 14-16 December 2018
2. **Journée d'Econométrie**, 17th conference "Développements Récents de l'Econométrie Appliquée à la Finance", University of Nanterre, France, November 7, 2018
3. **IAAE**, Annual Conference of the International Association for Applied Econometrics, Montréal, Canada, June 26-29, 2018
4. **2018 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics**, RiskLab at Arcada, Bank of Finland and European Systemic Risk Board, Helsinki, Finland, 28-30 May, 2018
5. **AFFI**, 35th Spring International Conference of the French Finance Association, ESCP Europe, Paris, May 22-24, 2018
6. **AFSE**, 67th Annual Meeting of the French Economic Association, Paris School of Economics, Paris, May 14-16, 2018
7. **7th PhD Student Conference in International Macroeconomics and Financial Econometrics**, University of Nanterre, Paris, March 16, 2018
8. **2nd International Bordeaux Workshop in Quantitative Finance, Risk, and Decision Theory**, Unversity of Bordeaux, France, November 24, 2017
9. **Journée d'Econométrie**, 16th conference "Développements Récents de l'Econométrie Appliquée à la Finance", University of Nanterre, France, November 8, 2017
10. **JEAM**, 5th applied macroeconometric workshop "Journée d'Econométrie Appliquée à la Macroéconomie", University of Paris 13, France, October 13, 2017
11. **CREDIT**, 16th International Conference on Credit Risk Evaluation, Italy, Venice, September 28-29, 2017 (poster)
12. **Bank Colloquium for Junior Researchers : The Future of Bank Regulation**, Limoges, France, June 28, 2017
13. **MFS**, 24th Annual Conference of the Multinational Finance Society, Bucharest, Romania, June 25-28, 2017
14. **INFER**, 19th Annual Conference of the International Network for Economic Research, Bordeaux, France, June 7-9, 2017
15. **ACDD**, 14th Augustin Cournot Doctoral Days, Strasbourg, France, April 27-28, 2017
16. **SNDE**, 25th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, Paris, France, March 30-31, 2017

17. **SMYE**, 22nd Spring Meeting of Young Economists, Halle (Saale), Germany, March 23-25, 2017
18. **VieCo**, Vienna–Copenhagen Conference on Financial Econometrics, Vienna, Austria, March 9-11, 2017
19. **CFE**, 10th International Conference on Computational and Financial Econometrics, Seville, Spain, December 9-11, 2016
20. **FEC**, 8th French Econometrics Conference, ESSEC, Cergy, France, November 17-18, 2016
21. **Journée d'Econométrie**, 15th conference "Développements Récents de l'Econométrie Appliquée à la Finance", Paris Ouest - Nanterre La Défense, France, November 4, 2016
22. **MIFN**, 10th International Workshop of Methods in International Finance Network, Canterbury, UK, October 27-28, 2016
23. **NFR Seminar**, 5th New Financial Reality Seminar, University of Orléans and Kent University, France, October 3, 2016
24. **ERFIN**, Econometric Research in Finance Workshop, Warsaw School of Economics, Poland, September 16, 2016
25. **ESEM**, 69th European meeting of the Econometric Society, Geneva, Switzerland, August 22-26, 2016
26. **AFSE**, 65th Annual Meeting of the French Economic Association, Nancy, France, June 27-29, 2016
27. **FEBS**, 6th International Conference of the Financial Engineering and Banking Society, Malaga, Spain, June 10-12, 2016
28. **AFFI**, 33rd French Finance Association Conference, Liège, Belgium, May 23-25, 2016
29. **ACPR Chair** "Regulation and Systemic Risk", House of Finance Day Conference, Paris, France, March 24, 2016

Seminars

- **Econometric seminar**, Maastricht University, the Netherlands, November 15, 2018
- **Econometric seminar**, ENSAE-CREST, Paris, France, April 20, 2017
- **Ph.D. Seminar**, University of Orleans, Orleans, France, May 4, 2016

Scholarships and Grants

2018 – 2019 **Research grant sponsored by the Fondation Banque de France**
Awarded project : Backtesting Systemic Risk Measures
Participants : D. Banulescu (University of Orleans), C. Hurlin (University of Orleans), O. Scaillet (University of Geneva), Jérémy Leymarie (University of Orleans)

2017 **Award at the Augustin Cournot Doctoral Days conference**
3rd best paper

2016 – 2018 **Ph.D. Scholarship**
Regional Research Allocation

Participation in Funded Research Projects

2017 - 2020 **ANR MultiRisk**
Project : Econometric Methods for the Modelling of Multiple Risks
Coordinator : Christophe Hurlin (University of Orleans), Gaëlle LeFol (University Paris Dauphine), Jean-Michel Zakoian (CREST)

Refereeing activities

Referee for the reviews : Economic Modelling, Empirical Economics, Research in International Business and Finance, International Economics, International Journal of Forecasting.

Other research related activities

- Econometric Game contest, Amsterdam, Netherlands, 11-13 April, 2018. Captain of University of Orléans' team

- Oxmetrics summer school – Intensive Ph.D. courses (sept. 1-10, 2015) organized by the University of Aix-Marseille. Courses : financial econometrics, state-space models, econometrics for intra-day data, Ox programming. Teachers : Prof. D. Hendry, J.A. Doornik, S. Laurent, and S.J Koopman

- Member of the « Association des Doctorants et Docteurs Orléanais en Sciences de l'Homme et de la Société » (ADDOSHS) - PhD Students Association, Orléans, France (2016 - 2018)

- Training courses for research staff at the University of Orléans (2015 - 2017)

- Member of the development team for the project RunMyCode (2017)

Teaching

Master, first year – Finance, University of Orléans (more informations)

- Quantitative Techniques in Finance (2018,2019) - Tutorial, 30 TH

Master, first year – Econometrics and Applied Statistics, University of Orléans (more informations)

- Econometrics of Discrete Choice Models (2016, 2017, 2018,2019) - Tutorial, 120 TH

Undergraduate, third year – DEG, University of Orléans (more informations)

- Econometrics for Finance (2019) - Lecture, 24 LH
- Mathematical Statistics (2016, 2017, 2018) - Tutorial, 75 TH
- Mathematical Statistics (sept. - nov. 2016, nov. 2017) - Lecture, 31 LH

Undergraduate, second year – DEG, University of Orléans (more informations)

- Game Theory (2019) - Tutoral, 10 TH
- Personal and Professional Project (2016, 2017) - Tutorial, 8 TH

RunMyCode project

Objective : The RunMyCode project is a web service devoted to the worldwide dissemination of scientific computer codes in the fields of business and economics. The key element in this project is the novel concept of companion website associated with a scientific paper (working paper, journal, article, monograph or textbook). The project aims to contribute for changing the way researchers spread their research and improve the reproducibility of research. The RunMyCode project is sponsored by the CNRS, HEC Paris and Sloan Foundation.

Activity : Member of the development team for RunMyCode.

My companion websites :

1. "Loss Functions for Loss Given Default Model Comparison" with C. Hurlin, and A. Patin, 2018. *European Journal of Operational Research, Vol 268, Issue 1, 348-360* [companion website]
2. "Backtesting Expected Shortfall via Multi-Quantile Regression" with O. Couperier, 2018. Working paper. [companion website]

Computing skills

- SAS (SAS BASE PROGRAMMING version 9.2 Certification)
- Matlab, R, OxMetrics, E-views, GeoDa
- LaTeX, Microsoft Office

Information systems

- CRSP, Computstat, Datastream, Bankscope

Languages

- **French** – mother tongue
- **English** – fluent (spoken and written)

European Journal of Operational Research 268 (2018) 348–360

Contents lists available at ScienceDirect

European Journal of Operational Research

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

ELSEVIER

Innovative Applications of O.R.

Loss functions for Loss Given Default model comparison☆

Christophe Hurlin*, Jérémy Leymarie, Antoine Patin

University of Orléans (LEO, UMRS CNRS 7332), 11 rue de Blois. 45000 France

ARTICLE INFO

Article history:
Received 7 June 2017
Accepted 6 January 2018
Available online 11 January 2018

Keywords:
Risk management
Loss Given Default (LGD)
Credit risk capital requirement
Loss function
Forecasts comparison

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 sample of almost 10,000 defaulted credit and leasing contracts provided by an international bank. Our empirical findings clearly show that models' rankings based on capital charge losses differ from those based on the LGD loss functions currently used by regulators, banks, and academics.

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.

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 problem 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.[1] Indeed, the academic literature on LGD definition, measurement, and modelling is surprisingly underdeveloped 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 in case of default, or equivalently by one minus the recovery rate. While this definition is clear, the measurement and the modelling of LGD raise numerous issues in practice. Regarding the measurement, both the BCBS and the European Banking Authority (see, for instance, European Banking Authority, 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 for the LGD models. This may explain why there is such a large heterogeneity in the modelling approaches used by AIRB banks and academics (see Section 2.3 for a survey). Commonly used approaches include among many others, simple look-up (contingency) tables,

☆ We would like to thank for their comments Gilbert Colletaz, Denisa Banulescu, Sylvain Benoit, Elena Dumitrescu, John Kose, Rachidi Kotchoni, Patrick Meidine, Sébastien Michelis, Christophe Pérignon, Samy Taouri-Mouloud, Olivier Scaillet, Sessi Tokpavi, and two anonymous referees. We also thank the participants of the 16th International Conference on Credit Risk Evaluation (CREDIT, Venice), the Bank Colloquium for Junior Researchers "The Future of Bank Regulation" (University of Limoges), the 5th conference Econometrics Applied to Macroeconomics (MSH Nord, Paris), the 16th conference Recent Developments of Econometrics Applied to Finance (University of Paris Ouest Nanterre) and the 2nd International Workshop in Quantitative Finance, Risk, and Decision Theory (University of Bordeaux). We thank the Chair ACPR/Risk Foundation: Regulation and Systemic Risk, and the ANR MultiRisk (ANR-16-CE26-0015-01) for supporting our research.

* Corresponding author.
E-mail address: christophe.hurlin@univ-orleans.fr (C. Hurlin).

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

https://doi.org/10.1016/j.ejor.2018.01.020
0377-2217/

parametric regression models (linear regression, survival analysis, fractional response regression, inflated beta regression, or Tobit models, for instance), and non-parametric techniques (regression tree, random forest, gradient boosting, artificial neural network, support vector regression, etc.). Within an extensive benchmarking study based on six real-life datasets provided by major international banks, Loterman, Brown, Martens, Mues and Baesens (2012) evaluate 24 regression techniques. They found that the average prediction performance of the models in terms of R-square ranges from 4% to 43%. Similarly, 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 lies with the lack of economic interpretability 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%.

On the contrary, within 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 in terms of *regulatory 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 models ranking determined by a LGD-based loss function (MSE, MAE, etc.) may differ from the ranking based on the corresponding capital charge loss function. In particular, we demonstrate the conditions under which both rankings are consistent and show that these conditions are likely to be violated in practice. This theoretical analysis confirms the relevance of our comparison framework for the LGD models and the usefulness of the regulatory capital estimation errors as comparison criteria.

We apply our methodology using a sample of almost 10,000 defaulted credit and leasing contracts provided by the bank of a worldwide leader automotive company. The originality of our dataset lies in the fact that the LGD observations incorporate all expenses (with an appropriate discount rate) arising during the workout process, to meet the Basel II requirements. Hartmann-Wendels, Miller and Töws (2014) and Miller and Töws (2017) argue that workout costs are rarely considered in empirical studies, even if they are essential for LGD modelling. Indeed, Gürtler and Hibbeln (2013) show that neglecting the workout costs leads to underestimate the LGD. Given this dataset, we compare six competing LGD models which are among the most often used in the empirical literature, namely (1) the fractional response regression, (2) the regression tree, (3) the random forest, (4) the gradient boosting, (5) the artificial neural network, and (6) the least squares support vector regression. We find that the models ranking based on the LGD loss function is generally different from the models ranking obtained with the capital charge loss function. 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 models ranking which is very different from the ranking obtained with symmetric loss functions.

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 underestimation 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 method 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. Furthermore, our paper complements the nascent literature on the LGD model validation. Loterman, Debruyne, Vanden Branden, Van Gestel and Mues (2014) propose a backtesting framework for LGD models using statistical hypothesis tests. Kalotay and Altman (2017) show that variation in the composition of the defaulted debt pool at the time of default generate time variation in the LGD distribution. They quantify the importance of accounting for such time variation in out-of-sample comparisons of alternative LGD models.

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. The discussion continues thereafter with a brief survey of LGD models and the method currently 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 as the six competing LGD models. In Section 5, we conduct our empirical analysis and display our main takeaways. In Section 6, we discuss various robustness checks. We summarize and conclude our paper in Section 7.

2. Capital charge for credit risk portfolios

In this section, we propose a brief overview of the importance of the LGD within the AIRB approach. Then, we present the main issues related to LGD measurement and we summarize the existing literature on LGD models. Finally, we discuss the method which is currently used to compare LGD models.

2.1. Capital requirement, individual risk contributions, and LGD

Let us consider a portfolio of n credits indexed by $i = 1, \ldots, n$. 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 measured as the difference between the 99.9% value-at-risk of the portfolio loss and the expected loss $\mathbb{E}(L)$. In order to derive this unexpected credit loss, the Basel Committee proposes a framework based on 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 (Basel Committee on Banking Supervision, 2005). Under these assumptions, 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 (cf. Appendix A). The regulatory capital is then equal to

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

The risk contribution RC_i for the ith credit 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 (see Appendix B for more details).

These equations highlight the key role of LGD within the Basel II framework. Since LGD enters the capital requirement formula in a linear way, LGD forecast errors have necessarily a strong impact on the regulatory capital. Consequently, the LGD measurement and the choice of an efficient forecasting model are crucial for bank's solvability.

2.2. LGD measurement

The LGD measurement raises numerous practical issues. Schuermann (2004) identifies three 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 issued by large firms. On the contrary, the workout LGD can be measured for any type of instrument. The workout LGD is based on an economic notion of loss including all the relevant costs tied to the collection process. The Basel II Accord identifies three types of costs: (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, 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. So, the scope of necessary data for proper LGD measurement is very broad.[2] However, the workout approach is clearly preferred by the regulators. For instance, in its guidelines on LGD estimation, the EBA states that "*the workout LGD is considered to be the main, superior methodology that should be used by institutions.*" (European Banking Authority, 2016, page 11).

Whatever the measure considered, the LGD distribution across defaulted bank loans or bonds generally exhibits two main stylized facts. Firstly, 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 (Gürtler & Hibbeln, 2013; Miller & Töws, 2017; Schmit, 2004) show that when workout costs are incorporated, the LGD is sometimes larger than 100%. Secondly, many empirical studies show a bimodal LGD distribution (see Miller & Töws, 2017, for instance). Most of the LGD values of defaulted contracts are either concentrated around high values (typically 70–80%) or low values (typically 20–30%).

2.3. LGD models

The general purpose of the LGD (internal) models consists in providing an estimate 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 the non-defaulted credits.

Because of the specific nature of the LGD distribution, a large variety of LGD models are currently used by academics and practitioners. Within the empirical literature, we can distinguish parametric and non-parametric approaches. The simplest parametric approach consists in using linear regression models based on debt characteristics and macroeconomic variables (Bastos, 2010; Gupton & Stein, 2002; Khieu et al., 2012). However, the linear model generally yields poor out-of-sample predictive performances.[3] Consequently, many other parametric models have been considered for LGD forecasting. Since the LGD is theoretically defined over [0, 1], these models are generally based on various transformations of LGD data which are done prior to the modelling stage or within the model itself. The most often used transformations are either based on beta (Credit Portfolio View of McKinsey, Gupton and Stein (2002)), exponential-gamma (Gouriéroux, Monfort, & Polimenis, 2006), or logistic-Gaussian distributions. In a similar way, the fractional response regression or log-log models, which keep the predicted values in the unit interval, have also been used for LGD modelling by

[2] This may explain why most empirical academic studies neglect workout costs because of data limitations (cf. Miller & Töws, 2017 for a discussion), even if Khieu, Mullineaux and Yi (2012) found evidence that market LGDs are biased estimates of the workout LGD.

[3] Notice that Zhang and Thomas (2012) found that linear regression models yield better performance than survival analysis models.

Bastos (2010); Dermine and de Carvalho C. (2006); Qi and Zhao (2011) or Bellotti and Crook (2012). Calabrese (2014a) uses an inflated beta regression model based on a mixture of a continuous beta distribution on [0, 1] and a discrete Bernoulli distribution, in order to model the probability mass at the boundaries 0 and 1. Similarly, Calabrese (2014b) proposes a parametric mixture distribution approach for downturn LGD. More recently, Kalotay and Altman (2017) suggest conditional mixtures of distributions allowing time variation in the LGD distribution. Using a different approach, Tanoue, Kawada and Yamashita (2017) propose a parametric multistep approach for the LGD of bank loans in Japan.

The main advantage of parametric models is their interpretability, but they usually have weak predictive performances compared to non-parametric methods that do not assume a specific distribution for LGD. Qi and Zhao (2011) compare fractional response regression to other parametric and non-parametric methods, such as regression trees and neural networks. They conclude that non-parametric methods perform better than parametric ones when overfitting is properly controlled for. A similar result is obtained by Bastos (2010) who recommends the use of non-parametric regression trees. If the predictive performance of non-parametric techniques is largely documented, it is difficult to identify the best models given the great heterogeneity of datasets and benchmarks considered. Using data from Moody's Ultimate Recovery Database (MURD), Bastos (2014) recommends a bagging algorithm. Hartmann-Wendels et al. (2014) use three datasets from German leasing companies to compare hybrid finite mixture models, model trees and regression trees. Their conclusions depend on the sample size and differ according to out-of-sample or in-sample performance criteria. Yao, Crook and Andreeva (2015) compare the predictions of support vector regression techniques with thirteen other algorithms using data from MURD. They conclude that all support vector regression models substantially outperform other statistical models in terms of both model fit and out-of-sample predictive accuracy. The previously mentioned benchmarking study of Loterman et al. (2012) compares 24 parametric and non-parametric techniques, including ordinary least squares regression, beta regression, robust regression, ridge regression, regression splines, neural networks, support vector regressions, and regression trees. They conclude that non-linear techniques, and in particular support vector regressions and neural networks, perform significantly better than more traditional linear techniques.[4]

In addition to single-stage models, some studies implement two-stage models to forecast LGD. These methods have the advantage to model the extreme values concentrating on the boundaries at 0 and 1. Bellotti and Crook (2012) propose a two-stage model based on a decision tree algorithm (with two logistic regression sub-models) which is applied to split the whole sample into three groups according to the values of LGD (0, 1, or between 0 and 1). Then the values in]0, 1[are fitted by an OLS regression model. Yao, Crook and Andreeva (2017) improve this two-stage approach by considering a least squares support vector classifier rather than logistic regressions. They show that this two-stage model outperforms the single-stage support vector regression model in terms of out-of-sample R-square. Considering two datasets of home equity and corporate loans, Tobback, Martens, Van Gestel and Baesens (2014) also find that a two-stage model (which combines linear regression and support vector regression) outperforms the other techniques when forecasting out-of-time. But, they observe that non-parametric regression tree has better performance when forecasting out-of-sample. Miller and Töws (2017) propose an original multi-step estimation approach based on an economic separation of the LGD determined by the workout process. Nazemi, Fatemi Pour, Heidenreich and Fabozzi (2017) implement a fuzzy fusion model which uses a function to combine the results of several base models. They show that the fuzzy fusion model has higher predictive accuracy compared to support vector regression models.

This brief overview of the literature shows that there is no benchmark model emerging from the "zoo" of LGD models. Consequently, for each new real-life database, academics and practitioners have to consider several LGD models and compare them according to appropriate comparison criteria.

2.4. LGD models comparison

In this section, we briefly present the method currently used both by academics and banks to compare the predictive performances of LGD models. Consider a set of $\mathcal{M}$ LGD models indexed by $m = 1, \ldots, \mathcal{M}$ 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.[5] 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 value observed for the ith credit of the test set, for $i = 1, \ldots, n_v$ and by $\widehat{\text{LGD}}_{i,m}$ 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

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

where $L(.,.)$ is an integrable loss function, with $L : \Omega^2 \to \mathbb{R}^+$.[6] Since the LGD is a continuous variable defined over a subspace Ω of $\mathbb{R}^+$ (typically [0, 1] or [0, δ] with $\delta > 1$), the loss functions generally considered in academic literature 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 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

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

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

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

[4] Other studies aim to model the LGD distribution using non-parametric estimators. Renault and Scaillet (2004) or Hagmann, Renault and Scaillet (2005) propose different kernel estimators of the LGD density for defaulted loans. Calabrese and Zenga (2010) consider a mixture of beta kernels estimator to model the LGD density of a large dataset of defaulted Italian loans. These approaches have the common advantage to reveal a number of bumps which can be larger than those obtained with parametric distributions. We can also mention Krüger and Rösch (2017) who consider quantile regressions for modelling downturn LGD.

[5] For machine learning methods (regression trees, neural networks, etc.), the training set is sometimes further split into training and validation subsets. The validation set is used to select the optimal tuning parameters that provide the best in-sample predictive performance.

[6] If we denote by $e = x - \widehat{x}$ the error, the loss function is assumed to satisfy the following properties: (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 $e_1 < e_2 < 0$.

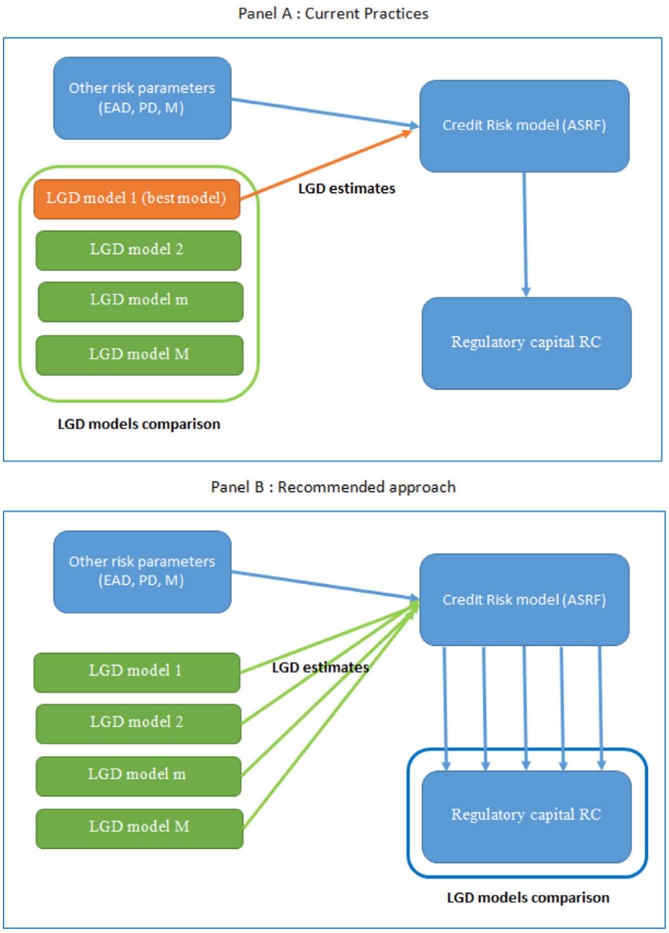

Fig. 1. Comparison of LGD models in the regulatory framework.

Other standard comparison criteria (deduced from these loss functions) can also be used for models comparison (see Yao et al., 2017; Nazemi et al., 2017, for instance), such as R-square, RMSE, etc. Whatever the criterion used, 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 $m = 1, \ldots, \mathcal{M}$. 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

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

This general approach has two main shortcomings. The first one is the lack of interpretability of 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 second 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 Fig. 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.).

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 for LGD models 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 Fig. 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 $L(.,.)$ is an integrable *capital charge* loss function with $L: \mathbb{R}^{+2} \rightarrow \mathbb{R}^+$, and

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

The variable RC_i denotes the risk contribution of the ith credit, defined by the regulatory formula (Eq. (3)). This risk contribution depends on the risk parameters, namely EAD_i, LGD_i , and M_i. 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, RMSE, R^2, or any usual 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

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

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

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).

The expected loss $\mathcal{L}_{CC,m}$ 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. However, the *ranking* of the LGD models based on the capital charge expected loss, does not depend on the choice of the PD value. Indeed, since $\delta(\text{PD})$ 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 $\mathcal{L}_{CC,m'}$.

Eq. (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 $\delta(\text{PD})$ and hence, the regulatory capital since RC_i is an increasing function of $\delta(\text{PD})$. The profile of the capital charge coefficient $\delta(\text{PD})$ depends on the type of exposure (cf. Appendix B). 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 $\mathbb{E}(L)$. 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.

3.2. Ranking consistency

The LGD models comparison can be based either on traditional LGD-based loss functions $\mathcal{L}_m$ or capital charge-based loss functions $\mathcal{L}_{CC,m}$. Suppose that both approaches lead to the same models ranking (e.g, in the case of two models m and m', $\mathcal{L}_m < \mathcal{L}_{m'}$ and $\mathcal{L}_{CC,m} < \mathcal{L}_{CC,m'}$). Then, one should favor the simplest approach that only focuses on LGD errors. In this case, it is useless to collect additional data for other risk parameters (EAD, PD, maturity, etc.) and to compute the capital charges for each credit in order to compare LGD models. However, nothing guarantees ex-ante that both approaches will necessarily lead to consistent models' rankings.

The goal of this section is twofold. First, we determine the conditions under which the models ranking induced by a LGD-based loss function and the models ranking obtained with a capital charge-based loss function are consistent. Second, we show that these conditions are very particular and are likely to be violated in practice. Hence, this theoretical analysis illustrates the relevance of our comparison framework for LGD models, which is based on regulatory capital estimation errors.

Consider the following assumptions on the LGD loss functions.

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

Assumption A2. *The function $g(.)$ is multiplicative: $\forall k \in R$, $g(k(x - \widehat{x})) = g(k)g(x - \widehat{x})$.*

Notice that Assumptions A1 and A2 are satisfied by the usual loss functions considered in the LGD literature.[7]

Consider a set of $\mathcal{M}$ LGD models, indexed by $m = 1, \ldots, \mathcal{M}$. 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

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

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

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

with $\eta_{i,m} = \text{RC}_i - \widehat{\text{RC}}_{i,m}$. By definition of the regulatory capital charge, we have[8]

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

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

$$\begin{aligned} &cov(g(\text{EAD}_i), g(\varepsilon_{i,m})) - cov(g(\text{EAD}_i), g(\varepsilon_{i,m+1})) \\ &\quad < \mathbb{E}(g(\text{EAD}_i))(\mathcal{L}_{m+1} - \mathcal{L}_m) \end{aligned} \tag{12}$$

The proof of Proposition 1 is reported in Appendix C. Since $\mathcal{L}_m < \mathcal{L}_{m+1}$, the consistency condition of Proposition 1 is satisfied as soon as $cov\left(g(\text{EAD}_i), g\left(\varepsilon_{i,m}\right)\right) < cov\left(g(\text{EAD}_i), g\left(\varepsilon_{i,m+1}\right)\right)$. 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 models' rankings. Proposition 1 has a direct

[7] 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 $L(kx, k\widehat{x}) = L(g(k(x - \widehat{x}))) = |k||x - \widehat{x}| = g(k)g(x - \widehat{x})$.

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

interpretation in the special case where the exposures are independent from the estimation errors of the LGD models.

Corollary 2. *As soon as the* EAD_i *and the LGD estimation errors* $\varepsilon_{i,m}$ *are independent, the models' 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, this independence assumption is likely to be violated in practice. First, 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 regression, or random forest, the forecast errors may be correlated with the explanatory variables.

4. Comparison framework

We now propose an empirical application of our comparison approach for LGD models. In this section, we describe our dataset, the experimental set-up, and the six competing LGD models.

4.1. Data description

Our dataset consists in a portfolio of retail loans (credit and leasing contracts) provided by an international bank specialized in financing, insurance, and related activities for a worldwide leader automotive company. The initial sample includes 23,933 loans that defaulted between January 2011 and December 2016. For the more recent defaults, the recovery processes are not necessarily completed and we don't observe the bank's final loss. As a consequence, we exclude these contracts and limit our analysis to the 9,738 closed recovery processes for which we observe the final workout LGD. This approach has also been used by Gürtler and Hibbeln (2013) and Krüger and Rösch (2017) who recommend restricting the observation period of recovery cash flows to avoid the under-representation of long workout processes, which might result in an underestimation of LGD and regulatory capital.

The final 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, original maturity, etc.) and the borrower (professional, individual, etc.), as well as the workout 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. Hence, we have all the information to compute the regulatory capital charge for each credit. Finally, we complete the database with three macroeconomic variables, namely (1) the quarterly Brazilian GDP growth rate, (2) the monthly unemployment rate and (3) the

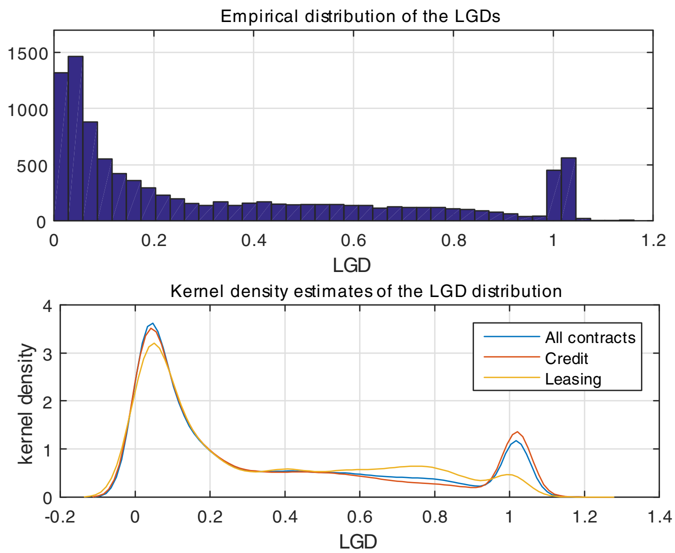

Fig. 2. Empirical distribution of the LGDs.

monthly average of the daily interbank rates.[9] For each contract, the macroeconomic variables are considered at the date of default. Their introduction in LGD models aims to capture the influence of the business cycles on the recovery process, as suggested by Bellotti and Crook (2012) and Tobback et al. (2014). The description of the dataset variables is reported in Appendix E.

Table 1 displays some descriptive statistics (mean, q25, and q75) about the LGD, PD, and EAD by year, exposure and customer type. The number of defaulted contracts per year ranges between 1,573 and 2,946. The mean of losses is equal to 33.12%, a similar value to that reported by Miller and Töws (2017) for a German leasing company, and tends to decrease between 2011 and 2014. The average PD is equal to 9.53%, but this figure hides a large heterogeneity since the PD values range from less than 1% to 71%, whereas 3/4 of the PD values are below 11.08%. Similarly, the EAD ranges from less than 1 BRL to 123,550 BRL, with an average exposure equal to 20,830 BRL. The credit and leasing contracts exhibit the same level of exposure, but the average PD is higher for leasing than for credit (10.24% against 9.25%). As often reported in the literature, the LGD for leasing contracts (32.01%) is slightly smaller than for credits (33.57%). We observe that the average PD is higher for the professional clients than for individuals, but their average LGD is smaller due to their highest collateral. Finally, 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.

The empirical distribution of the 9,738 workout LGDs is displayed on the top panel of Fig. 2. 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. Second, we also confirm that the kernel density estimate of the LGD distribution is bimodal (bottom panel of Fig. 2). Finally, the LGD distributions for the credit and leasing contracts are relatively close, except for the right part of the distribution. The outcome of a loss event is less severe for leasing than for credit. This difference illustrates the role of the collateral in the recovery

[9] The data have been collected from OECD.Stat databases: Monthly Monetary and Financial Statistics (MEI), Quarterly National Accounts (QNA) and Labour Market Statistics (LMS).

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

	Nb of obs	LGD (%)			PD (%)			EAD (thous. BRL)		
Panel A. All loans										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
	9,738	5.03	33.12	57.22	0.78	9.53	11.08	10.63	20.83	28.11
Panel B. By year										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
2011	1,573	6.82	40.95	77.29	0.80	10.56	16.93	11.12	20.80	27.78
2012	2,430	8.09	39.14	67.31	0.71	8.36	8.98	11.67	22.20	29.57
2013	2,946	7.16	36.43	61.98	0.72	9.07	11.00	11.02	21.43	28.50
2014	2,789	2.85	19.97	22.92	1.05	10.46	17.42	9.14	19.02	26.18
Panel C. By exposure										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
Credit	6,946	5.03	33.57	56.83	0.77	9.25	10.14	10.36	20.94	28.40
Leasing	2,792	5.02	32.01	58.15	0.84	10.24	14.79	11.28	20.57	27.39
Panel D. By customer type										
	–	q25	mean	q75	q25	mean	q75	q25	mean	q75
Individuals	6,521	5.15	33.80	59.34	0.70	8.75	9.89	11.08	20.39	27.61
Professionals	3,217	4.64	31.75	52.70	1.11	11.12	14.79	9.77	21.73	29.71

processes (in the case of leasing, the vehicle belongs to the bank and plays the same role as a collateral).

4.2. Competing LGD models

For our comparison, we consider six competing LGD models which are commonly used in academic literature, namely (1) the fractional response regression (FRR) model, (2) the regression tree (TREE), (3) the random forest (RF), (4) the gradient boosting (GB), (5) the artificial neural network (ANN), and (6) the least squares support vector regression (LS-SVR).[10]

The FRR model allows to estimate the conditional mean of a continuous variable defined over [0, 1]. It is often considered as a benchmark parametric model for LGD (see Bastos, 2010; Qi & Zhao, 2011 etc.). The TREE model consists in recursively partitioning the covariates space according to a prediction error and then, to fit a simple mean prediction within each partition. Here, we consider the CART algorithm which has been applied to LGD estimation by Bastos (2010, 2014); Matuszyk, Mues and Thomas (2010); Qi and Zhao (2011), and Loterman et al. (2012), among many others. The RF is a bootstrap aggregation method of regression trees, trained on different parts of the same training set, with the goal of reducing overfitting. This model has been used for LGD modelling by Bastos (2014) and Miller and Töws (2017), among others. The ANNs are a class of flexible non-linear models. 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. Here, we consider a multilayer perceptron similar to that used by Qi and Zhao (2011) or Loterman et al. (2012) for the LGD forecasts. Finally, we consider the LS-SVR model. Compared to other support regression techniques, the LS-SVR has a low computational cost as it is equivalent to solving a linear system of equations. Loterman et al. (2012); Yao et al. (2015, 2017) and Nazemi et al. (2017) illustrate the good predictive performance of LS-SVR for LGD modelling. For more details and references about these models, see Appendix F.

4.3. Experimental set-up

The six competing models are estimated on a training set of 7,791 loans (80% of the sample) and the out-of-sample LGD forecasts are evaluated on a test set of 1,947 loans. For each model, we consider the same set of explanatory variables including the exposure at default, the original maturity, the time to default, the relative duration (defined as the ratio between time to default and maturity), the interest or renting rate, the type of exposure (credit versus leasing), the customer type (individual or professional), the state of the car (new or second-hand), and the brand of the car.[11] Appendix E displays the description of these independent variables, as well as descriptive statistics. For each model and each contract within the test set, we compute the LGD forecast and the regulatory capital charge (based on the LGD forecast or the true LGD value), by using the other Basel risk parameters. The regulatory capital charges are computed with a PD of 40.45%, which corresponds to the maximal charge for the retail exposures.

The hyperparameters of the machine learning algorithms are tuned using five-fold cross validation on the training set. They were all selected based on the MSE criterion. For the TREE model, the procedure leads to select the optimal depth of the tree. For the GB, the cross validation procedure determines the optimal number of iterative training cycles (candidates 10, 50, 100, 250, 500, 1000 are considered). The same approach is applied for the RF for identifying the optimal number of trees in the forest. For the ANN, the five-fold cross validation procedure is used to select the number of hidden neurons (a value from 1 to 20 is considered). A logistic function is used as the activation function in hidden layer neurons. Finally, in order to implement the LS-SVR, we consider a radial basis function kernel. The radial basis function kernel parameter σ and the regularization parameter C, are tuned using five-fold cross validation on the training dataset. A grid search procedure firstly evaluates a large space of possible hyperparameter combinations to determine suitable starting candidates. The search limits are set to $[\exp(-10), \exp(10)]$. Then, given these starting values, the hyperparameters σ and C are optimized with a simplex routine so as to find the combination that minimizes the MSE.

5. Empirical results

5.1. LGD and RC estimation errors

Table 2 displays some figures about LGD and regulatory capital forecast errors, respectively defined by $\text{LGD}_i - \widehat{\text{LGD}}_{i,m}$ and

[10] We thank an anonymous referee for the suggestion of the LS-SVR model.

[11] An extended set of information with macroeconomic variables will be considered in Section 6.

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

	FRR	ANN	TREE	LS-SVR	RF	GB
	LGD errors					
Mean	0.011	0.011	0.010	0.011	0.009	0.010
Median	−0.136	−0.122	−0.142	−0.127	−0.114	−0.138
Variance	0.117	0.116	0.118	0.115	0.117	0.116
Skewness	0.824	0.804	0.817	0.828	0.791	0.830
Excess kurtosis	−0.618	−0.525	−0.605	−0.551	−0.473	−0.626
	Regulatory capital errors					
Mean	60	44	66	56	38	66
Median	−257	−231	−256	−233	−232	−257
Variance	3,813,596	3,799,691	3,730,561	3,698,999	3,737,503	3,717,518
Skewness	0.55	0.41	0.80	0.58	0.61	0.79
Excess kurtosis	2.52	2.83	2.01	2.54	2.73	1.91

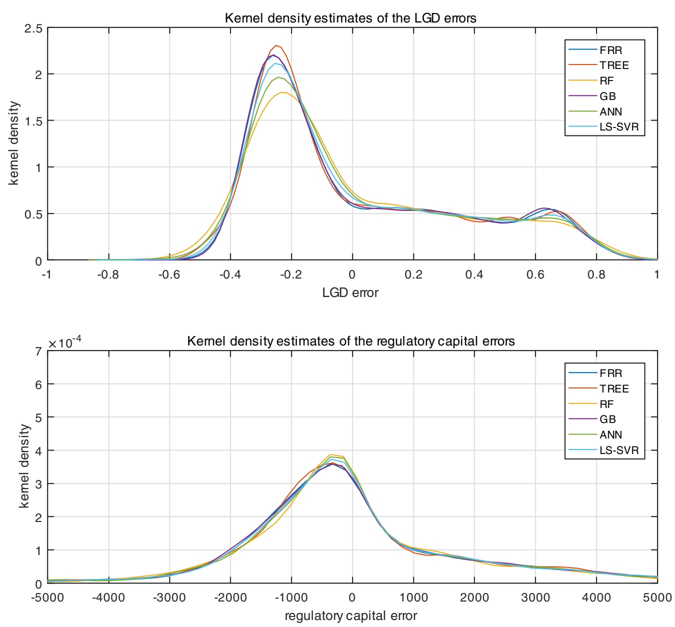

Fig. 3. Kernel density estimate of the estimation error.

$RC_i - \widehat{RC}_{i,m}$. Notice that, given this notation, a positive error implies an underestimation of the true value. 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. We also observe that the excess kurtosis for the regulatory capital are positive, indicating fat tails for the errors distribution. When one considers the LGD errors, the LS-SVR, ANN and the GB models have the smallest variance. However, it is no longer the case for the ANN when one considers the RC forecast errors. This result clearly illustrates the usefulness of our comparison approach.

The kernel density estimates of the forecast errors distributions displayed in Fig. 3 confirm the positive skewness of the errors' distributions. This figure shows that one can frequently observe capital requirement underestimates larger than 4,000 BRL, whereas similar overestimates are much rarer. Such a feature is clearly problematic within a regulatory perspective, and justifies the use of asymmetric loss functions for comparing LGD models.

Fig. 4 displays the scatter plot of the LGD forecast errors (*x*-axis) and the RC forecast errors (*y*-axis), obtained with the GB model. 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 61,271 BRL and 2,034 BRL, respectively. For the same level of LGD forecast error (64.3%), the GB slightly underestimates the capital requirement (278 BRL) in the case of the credit B, whereas the underestimation reaches 8,367 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 loss functions expressed in terms of capital charge. Furthermore, the scatter plot confirms the asymmetric pattern of the errors distribution associated to the GB model. This model leads to relatively small 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 severe underestimates than the GB should be preferred from a regulatory perspective. For this reason, we recommend the use of asymmetric loss functions to compare the LGD models. These features (heterogeneity and asymmetry) are not specific to the GB model, even if the skewness of the error distribution is more pronounced for this model compared to the other ones. The scatter plots of the LGD and RC errors are quite similar for the six competing models (cf. Appendix G). This similarity comes from two sources. First, it is due to the fact that we use the same set of covariates for all the models. Second, this similarity is also related to the definition of the regulatory capital that leads to increase the RC errors dispersion and induces a similar appearance for the different scatter plots (i.e. for the different LGD models).

5.2. LGD models' rankings

Table 3 displays the models' rankings issued from two usual loss functions, namely the MSE and MAE, associated to the LGD (column 1) and regulatory capital (column 2) forecast errors. We also report the rankings based on the asymmetric expected losses (columns 3 and 4) that only penalize the LGD forecast errors which lead to underestimating the regulatory capital. The values of the losses (MSE, MAE) are displayed in Appendix H, along with the corresponding R^2 and RMSE.

The models' rankings that we obtain with the MSE or MAE criteria computed with LGD estimation errors, are similar to those generally obtained in the literature. As in Loterman et al. (2012); Yao et al. (2015, 2017), and Nazemi et al. (2017), we observe that the LS-SVR model outperforms the five competing LGD models. As in Bastos (2010); Hartmann-Wendels et al. (2014); Loterman et al. (2012); Qi and Zhao (2011) or Miller and Töws (2017), we observe that non-parametric approaches such as LS-SVR, ANN, and RF generally yield better predictive performances than the FRR

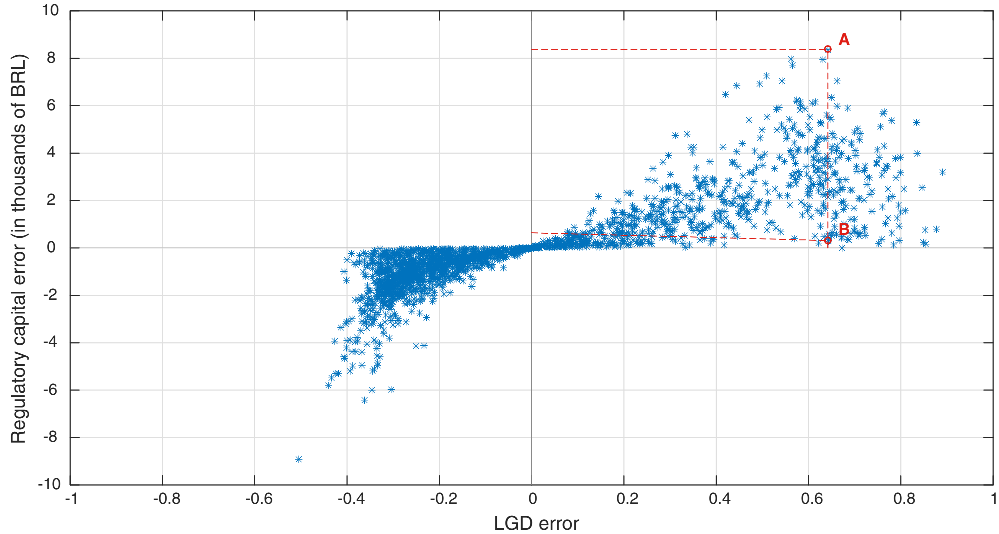

Fig. 4. Scatter plot of LGD versus regulatory capital forecast errors for the GB model.

Table 3
Models' 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.	LS-SVR	LS-SVR	RF	RF
2.	GB	GB	LS-SVR	ANN
3.	ANN	TREE	FRR	LS-SVR
4.	FRR	RF	ANN	FRR
5.	RF	ANN	TREE	TREE
6.	TREE	FRR	GB	GB
Mean absolute error				
1.	LS-SVR	RF	RF	RF
2.	RF	LS-SVR	LS-SVR	LS-SVR
3.	ANN	ANN	FRR	ANN
4.	GB	TREE	ANN	FRR
5.	FRR	GB	TREE	TREE
6.	TREE	FRR	GB	GB

Note: The two columns LGD loss and CC loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors. The columns Asym. LGD loss and Asym. CC loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

Table 4
Spearman's and Kendall's rank correlation coefficients.

	Spearman rank correlation		Kendall rank correlation	
	MSE			
	ρ	p-value	ρ	p-value
LGD vs. CC	0.4857	0.1778	0.3333	0.2347
LGD vs. Asym. LGD	−0.0286	0.5403	−0.0667	0.6403
LGD vs. Asym. CC	−0.0857	0.5986	−0.0667	0.6403
	MAE			
	ρ	p-value	ρ	p-value
LGD vs. CC	0.7714	0.0514	0.6000	0.0681
LGD vs. Asym. LGD	0.6571	0.0875	0.4667	0.1361
LGD vs. Asym. CC	0.7714	0.0514	0.6000	0.0681

Note: The ρ coefficients denote the Spearman's or Kendall's rank correlation coefficients. The p-values are computed under the null hypothesis $\rho = 0$ and are based on the exact permutation distributions for small sample sizes. LGD and CC respectively denote the rankings based on LGD or regulatory capital errors. Asym. LGD and Asym. CC respectively denote the rankings obtained with asymmetric loss functions based on LGD or regulatory capital errors.

parametric model. Furthermore, we observe similar values for the RMSE of the LS-SVR model (see Appendix H) as those reported in Yao et al. (2015) or Loterman et al. (2012).

However, considering the loss function based on RC estimation errors leads to different conclusions. Regarding the MSE criterion, the LS-SVR is still ranked as the best model, whatever the errors considered (LGD or RC). The GB is also consistently ranked as the second best model. But, the rest of the LGD models ranking is not consistent. For instance, ANN is identified as the third-best model with the LGD loss, while it holds the penultimate rank with the capital charge loss. Conversely, the TREE model is ranked third with the capital charge loss while it is ranked at the last 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 1 is not valid, at least in our sample, for some couples of models. Ranking the LGD models according to their LGD forecast errors or their RC forecast errors is not equivalent. The conclusions are similar for the MAE criteria. In this case, the LS-SVR is the best model when one considers LGD forecast errors, but the RF should be preferred when one considers RC forecast errors.

Furthermore, our results highlight the usefulness of asymmetric loss functions. These functions penalize more the models with the largest positive errors (underestimates), as the GB, for instance. Indeed, the GB is ranked as the worst model when the MSE is computed with asymmetric LGD or RC forecast errors. It also remains the worst model when one considers asymmetric MAE, due to the large skewness of its forecast errors. On the contrary, the RF exhibits the lowest asymmetric MSE and MAE whatever the type of errors considered. These findings clearly illustrate the fact that ranking the LGD models according to their estimation errors (whatever their signs) or to their underestimates, is not equivalent.

These conclusions are confirmed by rank correlation tests. Table 4 displays the Spearman's and Kendall's rank correlation coefficients ρ, with the p-values associated to the null hypothesis $\rho = 0$. These p-values are computed with the exact permutation distributions for small sample sizes. Our goal is to test if the models' rankings differ significantly. We compare the rankings obtained with (1) MSE (MAE) based on LGD errors, (2) MSE (MAE) based on RC errors, and (3) asymmetric MSE (MAE) based on LGD or RC errors. The conclusions are clear-cut: for a 5% significance level, the rank correlation coefficients are never statistically different from

Table 5
Models' rankings based on Basel PDs.

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

Note: The two columns LGD loss and CC loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors (computed with Basel PD values). The columns Asym. LGD loss and Asym. CC loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

Table 6
Models' rankings based on LGD and capital charge expected loss functions: LGD models with macroeconomic variables and common PD.

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

Note: The two columns LGD loss and CC loss correspond to the models' rankings obtained with loss functions (MSE or MAE) respectively defined in terms of LGD forecast errors and regulatory capital forecast errors. The columns Asym. LGD loss and Asym. CC loss display the rankings obtained with asymmetric loss functions either defined in terms of LGD or regulatory capital forecast errors.

0. These tests confirm that using a regulatory capital-based criterion (MSE or MAE) do not provide the same ranking as that obtained with a similar criterion only based on LGD estimation errors. Furthermore, the tests show that the rankings are sensitive to the choice of a symmetric or asymmetric criterion.

Beyond the rank correlations, we also investigate if the choice of the loss function may affect the pairwise models comparison. In Appendix I, we display the paired t-tests for comparisons of MSE and MAE, based on LGD estimation errors or regulatory capital estimation errors. The logic here is similar to Yao et al. (2015) and Yao et al. (2017) or Nazemi et al. (2017). The main takeaway of these pairwise tests is the following. Considering out-of-sample criteria (MSE or MAE) based on regulatory capital estimation errors sometimes change the conclusions of the pairwise comparison of LGD models performance. For instance, if we consider the MSE criterion based on LGD errors, the LS-SVR model outperforms all other models (except the GB), as the differences between the corresponding MSEs are always positive for a 5% significance level. However, when considering the MSE based on regulatory capital errors, the MSE difference between the LS-SVR and the TREE model is not significant, meaning that both models lead to similar regulatory capital estimation errors. Similarly, the LS-SVR does not make significant improvements compared to the RF when one considers regulatory capital errors.

6. Robustness checks

Our empirical results are robust to a variety of robustness checks. Firstly, instead of considering a common PD for the computation of the capital charges, we use the individual PD calculated by the internal bank's risk model for each credit one year before the default occurs. The corresponding LGD models' rankings are reported in Table 5. The rankings based on the MSE are similar to those obtained with a common PD (cf. Table 3). For the symmetric capital charge MSE, the only change concerns the RF and the TREE models. For the asymmetric MSE, the ranking changes for the LS-SVR, the ANN, the GB and TREE models. But, we still observe ranking inversions compared to the ranking based on the LGD loss functions.

Secondly, 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 models' 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 1, since we only consider non-linear LGD models in our application.

Finally, we extend the set of explanatory variables by considering three macroeconomic variables in order to capture the influence of the business cycles on the recovery process, as suggested by Bellotti and Crook (2012); Schuermann (2004), and Tobback et al. (2014). These variables are the Brazilian GDP growth, the unemployment and interbank rates. Table 6 displays the corresponding LGD models' rankings. With the MSE criterion, the RF outperforms all competing models whatever the loss function considered. It is also the case for the MAE criterion. As in the previous cases, we observe a ranking inconsistency for other models, meaning that the condition of Proposition 1 is not valid for these couples of models. The values of the losses (MSE, MAE) are displayed in Appendix H, along with the corresponding R^2 and RMSE. Our results are similar to those obtained in the literature. For instance, Hartmann-Wendels et al. (2014) who examine three leasing datasets, report a MAE that ranges from 0.2710 to 0.3370 for the TREE model (0.2768 in our case), while their RMSE takes values between 0.3462 and 0.3958 depending on the dataset (0.3343 in our case). We also get similar results for the RF as those reported in Miller and Töws (2017). Within their sample, the authors obtain a MAE of 0.3272 (0.2705 in our case) and a MSE of 0.1722 (0.1092 in our case). We obtain relatively low R^2 values (around 10%) when we consider LGD errors, but the R^2 reaches higher values (around 35%) when considering RC errors.

Beyond the rankings analysis, we report in Appendix J the marginal effects of the debt characteristics and macroeconomic variables on the LGD estimates obtained within the FRR model. Our qualitative results are similar to those obtained in the literature. As in Bastos (2010), the credit interest rate (fixed at the beginning of the credit contract) positively affects the LGD. This positive effect reflects the fact that the risky clients, who have the lowest collateral, have generally also the highest interest rates and in fine the lowest recovery rates. The original maturity has also a positive and significant effect on LGD. Indeed, for a given retail credit or leasing contract, longer maturities are generally negotiated by riskiest clients with the lowest collateral and revenues, and as a consequence the highest LGD. Contrary to Schuermann (2004), we observe a significant and positive impact of the EAD. The brand and the characteristics (new or second hand) of the car, the customer type (professional or individual), and the credit type (leasing versus standard credit) do not significantly impact the LGD. Finally, the time to default has a negative impact meaning that bank generally suffers limited losses for contracts that default close to their maturity. This result is similar to that obtained by Bellotti and Crook (2012) who found a negative and significant impact for the date of default.

Concerning macroeconomic variables, we observe that the interbank interest rate (measured at the date of default) has a negative and significant coefficient. Tobback et al. (2014) also find a negative impact of the Federal Funds rate and explain it by the fact that a higher Federal Funds rate decreases the ability of the borrowers to pay off already defaulted loans. We observe that the unemployment rate has a significant positive impact on LGDs, as in Tobback et al. (2014). Finally, we observe that the GDP growth rate coefficient is negative, but non-significant at the 5% level. During a period of economic boom, banks are willing to issue loans to more risky borrowers against a high return. Therefore, the expansion and peak phases of the business cycle are accompanied by an accumulation of risks which result in greater losses once the growth starts slowing down. However, as pointed out by Tobback et al. (2014), one should be very cautious about interpreting the GDP growth effect, as the peak phase of the business cycle generally corresponds to a low growth percentage of GDP.

7. Conclusion

LGD is one of the key modelling 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 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 functions 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 generally different from the models ranking obtained with the capital charge symmetric (or asymmetric) loss.

A natural extension of our work includes the identification of a Model Confidence Set (Hansen, Lunde, & Nason, 2011) that contains the "best" LGD models for a given level of confidence and a given criterion. This method of "models clustering", based on pairwise t-tests and an iterative algorithm, have been recently used to compare conditional risk measures (Hurlin, Laurent, Quaedvlieg, & Smeekes, 2017) and could be adapted to compare LGD models.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ejor.2018.01.020.

References

Basel Committee on Banking Supervision. (2005). An explanatory note on the Basel II IRB risk weight functions. Consultation Paper, Basel Committee on Banking Supervision.

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

Bastos, J. A. (2014). Ensemble predictions of recovery rates. *Journal of Financial Services Research, 46*(2), 177–193.

Bellotti, T., & Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting, 28*(1), 171–182.

Calabrese, R. (2014a). Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry, 30*(2), 99–114.

Calabrese, R. (2014b). Downturn loss given default: Mixture distribution estimation. *European Journal of Operational Research, 237*(1), 271–277.

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

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance, 66*(4), 1047–1108.

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

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*(3), 253–263.

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

Gouriéroux, C., Monfort, A., & Polimenis, V. (2006). Affine models for credit risk analysis. *Journal of Financial Econometrics, 4*(3), 494–530.

Gupton, G. M., & Stein, R. M. (2002). LOSSCALC: Moody's model for predicting loss given default. New York, Moody's Investors Service.

Gürtler, M., & Hibbeln, M. (2013). Improvements in loss given default forecasts for bank loans. *Journal of Banking and Finance, 37*(7), 2354–2366.

Hagmann, M., Renault, O., & Scaillet, O. (2005). Estimation of recovery rate densities: Nonparametric and semi-parametric approaches versus industry practice. In Altman, Resti, & Sironi (Eds.), *Recovery risk: The next challenge in credit risk management* (pp. 323–346).

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.

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.

Hurlin, C., Laurent, S., Quaedvlieg, R., & Smeekes, S. (2017). Risk measure inference. *Journal of Business and Economic Statistics, 35*(4), 499–512.

Kalotay, E. A., & Altman, E. I. (2017). Intertemporal forecasts of defaulted bond recoveries and portfolio losses. *Review of Finance, 21*(1), 433–463.

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.

Krüger, S., & Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking and Finance, 79*, 42–56.

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.

Loterman, G., Debruyne, M., Vanden Branden, K., Van Gestel, T., & Mues, C. (2014). A proposed framework for backtesting loss given default models. *Journal of Risk Model Validation, 8*(1), 69–90.

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.

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

Miller, P., & Töws, E. (2017). Loss given default adjusted workout processes for leases. *Journal of Banking and Finance*. doi:10.1016/j.jbankfin.2017.01.020.

Nazemi, A., Fatemi Pour, F., Heidenreich, K., & Fabozzi, F. J. (2017). Fuzzy decision fusion approach for loss-given-default modeling. *European Journal of Operational Research, 262*(2), 780–791.

Qi, M., & Zhao, X. (2011). Comparison of modeling methods for loss given default. *Journal of Banking and Finance, 35*(11), 2842–2855.

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.

Schmit, M. (2004). Credit risk in the leasing industry. *Journal of Banking and Finance, 28*(4), 811–833.

Schuermann, T. (2004). What do we know about loss given default?. In Shimko (Ed.), *Credit risk models and management*.
Tanoue, Y., Kawada, A., & Yamashita, S. (2017). Forecasting loss given default of bank loans with multi-stage model. *International Journal of Forecasting, 33*(2), 513–522.
Tobback, E., Martens, D., Van Gestel, T., & Baesens, B. (2014). Forecasting loss given default models: Impact of account characteristics and the macroeconomic state. *Journal of the Operational Research Society, 65*(3), 376–392.
Vasicek, O. A. (2002). The distribution of loan portfolio value. *Risk, 15*(12), 160–162.
Yao, X., Crook, J., & Andreeva, G. (2015). Support vector regression for loss given default modelling. *European Journal of Operational Research, 240*(2), 528–538.
Yao, X., Crook, J., & Andreeva, G. (2017). Enhancing two-stage modelling methodology for loss given default with support vector machines. *European Journal of Operational Research, 263*(2), 679–689.
Zhang, J., & Thomas, L. C. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting, 28*(1), 204–215.