

Vrins, Frédéric (Ed.)

Book — Published Version

Advances in credit risk modeling and management

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Vrins, Frédéric (Ed.) (2020) : Advances in credit risk modeling and management, ISBN 978-3-03928-761-1, MDPI, Basel, <https://doi.org/10.3390/books978-3-03928-761-1>

This Version is available at:

<https://hdl.handle.net/10419/230539>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



risks

Advances in Credit Risk Modeling and Management

Edited by
Frédéric Vrins

Printed Edition of the Special Issue Published in *Risks*

Advances in Credit Risk Modeling and Management

Advances in Credit Risk Modeling and Management

Special Issue Editor

Frédéric Vrans

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Special Issue Editor

Frédéric Vrins

Université catholique de Louvain

Belgium

Editorial Office

MDPI

St. Alban-Anlage 66

4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Risks* (ISSN 2227-9091) (available at: [https://www.mdpi.com/journal/risks/special issues/Credit Risk Modeling](https://www.mdpi.com/journal/risks/special%20issues/Credit%20Risk%20Modeling)).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , Article Number, Page Range.

ISBN 978-3-03928-760-4 (Pbk)

ISBN 978-3-03928-761-1 (PDF)

© 2020 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Special Issue Editor	vii
Preface to “Advances in Credit Risk Modeling and Management”	ix
Hui Ye and Anthony Bellotti	
Modelling Recovery Rates for Non-Performing Loans Reprinted from: <i>Risks</i> 2019 , 7, 19, doi:10.3390/risks7010019	1
Pascal François	
The Determinants of Market-Implied Recovery Rates Reprinted from: <i>Risks</i> 2019 , 7, 57, doi:10.3390/risks7020057	19
Dan Cheng and Pasquale Cirillo	
An Urn-Based Nonparametric Modeling of the Dependence between PD and LGD with an Application to Mortgages Reprinted from: <i>Risks</i> 2019 , 7, 76, doi:10.3390/risks7030076	35
Rasa Kanapickiene and Renatas Spicas	
Credit Risk Assessment Model for Small and Micro-Enterprises: The Case of Lithuania Reprinted from: <i>Risks</i> 2019 , 7, 67, doi:10.3390/risks7020067	57
Marc Chataigner and Stéphane Crépey	
Credit Valuation Adjustment Compression by Genetic Optimization Reprinted from: <i>Risks</i> 2019 , 7, 100, doi:10.3390/risks7040100	81
Ioannis Anagnostou and Drona Kandhai	
Risk Factor Evolution for Counterparty Credit Risk under a Hidden Markov Model Reprinted from: <i>Risks</i> 2019 , 7, 66, doi:10.3390/risks7020066	103
Tolulope Fadin and Thorsten Schmidt	
Default Ambiguity Reprinted from: <i>Risks</i> 2019 , 7, 64, doi:10.3390/risks7020064	125
Delphine Boursicot, Geneviève Gauthier and Farhad Pourkalbassi	
Contingent Convertible Debt: The Impact on Equity Holders Reprinted from: <i>Risks</i> 2019 , 7, 47, doi:10.3390/risks7020047	143

About the Special Issue Editor

Frédéric Vrins was awarded his PhD from the Ecole Polytechnique de Louvain (UCLouvain) in 2007 in the field of machine learning and adaptive signal processing, where he worked on signal separation techniques with a focus on biomedical applications. His contributions of both theoretical and empirical nature have been published in top journals in the field, such as IEEE Trans. Neural Networks, IEEE Trans. Signal Processing, and IEEE Trans. Information Theory. After his PhD, F. Vrins moved to the banking sector, where he spent 7 years as a front office quant, working in the trading room of a major European Bank. He was in charge of developing pricing and hedging models related to credit-sensitive derivatives products. He has served as a full time tenured professor of quantitative finance at the Louvain School of Management (UCLouvain) since his appointment in 2014. He is member of the Louvain Institute for Data Analysis and Modelling in statistics and economics (LIDAM) and chairman of the Louvain Finance research center (LFIN). He publishes his research in both practitioner and academic journals (Risk Magazine, Journal of Credit Risk, European Journal of Operations Research, Mathematical Finance, or Journal of Banking Finance, to name but a few).

Preface to "Advances in Credit Risk Modeling and Management"

Correctly assessing credit risk still represents an important challenge for both practitioners and scholars. On the one hand, credit risk measures play a central role in the banking sector's regulations, governing the profitability of financial institutions which remain at the heart of our economic system. On the other hand, effectively computing such measures in a sound and rigorous way triggers important challenges because of the lack of relevant information and/or models. It is therefore important that academics pursue efforts to improve their models. This book presents some recent advances which methodologically and/or computationally contribute to the more rigorous and reliable management of credit risk of firms. The book covers default and recovery rate models, trade credit, counterparty credit risk, and hybrid product pricing.

Frédéric Vrins
Special Issue Editor

Modelling Recovery Rates for Non-Performing Loans

Hui Ye * and Anthony Bellotti *

Department of Mathematics, Imperial College London, London SW7 2AZ, UK

* Correspondence: hui.ye16@alumni.imperial.ac.uk (H.Y.); a.bellotti@imperial.ac.uk (A.B.)

Received: 12 February 2019; Accepted: 15 February 2019; Published: 20 February 2019

Abstract: Based on a rich dataset of recoveries donated by a debt collection business, recovery rates for non-performing loans taken from a single European country are modelled using linear regression, linear regression with Lasso, beta regression and inflated beta regression. We also propose a two-stage model: beta mixture model combined with a logistic regression model. The proposed model allowed us to model the multimodal distribution we found for these recovery rates. All models were built using loan characteristics, default data and collections data prior to purchase by the debt collection business. The intended use of the models was to estimate future recovery rates for improved risk assessment, capital requirement calculations and bad debt management. They were compared using a range of quantitative performance measures under K -fold cross validation. Among all the models, we found that the proposed two-stage beta mixture model performs best.

Keywords: recovery rates; beta regression; credit risk

1. Introduction

In Basel II, an internal ratings-based (IRB) approach was proposed by the Basel Committee in 2001 to determine capital requirements for credit risk ([Bank for International Settlements 2001](#)). This IRB approach grants banks permission to use their own risk models or assessments to calculate regulatory capital. Under the IRB approach, banks are required to estimate the following risk components: probability of default (PD), loss given default (LGD), exposure at default (EAD) and maturity (M) ([Bank for International Settlements 2001](#)). Since Basel II's capital requirement calculation depends heavily on LGD, financial institutions have put more emphasis on modelling LGD in recent years. Unlike the estimation of PD, which is well-established, LGD is not so well-understood and still subject to research. Improving LGD modelling can help financial institutions assess their risk and regulatory capital requirement more precisely, as well as improving debt management.

LGD is defined as the proportion of money financial institutions fail to collect during the collection period, given the borrower has already defaulted. Conversely, Recovery Rate (RR) is defined as the proportion of money financial institutions successfully collected minus the administration fees during the collection period, given the borrower has already defaulted. Equations (1) and (2) give formal definitions of RR and LGD, respectively:

- Suppose individual i has already defaulted on a loan, let EAD_i be the exposure at default for this individual i .
- Let A_i be the administration costs (e.g., letters, phone calls, visits, lawyers and legal work) incurred for individual i .
- Let R_i be the amount recovered for individual i .

Then,

$$\text{Recovery Rate} = \frac{R_i - A_i}{EAD_i} = \frac{\sum \text{Collections} - \sum \text{Admin Fee}}{\text{Outstanding Balance at Default}} \quad (1)$$

and

$$\text{Loss Given Default} = 1 - \text{Recovery Rate} = 1 - \frac{R_i - A_i}{EAD_i} \quad (2)$$

RR mainly lies in the interval $[0, 1]$ and typically has high concentrations at the boundary points 0 and 1. It is possible for RR to be negative if recoveries are less than administration costs, $A_i > R_i$, and greater than 1 if recoveries exceed exposure plus administration costs, $R_i > EAD_i + A_i$. Typically, however, RR is truncated within the interval $[0, 1]$ when developing LGD models.

The main challenge in estimating LGD is the bimodal property with high concentrations at 0 and 1 typically present in LGD empirical distributions, where people either repay in full or repay nothing. For the dataset we used in this study, we found our LGD distribution is actually tri-modal. Therefore, regression models have been studied that specifically deal with this problem. For example, Bellotti and Crook (2012) built Tobit and decision tree models along with beta and fractional logit transformation of the RR response variable to forecast the LGD based on a dataset of 55,000 defaulted credit cards in the UK from 1999 to 2005. They concluded that ordinary least squares regression with macroeconomic variables performed the best in terms of forecast performance. Calabrese (2012) proposed a mixed continuous-discrete model, where the boundary values 0 and 1 are modelled by Bernoulli random variables and the continuous part of the RR is modelled by a Beta random variable. This model is then applied to predict RR of Bank of Italy's loans from 1985 to 1999. The result is compared with Papke and Wooldridge's fractional response model with log-log, logistic and complementary log-log link functions (Papke and Wooldridge 1996) and linear regression. The mixed continuous and discrete model achieves the best performance. Qi and Zhao (2011) applied four linear models, namely ordinary least squares regression, fractional response regression, inverse Gaussian regression, and inverse Gaussian regression with beta transformation, and two non-linear models, namely regression tree and neural network, to model the LGD of 3751 defaulted bank loans and bonds in the US from 1985 to 2008. They concluded that fractional response regression is slightly better than the ordinary least squares regression. Moreover, they reported that non-linear models perform best. Loterman et al. (2012) performed a benchmark study of LGD by comparing twenty-four different models using six datasets extracted from international banks. They concluded that non-linear models, such as neural network, support vector machine and mixture models perform better than linear models.

For this project, we specifically modelled and predicted RR for data from a single European country provided by a debt collection company. Due to reasons of commercial confidentiality and data protection, the debt collection company will remain anonymous and some aspects of the data were also anonymised, including the country of origin. Consequently, the data cannot be made publicly available. We applied some of the models that have already been studied previously and also extended the existing models, proposing a new beta mixture model to improve the accuracy of RR prediction. A good prediction of RR would help the debt collection company to determine collection policy for new debt portfolios. It is important to note that the RR we modelled is different from most RR, as the data only contain positive repayments and no administration fee was recorded. Therefore, all the RRs in our data lie in the range $(0, 1]$ instead of $[0, 1]$. Figure 1 shows a histogram of RR for the data. We can clearly see that there are modes at 0, 0.55 (approximately) and a high spike at boundary value 1. Since the shape of the empirical RR distribution demonstrates a trimodal feature, it is reasonable to assume that the recovery rate is a mixed type random variable. The multi-modality of RR is a natural consequence of different groups of bad debts being serviced using different strategies; e.g., one strategy may be that some bad debts are allowed to be written off if the debtor paid back some agreed fixed percentage of the outstanding balance. Having outcome RR within $(0, 1]$ motivated the use of the beta regression model and the multi-modal nature of RR motivates the use of a mixture model within this context.

The beta mixture model has been applied successfully within several other application domains. Ji et al. (2005) showed how to apply the beta mixture regression model in several bioinformatics

applications such as meta-analysis of gene expression data and to cluster correlation coefficients between gene expressions. Laurila et al. (2011) used a beta mixture model to describe DNA methylation patterns, helping to reduce the dimensionality of microarray data. Moustafa et al. (2018) used a beta mixture model as the basis of an anomaly detection system. Their network data are typically bounded, which suggests a beta distribution, and the use of the beta mixture allowed them to identify latent clusters in normal network use.

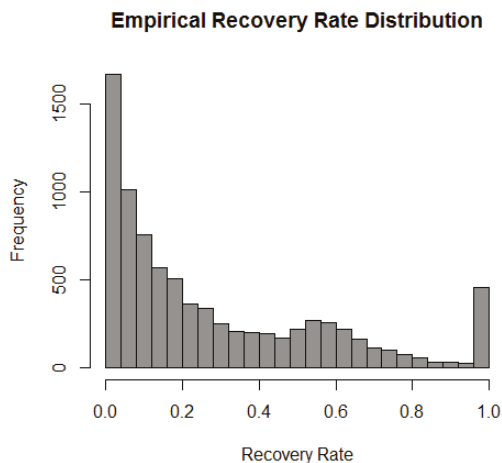


Figure 1. Histogram of recovery rates for 8237 loans after pre-preprocessing described in Section 2. The stack of 1s shows frequency of $RR = 1$, but the stack at 0 shows frequency for small $RR > 0$.

Inspired by Calabrese’s mixed continuous-discrete model (Calabrese 2012), we propose a two-stage model composed of:

- A beta mixture model is parameterised by mean and precision based on two sets of predictor variables on the interval of $(0, 1)$ in order to model the two modes located at just after 0 and around 0.55.
- A logistic regression model is used for the mode at boundary value 1.

The above proposed model allows representation of the trimodal feature of the data. The beta mixture component groups the clients into two clusters for $RR < 1$, based on their personal information, debt conditions and repayment history, which may become useful information for other business analysis and decision-making, and then uses logistic regression to model the third case $RR = 1$. In addition, we also used linear regression, linear regression with Lasso, beta regression and inflated beta regression to model RR. Model performance was measured by mean squared error, mean absolute error and mean aggregate absolute error under K -fold cross validation.

To our knowledge, this is the first study for estimating RR for portfolios of non-performing loans using a statistical model, and the first use of a beta mixture model for LGD. We also developed a novel procedure for predicting an expected value of outcome from a beta mixture model based on assigning a new observation to one of the clusters in the mixture. The remainder of the article is organised as follows: Section 2 provides a detailed data overview. Section 3 introduces the modelling methodology with great emphasis on the proposed beta mixture model combined with logistic regression model. Section 4 analyses some important features of the models and reports the model performance and Section 5 concludes with key findings and future recommendations.

2. Data

Three datasets were provided by the debt collection company:

- Dataset 1** provides 48 predictor variables of personal information including socio-demographic variables, Credit Bureau Score and debt status for 120,699 individuals for loans originating between January 1998 and May 2014 from several different financial institutions. Overall, 97.5% of them have credit card debt and 2.5% are refinanced credit cards (product = "R"). Partial information was extracted from a Bad Debt Bureau. Each record corresponds to a bad loan and has a unique key Loan.Ref.
- Dataset 2** records all the recoveries made by the bank before the debt collection company purchased the debt portfolio. It contains 15 predictor variables about historical collection information, which includes number of calls, contacts and visits made by the bank to collect the debt. It also includes repayments in the format of monthly summary. In total, there are 42,832 individuals' records in Dataset 2, among which only 34,807 individuals can be matched to Dataset 1 by Loan.Ref. Numbers of calls, contacts, visits, repayment and some other monthly activities are aggregated by summing for each loan identified by Loan.Ref.
- Dataset 3** records all the recoveries made by the debt collection company after they purchased the debt portfolio from the bank. It includes 12 predictor variables about the ongoing collection information. There are 8281 individuals in total, among which only 8237 individuals are from Dataset 1. Since only positive repayments are recorded, all the recovery rates we calculated are strictly greater than 0. Therefore, in the modelling section, we only focus on the recovery modelling in the interval $(0, 1]$, which is slightly different from the usual RR defined in $[0, 1]$. The debt collection period recorded in this dataset is from January 2015 to end of November 2016.

Figure 2 shows how the data were joined. There are 8237 data points presented in Dataset 3, but only 7161 individual historical collection information are recorded in Dataset 2. In these cases, there are no historical recoveries by bank, i.e., no calls, contacts, visits or payments for the remaining 1076 individuals. Therefore, a value of 0 was assigned to aggregate recoveries in Dataset 2 for the remaining 1076 individuals. The modified Dataset 2 was then joined to Datasets 1 and 3 by the unique key Loan.Ref and we obtained a table of 8237 data points with 61 variables.

Table A1 gives descriptive statistics for each of the variables in the joined dataset used in the statistical modelling. The predictor variable Pre-Recovery Rate is the bank's RR before the debt portfolio was purchased. The minimum value is -0.130 , which is negative due to the substantial amount of administration fee exceeding repayments incurred during the collection period. The predictor variable Credit Bureau Score is a generic credit score provided by a credit bureau.

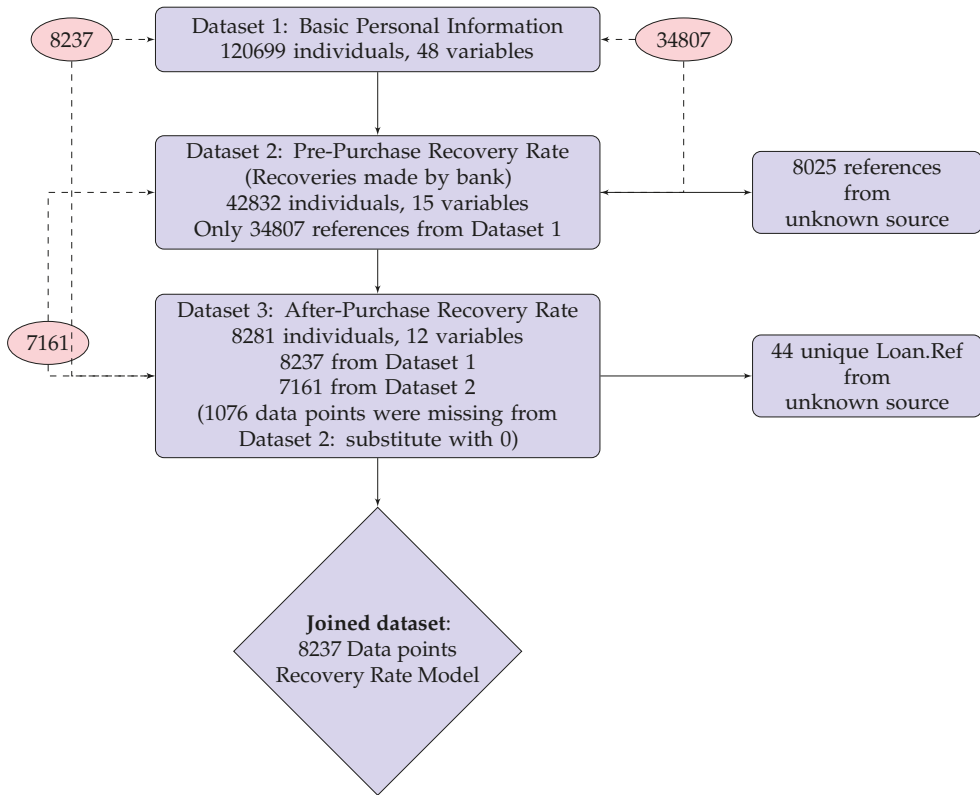


Figure 2. Joining the three datasets.

Recovery Rate Calculation

Since the repayments in Datasets 2 and 3 were recorded in the format of monthly activity summaries, each individual may have several repayments for the same loan. Therefore, we defined the recovery rate as the sum of repayments minus the administration fee (if available) over the original balance of the loan, which is also equivalent to the difference between original balance and ending balance over the original balance. For each individual i , RR is calculated using:

$$\text{Recovery Rate}_i = \frac{\sum \text{Repayments}_i - \text{AdminFee}_i}{\text{Original Balance}_i} = \frac{\text{Original Balance}_i - \text{Ending Balance}_i}{\text{Original Balance}_i} \quad (3)$$

Figure 1 is the empirical RR histogram calculated based on Equation (3), for the 8237 data points after pre-processing. The remaining 112,462 data points not included in the analysis essentially have $RR = 0$, but we do not know whether they have been serviced or not, thus they were not included in the analysis. Essentially, the goal of our model is to estimate RR computed from Dataset 3 (post-purchase), based on pre-purchase information given in Datasets 1 and 2.

3. Modelling Methodology

We applied various models to estimate RR. In all cases, model performance was measured within a K -fold cross validation framework. We first tried using ordinary least squares linear regression, with and without stepwise backward variable selection using the AIC criterion. In the following

sub-sections, we list the other modelling approaches we explored. Let y indicate the outcome variable, recovery rate, and X is a corresponding vector of predictor variables.

3.1. Linear Regression with Lasso

We applied linear regression with a Lasso (Least Absolute Shrinkage and Selection Operator) penalty. The model structure is

$$y = \beta_0 + \boldsymbol{\beta}^T X + \epsilon$$

where β_0 and $\boldsymbol{\beta}$ are intercept and coefficients to be estimated and ϵ is the error term. Then, estimation using least squares error with Lasso is given by the optimisation problem on a training dataset of N observations:

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \underset{\beta_0, \boldsymbol{\beta}}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - \boldsymbol{\beta}^T X_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \right], \quad (4)$$

where $\lambda > 0$ is a tuning parameter controlling the size of regularisation. Regression with Lasso will tend to shrink coefficient estimates to zero and hence is a form of variable selection (Friedman et al. 2010). The value of λ is chosen using K -fold cross validation. For this project, the R packages “lars” (Hastie and Efron 2013) and “glmnet” (Friedman et al. 2010) were used to estimate linear regression with Lasso.

3.2. Multivariate Beta Regression

The problem with linear regression is that it does not take account of the particular distribution of RR, which is between 0 and 1. The beta distribution, with two shape parameters α and β , allows us to model RR in the open interval (0, 1):

$$f(y_i; \alpha_i, \beta_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} y_i^{\alpha_i-1} (1 - y_i)^{\beta_i-1}, \quad 0 < y_i < 1, \quad (5)$$

where $\alpha, \beta > 0$ are the shape parameters and $\Gamma(\cdot)$ is the Gamma function. The beta distribution is reparameterised by mean and precision parameters, denoting by μ and ϕ , respectively, following Ferrari and Cribari-Neto (2004), since this parameterisation meaningfully express the expected value and variance:

$$\phi_i = \alpha_i + \beta_i, \quad E(y_i) = \mu_i = \frac{\alpha_i}{\alpha_i + \beta_i}, \quad \operatorname{Var}(y_i) = \frac{\mu_i(1 - \mu_i)}{\phi_i + 1}, \quad (6)$$

The reparameterised beta distribution is then

$$f(y_i; \mu_i, \phi_i) = \frac{\Gamma(\phi_i)}{\Gamma(\mu_i)\Gamma((1 - \mu_i)\phi_i)} y_i^{\mu_i\phi_i-1} (1 - y_i)^{(1-\mu_i)\phi_i-1}, \quad 0 < y_i < 1, \quad (7)$$

with $0 < \mu_i < 1$ and $\phi_i > 0$. Figure 3a demonstrates three examples of the beta distribution with fixed $\phi = 5$ and different μ . The variance is maximised at $\mu = 0.5$. Figure 3b demonstrates another three examples of beta distribution with fixed $\mu = 0.5$ and different ϕ .

The precision parameter ϕ is negatively correlated with $\operatorname{Var}(y_i)$, given a fixed μ . Furthermore, the variance of Y is a function of μ , which enables the regression to model heteroskedasticity. RR is modelled as $y_i \sim B(\mu_i, \phi_i)$ for $i \in (1, \dots, N)$ for sample size N . The multivariate beta regression model (Cribari-Neto and Zeileis 2010) is defined as:

$$F_1(\mu_i) = \eta^T X_i = \xi_{1i},$$

$$F_2(\phi_i) = \gamma^T W_i = \xi_{2i},$$

where η is a vector of parameters which needs to be estimated corresponding to predictor variables X and γ is a vector of parameters which needs to be estimated corresponding to predictor variables W .

The predictor variables in W may be the same as in X , or a subset, or contain different variables. For this study, W will have a subset of predictor variables determined using stepwise variable selection. The link function ensures that $\mu_i \in (0, 1)$ and $\phi_i > 0$. We applied Logit and Log link function to μ_i and ϕ_i , respectively:

$$\mu_i = \frac{1}{1 + e^{-\eta^T X_i}}, \quad \phi_i = e^{-\gamma^T W_i}.$$

With this multivariate beta regression model, η and γ can be estimated by maximum likelihood estimation, where the log-likelihood function is

$$L(\eta, \gamma) = \sum_{i=1}^N \left[\log \Gamma(\phi_i) - \log \Gamma(\mu_i \phi_i) - \log \Gamma((1 - \mu_i) \phi_i) + (\mu_i \phi_i - 1) \log y_i + ((1 - \mu_i) \phi_i - 1) \log(1 - y_i) \right]. \tag{8}$$

By substituting $\mu_i = F_1^{-1}(\eta^T X_i)$ and $\phi_i = F_2^{-1}(\gamma^T W_i)$ into Equation (8), the log-likelihood is obtained as a function of η and γ . The parameters can be estimated using Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton method, which is considered to be the most appropriate method (Mittelhammer et al. 2000; Nocedal and Wright 1999).

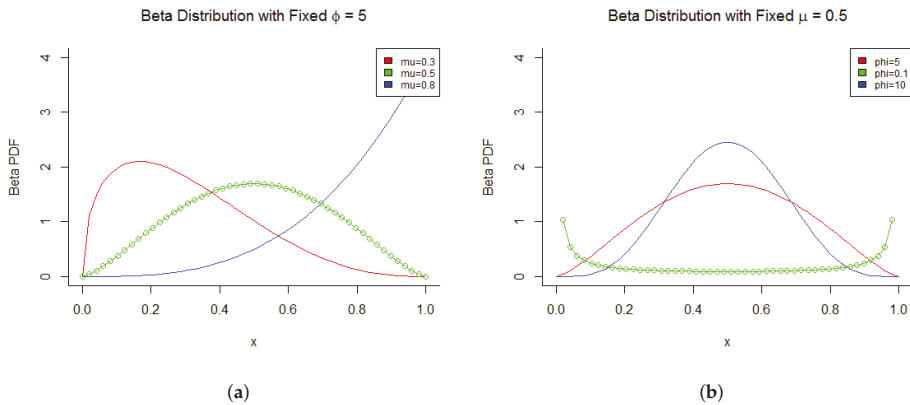


Figure 3. Beta distribution. (a) Beta Distribution with Fixed ϕ ; (b) Beta Distribution with Fixed μ .

3.3. Inflated Beta Regression

The disadvantage of beta regression is that it does not include the boundary values 0 or 1. Therefore, a modification is required before fitting the model. To better represent RR on the boundaries 0 and 1, Calabrese (2012) suggested considering RR as a mixture of Bernoulli random variables for the boundary 0 and 1, and a Beta random variable for the open interval (0, 1). The distribution for this inflated beta regression on $[0, 1]$ is then defined as

$$f_Y(y) = \begin{cases} p_0, & \text{if } y = 0 \\ (1 - p_0 - p_1)f_B(y; \alpha, \beta), & \text{if } 0 < y < 1 \\ p_1, & \text{if } y = 1 \end{cases} \tag{9}$$

for $y \in [0, 1]$, $p_0 = P(y = 0)$, $p_1 = P(y = 1)$, $0 < p_0 + p_1 < 1$ and $f_B(y)$ is the beta distribution defined in Section 3.2. Moreover, if $RR\ y \in (0, 1]$, i.e., it only inflates at one, as our data do, then the distribution is just

$$f_Y(y) = \begin{cases} (1 - p_1)f_B(y; \alpha, \beta), & \text{if } 0 < y < 1 \\ p_1, & \text{if } y = 1 \end{cases} \tag{10}$$

We used maximum likelihood estimation to estimate parameters for Bernoulli random variable and Beta random variables, parameterising the discrete part in the following way (Calabrese 2012):

$$s_i = \frac{p_1}{p_1 + p_0}, \quad d_i = p_0 + p_1,$$

The log-likelihood function is then

$$L(s, d, \alpha, \beta) = \sum_{y_i=0} \log(1 - s_i) + \sum_{y_i=0} \log(d_i) + \sum_{y_i=1} \log(s_i) + \sum_{y_i=1} \log(d_i) + \sum_{0 < y_i < 1} \log(1 - d_i) + \sum_{0 < y_i < 1} \log(f_B(y; \alpha_i, \beta_i)). \tag{11}$$

The continuous beta random variables can be parameterised in the same way as described in Section 3.2.

3.4. Beta Mixture Model combined with Logistic Regression

Examining the distribution of RR shown in Figure 1, it can be seen that the distribution between 0 and 1 is bimodal. For this reason, we consider a beta mixture model to deal with what appears to be two different groups of recoveries. We propose a two-stage model: beta mixture model combined with logistic regression. The beta mixture model allows us to model the multimodality of RR in the interval (0, 1). This is similar to the two-stage (decision tree) model used by Bellotti and Crook (2012), but with a beta mixture used for regression.

Firstly, RR is classified into ones and non-ones using logistic regression. Secondly, within the non-ones group, a mixture of beta distributions is used to model RR in the range (0, 1). In general, a mixture of beta distribution consists of m components where each component follows a parametric beta distribution. The prior probability of component j is denoted as π_j , where $j \in (1, \dots, m)$. Let M_j denote the j th component/cluster in the beta mixture model. The beta mixture model with m components is defined as:

$$\begin{aligned} g(y; \mu, \phi) &= \sum_{j=1}^m \pi_j f_j(y; X, \mu_j, \phi_j) \\ &= \sum_{j=1}^m \pi_j f_j(y; X, W, F_1^{-1}(\eta_j^T X_i), F_2^{-1}(\gamma_j^T W_i)) \\ &= \sum_{j=1}^m \pi_j f_j(y; X, W, \eta_j, \gamma_j), \end{aligned}$$

where f_j is the beta distribution corresponding to the j th component with separate parameter vectors η_j and γ_j . The same link functions are used as in Section 3.2. The prior probabilities, π_j , need to satisfy the following conditions:

$$\sum_{j=1}^m \pi_j = 1, \quad \pi_j \geq 0.$$

The iterative Expectation-Maximisation (EM) algorithm was used to estimate the parameters of the beta mixture model, as described by (Leisch 2004). In particular, R package “flexmix” (Leisch 2004; Gruen and Leisch 2007, 2008) embedded in R package “betareg” (Cribari-Neto and Zeileis 2010; Gruen

et al. 2012) was applied to estimate the model. Figure 4 illustrates the two-stage mixture model as a decision tree.

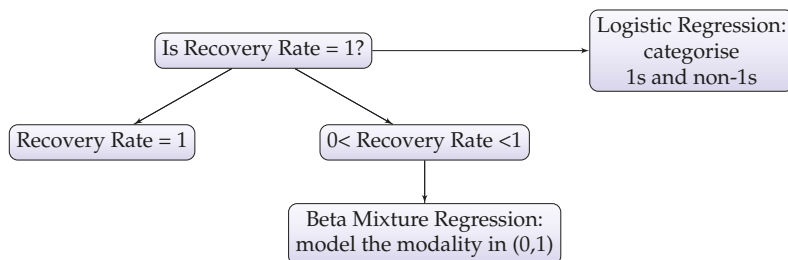


Figure 4. Estimate the expected value of RR using two-stage decision tree model.

The choice of m in the model depends on the number of clusters expected in the data. Based on our analysis of the recoveries for the dataset we used, $m = 2$ was used since this corresponded to the two modes we see in the RR distribution for $RR < 1$, as shown in Figure 1. If it is not clear how many clusters may exist, approaches based on AIC can be used.

Predictions Using the Beta Mixture Model

Given the beta mixture model, we need to predict the RR for new clients based on their information, i.e., X_{new} and W_{new} . Figure 5 shows a flowchart explaining how to calculate the estimated RR from the beta mixture model. This gives an expected value of RR y conditional on the cluster M_j . Therefore, we need to first identify which cluster the new observation belongs to. Even though the R package “betareg” (Cribari-Neto and Zeileis 2010; Gruen et al. 2012) can compute the conditional expectation for us, it does not identify which cluster the new points should be assigned to. Therefore, we propose a method to do this. In general, there are two feasible approaches to assign a new observation to M_j :

1. Assign the new observation to the cluster that achieves the highest log-likelihood. This is a hard clustering approach, which assigns the observation to exactly one cluster (Fraley and Raftery, 2002).
2. Assign the new observation to each cluster j with probability $P(M_j)$. This is a soft clustering approach, which assigns the observation to a percentage weighted cluster (Leisch 2004).

Decomposing the expected value of y using the Law of Total Expectation, we get

$$E(y | x_i) = \sum_{j=1}^m P(M_j | x_i) E(y | x_i, M_j) \tag{12}$$

where $E(y | x_i, M_j)$ is calculated from the beta mixture model prediction (refer to Figure 5). We can replace $P(M_j | x_i) = \frac{f(x_i | M_j) P(M_j)}{f(x_i)}$ where $f(x_i) = \sum_{j=1}^m f(x_i | M_j) P(M_j)$, to get

$$E(y | x_i) = \frac{\sum_{j=1}^m f(x_i | M_j) P(M_j) E(y | x_i, M_j)}{f(x_i)} \tag{13}$$

where $P(M_j)$ is the prior probability of belonging to cluster M_j . The density $f(x_i | M_j)$ is estimated using kernel density estimation,

$$\hat{f}(x^{new}) = \sum_{i=1}^n \frac{1}{n \prod_{k=1}^d h_{i,k}} \prod_{k=1}^d K\left(\frac{x_k^{new} - x_{i,k}}{h_{i,k}}\right)$$

where $K(\cdot)$ is the Gaussian kernel (Azzalini and Menardi 2014) and d is the number of dimensions in data x . In addition, x_i may be high-dimensional, which makes the kernel density estimation computationally expensive. As a remedy, we applied Principal Component Analysis (PCA) to reduce the dimension of x_i , and then kernel density estimation was performed in the reduced dimension space.

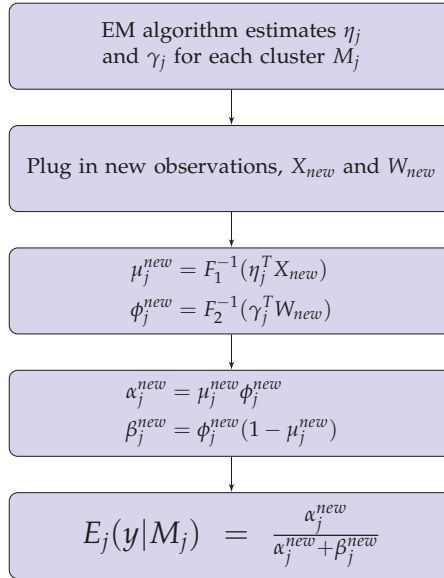


Figure 5. Prediction of RR conditional on each cluster M_j .

Approach 1: Maximum log-likelihood.

Given a new observation x_i , choose j that maximises the density:

$$\arg \max_j \log f(y|x_i, M_j),$$

which is computed using the log-likelihood function. If the objective function is maximised with respect to Cluster M_j , then set

$$P(M_j|x_i) = 1 \text{ and } P(M_k|x_i) = 0 \text{ for all } k \neq j$$

and hence, from Equation (12), the expected value of y is given by $E(y|x_i, M_1)$.

Approach 2: Prior Probability.

Treat $P(M_j)$ as a prior estimated using methods given in Table 1 and use in Equation (13) for soft clustering. By substituting $P(M_j)$ given in Table 1 into Equation (13), we can compute $E(y | x)$ for $y \in (0, 1)$.

Table 1. Determining $P(M_j)$ in Approach 2.

Approach 2	$P(M_1)$	$P(M_2)$
π_j prior	Extract π_j from the EM algorithm $\pi_j[1]$	Extract π_j from the EM algorithm $\pi_j[2]$
Prior based on training set cluster size ratio	$\frac{\text{Cluster 1 size}}{\text{Total sample size}}$	$\frac{\text{Cluster 2 size}}{\text{Total sample size}}$
Indifferent Prior	$\frac{1}{2}$	$\frac{1}{2}$

After calculating $E(y | x)$ for the interval $(0, 1)$ using the beta mixture model, the boundary 1 needs to be taken into consideration using a logistic regression model. From the decision tree defined in Figure 4, the logistic regression can provide the estimates at the first leaf node: $P(y = 1 | X = x)$. Then, the overall expectation of RR $y \in (0, 1]$ is

$$\begin{aligned}
 E(y | x) &= P(y = 1 | x)E(y | x, y = 1) + P(0 < y < 1 | x)E(y | x, 0 < y < 1) \quad (14) \\
 &= P(y = 1 | x) \times 1 + (1 - P(y = 1 | x)) E(y | x, 0 < y < 1) \\
 &= P(y = 1 | x) + (1 - P(y = 1 | x)) E(y | x, 0 < y < 1)
 \end{aligned}$$

where $E(y | x, 0 < y < 1)$ is the predicted RR from the beta mixture model using Approach 1 or 2.

4. Results

The linear model had an adjusted R^2 of 0.69, which was considerably higher than most models of RR (e.g., see Bellotti and Crook (2012); Loterman et al. (2012)), which could be explained by the richness of data, especially collections information. We expected that the linear regression model was misspecified, due to the range of the outcome variable and this is confirmed in the residual vs. fitted plot for the model and a Breusch-Pagan test for heteroscedasticity ($p < 0.0001$).

For the beta mixture model, we used all variables for X , but variable selection for W based firstly on the output of stepwise selection using AIC in linear regression and then on a series of likelihood ratio tests. The result was the selection of four variables for W : pre-recovery rate, post balance, customer payment frequency and credit bureau score. Table 2 shows parameter estimates for η and γ for the two clusters, along with coefficient estimates under standard beta regression in the interval $(0, 1)$ for comparison.

In Table 2, there are “NA” values for some of the p-values in the beta mixture model. This is because the estimation algorithm could not produce reliable standard errors in these cases. We can see that the significance of variables was diluted by the two clusters. For instance, credit bureau score was significant in the standard beta regression with a p-value of 0.0022, but in the beta mixture model, it was not significant for either of the clusters, taking a significance level of 5%. The direction of association of coefficient estimates in beta mixture model for both clusters were mostly consistent, where the estimates were significant (at 5% level), although magnitude of association differed. Pre-Recovery Rate for γ component was the only exception to this observation. The model also demonstrated some interesting significant associations between some variables and RR: taking insurance showed higher recoveries and having a record at the bad debt bureau was associated with lower recovery rates. In addition, the recoveries, pre-purchase, were positively correlated with future RR, although total number of calls to customer had a negative association, perhaps because these were difficult customers from whom to collect, hence requiring more intervention.

Table 2. η and γ estimated by EM algorithm. M1 and M2 represent Clusters 1 and 2.

Variables	Beta Mixture Model in (0, 1)				Beta Regression in (0, 1)	
	M1 Estimate	Pr(> z)	M2 Estimate	Pr(> z)	Betareg Estimate	Pr(> z)
η						
(Intercept)	-0.67015	<0.0001	-2.62862	<0.0001	-1.80064	<0.0001
Product R	-0.03376	0.47711	-0.00766	0.59733	0.02270	0.41766
Principal	0.00056	NA	0.00114	NA	0.00081	0.00000
Interest	0.00065	<0.0001	0.00118	NA	0.00097	0.00000
Insurance	0.00082	<0.0001	0.00116	<0.0001	0.00086	<0.0001
Late Charges	0.00042	0.00578	0.00115	<0.0001	0.00072	<0.0001
Overlimit Fees	-0.00105	0.07594	0.00145	<0.0001	0.00018	0.52533
Credit limit	0.00004	NA	-0.00001	NA	-0.00003	<0.0001
Sex = Male	0.03659	0.17453	-0.01412	0.13364	0.00969	0.43796
Marital status =						
Divorced	-0.01175	0.85305	-0.01427	0.47359	-0.03144	0.25840
Married	-0.06356	0.10819	-0.01476	0.16836	-0.03850	0.01957
Single	0.00982	0.83178	0.00695	0.63324	0.00332	0.86926
Widow	-0.14627	0.19497	0.02311	0.51404	-0.03869	0.45314
Other	0.12328	0.17954	-0.03476	0.22384	0.04570	0.24125
Age	-0.00273	0.05378	-0.00038	0.42389	-0.00115	0.07159
Credit Bureau Score	0.00059	0.10337	0.00007	0.07890	0.00038	0.00222
Bureau bad debt	-0.32990	0.01290	-0.06936	<0.0001	-0.24123	0.00000
Cust Payment Freq	0.06530	<0.0001	0.03506	<0.0001	0.05046	<0.0001
Post Balance	-0.00106	NA	-0.00127	NA	-0.00103	0.00000
Total Paid Amount	0.00004	NA	-0.00038	NA	-0.00014	<0.0001
Total Calls	-0.00044	0.00515	-0.00023	0.00275	-0.00032	<0.0001
Total Contacts	-0.00136	0.03257	0.00040	0.08116	-0.00031	0.28402
Bank report Freq	-0.01719	<0.0001	-0.00407	<0.0001	-0.01117	<0.0001
Pre recovery Rate	0.56850	<0.0001	3.63447	<0.0001	2.26212	<0.0001
EmployerNoInfo	-0.04457	0.63820	-0.01277	0.65487	0.03439	0.38375
Total Number	-0.00949	0.16151	-0.00169	0.42951	-0.00776	0.00651
γ						
(Intercept)	1.60514	<0.0001	2.64737	<0.0001	1.45450	0.00000
Pre recovery Rate	0.49096	0.00025	-2.11510	<0.0001	-0.18488	0.01538
Post Balance	0.00039	<0.0001	0.00018	NA	0.00031	0.00000
Cust Payment Freq	0.02949	<0.0001	0.17612	<0.0001	0.07759	0.00000
Credit Bureau Score	-0.00058	0.00458	-0.00033	0.09534	-0.00028	0.01388

Following the procedure in Figure 5, the expected value of RR conditional on Cluster M_j was calculated based on the parameters η and γ estimated in Table 2. Since it was too time consuming to perform kernel density estimation on 29 variables, we reduced the dimension to six by employing PCA analysis, which greatly shortened the running time for two clusters' density estimations. Nevertheless, it is inevitable that information is lost during the dimension reduction process, which may result in weaker estimates. Figure 6 shows histograms of expected value of RR conditional on each j th cluster for the test dataset. The shapes of the two clusters are similar, except Cluster 2 has more estimates in the range 0.2 to 0.6.

Figure 7 shows four histograms of predicted RR corresponding to the four different priors defined in Table 1, in contrast to the true RR. The predicted value of beta mixture model combined with logistic regression model was calculated by applying the formula derived in Equation (14). Models with the different priors performed in a similar way. Importantly, they were all able to model the bimodal nature of the RR. The figure shows that none of the models were good at predicting the extreme values of RR close to 0 or 1, but this naturally follows from the fact that these predictions are estimates of expected values of RR, through Equation (12), albeit conditional on predictor variables, and thus do not represent the extremes in the distribution well. Further detail can be seen in Figure 8, which shows predicted RR against true RR. The strong correlation between predicted and true RR is clear. However, it is noticeable that, when true RR was around 0.6, the model tended to under-estimate for some observations. This was because the model was not perfect at detecting observations in Cluster 2. This suggests future improvements to the model to enhance its capacity to predict the correct latent cluster.

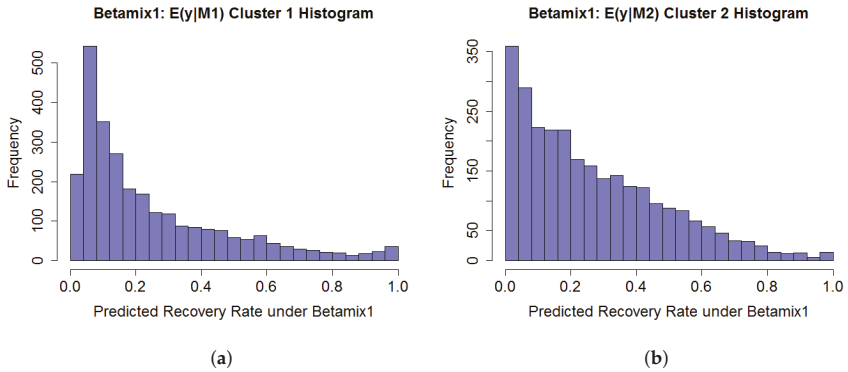


Figure 6. $E(y|M_i, X_{new})$ based on the Test dataset, for the two clusters ($m = 2$). (a) $E(y|M_1, X_{new})$; (b) $E(y|M_2, X_{new})$.

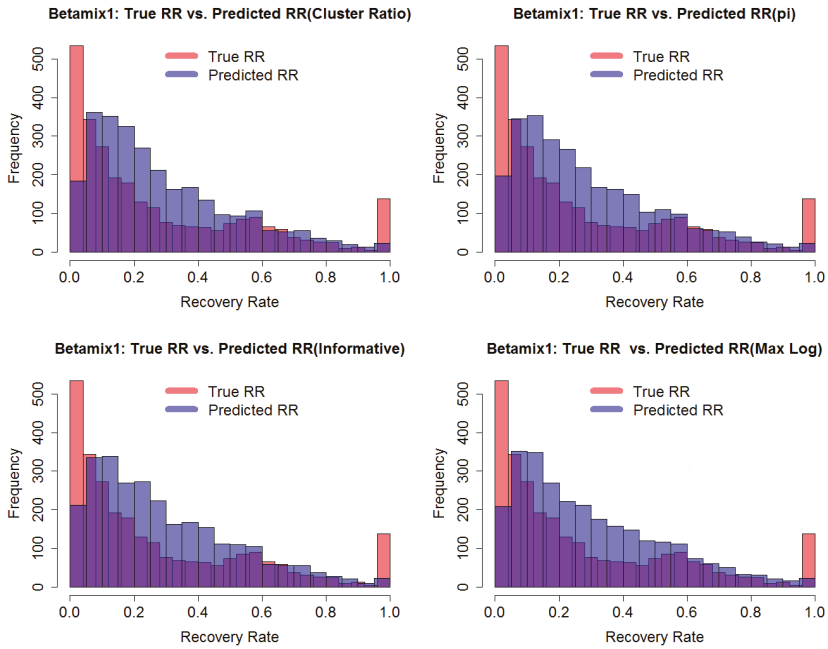


Figure 7. Predicted RR on test data ($n = 2746$) using beta mixture with four different priors, combined with logistic regression.

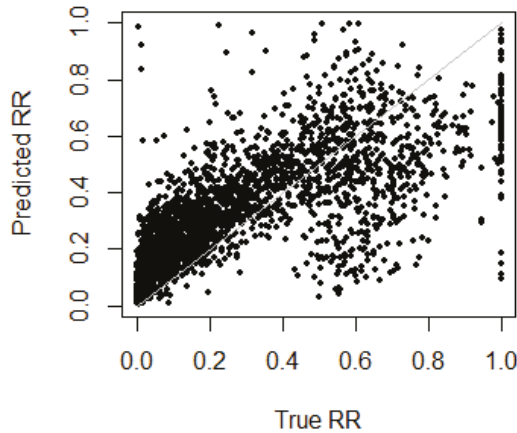


Figure 8. Predicted RR against true RR on test data ($n = 2746$) using beta mixture with the indifferent prior, combined with logistic regression.

Model Performance

Predictive performance was measured using K -fold cross validation with three performance measures popular in the literature on RR estimation: mean squared error (MSE), mean absolute error (MAE) and mean aggregate absolute error (MAAE). Since the sample size (8237) was relatively large and model estimation time was long, $K = 3$ was chosen. Let n be the sample size. Then,

$$MSE = \frac{1}{K} \sum_{k=1}^K \frac{1}{n/K} \sum_{i=1}^{n/K} (\hat{y}_{ki} - y_i)^2 = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n/K} (\hat{y}_{ki} - y_i)^2,$$

$$MAE = \frac{1}{K} \sum_{k=1}^K \frac{1}{n/K} \sum_{i=1}^{n/K} |\hat{y}_{ki} - y_i| = \frac{1}{n} \sum_{k=1}^K \sum_{i=1}^{n/K} |\hat{y}_{ki} - y_i|.$$

MAAE is the MAE at segment level (Thomas and Bijak 2015) and defined here as

$$MAAE = \frac{1}{V} \sum_{i=1}^V \frac{1}{|S_i|} \left| \sum_{j \in S_i} (y_j - \hat{y}_j) \right|$$

where V is the number of segments expressed as disjoint index sets S_1, \dots, S_V . The segments could express different characteristics, e.g., risk bands. However, for this study, each segment was a different random sample from the test data with approximately the same sample size and jointly exhaustive. We used $V = 100$ since this gave a balance of number of segments approximately equal to number of observations in each segment. MSE, MAE and MAAE are all penalty measures, thus the smaller the value, the better the model. Since the RR is a financial ratio between 0 and 1, the MAE can reflect the size of the error in a more intuitive and direct way. If one is interested in the segment portfolio level, then the MAAE should be used.

All models were trained on the same partitions of data into cross validation folds, to avoid bias being introduced due to different samples. Table 3 shows the results. There was little difference between

results for the various linear regressions, with or without variable selection or Lasso penalty, in terms of predictive performance. The last linear model, “excluding Dataset 2”, was built without predictor variables from Dataset 2. This showed noticeably worse performance than the other linear models, especially for MAE, which demonstrates that including past recoveries data (i.e., Dataset 2) improved performance. The standard beta regression model, with and without zero-inflation, performed much worse than linear regression, but the beta mixture model with logistic regression gave the best performance on all three measures. The different priors gave slightly different performances but are not very much different, although Approach 1 method for selecting cluster assignment (max log-likelihood) was slightly worse than Approach 2, soft clustering methods.

Table 3. Predictive results using three-fold cross validation.

Model	MSE	MAE	MAAE
Linear Regression			
Linear regression	0.024984	0.114268	0.025894
Stepwise linear regression	0.024752	0.113621	0.025700
Linear regression with Lasso	0.025228	0.114847	0.023739
Linear regression, excluding Dataset 2	0.026822	0.121385	0.026303
Beta regression			
Standard beta regression	0.085630	0.260459	0.161366
Inflated beta regression	0.076650	0.216374	0.048466
Beta mixture model combined with logistic regression			
Max log-likelihood	0.018750	0.095432	0.030629
Prior based on R Flexmix π_j	0.018460	0.091833	0.023991
Prior based on training set cluster size ratio	0.019325	0.092225	0.022594
Indifferent Prior	0.018030	0.092399	0.026298

5. Conclusions

Linear regression, beta regression, inflated beta regression, and a beta mixture model combined with logistic regression were applied to model the recovery rate of non-performing loans. The models’ predictive performances were measured using mean squared error, mean absolute error and mean aggregate absolute error under three-fold cross validation. To produce predictions from the beta mixture model, methods of hard and soft clustering were developed and the soft clustering approaches gave marginally better predictive performance. Theoretically, the proposed model, beta mixture model combined with logistic regression model, should be a suitable model to predict recovery rate for this data since it allows us to model the multimodality in the dataset and takes extra consideration of the boundary value. Indeed, we found that it achieved the best results amongst the models. Stepwise linear regression also achieved relatively good performance; however, the normality and homoscedasticity assumptions did not hold. In our experiments, we also found that inclusion of previous collections data boosted predictive performance.

We believe the beta mixture model is useful for modelling RR because it is explaining different servicing strategies. In the case of our study, the cluster with mode around 0.55 is likely expressing those loans for which the debt servicer has agreed with the borrower to repay just a proportion of the outstanding debt. There may be servicing strategies in other bad debt portfolios that could be discovered using a similar mixture model or clustering approach. We developed a technique to predict the correct latent cluster for new observations and this works well. However, results suggest that further work to refine this aspect of the use of the model could yield improved performance.

Author Contributions: Conceptualisation, H.Y. and T.B.; methodology, H.Y. and T.B.; validation, T.B.; investigation and statistical modelling, H.Y.; data analysis, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, T.B.; and supervision, T.B.

Funding: This research received no external funding.

Acknowledgments: We wish to thank the anonymous debt collection company for use of their data and their expertise, which was essential to understand the meaning and context of the data. We would also like to thank

Tommaso Pappagallo who did preliminary data analysis as part of his MSci project, which was useful in taking this project work forward.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IRB	Internal ratings based
RR	Recovery rate
LGD	Loss given default
PD	Probability of default
EAD	Exposure at default
EM	Expectation-Maximisation (algorithm)
MSE	Mean square error
MAE	Mean absolute error
MAAE	Mean absolute aggregate error

Appendix A

Table A1. Descriptive statistics. $n = 8237$. For numeric variables: min, mean (standard deviation), max. For factors, frequency (%age) for each level. All predictive data is collected prior to servicing.

Variable	Type	Description	Statistics
RR post	numeric	Recovery rate (outcome variable)	0.000508, 0.280 (0.283), 1
Product	factor	Type of loan	C:7468 (90.7%), R:769 (9.3%)
Principal	numeric	Original loan amount	0, 3120 (2330), 15000
Interest	numeric	Interest payments	0, 551 (439), 3380
Insurance	numeric	Insurance fees	0, 42 (84.6), 953
Late charges	numeric	Late charge fees	0, 269 (109), 1470
Overlimit fees	numeric	Over credit limit fees	0, 13.3 (24.6), 315
Creditlimit	numeric	Credit limit	0, 4560 (2660), 13800
Sex	factor	Sex	F:3196 (38.8%), M:5041 (61.2%)
Married	factor	Marriage status	0:1201 (14.6%), D:518 (6.3%), M:3929 (47.7%), O:217 (2.6%), S:2230 (27.1%), W:142 (1.7%)
Age	numeric	Age	1, 48.7 (11.1), 87
DelphiScore	integer	Credit bureau score	0, 298 (138), 443
Bureau Sub 1	factor	Loan is in the servicer's bureau (1 = True)	0: 1520 (18.5%), 1: 6717 (81.5%)
CustPaymentFreq	integer	Customer repayment frequency	1, 7.56 (5.59), 29
Post Balance	numeric	Exposure amount at start of servicing	0, 3130 (2630), 15900
Total paid amount	numeric	Total net paid amount	−275, 1200 (1100), 11200
Total calls	numeric	Total number of calls	0, 104 (106), 911
Total contacts	numeric	Total number of contacts (except calls)	0, 28.5 (26.5), 196
Bankreport Freq	numeric	Bank reporting frequency	0, 11.6 (7.92), 26
Pre recovery rate	numeric	Recovery rate	−0.130, 0.258 (0.217), 2.89
Employer	factor	Employer known	EmployerProvided:8053 (97.8%), NoInfo:184 (2.2%)
Total number	integer	Total number of loan accounts	0, 2.3 (2.43), 68

References

- Azzalini, Adelchi, and Giovanna Menardi. 2014. Clustering via nonparametric density estimation: The R package pdfCluster. *Journal of Statistical Software* 57: 1–26. [CrossRef]
- Bank for International Settlements. 2001. *The Internal Ratings-based Approach*. Basel: Bank for International Settlements.
- Bellotti, Tony, and Jonathan Crook. 2012. Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting* 28: 171–82. [CrossRef]
- Calabrese, Raffaella. 2012. Predicting bank loan recovery rates with a mixed continuous-discrete model. *Applied Stochastic Models in Business and Industry* 30: 99–114. [CrossRef]

- Cribari-Neto, Francisco, and Achim Zeileis. 2010. Beta regression in R. *Journal of Statistical Software* 34: 1–24. [CrossRef]
- Ferrari, Silvia, and Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31: 799–815. [CrossRef]
- Fraley, Chris, and Adrian E Raftery. 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97: 611–31. [CrossRef]
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1–22. [CrossRef] [PubMed]
- Gruen, Bettina, Ioannis Kosmidis, and Achim Zeileis. 2012. Extended beta regression in R: Shaken, stirred, mixed, and partitioned. *Journal of Statistical Software* 48: 1–25. [CrossRef]
- Gruen, Bettina, and Friedrich Leisch. 2007. Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics & Data Analysis* 51: 5247–52.
- Gruen, Bettina and Friedrich Leisch. 2008. Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software* 28: 1–35. [CrossRef]
- Hastie, Trevor, and Brad Efron. 2013. *lars: Least Angle Regression, Lasso and Forward Stagewise*. R package version 1.2. Available online: <https://cran.r-project.org/web/packages/lars/lars.pdf> (accessed on 18 February 2019).
- Ji, Yuan, Chunlei Wu, Ping Liu, Jing Wang, and Kevin R. Coombes. 2005. Applications of beta-mixture models in bioinformatics. *Bioinformatics* 21: 2118–22. [CrossRef] [PubMed]
- Laurila, Kirsti, Bodil Oster, Claus L. Andersen, Philippe Lamy, Torben Orntoft, Olli Yli-Harja, and Carsten Wiuf. 2011. A beta-mixture model for dimensionality reduction, sample classification and analysis. *BMC Bioinformatics* 12: 215. [CrossRef] [PubMed]
- Leisch, Friedrich. 2004. Flexmix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 11: 1–38. [CrossRef]
- Loterman, Gert, Iain Brown, David Martens, Christophe Mues, and Bart Baesens. 2012. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting* 28: 161–70. [CrossRef]
- Mittelhammer, Ron C., George Judge, and Douglas Miller. 2000. *Econometric Foundations*, 1st ed. Cambridge: Cambridge University Press.
- Moustafa, Nour, Gideon Creech, and Jill Slay. 2018. Anomaly Detection System using Beta Mixture Models and Outlier Detection. In *Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing*. Edited by Prasant Kumar Pattnaik, Siddharth Swarup Rautaray, Himansu Das and Janmenjoy Nayak. Singapore: Springer, vol. 710. [CrossRef]
- Nocedal, Jorge, and Stephen J. Wright. 1999. *Numerical Optimization*, 1st ed. Berlin: Springer.
- Papke, Leslie, and Jeffrey Wooldridge. 1996, 11. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11, 619–32. [CrossRef]
- Qi, Min, and Xinlei Zhao. 2011. Comparison of modeling methods for loss given default. *Journal of Banking & Finance* 35: 2842–55.
- Thomas, Lyn, and Katarzyna Bijak. 2015. Impact of Segmentation on the Performance Measures of LGD Models. Available online: <https://crc.business-school.ed.ac.uk/wp-content/uploads/sites/55/2017/02/Impact-of-Segmentation-on-the-Performance-Measures-of-LGD-Models-Lyn-Thomas-and-Katarzyna-Bijak.pdf> (accessed on 18 February 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

The Determinants of Market-Implied Recovery Rates

Pascal François

Department of Finance, HEC Montréal, 3000 Chemin de la Côte-Ste-Catherine, Montreal, QC H3T 2A7, Canada; pascal.francois@hec.ca

Received: 1 April 2019; Accepted: 10 May 2019; Published: 18 May 2019

Abstract: In the presence of recovery risk, the recovery rate is a random variable whose risk-neutral expectation can be inferred from the prices of defaultable instruments. I extract market-implied recovery rates from the term structures of credit default swap spreads for a sample of 497 United States (U.S.) corporate issuers over the 2005–2014 period. I analyze the explanatory factors of market-implied recovery rates within a linear regression framework and also within a Tobit model, and I compare them with the determinants of historical recovery rates that were previously identified in the literature. In contrast to their historical counterparts, market-implied recovery rates are mostly driven by macroeconomic factors and long-term, issuer-specific variables. Short-term financial variables and industry conditions significantly impact the slope of market-implied recovery rates. These results indicate that the design of a recovery risk model should be based on specific market factors, not on the statistical evidence that is provided by historical recovery rates.

Keywords: recovery rate; credit risk; loss given default

1. Introduction

The management and pricing of credit risk rely on the accurate assessment of the likelihood of default and of the recovery upon default. To this end, credit risk models extract market information from the term structure of credit spreads, which can be estimated from corporate bond prices and, more recently, also from credit default swap (CDS) premiums. Early literature regarding credit risk (Merton 1974) established that, in the absence of arbitrage, the credit spread should equal the product of the default probability and the loss given default under the equivalent martingale measure. For several decades, much of the research effort has been dedicated to modelling the default event, while recovery was assumed to be constant. For example, Giesecke et al. (2011) estimate the relation between default rates and credit spreads while assuming a fixed 50% recovery in their 150-year sample period.

With the accumulated evidence of time-series and cross-sectional variations of historical recovery rates, more advanced credit risk models incorporate non-constant loss given default. However, mere calibration on historical recovery rates ignores the premium that the market may assign to recovery risk. Recent empirical literature has extracted recovery rates from the joint observation of different defaultable instruments issued by the same entity (Ünal et al. 2003; Das and Hanouna 2009; Doshi et al. 2018). These market-implied recovery rates lead to a substantial reevaluation of credit risk exposures (François and Jiang 2019). However, an analysis of the factors driving them and a comparison with the determinants of historical recovery rates remains, to the best of my knowledge, to be carried out. The paper aims at filling this gap.

Intuitively, one should expect the factors driving historical recovery rates and those driving the market-implied recovery rates to be different. Indeed, recovery rates are random variables and historical recoveries are their *ex post* realizations, whereas market-implied recoveries are their risk-neutral expectations. The way that the market assigns a recovery risk premium, on one hand, and the way that financial distress is eventually resolved, on the other, should induce distinct patterns for *ex ante* and *ex post* recovery.

I extract market-implied recovery rates from the term structure of CDS spreads for 497 United States (U.S.) corporate issuers over the 2005–2014 period. They exhibit a fairly symmetric distribution with clustered observations at zero—in sharp contrast with the typical bimodal distribution of historical recovery rates. I then proceed to a factor analysis using firm-specific, industry-specific, and macroeconomic determinants that are commonly identified in the bond recovery literature.

My contribution includes the following findings. Historical recovery rates have been shown to be affected by short-term, firm-specific variables, such as liquidity and profit margin. By contrast, I show that market-implied recovery rates are impacted by corporate factors that are more informative about the long run, such as leverage. In addition, while the effect of size is unclear on historical recovery rates, I unambiguously document a positive relation between size and market-implied recovery within a non-linear model. Similarly, the role of macroeconomic factors on historical recovery has received mixed support, except for default rates. I show that the strong negative relation with default rates also prevails for market-implied recovery rates. However, in contrast to their historical counterparts, macroeconomic factors, such as GDP growth and unemployment rate, also impact market-implied recovery rates. The slope of the term structure of market-implied recovery rates is affected by the same variables, short-term financial variables (such as liquidity and profit margin), and by the industry stock performance. The main implication of my results is the following. The design of a recovery risk model should be based on specific market factors, not on the statistical evidence that historical recovery rates provide.

I review the related literature in the next section. Section 3 addresses all aspects of the methodology (including the recovery model, the data and the model calibration, and a comparison with historical recovery rates). Section 4 presents the factor analysis and its results. I conclude in Section 5.

2. Related Literature

My study stands at the intersection of two streams of literature: (i) The determinants of historical recovery rates and (ii) the inference of recovery rates from market data. I shall briefly review these two research areas.

2.1. Factors Driving Historical Recovery Rates

Significant literature has examined the determinants of historical recovery rates. Two types of contributions can be distinguished. Most of the known determinants have been identified using standard linear regression models. These works include [Frye \(2000\)](#), [Düllmann and Trapp \(2004\)](#), [Altman et al. \(2005\)](#), [Varma and Cantor \(2005\)](#), [Chava et al. \(2011\)](#), among others. Subsequent contributions aim at improving the fit of original models using more sophisticated specifications. These include parametric models, such as fractional response regressions ([Bastos 2009](#)) or inverse Gaussian regressions ([Qi and Zhao 2011](#)), as well as non-parametric models, such as regression trees ([Bastos 2009](#); [Nazemi and Fabozzi 2018](#)), neural networks ([Qi and Zhao 2011](#); [Loterman et al. 2012](#)), and support vector regressions ([Loterman et al. 2012](#); [Nazemi et al. 2018](#)).

The determinants identified by all these studies can be grouped into four categories.

- Debt contract-specific variables: Coupon rate, seniority, collateral.
- Firm-specific variables: Size, asset tangibility, market-to-book ratio, liquidity ratio, interest coverage ratio, profit margin, leverage, firm age.
- Industry-specific variables: industry dummy, utilities dummy, industry sales growth, industry stock return.
- Macroeconomic variables: Bond default rate, GDP growth, S&P500 index return, S&P500 index volatility, unemployment rate, Fama-French factors, economic uncertainty.

Table 1 provides a summary of the findings, relating the determinants and the sign of their effect to the various empirical studies. The seminal study of [Altman and Kishore \(1996\)](#) supports the economic intuition that secured debt and senior debt are associated with higher recovery. These relations have

been documented in early studies using linear regressions, and they have been confirmed by recent work using more advanced models (Qi and Zhao 2011; Siao et al. 2016).

Table 1. Review of the determinants of historical recovery rates. For each identified determinant, the table lists some studies documenting a significant relation with historical recovery rates. The next column indicates whether that relation is positive or negative. The method refers to linear regression (LR), probit regression (PR), logistic quantile regression (LQ), support vector regression (SVR), or regression trees (RT).

Determinant	Examples of Studies	Effect	Method
Panel A: Debt contract-specific variables			
Coupon rate	Chava et al. (2011)	+	PR
Collateral	Frye (2000)	+	LR
	Qi and Zhao (2011)	+	RT
Seniority	Varma and Cantor (2005)	+	LR
	Acharya et al. (2007)	+	LR
	Siao et al. (2016)	+	LQ
Rating	Jankowitsch et al. (2014)	+	LR
Panel B: Firm-specific variables			
Size	Acharya et al. (2007); Chava et al. (2011)	+/-	LR, PR
Market-to-book	Chava et al. (2011)	-	PR
Asset tangibility	Varma and Cantor (2005)	+	LR
	Chava et al. (2011)	+	PR
Liquidity	Varma and Cantor (2005)	+	LR
Profit margin	Acharya et al. (2007)	+	LR
Leverage	Varma and Cantor (2005)	-	LR
Default event severity	Franks and Torous (1994); Altman and Karlin (2009)	-	LR
Panel C: Industry-specific variables			
Industry dummies	Acharya et al. (2007); Chava et al. (2011)	+/-	LR, PR
Industry sales growth dummy	Acharya et al. (2007)	+	PR
Industry stock return dummy	Acharya et al. (2007)	+	PR
Industry default rate	Jankowitsch et al. (2014)	-	LR
Panel D: Macroeconomic variables			
Default rate	Frye (2000); Altman et al. (2005)	-	LR
GDP growth	Altman et al. (2005); Chava et al. (2011)	+	LR, PR
Fed fund rate	Jankowitsch et al. (2014)	+	LR
Stock index return	Nazemi et al. (2018)	+	SVR, RT
Corporate bond spread	Nazemi et al. (2018)	-	SVR, RT
Unemployment rate	Nazemi et al. (2018)	-	SVR, RT

It is also natural to think that the characteristics of the issuing firm affect the eventual recovery on its distressed debt. All else equal, when assets are big (firm size), when they are tangible (mainly property, plant and equipment), and when they are mostly in place (low market-to-book ratio), creditors will obtain a higher recovery (Varma and Cantor 2005; Chava et al. 2011). In addition, the firm's ability to repay (captured by various liquidity ratios, by profit margin, or by leverage) is also found to positively impact recovery (Acharya et al. 2007). It is also documented that the type of default (private workout versus formal bankruptcy) impacts on historical recovery rates (Franks and Torous 1994; Altman and Karlin 2009). All else equal, when the firm cannot resolve financial distress by means of private renegotiations (as in distressed exchanges), it is an indication that the default event is particularly severe and this negatively affects recovery.

Acharya et al. (2007) particularly put the role that is played by industry factors. In line with the argument made by Shleifer and Vishny (1992), recovery is negatively impacted when the industry of the defaulting firm performs poorly. This is because the specific assets of the defaulting firm cannot be sold at their fair price, since potential buyers are themselves under financial stress.

It is a well-established stylized fact that recoveries are inversely related to default rates (Frye 2000; Altman et al. 2005). The importance of additional macroeconomic variables has been emphasized

in more recent studies. [Nazemi and Fabozzi \(2018\)](#) and [Nazemi et al. \(2018\)](#) mainly extract significant macroeconomic variables using Lasso and variable importance metrics from gradient boosting, respectively. Through different model specifications, including linear regressions, regression trees, and support vector regressions, they conclude that the stock index return, the corporate bond spread, and the unemployment rate are among the most salient determinants of historical recovery rates. [Gambetti et al. \(2019\)](#) identify economic uncertainty as an additional factor. They measure economic uncertainty using several proxies, including the VIX, survey-based, and news-based variables. They show that higher economic uncertainty lowers recovery rates, but it also affects their distribution (assuming they are generated from a beta distribution).

2.2. Inferring Recovery Rates from Market Data

The credit spread on a defaultable instrument reflects the product of the default probability and the loss given default, both under the equivalent martingale measure. Therefore, observing the credit spread leads to the identification problem of disentangling the default probability from the recovery rate. When discussing the implementation of their reduced-form credit risk model, [Duffie and Singleton \(1999\)](#) initially suggested working with defaultable instruments written by the same issuer but distinct in recovery.

In this line of reasoning, [Ünal et al. \(2003\)](#) apply a reduced-form credit risk model on the prices of senior and junior debts. Another possibility is to work with the term structure of CDS spreads. [Jaskowski and McAleer \(2012\)](#) estimate a reduced-form credit risk model with the Bayesian MCMC method. [Doshi et al. \(2018\)](#) extract the term structure of recovery rates from the joint observation of senior and subordinated CDS. [Das and Hanouna \(2009\)](#) and [François and Jiang \(2019\)](#) rely on the term structure of CDS spreads and on stock prices. In this latter approach, the default hazard rate is inversely related to the level of the stock price.

3. Methodology

3.1. The Recovery Model

Consider a reference firm with traded equity (stock price process is denoted by S_t) and a traded term structure of CDS contracts (the j -maturity CDS premium process is denoted by $\pi_t(j)$). Assuming that the absolute priority rule is enforced, the equity-holders obtain zero recovery upon default. By contrast, the CDS offers compensation for the loss given default on the underlying reference bond. Consequently, the (risk-neutral) default probability of the issuing firm and the recovery rate on the reference bond affects the CDS premium. Therefore, it is possible to infer market-implied recovery rates from the joint observation of S_t and the collection of $\pi_t(j)$, since equity and CDS contracts are driven by the same default process, but exhibit distinct recovery rates.

I follow [Das and Hanouna \(2009\)](#) and [François and Jiang \(2019\)](#) to extract market-implied recovery rates. Specifically, the stock price process dynamics is modelled by [Cox et al. \(1979\)](#) binomial tree with a state of jumping to default. At each node on the tree, S_i^k denotes the stock price, where superscript k indexes time (from 0 to N periods) and subscript i indexes the level of the node at time k . In the k th period, i takes value 0 at the top and value k at the bottom.

In one time step, the stock price with return volatility σ can take on the following values

$$\begin{aligned} S_i^{k+1} &= S_i^k \exp(\sigma \sqrt{h}) \text{ with probability } q_i^k (1 - \lambda_i^k), \\ S_{i+1}^{k+1} &= S_i^k \exp(-\sigma \sqrt{h}) \text{ with probability } (1 - q_i^k) (1 - \lambda_i^k), \\ &0 \text{ with probability } \lambda_i^k, \end{aligned}$$

where h is the time step, λ_i^k is the one-period probability of jumping to default (default intensity), and q_i^k is the risk-neutral probability of an up move if the firm survives.

In the absence of arbitrage opportunities, the risk-neutral probability q_i^k is given by

$$q_i^k = \frac{\frac{R^k}{1-\lambda_i^k} - \exp(-\sigma \sqrt{h})}{\exp(\sigma \sqrt{h}) - \exp(-\sigma \sqrt{h})},$$

where R^k is the one-period compound factor.

Das and Hanouna (2009) and François and Jiang (2019) posit a non-linear inverse relation between the default hazard rate ε_i^k and the level of the stock price

$$\lambda_i^k = 1 - \exp(-\varepsilon_i^k h) = 1 - \exp\left(-\left(S_i^k\right)^{-b} h\right),$$

where parameter b is the elasticity of the hazard rate with respect to the stock price. Such a specification results in time- and state-dependent default intensity.

In addition, they work with a logistic specification for the recovery rate φ_i^k on the defaultable reference bond

$$\varphi_i^k = \frac{1}{1 + \exp(a_0 + a_1 \lambda_i^k)},$$

where a_0 is a scale parameter and a_1 captures the dependence between the default intensity and the recovery rate.

Consider the CDS with settlement dates $T_u, u = 1, \dots, j$. For clarity of exposure, $\lambda^k \equiv \lambda^k(T_{k-1}, T_k)$ denotes the probability of jumping to default between settlement dates T_{k-1} and T_k . Likewise, $\varphi^k \equiv \varphi^k(T_{k-1}, T_k)$ is the recovery rate prevailing between the settlement dates T_{k-1} and T_k . The λ^k and φ^k are obtained from aggregation across all states of the binomial tree.¹ In this framework, the no-arbitrage premium $\pi(j)$ equalizes the two legs of the CDS²

$$\pi(j) = \frac{\sum_{u=1}^j \left(\prod_{k=1}^{u-1} (1 - \lambda^k) \lambda^u (1 - \varphi^u) B(T_u) \right)}{\sum_{u=1}^j \left(\prod_{k=1}^u (1 - \lambda^k) B(T_u) \right)},$$

where $B(t)$ is the risk-free discount factor for horizon t .

I calibrate the model-implied CDS premiums on the observed term structure $\pi_{obs}(j)$ to extract the term structure of the market-implied recovery rates (φ^k). That is, for J available CDS maturities, I solve for

$$\min_{a_0, a_1, b, \sigma} \sum_{j=1}^J (\pi(j) - \pi_{obs}(j))^2,$$

which yields the quadruplet of calibrated parameters $\Omega = (a_0, a_1, b, \sigma)$.

3.2. Data and Model Calibration

The initial sample contains monthly stock prices and seven-maturity CDS term structures for 542 firms traded on the U.S. market (NYSE or NASDAQ).³ The sample period begins in January 2005 and ends in June 2014 (114 months),⁴ which results in 52,021 terms structures and stock price observations. The sample period spans an entire business cycle with the 2008–2009 financial crisis being located in

¹ The aggregation is performed through a recursive algorithm. See Das and Hanouna (2009) for details.

² The denominator is the expected present value of all the premiums to be paid. The numerator is the expected present value of the compensation for the loss given default.

³ All CDS are U.S. dollar denominated and senior unsecured single-name contracts.

⁴ Before September 2010, the data was provided by CMA via Datastream. After that date, the data is provided by Thomson Reuters. The data is combined from the two providers by using the function “SPLC” of Datastream.

the middle. All of the discount factors are obtained from the smoothed Treasury yield curve reported by the Fed St-Louis website.

Table 2 reports statistics regarding the quality of the calibration on the CDS term structures. As far as the whole sample is concerned, the average relative fitting error is 10.2%. However, in some extreme cases, the fit is poor and the inferred market-implied recovery rate might not be accurate. Therefore, I restrict the analysis by excluding the 5% of cases of worst calibrations. Working with that restricted sample, I obtain an average fitting error of 9 bps (median is 2 bps) and an average relative fitting error of 7.2% (median is 4.4%).

Table 2. Fitting the credit default swap (CDS) term structure. The table reports the calibration results on monthly term structures of CDS spreads. The root mean squared errors (RMSE) and the relative root mean squared errors (RRMSE) are expressed in basis points and in percentage points, respectively. The initial sample contains 52,021 firm-month observations. The restricted sample is obtained by excluding the 5% of observations with the worst fit.

Statistic	Initial Sample		Restricted Sample	
	RMSE	RRMSE	RMSE	RRMSE
Mean	23	10.23	9	7.19
Median	3	4.76	2	4.44
Standard deviation	222	17.12	17	7.58
Maximum	9389	440.08	93	39.34
95% percentile	93	39.34	51	24.59

I focus on the 10-year recovery rate, since long-term recovery rates more adequately capture all the impacts of a lengthy resolution of financial distress to analyze the level of market-implied recovery rates.⁵ However, the calibrated term structure also provides information about how the market expects the recovery rate to evolve. Consequently, I also examine the slope of market-implied recovery rates defined as the difference between the 10-year rate and the five-year rate.

From here on, I only retain quarterly observations of recovery rates (March, June, September, and December) to match with the highest frequency that is available for explanatory variables. The final sample comprises 16,062 observations. Figure 1 presents the distributions of the level and the slope of market-implied recovery rates. Not surprisingly, the level exhibits a unimodal distribution that is truncated at zero. That distribution is fairly symmetric, judging by the skewness (−0.05) and by the mean (33.60%) and the median (33.27%) being very close to one another. I find that the term structures of recovery rates are decreasing (mean slope is −0.17). Das and Hanouna (2009) and Doshi et al. (2018) also find downward sloping curves for the recovery rates. This indicates that firms experiencing a sudden default may obtain a greater resale value for their assets. By contrast, firms defaulting in the medium or long run are associated with slow financial health deterioration and the gradual loss of asset resale value.

Figure 2 represents the time series behavior of the level and the slope of market-implied recovery rates. Just like their historical counterparts, the market-implied recovery rates appear to be pro-cyclical, at least judging by the business cycle spanned by the sample period. One can also note that the dispersion around the median (i.e., the gap between the third and first quartiles) increased during the recession. The slope appears to be strongly negatively correlated with the level. Thus, the whole term structure of market-implied recovery rates went down and became more flat during the crisis.

⁵ For instance, empirical studies on U.S. bankruptcy filings (Bris et al. 2006; Denis and Rodgers 2007) report that firms in financial distress spend between two and three years on average under bankruptcy.

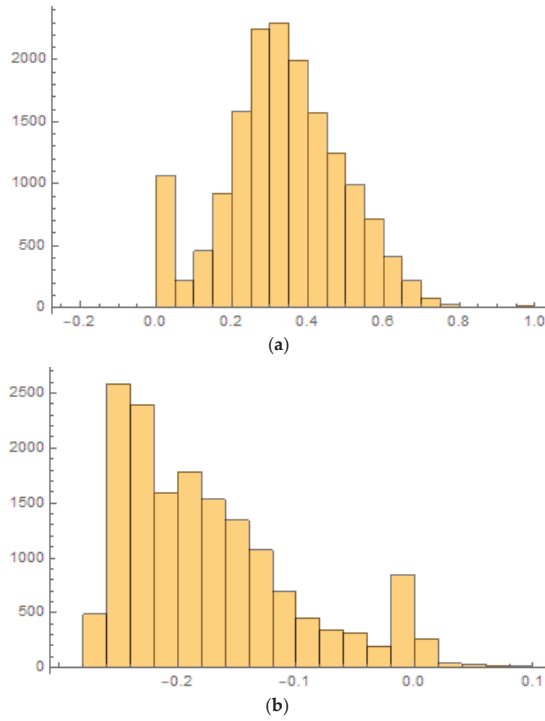


Figure 1. Histograms for the level and the slope of market-implied recovery rates. Panel (a) shows the distribution of the level (10-year recovery rate). Panel (b) shows the distribution of the slope (10-year rate minus five-year rate). The sample contains 16,062 quarterly observations and it spans the January 2005–June 2014 period.

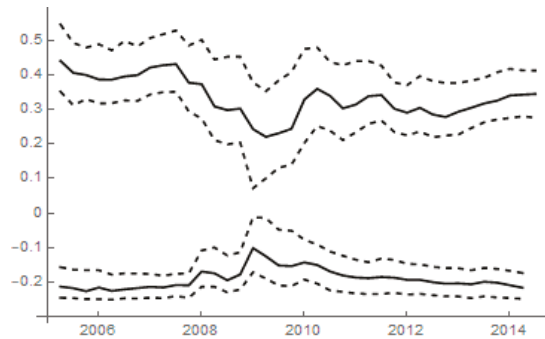


Figure 2. Time series of market-implied recovery rate level and slope. The level (slope) is shown in the top (bottom) curves. The solid line represents the median, the two dashed lines represent the first and third quartiles.

3.3. Comparison with Historical Recoveries

The recovery rates that are inferred from the sample of CDS contracts reflect market anticipations about recoveries on senior unsecured bonds. Therefore, it is interesting to compare them with historical recoveries on the same bond category. However, that comparison is subject to the following caveats.

- The actual recovery rates from CDS auctions are slightly lower than actual recovery rates on the underlying bonds. [Chernov et al. \(2013\)](#) and [Gupta and Sundaram \(2015\)](#) find a downward bias of about 15% of the bond price (which represents a smaller fraction of par) and attribute it to a liquidity premium.
- Market-implied recovery rates from CDS might contain a premium for the CDS writer's counterparty risk. That being said, [Arora et al. \(2012\)](#) find that the negative relation between CDS spreads and CDS writer credit quality is economically very small because of risk mitigation techniques, such as overcollateralization and bilateral netting.
- Historical recovery rates (as reported by Moody's for instance) are, by definition, calculated on a sample of defaulting firms. The market-implied recoveries are extracted from a sample of firms underlying a CDS contract. The difference in the two populations of firms could generate a bias in the comparison of recoveries.⁶
- Most importantly, market-implied recoveries are risk-neutral expectations of random recovery rates. By contrast, historical recoveries are calculated once the default event has materialized. In some studies, the historical recovery rates are computed at the resolution of financial distress and they can be viewed as the realizations of random recovery. In other studies, the historical recovery rates are computed using 30-day post default bond prices. Such historical recovery rates are still expectations about ultimate recovery, but they are conditionally calculated on default having occurred.

While most credit risk models have been typically calibrated using historical recoveries, pricing credit risky instruments should require forward-looking recoveries (*ex ante* measures, i.e., prior to default) that may incorporate a risk premium. The aim of this paper is precisely to show how these *ex ante* and *ex post* measures exhibit distinct patterns, and how they are affected by different drivers.

Figure 3 shows a histogram of historical recovery rates that were reported by Moody's over the 2005–2017 period. The distribution is clearly bimodal, which suggests that the resolution of financial distress leads to either limited losses (one mode is around 75%) or a severe write-off for creditors (the other mode is around 25%).⁷ The average rate of 43% is not representative of the heterogeneity in historical recoveries, as previously documented by [Altman and Kishore \(1996\)](#), [Düllmann and Trapp \(2004\)](#), [Altman et al. \(2005\)](#), or [Bruche and Gonzalez-Aguado \(2010\)](#), among others.

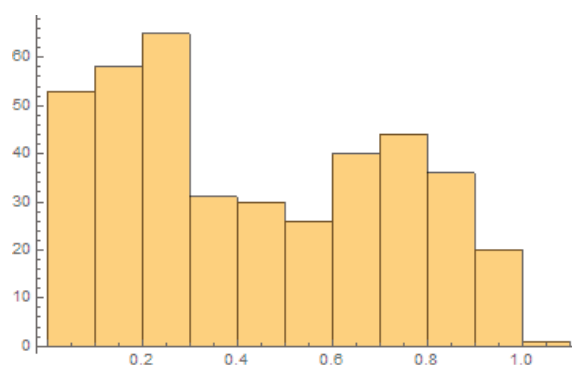


Figure 3. Distribution of historical recovery rates. Data is obtained from Moody's credit reports and spans the 2005–2017 period.

⁶ When regressing historical recovery rates, [Jankowitsch et al. \(2014\)](#) find a positive and significant coefficient for CDS availability.

⁷ [Altman and Kalotay \(2014\)](#) propose a mixture of normals to model the bimodal distribution of historical recovery rates. [Siao et al. \(2016\)](#) opt for a quantile-based regression.

4. Factor Analysis

I retain the following explanatory variables, in line with the previous literature.⁸ Firm-specific variables are size, asset tangibility, liquidity, profit margin, leverage, and a dummy for investment grade firms. Industry-specific variables include dummies for the NAICS first digit and a dummy for industry stock underperformance. GDP growth, unemployment rate, S&P500 index return, and default rate are the macroeconomic variables. Table 3 presents the definitions of variables in detail.

Table 3. Definition of explanatory variables.

Variable	Description
Size	Logarithm of total assets.
Asset tangibility	Property, plant and equipment/total assets.
Liquidity	(Cash plus short-term investments)/total assets.
Profit margin	EBITDA/sales.
Leverage	Long-term debt/total assets.
Rating	Dummy = 1 if issuer is investment grade.
Industry stress	Dummy = 1 if quarterly industry index return is below −30%.
GDP growth	Seasonally adjusted, quarterly growth rate of U.S. GDP.
Unemployment	Seasonally adjusted, quarterly U.S. unemployment rate.
Stock index	S&P500 index adjusted, quarterly return.
Default rate	Quarterly default rate reported by Moody's and S&P.

Table 4 shows the correspondence between NAICS industries and their sectorial stock indices, which is used to assess industry stock performance and to ultimately build the dummy related to industry stress.

Table 4. NAICS industries and corresponding stock indices.

NAICS	Industry	GICS Stock Index	Ticker
11	Agriculture, forestry and fishing	Agriculture	S5AGRI
21	Minerals and gases	Energy	SPN
22	Utilities	Utilities	S5UTIL
23	Construction	Construction and engineering	S5CSTEX
31	Food manufacturing	Food and beverage	SPSIFBUP
32	Wood and concrete manufacturing	Materials	S5MATR
33	Metal manufacturing	Metal and mining	SPSIMM
42	Wholesale trade	Retail	SPSIRE
44	Retail trade	Retail	SPSIRE
45	Sporting goods and book stores	Retail	SPSIRE
48	Transportation and warehousing	Transportation	SPSITN
49	Postal service	Transportation	SPSITN
51	Information and newspaper	Media and entertainment	S5MEDA
52	Finance and insurance	Financials	SPF
53	Real estate, rental and leasing	Real estate	S5RLST
54	Professional and technical services	Commercial and professional services	S5COMS
56	Administrative and support services	Consumer services	S5HOTR
62	Health care	Health care	S5HLTH
72	Food services	Restaurants	S5REST
81	Other non-public services	Consumer services	S5HOTR

Observations with missing firm characteristics are simply removed from the analysis. The final sample contains 497 firms and 12,598 observations. After reporting descriptive statistics in the next subsection, I proceed to the factor analysis of market-implied recovery rates using the linear regression model.

⁸ Since CDS in the sample are written on the same type of bonds (senior unsecured), the factor analysis precludes those variables that are specific to the debt contract such as coupon, seniority, or collateral.

That specification is simple and allows for a direct comparison with the literature on historical recovery rates. It would be worth investigating more sophisticated approaches to improve the goodness-of-fit and possibly highlight additional significant relations in the same spirit that this literature evolved. This is left for future research. However, given the censored distribution of market-implied recovery rates, I complement the factor analysis with Tobit regressions in Section 4.3. These regressions can be viewed as a robustness check for the findings that are highlighted in the linear regression approach.

4.1. Descriptive Statistics

Table 5 reports the main descriptive statistics of the explanatory variables (excluding dummies).

Table 5. Descriptive statistics for explanatory variables.

Variable	Mean	Std Dev	Min	Q1	Q2	Q3	Max
Size	9.5283	1.3598	4.7791	8.5894	9.3755	10.2807	14.8302
Asset tangibility	0.3120	0.2461	0.0000	0.1009	0.2534	0.5050	0.9530
Liquidity	0.1619	0.2233	0.0000	0.0336	0.0877	0.1975	3.4089
Profit margin	0.2015	0.3369	−7.7214	0.1043	0.1735	0.2771	24.1566
Leverage	0.2767	0.1973	0.0000	0.1445	0.2428	0.3714	2.3640
GDP growth	0.0177	0.0267	−0.0840	0.0050	0.0225	0.0360	0.0540
Unemployment	0.0688	0.0187	0.0440	0.0500	0.0655	0.0880	0.1000
Stock index	0.0153	0.0688	−0.2356	−0.0207	0.0335	0.0605	0.1314
Default rate	0.0042	0.0039	0.0008	0.0020	0.0022	0.0047	0.0174

By construction, studies regarding historical recovery rates report characteristics of defaulting firms. As expected, my sample of firms exhibits much different characteristics: mean profitability (20%), liquidity (16%), and leverage (28%) reflect an average firm in good financial health. As a matter of fact, the majority of sample firms are well rated (the rating dummy equals 1 for 53.9% of observations). In addition, the distribution of firms across industries is fairly representative of the economy, as shown in Table 6. The industry stress dummy only equals 1 for 3.6% of observations. This is because the threshold for industry underperformance (−30%) is meant to capture extreme liquidity shortage among competitors in the spirit of Shleifer and Vishny (1992). In my *ex ante* study of recovery rates, where most firms are far from the default state, such a scenario has much lower likelihood.

Table 6. Distribution of firms across industries (NAICS first digit).

Industry	Count	Industry	Count
Agriculture	1 (0.2%)	Professional services	149 (30.0%)
Minerals and gases	80 (16.1%)	Health care	9 (1.8%)
Manufacturing	186 (37.4%)	Food services	9 (1.8%)
Transportation and trade	62 (12.5%)	Other non-public services	1 (0.2%)

4.2. Linear Regression Results

I first investigate the determinants of market-implied recovery rates within a simple linear regression framework. Table 7 reports the results for the level using five different specifications. The first model includes all of the contemporaneous variables with industry fixed effects. The second model uses firm fixed effects instead. The third model considers macroeconomic variables lagged by one-quarter, as previous studies on historical recovery rates documented their superior explanatory power (Chava et al. 2011). To control for multicollinearity issues, the fourth model only retains those variables with a variance inflation factor (VIF) that is below 10. The fifth model is the same as the first one, but it is estimated by maximum likelihood. All of the specifications include time fixed effects and robust standard errors are adjusted for heteroscedasticity.

In previous empirical studies (Acharya et al. 2007; Jankowitsch et al. 2014; Siao et al. 2016), the standard determinants of historical recovery rates achieve a relatively high goodness-of-fit, with R^2

typically ranging between 40% and 60%. Table 7 indicates that their explanatory power is much lower when applied to market-implied recovery rates. Macroeconomic variables keep playing an important role, which confirms the pro-cyclical behavior of recovery rates. GDP growth rate stands as the macroeconomic variable that passes the multicollinearity test while keeping a high significance level. Among the firm specific variables, only leverage significantly contributes across all specifications. A possible interpretation is that the market mostly relies on a long-term financial analysis to estimate forward-looking recovery rates. Thus, financial indicators, such as liquidity or profit margin, which may affect immediate recovery, lose their explanatory power when it comes to assessing the consequences of default in a distant future. By contrast, leverage is a financial decision that impacts investment and operating policies in the long run. Similar to the findings regarding historical recovery rates, the relation between firm size and market-implied recovery rates is ambiguous.

Table 7. Regression results for the level of market-implied recovery rates. Robust standard errors are reported in parentheses. Variables preceded by “L_” are lagged by one quarter. Statistical significance at the 10%, 5%, and 1% level is indicated by superscript *, ** and ***, respectively.

Variable	1	2	3	4	5
Intercept	0.5762 *** (−0.0366)	0.5762 *** (0.1025)	0.4369 *** (0.1021)	0.3977 *** (0.0209)	0.3850 *** (0.0656)
Size	−0.0137 *** (0.0021)	−0.0137 (0.0104)	−0.0125 (0.0104)	− (0.0030)	0.0077 ** (0.0030)
Asset tangibility	0.0286 (0.0589)	0.0286 (0.0486)	0.0254 (0.0486)	0.0324 (0.0491)	0.0210 (0.0174)
Liquidity	0.0214 (0.0142)	0.0214 (0.0186)	0.0181 (0.0186)	0.0258 (0.0189)	0.0201 ** (0.0103)
Profit margin	0.0032 (0.0070)	0.0032 (0.0065)	0.0031 (0.0067)	0.0035 (0.0068)	0.0037 (0.0034)
Leverage	−0.1214 ** (0.0449)	−0.1214 *** (0.0326)	−0.1220 *** (0.0326)	−0.1223 *** (0.0328)	−0.1202 *** (0.0140)
Rating	0.0091 (0.0049)	0.0091 (0.0067)	0.0098 (0.0073)	0.0085 (0.0074)	0.0220 *** (0.0043)
Industry stress	0.0046 (0.0058)	0.0046 (0.0067)	−0.0068 (0.0064)	−0.0043 (0.0063)	0.0039 (0.0062)
GDP growth	0.1339 * (0.0595)	0.1339 *** (0.0485)	− (0.0064)	0.1497 *** (0.0483)	0.1362 ** (0.0621)
Unemployment	−0.8287 (0.6827)	−0.8287 *** (0.2794)	− (0.2794)	− (0.2827)	−0.7864 *** (0.2827)
Stock index	−0.0353 (0.0208)	−0.0353 * (0.0182)	− (0.0184)	−0.0304 * (0.0184)	−0.0356 (0.0249)
Default rate	−3.5229 *** (0.6149)	−3.5229 *** (0.6956)	− (0.6956)	− (0.6956)	−3.4103 *** (0.6752)
L_ GDP growth	−	−	0.1401 *** (0.0470)	−	−
L_ Unemployment	−	−	1.2948 *** (0.2692)	−	−
L_ Stock index	−	−	−0.0183 (0.0161)	−	−
L_ Default rate	−	−	3.7581 *** (0.4939)	−	−
Industry fixed effects	Yes	No	No	No	Yes
Firm fixed effects	No	Yes	Yes	Yes	No
R ² within	0.1182	0.1182	0.1198	0.1142	−
R ² between	0.0458	0.0458	0.0598	0.1567	−
R ² overall	0.0928	0.0928	0.1013	0.1447	−

Table 8 presents the linear regression results for the slope of market-implied recovery rates using the same five different specifications. The slope is negative in most cases. Therefore, its steepness measures how fast the market believes the recovery will deteriorate in the future. Thus, a variable that is associated with a positive coefficient contributes to flattening the slope. By contrast, a negative relation indicates that the variable aggravates the decline in recovery.

Table 8. Regression results for the slope of market-implied recovery rates. Robust standard errors are reported in parentheses. Variables preceded by “L_” are lagged by one quarter. Statistical significance at the 10%, 5% and 1% level is indicated by superscript *, ** and ***, respectively.

Variable	1	2	3	4	5
Intercept	-0.1164 ** (0.0435)	-0.1164 ** (0.0585)	-0.0740 ** (0.0335)	-0.1980 *** (0.0116)	-0.1605 *** (0.0340)
Size	-0.0173 *** (0.0044)	-0.0173 *** (0.0059)	-0.0077 *** (0.0015)		-0.0077 *** (0.0016)
Asset tangibility	0.0721 ** (0.0240)	0.0721 *** (0.0223)	0.0404 *** (0.0089)	0.1149 *** (0.0260)	0.0392 *** (0.0090)
Liquidity	0.0031 (0.0114)	0.0031 (0.0129)	0.0053 (0.0053)	0.0140 (0.0124)	0.0044 (0.0052)
Profit margin	-0.0111 * (0.0050)	-0.0111 ** (0.0049)	-0.0113 *** (0.0017)	-0.0122 ** (0.0056)	-0.0111 *** (0.0017)
Leverage	0.0624 *** (0.0146)	0.0624 *** (0.0203)	0.0614 *** (0.0072)	0.0972 *** (0.0208)	0.0602 *** (0.0072)
Rating	-0.0181 *** (0.0040)	-0.0181 *** (0.0042)	-0.0198 *** (0.0022)	-0.0219 *** (0.0047)	-0.0194 *** (0.0022)
Industry stress	0.0085 (0.0064)	0.0085 ** (0.0038)	0.0200 *** (0.0030)	0.0246 *** (0.0036)	0.0087 *** (0.0031)
GDP growth	0.0143 (0.0287)	0.0143 (0.0264)	-	-0.2603 *** (0.0301)	0.0140 (0.0316)
Unemployment	1.0304 *** (0.2230)	1.0304 *** (0.1548)	-	-	1.0447 (0.1438)
Stock index	-0.0445 *** (0.0109)	-0.0445 *** (0.0090)	-	-0.0459 *** (0.0113)	-0.0444 *** (0.0127)
Default rate	2.0635 *** (0.4733)	2.0635 *** (0.3694)	-	-	2.1315 *** (0.3434)
L_ GDP growth	-	-	-0.1958 *** (0.0330)	-	-
L_ Unemployment	-	-	-0.4481 *** (0.1436)	-	-
L_ Stock index	-	-	0.0327 *** (0.0127)	-	-
L_ Default rate	-	-	0.8050 ** (0.3444)	-	-
Industry fixed effects	Yes	No	Yes	No	Yes
Firm fixed effects	No	Yes	No	Yes	No
R ² within	0.2142	0.2142	0.2048	0.1543	-
R ² between	0.0722	0.0722	0.1274	0.0601	-
R ² overall	0.1377	0.1377	0.1848	0.0984	-

Macroeconomic variables continue to be highly significant. They point towards a countercyclical behavior of the slope. That is, in times of recession, the term structure of market-implied recovery rates is low and flat, indicating high losses given default across all horizons. By contrast, in times of economic growth, all of the recovery rates increase, but those in the short run increase even more. This is consistent with the notion that immediate defaults during expansion lead to limited losses.

In contrast with the regression for the level, all of the firm characteristics (except for liquidity) significantly impact the slope. Leverage contributes to a low and flat term structure. However, size, profit margin, and the rating steepen the negative slope. Hence, the market believes that big, profitable firms with a good rating preserve a high recovery in the short run.

Another notable difference with results in Table 7 is that the industry distress variable is now significant. Consistent with the argument put forward by [Shleifer and Vishny \(1992\)](#), competitors in financial difficulty cannot offer the fair price to liquidated specific assets. The regression results suggest that this effect is only temporary, as it mainly lowers short-term recoveries and not long-term ones (thereby flattening the term structure). Additionally, this finding is in line with [Shleifer and Vishny \(1992\)](#), who argue that a delayed liquidation procedure helps in avoiding fire asset sales.

4.3. Tobit Regression Results

The empirical literature on historical recovery rates shows that, even if model sophistication improves the goodness-of-fit, it does not fundamentally affect the identification of factors. Indeed, the most relevant determinants have been detected in simple linear regressions.

However, the same result may not hold for market-implied recovery rates. As documented in Figure 1, the level of market-implied recovery rates exhibits a unimodal, symmetric distribution, with clustered observations at zero. Therefore, it is natural to analyze its determinants within a censored regression model. I run a Tobit model with random effects and a censor threshold at 0.01%, leaving 11,873 (94%) uncensored observations. Table 9 gathers the results. All specifications include time fixed effects. Specifications 1 and 2 only differ in the inclusion of industry fixed effects. Specification 3 uses lagged macroeconomic variables. Specification 4 only retains those variables with a VIF below 10.

Table 9. Tobit regression results for the level of market-implied recovery rates. Robust standard errors are reported in parentheses. Variables preceded by “L_” are lagged by one quarter. Statistical significance at the 10%, 5% and 1% level is indicated by superscript *, ** and ***, respectively.

Variable	1	2	3	4
Intercept	0.3853 *** (0.0710)	0.3775 *** (0.0350)	0.2449 *** (0.0709)	0.4329 *** (0.0589)
Size	0.0086 *** (0.0032)	0.0077 ** (0.0032)	0.0091 *** (0.0033)	-
Asset tangibility	0.0129 (0.0187)	0.0055 (0.0166)	0.0105 (0.0186)	0.0089 (0.0189)
Liquidity	0.0212 ** (0.0108)	0.0206 * (0.0109)	0.0182 * (0.0108)	0.0188 * (0.0108)
Profit margin	0.0050 (0.0036)	0.0049 (0.0036)	0.0049 (0.0036)	0.0058 (0.0036)
Leverage	-0.1348 *** (0.0150)	-0.1386 *** (0.0150)	-0.1350 *** (0.0150)	-0.1406 *** (0.0150)
Rating	0.0221 *** (0.0046)	0.0228 *** (0.0046)	0.0227 *** (0.0046)	0.0241 *** (0.0045)
Industry stress	0.0047 (0.0065)	0.0051 (0.0065)	-0.0072 (0.0061)	-0.0049 (0.0061)
GDP growth	0.1268 * (0.0656)	0.1264 * (0.0656)	-	0.1414 ** (0.0656)
Unemployment	-0.9115 *** (0.2998)	-0.9145 *** (0.2998)	-	-
Stock index	-0.0342 (0.0263)	-0.0342 (0.0263)	-	-0.0273 (0.0252)
Default rate	-3.4702 *** (0.7143)	-3.4672 *** (0.7143)	-	-
L_ GDP growth	-	-	0.1497 ** (0.0681)	-
L_ Unemployment	-	-	1.2966 *** (0.2968)	-
L_ Stock index	-	-	-0.0192 (0.0262)	-
L_ Default rate	-	-	3.9085 *** (0.7104)	-
Industry fixed effects	Yes	No	Yes	Yes
Log-likelihood	7047.49	7043.20	7058.31	7020.83

The results are very similar to the ones from the linear regressions. Yet, the Tobit specification provides additional information regarding the factors driving market-implied recovery rates. It allows, in particular, to pick up the positive effects of size, rating, and, to a lesser extent, liquidity. Furthermore, the Tobit specification clarifies the firm size effect. Bigger firms are now unambiguously associated with larger market-implied recovery rates.

5. Conclusions

The goal of this study was to shed light on the determinants of market-implied recovery rates and to examine, in particular, the way that they differ with the well-identified drivers of historical recovery rates.

The link between historical recovery rates and the business cycle has clearly been established through the negative relation with default rates. However, the explicit contribution of macroeconomic factors has received mixed support (see [Mora 2012](#)). By contrast, this study shows that the forward-looking nature of market-implied recovery rates makes them very sensitive to macroeconomic variables, such as the GDP growth rate and the unemployment rate. A strong negative relation with default rates also prevails.

Other important differences arise when it comes to firm specific factors. The historical recovery rates are mostly affected by short-term financials, such as cash levels and profit margin, while size and leverage fail to be significant. Market-implied recovery rates are rather impacted by corporate factors that are more informative about the long run. They are strongly negatively related to leverage. The positive relation with size and rating quality can be captured in a non-linear model.

The [Shleifer and Vishny \(1992\)](#) industry effect also applies to historical and to market-implied recovery rates, but since this is a transitory effect, its impact mainly consists in flattening the term structure of market-implied recovery rates.

The major implication of this study is that the modelling of recovery risk in a credit risk pricing framework calls for specific factors and should not rely on the statistical evidence that is reported for historical recovery rates.

Funding: This research received no external funding.

Acknowledgments: The paper has benefitted from comments by Gunnar Grass, Weiyu Jiang, and three anonymous referees. Excellent research assistance by Jean-Michel Ostiguy is gratefully acknowledged. I am solely responsible for any remaining error.

Conflicts of Interest: The author declares no conflict of interest.

References

- Acharya, Viral, Sreedhar Bharath, and Anand Srinivasan. 2007. Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics* 85: 787–821. [[CrossRef](#)]
- Altman, Edward, and Egon Kalotay. 2014. Ultimate recovery mixtures. *Journal of Banking and Finance* 40: 116–29. [[CrossRef](#)]
- Altman, Edward, and Brenda Karlin. 2009. The re-emergence of distressed exchanges in corporate restructurings. *Journal of Credit Risk* 5: 43–55. [[CrossRef](#)]
- Altman, Edward, and Vellore Kishore. 1996. Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analyst Journal* 52: 57–64. [[CrossRef](#)]
- Altman, Edward, Brooks Brady, Andrea Resti, and Andrea Sironi. 2005. The link between default and recovery rates: Theory, empirical evidence, and implications. *The Journal of Business* 78: 2203–28. [[CrossRef](#)]
- Arora, Navneet, Priyank Gandhi, and Francis Longstaff. 2012. Counterparty credit risk and the credit default swap market. *Journal of Financial Economics* 103: 280–93. [[CrossRef](#)]
- Bastos, Joao. 2009. Forecasting bank loan loss-given-default. *Journal of Banking and Finance* 34: 2510–17. [[CrossRef](#)]
- Bris, Arturo, Iwo Welch, and Ning Zhu. 2006. The costs of bankruptcy: Chapter 7 liquidation versus Chapter 11 reorganization. *Journal of Finance* 61: 1253–303. [[CrossRef](#)]
- Bruche, Max, and Carlos Gonzalez-Aguado. 2010. Recovery rates, default probabilities and the credit cycle. *Journal of Banking and Finance* 34: 713–23. [[CrossRef](#)]
- Chava, Sudheer, Catalina Stefanescu, and Stuart Turnbull. 2011. Modeling the loss distribution. *Management Science* 57: 1267–87. [[CrossRef](#)]
- Chernov, Mikhail, Alexander Gorbenko, and Igor Makarov. 2013. CDS auctions. *Review of Financial Studies* 26: 768–805. [[CrossRef](#)]

- Cox, John, Steven Ross, and Mark Rubinstein. 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7: 229–63. [CrossRef]
- Das, Sanjiv, and Paul Hanouna. 2009. Implied recovery. *Journal of Economic Dynamics and Control* 33: 1837–57. [CrossRef]
- Denis, Diane, and Kimberly Rodgers. 2007. Chapter 11: Duration, outcome, and post-reorganization performance. *Journal of Financial and Quantitative Analysis* 42: 101–18. [CrossRef]
- Doshi, Hitesh, Redouane Elkamhi, and Chayawat Ornthanalai. 2018. The term structure of expected recovery rates. *Journal of Financial and Quantitative Analysis* 53: 2619–61. [CrossRef]
- Duffie, Darrell, and Kenneth Singleton. 1999. Modeling term structures of defaultable bonds. *Review of Financial Studies* 12: 687–720. [CrossRef]
- Düllmann, Klaus, and Monika Trapp. 2004. *Systematic Risk in Recovery Rates—An Empirical Analysis of U.S. Corporate Credit Exposures*. Discussion Paper, Series 2: Banking and Financial Supervision, No. 2004,02; Frankfurt am Main: Deutsche Bundesbank Research Centre.
- François, Pascal, and Weiyu Jiang. 2019. Credit value adjustment with market-implied recovery. *Journal of Financial Services Research*. forthcoming.
- Franks, Julian, and Walter Torous. 1994. A comparison of financial restructuring in distressed exchanges and Chapter 11 reorganizations. *Journal of Financial Economics* 35: 349–70. [CrossRef]
- Frye, Jon. 2000. Collateral damage. *RISK* 13: 91–94.
- Gambetti, Paolo, Geneviève Gauthier, and Frédéric Vrans. 2019. Recovery Rates: Uncertainty Certainly Matters. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3366889 (accessed on 15 April 2019).
- Giesecke, Kay, Francis Longstaff, Stephen Schaefer, and Ilya Strebulaev. 2011. Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics* 102: 233–55. [CrossRef]
- Gupta, Sudip, and Rangarajan Sundaram. 2015. Mispricing and arbitrage in CDS auctions. *Journal of Derivatives* 22: 79–91. [CrossRef]
- Jankowitsch, Rainer, Florian Nagler, and Marti Subrahmanyam. 2014. The determinants of recovery rates in the U.S. corporate bond market. *Journal of Financial Economics* 114: 155–77. [CrossRef]
- Jaskowski, Marcin, and Michael McAleer. 2012. *Estimating Implied Recovery Rates from the Term Structure of CDS Spreads*. Working paper. Amsterdam, The Netherlands: Tinbergen Institute.
- Loterman, Gert, Iain Brown, David Martens, Christophe Mues, and Bart Baesens. 2012. Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting* 28: 161–70. [CrossRef]
- Merton, Robert. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29: 449–70.
- Mora, Nada. 2012. What determines creditor recovery rates? *Federal Reserve Bank of Kansas City Economic Review* 97: 79–109.
- Nazemi, Abdolreza, and Frank Fabozzi. 2018. Macroeconomic variable selection for creditor recovery rates. *Journal of Banking and Finance* 89: 14–25. [CrossRef]
- Nazemi, Abdolreza, Konstantin Heidenreich, and Frank Fabozzi. 2018. Improving corporate bond recovery rate prediction using multi-factor support vector regressions. *European Journal of Operational Research* 271: 664–75. [CrossRef]
- Qi, Min, and Xinlei Zhao. 2011. Comparison of modeling methods for loss given default. *Journal of Banking and Finance* 35: 2842–55. [CrossRef]
- Shleifer, Andrei, and Robert Vishny. 1992. Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance* 47: 1343–66. [CrossRef]
- Siao, Jhao-Siang, Ruey-Ching Hwang, and Chih-Kang Chu. 2016. Predicting recovery rates using logistic quantile regression with bounded outcomes. *Quantitative Finance* 16: 777–92. [CrossRef]
- Ünal, Haluk, Madan Dilip, and Levent Güntay. 2003. Pricing the risk of recovery in default with APR violation. *Journal of Banking and Finance* 27: 1001–25. [CrossRef]
- Varma, Praveen, and Richard Cantor. 2005. Determinants of recovery rates on defaulted bonds and loans for North American corporate issuers: 1983–2003. *Journal of Fixed Income* 14: 29–44. [CrossRef]



Article

An Urn-Based Nonparametric Modeling of the Dependence between PD and LGD with an Application to Mortgages

Dan Cheng [†] and Pasquale Cirillo ^{*,†,‡}

Applied Probability Group, Delft Institute of Applied Mathematics (DIAM), Delft University of Technology, 2628 XE Delft, The Netherlands

* Correspondence: P.Cirillo@tudelft.nl; Tel.: +31-152-782-589

† These authors contributed equally to this work.

‡ Current address: E1.260 Building 28, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands.

Received: 30 March 2019; Accepted: 1 July 2019; Published: 7 July 2019

Abstract: We propose an alternative approach to the modeling of the positive dependence between the probability of default and the loss given default in a portfolio of exposures, using a bivariate urn process. The model combines the power of Bayesian nonparametrics and statistical learning, allowing for the elicitation and the exploitation of experts' judgements, and for the constant update of this information over time, every time new data are available. A real-world application on mortgages is described using the Single Family Loan-Level Dataset by Freddie Mac.

Keywords: probability of default; loss given default; wrong-way risk; dependence; urn model

1. Introduction

The ambition of this paper is to present a new way of modeling the empirically-verified positive dependence between the Probability of Default (PD) and the Loss Given Default (LGD) using a Bayesian nonparametric approach, based on urns and beta-Stacy processes (Walker and Muliere 1997). The model is able to learn from the data, and it improves its performances over time, compatibly with the machine/deep learning paradigm. Similarly to the recent construction of Cheng and Cirillo (2018) for recovery times, this learning ability is mainly due to the underlying urn model.

The PD and the LGD are two fundamental quantities in modern credit risk management. The PD of a counterparty indicates the likelihood that such a counterparty defaults, thus not fulfilling its debt obligations. The LGD represents the percent loss. In terms of the notional value of the exposure known as the Exposure-at-Default or EAD, one actually experiences when a counterparty defaults and every possible recovery process is over. Within the Basel framework (BCBS 2000, 2011), a set of international standards developed by the Basel Committee on Banking Supervision (BCBS) to harmonize the banking sector and improve the way banks manage risk, the PD and the LGD are considered pivotal risk parameters for the quantification of the minimum capital requirements for credit risk. In particular, under the so-called Internal Rating-Based (IRB) approaches, both the PD and the LGD are inputs of the main formulas for the computation of the risk-weighted assets (BCBS 2005, 2006).

Surprisingly, in most theoretical models in the literature, and most of all in the formulas suggested by the BCBS, the PD and the LGD are assumed to be independent, even though several empirical studies have shown that there is a non-negligible positive dependence between them (Altman 2006; Altman et al. 2001; Frye 2005; Miu and Ozdemir 2006; Witzany 2011). Borrowing from the terminology developed in the field of credit valuation adjustment (CVA) (Hull 2015), this dependence is often referred to as wrong-way risk (WWR). Simply put, WWR is the risk that the possible loss generated by

a counterparty increases with the deterioration of its creditworthiness. Ignoring this WWR can easily lead to an unreliable estimation of credit risk for a given counterparty (Witzany 2011).

The link between the PD and the LGD is particularly important when dealing with mortgages, as shown by the 2007–2008 financial crisis, which was triggered by an avalanche of defaults in the US subprime market (Eichengreen et al. 2012), with a consequent drop in estate prices, and thus in the recovery rates of the defaulted exposures (Hull 2015). The financial crisis was therefore one of the main drivers for the rising interest in the joint modeling of PD and LGD, something long ignored both in the academia and in the practice, including regulators. In the paper, we show how the model we propose can be used in the field of mortgages with a real-world application.

The first model implicitly dealing with the dependence between the PD and the LGD was Merton's one. In his seminal work, Merton (1974) introduced the first structural model of default, developing what we can consider the Black and Scholes model for credit risk. In that model, the recovery rate RR (with $RR = 1 - LGD$), conditionally on the credit event, follows a lognormal distribution (Resti and Sironi 2007), and it is negatively correlated with the PD. While not consistent with later empirical findings (Altman et al. 2005; Jones et al. 1984) on WWR, Merton's model remained for long time the only model actually dealing with this problem.

Many models have been derived from Merton's original construction, both in the academic and in the industrial literature, think for example of Geske (1977); Kim et al. (1993); Longstaff and Schwartz (1995); Nielsen et al. (2001) and Vasicek (1984). During the 1990s, on the wake of the—at the time—forthcoming financial regulations (Basel I and later Basel II), several new approaches were introduced, like the so-called reduced-form family (Duffie 1998; Duffie and Singleton 1999; Lando 1998) and the VaR (Value-at-Risk) methodology (JP Morgan 1997; Wilde 1997; Wilson 1998). However, and somehow surprisingly, the great majority of these models, while solving other problems of Merton's original contribution—like the fact that default could only happen at predetermined times—often neglected the dependence between PD and LGD, sometimes even explicitly assuming independence, as in the Credit Risk Plus case (Wilde 1997), where LGD is mainly assumed deterministic.

In the new century, given the rising interest of regulators and investors, especially after the 2007 crisis, a lot of empirical research has definitely shown the positive dependence—in most cases, positive correlation, hence linear dependence—between PD and LGD. For example, Frye (2000) proposed a standard one-factor model, becoming a pivotal reference in the modeling of the WWR between PD and LGD. The paper focuses on the PD and the Market LGD of corporate bonds on a firm level, assuming dependence on a common systematic factor, plus some other independent idiosyncratic components to deal with marginal variability.

Witzany (2011) moved forward, proposing a two-factor model on retail data, looking at the overall economic environment as the source of dependence between PD and LGD. Then, for the LGD only, a second component accounts for the conditions of the economy during the workout process. Other meaningful constructions in the common factor framework are Hamerle et al. (2011) and Yao et al. (2017). The latter is also interesting for the literature review, together with Altman et al. (2005).

Using a two-state latent variable construction, Bruche and Gonzalez-Aguado (2010) introduced another valuable model. The time series of the latent variable is referred to as credit cycle, and it is represented by a simple Markov chain. Interestingly, this credit cycle variable proves to be able to better capture the time variation in the joint distribution of the PD and the LGD, with respect to observable macroeconomic factors.

Notwithstanding the importance of the topic, and despite the notable efforts cited above, it seems that there is still a lot to do in the modeling of the PD/LGD dependence, which remains a yet-to-be-developed part of credit risk management; see the discussions in Altman (2006) and Maio (2017), and the references therein.

Our contribution to this open problem is represented by the present paper, which finds its root in the recent literature about the use of urn models in credit risk management (Amerio et al. 2004; Cheng and Cirillo 2018; Cirillo et al. 2010; Peluso et al. 2015), and more in general in the Bayesian modeling

of credit risk—see, for example, [Baesens et al. \(2016\)](#); [Cerchiello and Giudici \(2014\)](#); [Giudici \(2001\)](#); [Giudici et al. \(2003\)](#); [McNeil and Wendin \(2007\)](#).

For the reader’s convenience, we here summarise the main findings of the paper:

- An intuitive bivariate model is proposed for the joint modeling of PD and LGD. The construction exploits the power of Polya urns to generate a Bayesian nonparametric approach to wrong-way risk. The model can be interpreted as a mixture model, following the typical credit risk management classification ([McNeil et al. 2015](#)).
- The proposed model is able to combine prior beliefs with empirical evidence and, exploiting the reinforcement mechanism embedded in Polya urns, it learns, thus improving its performances over time.
- The ability of learning and improving gives the model a machine/deep learning flavour. However, differently from the common machine/deep learning approaches, the behavior of the new model can be controlled and studied in a rigorous way from a probabilistic point of view. In other words, the common “black box” argument ([Knight 2017](#)) of machine/deep learning does not apply.
- The possibility of eliciting an a priori allows for the exploitation of experts’ judgements, which can be extremely useful when dealing with rare events, historical bias and data problems in general ([Cheng and Cirillo 2018](#); [Derbyshire 2017](#); [Shackle 1955](#)).
- The model we propose can only deal with positive dependence. Given the empirical literature, this is not a problem in WWR modeling; however, it is important to be aware of this feature, if other applications are considered.

The paper develops as follows: Section 2 is devoted to the description of the theoretical framework and the introduction of all the necessary probabilistic tools. In Section 3, we briefly describe the Freddie Mac data we then use in Section 4, where we show how to use and the performances of the model, when studying the dependence between PD and LGD for residential US mortgages. Section 5 closes the paper.

2. Model

The main idea of the model we propose was first presented in [Bulla \(2005\)](#), in the field of survival analysis, and it is based on powerful tools of Bayesian nonparametrics, like the beta-Stacy process of [Walker and Muliere \(1997\)](#). Here, we give a different representation in terms of Reinforced Urn Processes (RUP), to bridge towards some recent papers in the credit risk management literature like [Peluso et al. \(2015\)](#) and [Cheng and Cirillo \(2018\)](#).

2.1. The Two-Color RUP

An RUP is a combinatorial stochastic process, first introduced in [Muliere et al. \(2000\)](#). It can be seen as a reinforced random walk over a state space of urns, and depending on how its parameters are specified, it can generate a large number of interesting models. Essential references on the topic are ([Muliere et al. 2000, 2003](#)) and [Fortini and Petrone \(2012\)](#). In this paper, we need to specify an RUP able to generate a discrete beta-Stacy process, a particular random distribution over the space of discrete distributions.

Definition 1 (Discrete beta-Stacy Process ([Walker and Muliere 1997](#))). *A random distribution function F is a beta-Stacy process with jumps at $j \in \mathbb{N}_0$ and parameters $\{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0}$, if there exist mutually independent random variables $\{V_j\}_{j \in \mathbb{N}_0}$, each beta distributed with parameters (α_j, β_j) , such that the random mass assigned by F to $\{j\}$, written $F(\{j\})$, is given by $V_j \prod_{i < j} (1 - V_i)$.*

Consider the set of the natural numbers \mathbb{N}_0 , including zero. On each $j \in \mathbb{N}_0$, place a Polya urn $U(j)$ containing balls of two colors, say blue and red, and reinforcement equal to 1. A Polya urn is the prototype of urn with reinforcement: every time a ball is sampled, its color is recorded, and the

ball is put back into the urn together with an extra ball (i.e., the reinforcement) of the same color (Mahmoud 2008). This mechanism clearly increases the probability of picking the sampled color again in the future. We assume that all urns $U(j)$ contain at least a ball of each color, but their compositions can be actually different. As the only exception, the urn centered on 0, $U(0)$, only contains blue balls, so that red balls cannot be sampled.

Let $n = 0, 1, 2, \dots$ represent time, which we take to be discrete. A two-color reinforced urn process $\{Z_n\}_{n \geq 0}$ is built as follows. Set $Z_0 = 0$ and sample urn $U(0)$: given our assumptions, the only color we can pick is blue. Put the ball back into $U(0)$, add an extra blue ball, and set $Z_1 = 1$. Now, sample $U(1)$: this urn contains both blue and red balls. If the sampled ball is blue, put it back, add an extra blue ball and set $Z_2 = 2$. If the sampled balls is red, put it back in $U(1)$, add an extra red ball and set $Z_2 = 0$. In general, the value of Z_n will be decided by the sampling of the urn visited by Z_{n-1} . If a blue ball is selected at time $n - 1$, then $Z_n = Z_{n-1} + 1$, otherwise $Z_n = 0$. Blue balls thus make the process $\{Z_n\}_{n \geq 0}$ go further in the exploration of the natural numbers, while every red ball makes the process restart from 0.

If one is interested in defining a two-color RUP only on a subset of \mathbb{N}_0 , say $\{0, 1, \dots, k\}$, it is sufficient to set to zero the number of blue balls in all urns $U(k), U(k + 1), U(k + 2)$ and so on. For the rest, the mechanism stays unchanged.

It is not difficult to verify that, if all urns $U(j), j = 1, 2, \dots$, contain a positive number of red balls, the process $\{Z_n\}_{n \geq 0}$ is recurrent, and it will visit 0 infinitely many times for $n \rightarrow \infty$.

A two-color recurrent RUP clearly generates sequences of nonnegative natural numbers, like the following

$$\{Z_0, Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, \dots, Z_{10}, Z_{11}, Z_{12}, \dots, Z_{15}, \dots\} = \{\underbrace{0, 1, 2}_{Block\ 1}, \underbrace{0, 1, 2, 3, \dots, 7}_{Block\ 2}, \underbrace{0, 1, \dots, 4}_{Block\ 3}, \dots\}. \quad (1)$$

These sequences are easily split into blocks, each starting with 0, as in Equation (1). In terms of sampling, those 0s represent the times the process has been reset after the extraction of a red ball from the urn centered on the natural number preceding that 0. Notice that by construction no sequence can contain two contiguous 0s. In the example above, the process is reset to 0 after the sampling of $U(2)$ in $n = 2$, $U(7)$ in $n = 10$, and $U(4)$ in $n = 15$. Muliere et al. (2000) have shown that the blocks—or the 0-blocks in their terminology—generated by a two-color RUP are exchangeable, and that their de Finetti measure is a beta-Stacy process.

Exchangeability means that, if we take the joint distribution of a certain number of blocks, this distribution is immune to a reshuffling of the blocks themselves, i.e., we can permute them, changing their order of appearance, and yet the probability of the sequence containing them will be the same; for instance $P(Block\ 1, Block\ 2, Block\ 3) = P(Block\ 1, Block\ 3, Block\ 2)$ in Equation (1). Notice that the exchangeability of the blocks does not imply the exchangeability of the single states visited; in fact, one can easily check that each block constitutes a Markov chain. This implies that the RUP can be seen as a mixture of Markov chains, while the sequence of visited states is partially exchangeable in the sense of Diaconis and Freedman (1980).

Given their exchangeability, de Finetti’s representation theorem guarantees that there exists a random distribution function F such that, given F , the blocks generated by the two-color RUP are i.i.d. with distribution F . Theorem 3.26 in Muliere et al. (2000) tells us that such a random distribution is a beta-Stacy process with parameters $\{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0}$, where α_j and β_j are the initial numbers of red, respectively blue, balls in urn $U(j)$, prior to any sampling. In other words, $F(0) = 0$ with probability 1 and, for $j \geq 1$, the increment $[F(j) - F(j - 1)]$ has the same distribution as $V_j \prod_{i=1}^{j-1} (1 - V_i)$, where $\{V_j\}$ is a sequence of independent random variables, such that $V_j \sim \text{beta}(\alpha_j, \beta_j)$. For a reader familiar with urn processes, each beta distributed V_j is clearly the result of the corresponding Polya urn $U(j)$ (Mahmoud 2008).

Let B_1, B_2, \dots, B_m be the first m blocks generated by a two-color RUP $\{Z_n\}_{n \geq 0}$. With T_i , we indicate the last state visited by $\{Z_n\}$ within block i . Coming back to Equation (1), we would have $T_1 = 2$,

$T_2 = 7$ and $T_3 = 4$. Since the random variables T_1, \dots, T_m are measurable functions of the exchangeable blocks, they are exchangeable as well, and their de Finetti measure is the same beta-Stacy governing B_1, B_2, \dots, B_m . In what follows, a sequence $\{T_i\}_{i=1}^m$ is called LS (Last State) sequence. In terms of probabilities, for T_1, \dots, T_m , one can easily observe that

$$P[T_1 = j] = \frac{\alpha_j}{\alpha_j + \beta_j} \prod_{i=0}^{j-1} \frac{\beta_i}{\alpha_i + \beta_i}, \tag{2}$$

and, for $m \geq 1$,

$$P[T_{m+1} = j | T_1, T_2, \dots, T_m] = \frac{\alpha_j + r_j}{\alpha_j + \beta_j + r_j + s_j} \prod_{i=0}^{j-1} \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_i + s_i}, \tag{3}$$

while

$$P[T_{m+1} \geq j | T_1, T_2, \dots, T_m] = \prod_{i=0}^j \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_j + s_i}, \tag{4}$$

where $r_j = \sum_{i=1}^m 1_{\{T_i=j\}}$ and $s_j = \sum_{i=1}^m 1_{\{T_i>j\}}$. From Equations (3) and (4), we see that, every time it is reset to 0 creating a new block, an RUP remembers what happened in the past thanks to the Polya reinforcement mechanism of each visited urn. The process thus learns, combining its initial knowledge, as represented by the quantities $\{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0}$, with the additional balls that are introduced in the system, to obtain the predictives in Equations (3) and (4).

Regarding the initial knowledge, notice that, when we choose the quantities α_j and β_j , for $j = 0, 1, 2, \dots$, from a Bayesian point of view, we are eliciting a prior. In fact, by setting

$$\alpha_j = c_j(G(\{j\})) \quad \text{and} \quad \beta_j = c_j \left(1 - \sum_{i=0}^j (G(\{i\})) \right), \quad j \in \mathbb{N}_0, \tag{5}$$

we are just requiring $E[F(\{j\})] = G(\{j\})$, so that the beta-Stacy process F —a random distribution on discrete distributions—is centered on the discrete distribution G , which we guess may correctly describe the phenomenon we are modeling. The quantity $c_j \geq 0$ is called strength of belief, and it represents how confident we are in our a priori. Given a constant reinforcement, as the one we are using here (+1 ball of the same color), a $c_j > 1$ reduces the speed of learning of the RUP, making the evidence emerging from sampling less relevant in updating the initial compositions. In other terms, c_j helps in controlling the stickiness of $F(\{j\})$ to $G(\{j\})$. For more details, we refer to (Muliere et al. 2000, 2003).

2.2. Modeling Dependence

Consider a portfolio \mathcal{P} containing m exposures. For $i = 1, \dots, m$, let X_i and Y_i represent the PD, respectively the LGD, of the i -th counterparty, when discretised and transformed into levels, in a way similar to what Cheng and Cirillo (2018) propose in their work. In other terms, one can split the PD and the LGD into $l = 0, \dots, L$ levels, such that for example $l = 0$ indicates a PD or LGD of 0%, $l = 1$ something between 0% and 5%, $l = 2$ a quantity in (5%, 17%], and so on until the last level L . The levels do not need to correspond to equally spaced intervals, and this gives flexibility to the modeling. Clearly, the larger L , the finer the partition we obtain. As we will see in Section 4, convenient ways of defining levels are through quantiles, via rounding and, when available, thanks to experts' judgements.

As observed in Altman et al. (2005), discretisation is a common and useful procedure in risk management, as it reduces the noise in the data. The unavoidable loss of information is more than compensated by the gain in interpretability, if levels are chosen in the correct way. From now on, the bivariate sequence $\{(X_i, Y_i)\}_{i=1}^m$ is therefore our object of interest in studying the dependence between the PD and the LGD. Both X_i and Y_i will take values in $\{0, 1, \dots, L\}$.

Let $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$ be three independent LS sequences generated by three independent two-color RUPs $\{Z_j^A\}_{j \geq 0}$, $\{Z_j^B\}_{j \geq 0}$, $\{Z_j^C\}_{j \geq 0}$. As we know, the sequence $\{A_i\}_{i=1}^m$ is exchangeable, and its de Finetti measure is a beta-Stacy process F_A of parameters $\{\alpha_j^A, \beta_j^A\}_{j \in \mathbb{N}_0}$. Similarly, for $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, we have F_B with $\{\alpha_j^B, \beta_j^B\}_{j \in \mathbb{N}_0}$, and F_C with $\{\alpha_j^C, \beta_j^C\}_{j \in \mathbb{N}_0}$. Now, as in Bulla et al. (2007), let us assume that, for each exposure $i = 1, \dots, m$, we have

$$\begin{aligned} X_i &= A_i + B_i, \\ Y_i &= A_i + C_i. \end{aligned} \tag{6}$$

This construction builds a special dependence between the discretised PD and the discretised LGD: we are indeed assuming that, for each counterparty i , there exists a common factor A_i influencing both, while B_i and C_i can be seen as idiosyncratic components. Observe that, conditionally on A_i , X_i and Y_i are clearly independent. Since both X and Y are between 0 and 100%¹, the compositions of the urns defining the processes $\{Z_j^A\}_{j \geq 0}$, $\{Z_j^B\}_{j \geq 0}$, $\{Z_j^C\}_{j \geq 0}$ can be tuned so that it is not possible to observe values larger than 100%.

From Equation (6), we can derive important features of the model we are proposing. Since we can write $Y_i = X_i - B_i + C_i$, it is clear that we are assuming a linear dependence between X_i and Y_i . This is compatible with several empirical findings, like Altman et al. (2005) or Miu and Ozdemir (2006).

Furthermore, given Equation (6) and the properties of the sequences $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, we can immediately observe that

$$\begin{aligned} \text{Cov}(X_1, Y_1) &= \text{Var}(A_1) \geq 0, \\ \text{Cov}(X_{m+1}, Y_{m+1} | \mathbf{A}_m, \mathbf{B}_m, \mathbf{C}_m) &= \text{Var}(A_{m+1} | \mathbf{A}_m) \geq 0, \text{ for } m \geq 2, \end{aligned} \tag{7}$$

where Var is the variance, Cov the covariance, $\mathbf{A}_m = [A_1, \dots, A_m]$, and similarly $\mathbf{B}_m, \mathbf{C}_m$. Therefore, with the bivariate urn construction, we can only model positive dependence. Again, this is totally in line with the purpose of our analysis—the study of wrong-way risk that, by definition, is a positive dependence between PD and LGD—and, with the empirical literature, as discussed in Section 1. However, it is important to stress that the model in Equation (6) cannot be used when negative dependence is possible.

Always from Equation (6), we can verify that the sequence $\{(X_i, Y_i)\}_{i=1}^m$ is exchangeable². This comes directly from the exchangeability of $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$ and the fact that (X_i, Y_i) is a measurable function of (A_i, B_i, C_i) . An implicit assumption of our model is therefore that the m counterparties in \mathcal{P} are exchangeable. As observed in McNeil et al. (2015), exchangeability is a common assumption in credit risk, for it is seen as a relaxation of the stronger hypothesis of independence (think about Bernoulli mixtures and the beta-binomial model). All in all, what we ask is that the order in which we observe our counterparties is irrelevant to study the joint distribution of their PDs and LGDs, which is therefore immune to changes in the order of appearance of each exposure. Exchangeability and the fact that X_i and Y_i are conditionally independent given A_i suggest that the methodology we are proposing falls under the larger umbrella of mixture models (Duffie and Singleton 2003; McNeil et al. 2015).

Since $\{(X_i, Y_i)\}_{i=1}^m$ is exchangeable, de Finetti’s representation theorem guarantees the existence of a bivariate random distribution F_{XY} , conditionally on which the couples are i.i.d. with distribution F_{XY} . The properties of F_{XY} have been studied in detail in Bulla (2005).

¹ In reality, as observed in Zhang and Thomas (2012), the LGD can be slightly negative or slightly above 100% because of fees and interests; however, we exclude that situation here. In terms of applications, all negative values can be set to 0, and all values above 100 can be rounded to 100.

² Please observe that exchangeability only applies among the couples $\{(X_i, Y_i)\}_{i=1}^m$, while within each couple there is a clear dependence, so that X_i and Y_i are not exchangeable.

Let F_X and F_Y be the marginal distributions of X_i and Y_i . Clearly, we have

$$\begin{aligned} F_X &= F_A \times F_B, \\ F_Y &= F_A \times F_C, \end{aligned}$$

so that both F_X and F_Y are convolutions of beta-Stacy processes. The dependence between X and Y , given F_{XY} and F_A is thus simply

$$\text{Cov}_{F_{XY}}(X, Y) = \text{Var}_{F_A}(A) = \sigma_A^2. \tag{8}$$

Furthermore, if P is the probability function corresponding to F , one has

$$P_{XY}(x, y) = \sum_{a=0}^{\min(x,y)} P_A(a)P_B(x-a)P_C(y-a), \quad \forall x, y \in \mathbb{N}_0^2.$$

Assume now that we have observed m exposures, and we have registered their actual PD and LGD, which we have discretised to get $\{(X_i, Y_i)\}_{i=1}^m$. The construction of Equation (6), together with the properties of the beta-Stacy processes involved, allows for a nice derivation of the predictive distribution for a new exposure (X_{m+1}, Y_{m+1}) , given the observed couples $(\mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m)$. This can be extremely useful in applications, when one is interested in making inference about the PD, the LGD and their relation.

In fact,

$$P[X_{m+1} = x, Y_{m+1} = y | \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] = \frac{P[X_{m+1} = x, Y_{m+1} = y, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]}{P[\mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]}. \tag{9}$$

Given Equation (6), Equation (9) can be rewritten as follows:

$$\begin{aligned} &P[X_{m+1} = x, Y_{m+1} = y | \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] \\ &= \sum_{\mathbf{a}_m} P[X_{m+1} = x, Y_{m+1} = y | \mathbf{A}_m = \mathbf{a}_m, \mathbf{B}_m = \mathbf{b}_m, \mathbf{C}_m = \mathbf{c}_m] \\ &\quad \times P[\mathbf{A}_m = \mathbf{a}_m | \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m], \end{aligned} \tag{10}$$

where $\mathbf{b}_m = \mathbf{x}_m - \mathbf{a}_m$ and $\mathbf{c}_m = \mathbf{y}_m - \mathbf{a}_m$.

From a theoretical point of view, computing Equation (10) just requires counting the balls in the urns behind $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, and then to use formulas like those in Equations (2) and (3), something that for a small portfolio can be done explicitly. However, when m is large, it becomes numerically unfeasible to perform all those sums and products.

Luckily, developing an alternative Markov Chain Monte Carlo algorithm is simple and effective. It is sufficient to go through the following steps:

- (1) Given the observations $\mathbf{X}_m = \mathbf{x}_m$ and $\mathbf{Y}_m = \mathbf{y}_m$, the sequence $\mathbf{A}_m = (A_1, \dots, A_m)$ is generated via a Gibbs sampling. The full conditional of A_m , $P[A_m = a_m | \mathbf{A}_{m-1} = \mathbf{a}_{m-1}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]$, is such that

$$\begin{aligned} P[A_m = a_m | \mathbf{A}_{m-1} = \mathbf{a}_{m-1}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] &\propto P[A_m = a_m | \mathbf{A}_{m-1} = \mathbf{a}_{m-1}] \\ &\quad \times P[X_m - A_m = x_m - a_m | \mathbf{B}_{m-1} = \mathbf{b}_{m-1}] \\ &\quad \times P[Y_m - A_m = y_m - a_m | \mathbf{C}_{m-1} = \mathbf{c}_{m-1}]. \end{aligned}$$

Since $\{A_j\}_{j=1}^m$ is exchangeable, all the other full conditionals, $P[A_j = a_j | \mathbf{A}_{-j} = \mathbf{a}_{-j}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]$, where $\mathbf{A}_{-j} = (A_1, \dots, A_{j-1}, A_{j+1}, \dots, A_m)$, have an analogous form.

- (2) Once \mathbf{A}_m is obtained, compute $\mathbf{B}_m = \mathbf{X}_m - \mathbf{A}_m$ and $\mathbf{C}_m = \mathbf{Y}_m - \mathbf{A}_m$.

- (3) The quantities A_{m+1} , B_{m+1} , and C_{m+1} are then sampled according to their beta-Stacy predictive distributions $P(A_{m+1} | \mathbf{A}_m)$, $P(B_{m+1} | \mathbf{B}_m)$, and $P(C_{m+1} | \mathbf{C}_m)$ as per Equation (3).
- (4) Finally, set $X_{m+1} = A_{m+1} + B_{m+1}$ and $Y_{m+1} = A_{m+1} + C_{m+1}$.

3. Data

We want to show the potentialities of the bivariate urn construction of Section 2 in modeling the dependence between PD and LGD in a large portfolio of residential mortgages. The data we use come from [Maio \(2017\)](#).

Maio's dataset is the result of cleaning operations (treatment of NA's, inconsistent data, etc.) on the well-known Single Family Loan-Level Dataset Sample by Freddie Mac, freely available online ([Freddie Mac 2019a](#)). Freddie Mac's sample contains 50,000 observations per year over the period 1999–2017. Observations are randomly selected, on a yearly basis, from the much larger Single Family Loan-Level Dataset, covering approximately 26.6 million fixed-rate mortgages. Extensive documentation about Freddie Mac's data collections can be found in [Freddie Mac \(2019b\)](#).

Maio's dataset contains 383,465 loans over the period 2002–2016. Each loan is uniquely identified by an alphanumeric code, which can be used to match the data with the original Freddie Mac's source.

For each loan, several interesting pieces of information are available, like its origination date, the loan age in months, the geographical location (ZIP code) within the US, the FICO score of the subscriber, the presence of some form of insurance, the loan to value, the combined loan to value, the debt-to-income ratio, and many others. In terms of credit performance, quantities like the unpaid principal balance and the delinquency status up to termination date are known. Clearly, termination can be due to several reasons, from voluntary prepayment to foreclosure, and this information is also recorded, following [Freddie Mac \(2019b\)](#). A loan is considered defaulted when it is delinquent for more than 180 days, even if it is later repurchased ([Freddie Mac 2019c](#)). In addition to the information also available in the Freddie Mac's collection, Maio's dataset is enriched with estimates of the PD and the LGD for each loan, obtained via survival analysis³. It is worth noticing that both the PD and the LGD are always contained in the interval $[0, 1]$.

If one considers the pooled data, the correlation between the PD and the LGD is 0.2556. The average PD is 0.0121 with a standard deviation of 0.0157, while, for LGD, we have 0.1494 and 0.0937, respectively. Regarding the minima and the maxima, we have 2.14×10^{-5} and 0.3723 for PD, and 0 and 0.5361 for LGD.

In the parametric approach proposed by [Maio \(2017\)](#), the covariates which affect both the PD and LGD, possibly justifying their positive dependence, are the unpaid principal balance (UPB) and the debt-to-income ratio (DTI). Other covariates, from the age of the loan to the ZIP code, are then relevant in explaining the marginal behaviour of either the PD or the LGD. In modeling both the PD and the LGD, [Maio \(2017\)](#) proposes the use of two Weibull accelerated failure time (AFT) models, following a recent trend in modern credit risk management ([Narain 1992](#)). The dependence between PD and LGD is then modelled parametrically using copulas and a brand new approach involving a bivariate beta distribution. For more details, we refer to [Maio \(2017\)](#).

A correlation around 0.26 clearly indicates a positive dependence between PD and LGD, and it is in line with the empirical literature we have mentioned in Section 1. However, considering all mortgages together may not be the correct approach, as we are pooling together very different counterparties, possibly watering down more meaningful areas of dependence.

Given the richness of the dataset, there are many ways of disaggregating the data. For example, one can compute the correlation between PD and LGD for different FICO score classes. The FICO

³ To avoid any copyright problem with Freddie Mac, which already freely shares its data online, from Maio's dataset (here attached), we only provide the PD and the LGD estimates, together with the unique alphanumeric identifier. In this way, merging the data sources is straightforward.

score, originally developed by the Fair Isaac Corporation (<https://www.fico.com>), is a leading credit score in the US, and one of the significant covariates for the estimation of PD in Maio’s survival model (Maio 2017). In the original Freddie Mac’s sample, it ranges from 301 to 850.

Following a common classification (Experian 2019), we can define five classes of creditworthiness: Very poor, for a FICO below 579; Fair, for 580–669; Good, for 670–739; Very good, for 740–799; and Exceptional, with a score above 800. Table 1 contains the number of loans in the different classes, the corresponding average PD and LGD, and naturally their correlation ρ . As expected, the average PD is higher when the FICO score is lower, while the opposite is observable for the average LGD. This is probably due to the fact that, for less creditworthy counterparties, stronger insurances are generally required, as compensation for the higher risk of default. Moreover, in terms of recovery, in putting the collateral on the market, it is probably easier to sell a house with a lower value than a very expensive property, for which discounts on the price are quite common (Eichengreen et al. 2012). Interestingly, in disaggregating the data, we see that the PD–LGD correlation is always above 0.3, with the only exception being the most reliable FICO class (≈ 0.19).

Table 1. Some descriptive information about the data used in the analysis. Loans are collected in terms of FICO score.

Class	Number of Loans	Avg. PD	Avg. LGD	ρ
Very Poor	1627	0.0378	0.1013	0.3370
Fair	46,720	0.0238	0.1237	0.4346
Good	124,824	0.0138	0.1409	0.3159
Very good	177,891	0.0083	0.1574	0.3599
Exceptional	32,403	0.0080	0.1777	0.1858

As an example, Figure 1 shows two plots of the relation between PD and LGD for the “Very poor” class. On the left, a simple scatter plot of PD vs. LGD. On the right, to deal with the large number of overlapping points, thus improving interpretability, we provide a hexagonal heatmap with counts. To obtain this plot, the plane is divided into regular hexagons (20 for each dimension), the number of cases in each hexagon is counted, and it is then mapped to a color scale. This second plot tells us that most of the PD–LGD couples lie in the square $[0, 0.1]^2$.

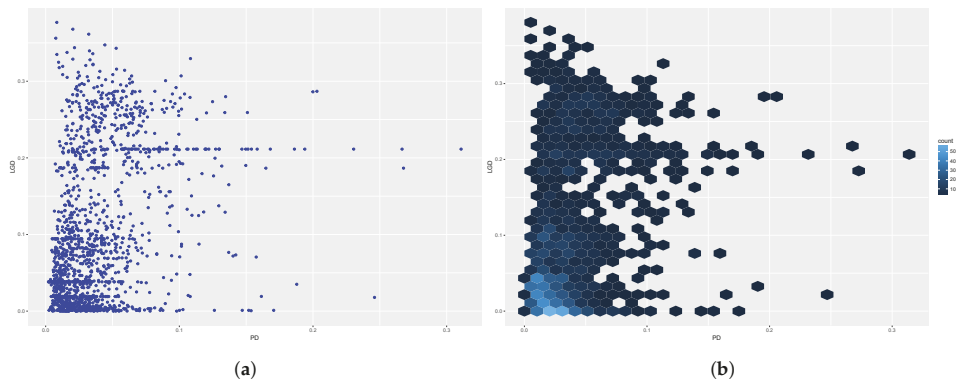


Figure 1. Plots of PD vs. LGD for mortgages in the “Very poor” (FICO score below 579) class—on the left, a simple scatter plot; on the right, a hexagonal heatmap with counts. (a) scatter plot; (b) hexagonal heatmap.

In Figure 2, the two histograms of the marginal distributions of PD and LGD in the “Very poor” rating class are shown. While for PD the distribution is unimodal, for LGD, we can clearly see bimodality (a second bump is visible around 0.25). These behaviours are present among the different classes and at the pooled level.

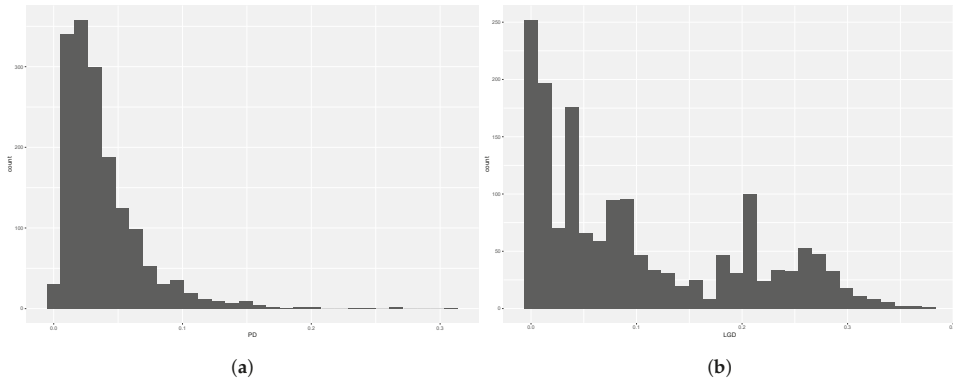


Figure 2. Histograms of the marginal distributions of PD and LGD in the “Very poor” rating class. (a) PD; (b) LGD.

4. Results

In this section, we discuss the performances of the bivariate urn model on the mortgage data described in Section 3. For the sake of space, we show the results for “Very poor” and the “Exceptional” FICO score classes, as per Table 1.

In order to use the model, we need (1) to transform and discretise both the PD and the LGD into levels, and (2) to define an a priori for the different beta-Stacy processes involved in the construction of Equation (6).

The results we obtain are promising and suggest that the bivariate urn model can represent an interesting way of modeling PD and LGD dependence for banks and practitioners.

Please notice that our purpose is to show that the bivariate urn model actually works. The present section has no ambition of being a complete empirical study on the PD–LGD dependence in the Freddie Mac’s or Maio’s datasets.

4.1. Discretisation

In order to discretise the PD and LGD into X and Y , it is necessary to choose the appropriate levels $l = 0, 1, \dots, L$. In the absence of specific ranges, possibly arising from a bank’s business practice or imposed by a regulator, a convenient way for defining the levels is through quantiles (Cheng and Cirillo 2018).

For example, let d_1^v, \dots, d_9^v be the deciles for the quantity $v \in \{PD, LGD\}$. We can set levels $l = 0, 1, \dots, 10$, where

$$L1 := \{0 : 0; 1 : (0, d_1^v]; 2 : (d_1^v, d_2^v]; \dots; 9 : (d_8^v, d_9^v]; 10 : > d_9^v\}. \tag{11}$$

Such a partition guarantees that each level, apart from $l = 0$, contains 10% of the values for both PD and LGD. Notice that the thresholds are not the same, for example, when focusing on the “Very poor” class, $d_1^{PD} = 0.0099$, while $d_1^{LGD} = 0.0030$. A finer partition can be obtained by choosing other percentiles. Clearly, in using a similar approach, one should remember that she is imposing a uniform behaviour on X and Y , in a way similar to copulas (Nelsen 2006). However, differently from copulas, the dependence between X and Y is not restricted to any particular parametric form (the copula function): dependence will emerge from the combination of the a priori and the data.

Another simple way for defining levels is to round the raw observations to the nearest largest integer (ceiling) or to some other value. For instance, we can consider:

$$L2 := \{0 : 0; 1 : (0, 1\%]; 2 : (1, 2\%]; \dots; 100 : (99, 100\%]\}. \tag{12}$$

Even if it is not a strong requirement, given the meaning of the value 0 in an RUP (recall the 0-blocks), we recommend to use 0 as a special level, not mapping to an interval. Moreover, as already observed, equally-spaced intervals are easier to implement and often to interpret, but again this is not a necessity: levels can represent intervals of different sizes.

Notice that, if correctly applied, discretisation maintains the dependence structure between the variables. For the “Very poor” class, using the levels in Equation (11) in defining X and Y , we find that $cor(X, Y) = 0.3348$, in line with the value 0.3370 in Table 1. For the “Exceptional” group, using the levels in Equation (12), we obtain 0.2080, still comparable.

In what follows, we discuss the results mainly using the levels in Equation (12). However, our findings are robust to different choices of the partition. In general, choosing a smaller number of levels improves fitting because the number of observations per level increases, reinforcing the Polya learning process, *ceteris paribus*. Moreover—and this is why partitions based on quantiles give nice performances—better results are obtained when intervals guarantee more or less the same number of observations per level.

Choosing a very large L may lead to the situation in which, for a specific level, no transition is observed, so that, for that level, no learning via reinforcement is possible, and, if our beliefs are wrong—or not meant to compensate some lack of information in the data—this naturally has an impact on the goodness of fit. Therefore, in choosing L , there is a trade-off between fitting and precision. The higher L , the lower the impact of discretisation and noise reduction.

The right way to define the number of levels is therefore to look for a compromise between precision, as required by internal procedures or the regulator, and the quantity and the quality of the empirical data. The more and the better the observations, the more precise the partition can be because the higher is the chance of not observing empty levels. In any case, each level should guarantee a minimum number of observations in order to fully exploit the Bayesian learning mechanism. To improve fitting, rarely visited levels should be aggregated. Once again, experts’ judgements could represent a possible solution. We refer to Cheng and Cirillo (2018) for further discussion on level setting.

4.2. Prior Elicitation

In order to use the bivariate urn model, it is necessary to elicit an a priori for all its components, namely $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$. This is fundamental for any computation and to initialise the Markov Chain Monte Carlo algorithm discussed in Section 2.2.

Prior elicitation can be performed (1) in a completely subjective way, on the basis of one’s knowledge of the phenomenon under scrutiny, (2) by just looking at the data in a fully empirical approach, or (3) by combining data and beliefs, as suggested for example by Figini and Giudici (2011). In reality, a fully data-driven approach based on the sole use of the empirical distribution functions, as for example the one suggested in Cheng and Cirillo (2018), is not advisable for the bivariate urn construction. While there is no problem in observing X and Y , thus exploiting their empirical distributions for prior elicitation, it is not immediate to do the same for the unobserved quantity A , necessary to obtain both B and C . A possibility could be to use the empirical Kendall’s function (Genest and Rivest 1993), but then one would end up with a model not different from the standard empirical copula approach (Rueschendorf 2009), especially if the number of observations used for the definition of the a priori is large. A compromise could thus be to elicit a subjective prior for A only, and to combine this information with the empirical distributions of X and Y , in order to obtain B and C . As common in Bayesian nonparametrics (Hjort et al. 2010), no unique best way exists: everything depends on personal preferences—this is the unavoidable subjectivity of every statistical model (Galavotti 2001)—and on exogenous constraints that, in credit risk management, are usually represented by the actual regulatory framework (BCBS 2000; Hull 2015).

Recalling Equation (8), it seems natural to choose the prior for $\{A_i\}_{i=1}^m$ so that its variance coincides with the empirical covariance between X and Y . When looking at the “Very poor” and “Exceptional” classes, this covariance is approximately 3 (i.e., 2.7345 and 3.2066, respectively), using the levels in Equation (11). For the situation in Equation (12), conversely, we have a covariance of 10.28

for the “Very poor” class, and 1.46 for the other. Apart from the reasonable constraint on the variance, the choice of the distribution for A can be completely free.

Given the ranges of variation of X and Y , one can choose the appropriate supports for the three beta-Stacy processes. Considering the levels in Equation (11), for $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, it is not rational to choose priors putting a positive mass above 10. Being the levels defined via deciles, it is impossible to observe a level equal to 11 or more. Similarly, using the levels in Equation (12) and recalling Figure 1, where neither the PD nor the LGD reach values above 40% (0.4), once can decide not to allow large values of X and Y , initialising the RUP’s urns only until level 40.

Since we do not have any specific experts’ knowledge to exploit, we have tried different prior combinations. Here, we discuss two possibilities:

- Independent discrete uniforms for $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, where the range of variation for B and C is simply inherited from X and Y (but extra conditions can be applied, if needed), while for A the range is chosen to guarantee $\sigma_A^2 = Cov(X, Y)$. For instance, if the covariance between X and Y is approximately 3, the interval $[0, 5]$ guarantees that $\sigma_A^2 \approx 3$ as well. We can simply use the formula for the variance of a discrete uniform, i.e.,

$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12}.$$

- Independent Poisson distributions, such that $A \sim Poi(\lambda_A = Cov(X, Y))$, while for B and C one sets $Poi(\bar{X} - \lambda_A)$ and $Poi(\bar{Y} - \lambda_A)$, where \bar{X} is the empirical mean of X . This guarantees, for example, that $X \sim Poi(\bar{X})$. Recall in fact that, in a Poisson random variable, the mean and the variance are both equal to the intensity parameter, and independent Poissons are closed under convolution. Given our data, where the empirical variances of X and Y are not at all equal to the empirical means, but definitely larger, the Poisson prior can be seen as an example of a wrong prior.

Notice that the possibility of eliciting an a priori is an extremely useful feature of every urn construction (Amerio et al. 2004; Cheng and Cirillo 2018; Cirillo et al. 2010). In fact, a good prior—when available—can compensate for the lack of information in the data, and it can effectively deal with extremes and rare events, a relevant problem in modern risk management (Calabrese and Giudici 2015; Hull 2015; Taleb 2007). For example, if an expert believes that her data under-represent a given phenomenon, like for example some unusual combinations of PD and LGD, she could easily solve the problem by choosing a prior putting a relevant mass on those combinations, so that the posterior distribution will always take into account the possibility of those events, at least remotely. This can clearly correct for the common problem of historical bias (Derbyshire 2017; Shackle 1955).

Once the priors have been decided, the urn compositions can be derived via Equation (5). Different values of c_j have been tried in our experiments. Here, we discuss the cases $c_j = 1$ and $c_j = 100$ for all values of j . The former indicates a moderate trust in our a priori, while the latter shows a strong confidence in our beliefs. Cheng and Cirillo (2018) have observed that, in constructions involving the use of reinforced urn processes, if the number of observations used to train the model is large, then priors become asymptotically irrelevant, for the empirical data prevail. However, when the number of data points is not very large, having a strong prior does make a difference. Given the set cardinalities, we shall see that a strong prior has a clear impact for the “Very poor” rating class (“only” 1627 observations), while no appreciable effect is observable for the “Exceptional” one (32,403 data points).

As a final remark, it is worth noticing that one could take all the urns behind $\{A_i\}_{i=1}^m$ to be empty. This would correspond to assuming that no dependence is actually possible between PD and LGD, so that $X_i = B_i$, $Y_i = C_i$ and $Cov(X, Y) = 0$. As discussed in Bulla et al. (2007), it can be shown that, when A is degenerate on 0—and no learning on dependence is thus possible, given the infringement of Cromwell’s rule (Lindley 1991)—the bivariate urn model simply corresponds to playing with the bivariate empirical distribution and the bivariate Kaplan–Meier estimator. While certainly not useful

to model a dependence we know is present, this possibility further props the flexibility of the bivariate urn model up.

4.3. Fitting

Figure 3 shows, for the exposures in the “Very Poor” group, the very good fitting performances of the model for the marginal distributions of X and Y , the discretised PD and LGD, respectively. Each subfigure shows the elicited prior, the empirical cumulative distribution function (ECDF) and the posterior, as obtained via learning and reinforcement. In the figure shown, the levels are given by Equation (12), while the priors are discrete uniforms with $c_j = 1$. Since the covariance between X and Y is 10.28, the discrete uniform on A is on $[0, 10]$. For both B and C , the support is $[0, 40]$. A two sample Kolmogorov–Smirnov (KS) test does not reject the null hypothesis of the same distribution between the ECDFs and the relative posteriors (p -values stably above 0.05).

It is worth noticing that, qualitatively, the results we discuss here are robust to different choices of the levels, until a sufficient number of observations is available for the update of the urns and hence learning, as already observed at the end of Section 4.1. For example, we have tested the partitions $\{0 : 0; 1 : (0, 2\%); 2 : (2, 4\%); \dots; 50 : (98, 100\%)\}$ and $\{0 : 0; 1 : (0, 5\%); 2 : (5, 10\%); \dots; 20 : (95, 100\%)\}$: all findings were consistent with what we see in these pages. An example of partition that does not work is conversely the one based on increments equal to 0.1% per level, which proves to be too fine, so that several urns are never updated and the posteriors do not pass the relative KS tests.

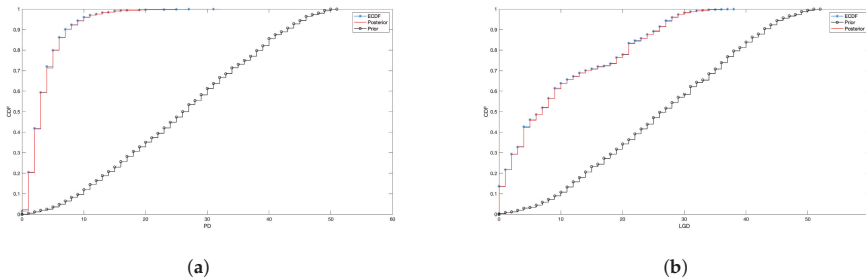


Figure 3. Prior, posterior and ECDF for the discretised PD and LGD in the “Very poor” rating class, when priors are uniform and levels are like those in Equation (12). The strength of belief is always set to 1. (a) PD; (b) LGD.

Figure 4 is generated in the same way as Figure 3. The only difference is that now $c_j = 100$, indicating a strong belief on the uniform priors. In this case, it is more difficult for the bivariate urn process to update the prior, and in fact the posterior distribution is not as close as before to the ECDF. The effect is more visible for PD than for LGD; however, in both cases, the KS test rejects the null (p -value 0.005 and 0.038). Please notice that this is not necessarily a problem, if one really believes that the available data do not contain all the necessary information, thus correcting for historical bias, or she wants to incorporate specific knowledge about future trends.

As anticipated, the number of observations available in the data plays a major role in updating—and, in case of a wrongly elicited belief, correcting—the prior. Focusing on the PD, Figure 5 shows that, when the “Exceptional” rating class is considered, no big difference can be observed in the obtained posterior, no matter the strength of belief c_j . In fact, for both cases, a KS test does not reject the null hypothesis with respect to the ECDF (p -values: 0.74 and 0.58). The reason is simple: with more than 30,000 observations, the model necessarily converges towards the ECDF, even if the prior distribution is clearly wrong and we put a strong belief on it (one needs $c_j > 500$ to see some difference). Similar results hold for the LGD. In producing the figure, A has a uniform prior on $[0, 3]$, while B and C on $[0, 45]$.

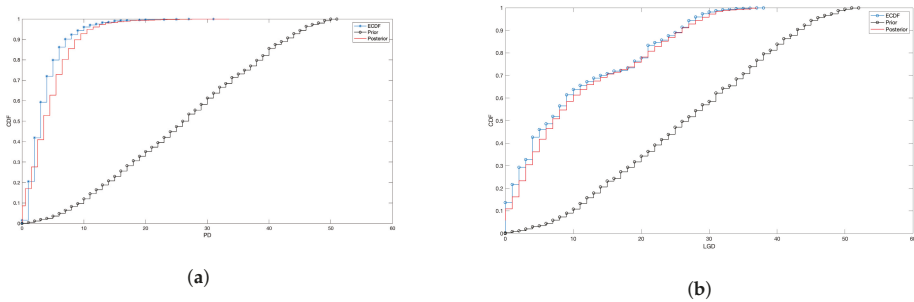


Figure 4. Prior, posterior and ECDF for the discretised PD and LGD in the “Very poor” rating class, when priors are uniform and levels are like those in Equation (12). The strength of belief is always set to 100. (a) PD; (b) LGD.

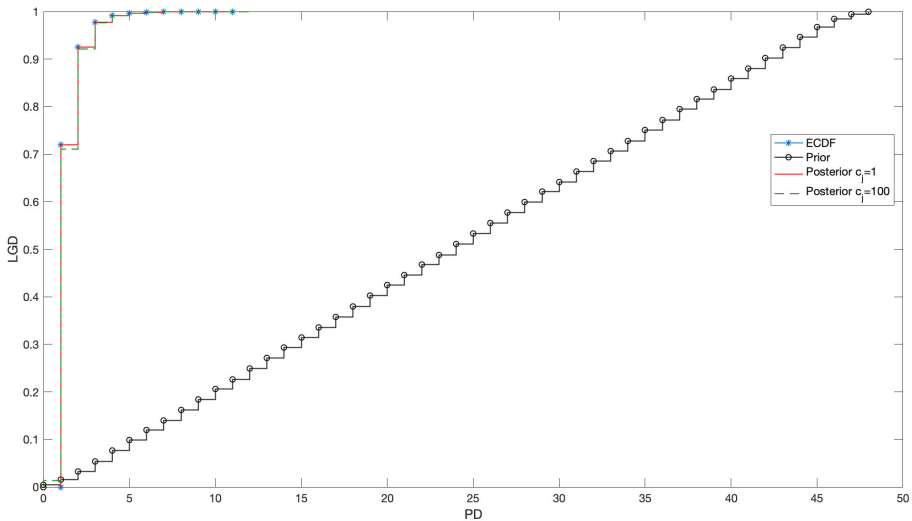


Figure 5. Prior, posteriors ($c_j = 1$ and $c_j = 100$) and ECDF for the discretised PD in the “Exceptional” rating class, when priors are uniform.

Figure 6 shows the bivariate distribution we obtain for the discretised PD and LGD for the “Very poor” FICO score group, when $c_j = 1$ and we use the Poisson priors on the levels of Equation (12). Figure 7 shows the case $c_j = 100$. As one would expect, the strong prior provides a smoother joint distribution, while the weak one tends to make the empirical data prevail, with more peaks, bumps and holes. In Figure 8, the equivalent of Figure 6 is given for the “Exceptional” rating class.

Regarding the numbers, all priors and level settings are able to model the dependence between X and Y . The correlation is properly captured (the way in which the prior on A is defined surely helps), as well as the mean and the variances of the marginals, especially when the strength of beliefs is small. The only problem is represented by the Poisson priors used on the “Very poor” class. In this case, the number of observations is not sufficient to correct the error induced by the initial use of a Poisson distribution, i.e., the same value for mean and variance, and the variance of both PD and LGD is underestimated, while the mean is correctly captured. In particular, while the estimated and actual means are 3.81 and 3.79 for X , and 10.61 and 10.12 for Y , in the case of the variances, the actual values 10.26 and 92.97 are definitely larger than the predicted ones, i.e., 5.88 and 22.50. This is a clear signal that more data would be necessary to properly move away from the wrong prior beliefs, forgetting the Poisson nature.

In the case of Figure 6, the actual correlation is 0.3370, and the model estimates 0.3366. However, under Poisson priors with strong degrees of belief, a certain overestimation is observable for the “Very poor” rating group, in line with the underestimation of the variances.

The results we have just commented are all in sample: we have indeed used all the observations available to verify how the model fits the data, and no out of sample performance was checked. Using the Bayesian terminology of Jackman (2009) and Meng (1994), we have therefore performed a posterior consistency check.

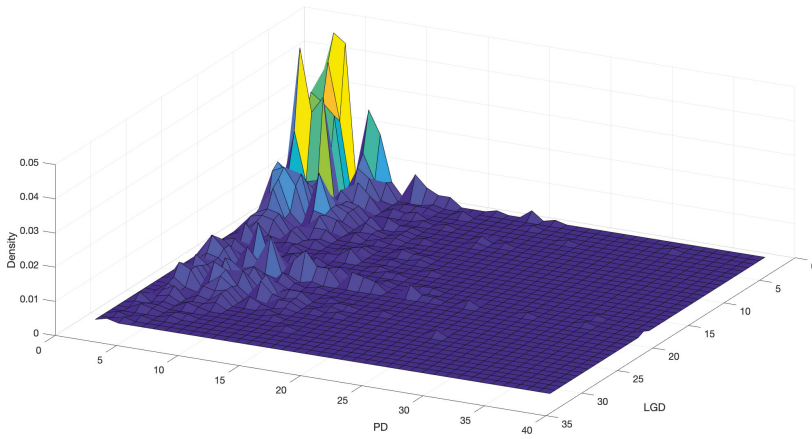


Figure 6. Bivariate density distribution of PD and LGD in the “Very poor” rating class, with Poisson priors and strength of belief equal to 1.

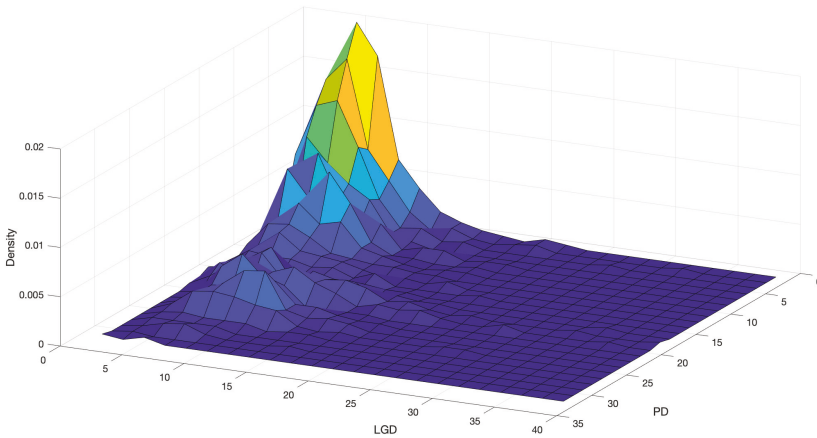


Figure 7. Bivariate density distribution of PD and LGD in the “Very poor” rating class, with Poisson priors and strength of belief equal to 100.

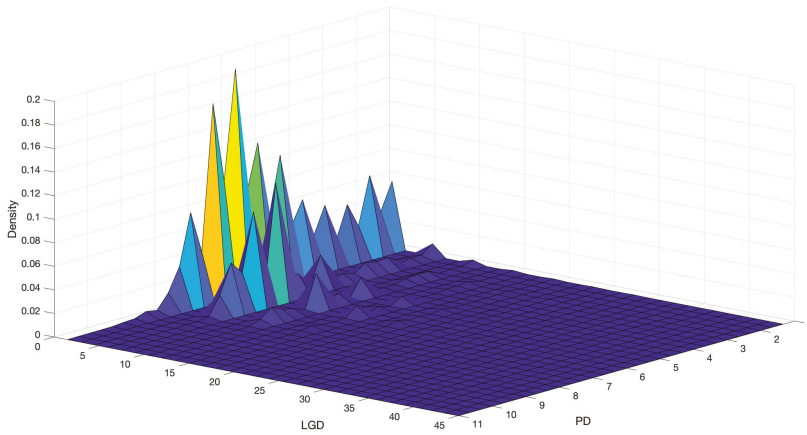


Figure 8. Bivariate density distribution of PD and LGD in the “Exceptional” rating class, with Poisson priors and strength of belief equal to 1.

Luckily, the out of sample validation of the bivariate urn model is equally satisfactory. To perform it, we have used the Freddie Mac’s sampling year to create two samples. The first one includes all the loans (362,104) sampled in the period 2002–2015 and we call it the training sample. The validation sample, conversely, includes all the loans sampled in 2016, for a total of 21,321 data points. The samples have then been split into FICO score groups, as before.

For the “Very Poor” class, Figure 9 shows the predictive distribution of the PD as obtained by training the bivariate urn construction (uniform priors) on the training sample (1590 loans) against the ECDF of the corresponding validation sample (37 loans). The fit is definitely acceptable, especially if we consider the difference in the sample sizes. The mean (level) of the predictive distribution is 3.85, while that of the validation sample is 3.81. The standard deviations are 3.24 and 2.53. The median 3 in both cases. Regarding dependence, a correlation of 0.3419 is predicted by the model, while the realised one in the validation sample is smaller and equal to 0.1963. This difference is probably due to the fact that in the validation set the PD never exceeds level 13, while in the training one the maximum level reached is 31. When the validation set is larger, as in the case of the “Exceptional” class, the dependence is better predicted: one has a correlation of 0.1855 against 0.1799. Similar or better results are obtained for the other classes, using the different prior sets, once again with the only exception of the Poisson priors with a large strength of belief, when applied to the “Very poor” class. In Table 2, we show the PD and correlation results for all classes under uniform priors, while Table 3 deals with the LGD.

Table 2. PD values (%) and correlation. Some descriptive descriptive statistics (mean, median, standard deviation, correlation for the joint) for the predictive distribution (P) and the validation set (V) for the different FICO classes, under uniform priors.

Class	$mean_P$	$mean_V$	$median_P$	$median_V$	SD_P	SD_V	ρ_P	ρ_V
Very Poor	3.85	3.81	3.02	3.03	3.24	2.53	0.34	0.20
Fair	2.41	2.32	1.50	1.44	2.36	2.33	0.45	0.51
Good	1.43	0.49	0.83	0.25	0.71	0.66	0.34	0.38
Very good	0.36	0.32	0.62	0.64	0.14	0.16	0.38	0.42
Exceptional	0.16	0.18	0.60	0.54	0.82	0.83	0.19	0.18

Table 3. LGD values (%). Some descriptive descriptive statistics (mean, median, standard deviation) for the predictive distribution (P) and the validation set (V) for the different FICO classes, under uniform priors.

Class	$mean_P$	$mean_V$	$median_P$	$median_V$	SD_P	SD_V
Very Poor	9.82	10.2	6.63	6.14	9.51	8.32
Fair	12.1	11.5	9.43	10.6	9.69	5.53
Good	13.8	19.8	14.5	19.2	5.70	5.26
Very good	15.6	18.5	18.7	18.0	5.27	5.13
Exceptional	17.8	17.2	20.4	19.7	8.00	4.99

It is important to stress that this simple out of sample validation does not guarantee that the predictive distribution would be able to actually predict new data, in the case of a major change in the underlying phenomenon—something not observed in Maio’s dataset—like a structural break, to use the econometric jargon. This would be exactly the case in which the clever use of the prior distribution—in the form of expert prior knowledge—could contribute in obtaining a better fit.

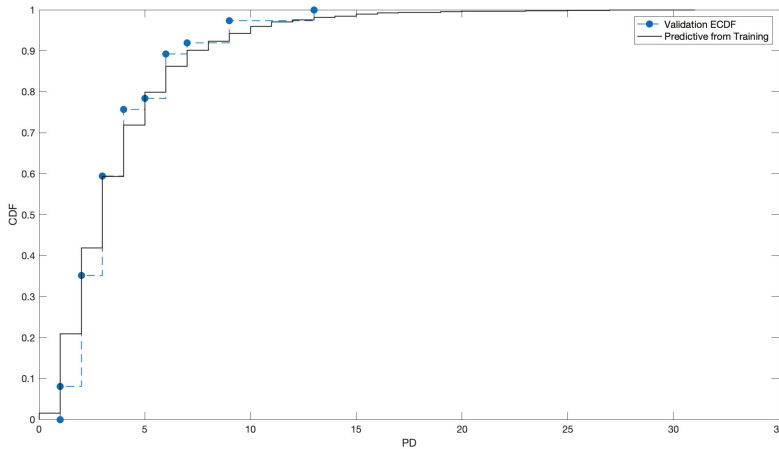


Figure 9. Predictive distribution generated by the bivariate urn model for the “Very poor” class when trained on the training sample (1590 loans), against the empirical ECDF from the validation set (37 loans).

4.4. What about the Crisis?

In line with findings in the literature (Eichengreen et al. 2012; Ivashina and Scharfstein 2010; Turlakov 2013; Witzany 2011), Maio’s dataset—as well as the original one by Freddie Mac—shows that the financial crisis of 2007–2008 had a clear impact on the dependence between PD and LGD. In particular, an increase in the strength of correlation is observable during the the crisis and in the years after. In fact, we can observe an overall value of 0.11 before 2008, which grows up to 0.24 in 2011 and remains pretty stable after. A compatible behavior is observed for the different FICO classes.

When evaluated in-sample, the bivariate urn model, given the large number of observations is always able to perform satisfactorily: this holds true for example when we restrict our attention on the intervals [2002–2007], [2007–2010] or [2010–2016]. Performances are conversely not good when the model trained on pre-crisis data is used to predict the post-crisis period, given the substantial difference in the strength of the dependence, which is persistently underestimated. Clearly, we are speaking about those situations in which the prior we elicited were not taking into consideration a strong increase in correlation. An ad hoc modification of the support of A could represent a viable solution, provided we could be aware of this ex ante, and not just ex post.

An interesting thing, worth mentioning, happens when we focus our attention on the loans originated (not defaulted) during the crisis and follow them. Here, the relation between PD and LGD tends to 0 or, for some FICO classes, it becomes slightly negative, probably because of the stricter selection the scared banks likely imposed on applicants, asking for more collateral and guarantees. While the model is still capable of dealing with no dependence, a negative correlation cannot be studied.

5. Conclusions

We have presented a bivariate urn construction to model the dependence between PD and LGD, also showing a promising application on mortgage data.

Exploiting the reinforcement mechanism of Polya urns and the conjugacy of beta-Stacy processes, the Bayesian nonparametric model we propose is able to combine experts' judgements, in the form of a priori knowledge, with the empirical evidence coming from data, learning and improving its performances every time new information becomes available.

The possibility of using prior knowledge is an important feature of the Bayesian approach. In fact, it can compensate for the lack of information in the data with the beliefs of experienced professionals about possible trends and rare events. For rare events in particular, the possibility of eliciting an a priori on something rarely or never observed before can mitigate, at least partially, the relevant problem of historical bias (Derbyshire 2017; Shackle 1955; Taleb 2007).

One could argue that nothing guarantees the ability of eliciting a reliable a priori: experts could be easily wrong. The answer to such a relevant observation is that the bivariate urn model learns over time, at every interaction with actual data. A sufficient amount of data can easily compensate unrealistic beliefs. Moreover, as already observed in Cirillo et al. (2013), thanks to reinforcement, urns are able to learn hidden patterns in the data, casting light on previously ignored relations and features. Such a capability bridges towards the paradigm of machine/deep learning (Murphy 2012). However, differently from standard machine/deep learning techniques, the combinatorial stochastic nature of the bivariate urn model allows for a greater control of its probabilistic features. There is no black box (Knight 2017).

Clearly, if an a priori cannot be elicited nor is desired, one can still use the model in a totally data-driven way, building priors based on ECDFs (Cheng and Cirillo 2018). In such a case, however, the results of the bivariate urn construction would not substantially differ from those of a more standard empirical copula approach (Rueschendorf 2009), notwithstanding the difficulty of dealing with non-directly (fully) observable quantities like the joint factor A . More interesting can therefore be the "merging" methodology developed by Figini and Giudici (2011), in which the combination of quantitative and qualitative information can (at least partially) solve the lack of experts' priors.

Furthermore, thanks to its nonparametric nature, the model we propose does not require the development of parametric assumptions, with the subsequent problem of choosing among them, as one should for instance do in using copulas (Nelsen 2006).

As observed in Section 2, an important limitation of the bivariate urn model is the possibility of only modeling positive linear dependence. Its use when dependence can be negative, or strongly nonlinear, is not advised. However, this seems not to be the case with mortgages, and PD/LGD in general.

Finally, the bivariate urn model can be computationally intensive. For large datasets and portfolios, a standard laptop may require from a few hours to a full day to obtain the posterior distribution. Clearly the quality of coding, which was not our main investigation path, has a big impact on the final performances.

Author Contributions: Conceptualization, P.C.; Data curation, D.C. and P.C.; Formal analysis, D.C. and P.C.; Investigation, D.C. and P.C.; Methodology, D.C. and P.C.; Supervision, P.C.; Visualization, P.C.; Writing the original draft, D.C. and P.C.; Review and editing, P.C. and D.C.

Funding: This research received no external funding.

Acknowledgments: The authors thank Vittorio Maio for sharing his data with them, and two anonymous referees for their suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Altman, Edward I. 2006. Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence. *New York University, Stern School of Business*. [CrossRef]
- Altman, Edward I., Andrea Resti, and Andrea Sironi. 2001. Analyzing and Explaining Default Recovery Rates. A Report Submitted to the International Swaps & Derivatives Association. Available online: <http://people.stern.nyu.edu/ealtman/Review1.pdf> (accessed on 14 January 2019).
- Altman, Edward I., Brooks Brady, Andrea Resti, and Andrea Sironi. 2005. The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications. *The Journal of Business* 78: 2203–27. [CrossRef]
- Amerio, Emanuele, Pietro Muliere, and Piercesare Secchi. 2004. Reinforced Urn Processes for Modeling Credit Default Distributions. *International Journal of Theoretical and Applied Finance* 7: 407–23. [CrossRef]
- Baesens, Bart, Daniel Roesch, and Harald Scheule. 2016. *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*. Hoboken: Wiley.
- BCBS. 2000. Principles for the Management of Credit Risk. Available online: <https://www.bis.org/publ/bcbs75.pdf> (accessed on 10 March 2019).
- BCBS. 2005. An Explanatory Note on the Basel II IRB Risk Weight Functions. Available online: <https://www.bis.org/bcbs/irbriskweight.pdf> (accessed on 10 March 2019).
- BCBS. 2006. *International Convergence of Capital Measurement and Capital Standards*. Number 30 June. Basel: Bank for International Settlements, p. 285.
- BCBS. 2011. *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Number 1 June. Basel: Bank for International Settlements, p. 69.
- Bruche, Max, and Carlos Gonzalez-Aguado. 2010. Recovery Rates, Default Probabilities, and the Credit Cycle. *Journal of Banking & Finance* 34: 754–64.
- Bulla, Paolo. 2005. Application of Reinforced Urn Processes to Survival Analysis. Ph.D. thesis, Bocconi University, Milan, Italy.
- Bulla, Paolo, Pietro Muliere, and Steven Walker. 2007. Bayesian Nonparametric Estimation of a Bivariate Survival Function. *Statistica Sinica* 17: 427–44.
- Calabrese, Raffaella, and Paolo Giudici. 2015. Estimating Bank Default with Generalised Extreme Value Regression Models. *Journal of the Operational Research Society* 28: 1–10. [CrossRef]
- Cerchiello, Paola, and Paolo Giudici. 2014. Bayesian Credit Rating Assessment. *Communications in Statistics: Theory and Methods* 111: 101–15.
- Cheng, Dan, and Pasquale Cirillo. 2018. A Reinforced Urn Process Modeling of Recovery Rates and Recovery Times. *Journal of Banking & Finance* 96: 1–17.
- Cirillo, Pasquale, Jürg Hüsler, and Pietro Muliere. 2010. A Nonparametric Urn-based Approach to Interacting Failing Systems with an Application to Credit Risk Modeling. *International Journal of Theoretical and Applied Finance* 41: 1–18. [CrossRef]
- Cirillo, Pasquale, Jürg Hüsler, and Pietro Muliere. 2013. Alarm Systems and Catastrophes from a Diverse Point of View. *Methodology and Computing in Applied Probability* 15: 821–39. [CrossRef]
- Derbyshire, James. 2017. The Siren Call of Probability: Dangers Associated with Using Probability for Consideration of the Future. *Futures* 88: 43–54. [CrossRef]
- Diaconis, Persi, and David Freedman. 1980. De Finetti's Theorem for Markov Chains. *The Annals of Probability* 8: 115–30. [CrossRef]
- Duffie, Darrell. 1998. *Defaultable Term Structure Models with Fractional Recovery of Par*. Technical Report. Stanford: Graduate School of Business, Stanford University.
- Duffie, Darrell, and Kenneth J. Singleton. 1999. Modeling Term Structures of Defaultable Bonds. *The Review of Financial Studies* 12: 687–720. [CrossRef]
- Duffie, Darrell, and Kenneth J. Singleton. 2003. *Credit Risk*. Cambridge: Cambridge University Press.
- Eichengreen, Barry, Ashoka Mody, Milan Nedeljkovic, and Lucio Sarno. 2012. How the Subprime Crisis Went Global: Evidence from Bank Credit Default Swap Spreads. *Journal of International Money and Finance* 31: 1299–318. [CrossRef]

- Experian. 2019. Blog: What Are the Different Credit Scoring Ranges? Available online: <https://www.experian.com/blogs/ask-experian/infographic-what-are-the-different-scoring-ranges/> (accessed on 18 January 2019).
- Figini, Silvia, and Paolo Giudici. 2011. Statistical Merging of Rating Models. *Journal of the Operational Research Society* 62: 1067–74. [CrossRef]
- Fortini, Sandra, and Sonia Petrone. 2012. Hierarchical Reinforced Urn Processes. *Statistics & Probability Letters* 82: 1521–29.
- Freddie Mac. 2019a. *Single Family Loan-Level Dataset*. Freddie Mac. Available online: http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page (accessed on 3 February 2019).
- Freddie Mac. 2019b. *Single Family Loan-Level Dataset General User Guide*. Freddie Mac. Available online: http://www.freddiemac.com/fmac-resources/research/pdf/user_guide.pdf (accessed on 3 February 2019).
- Freddie Mac. 2019c. *Single Family Loan-Level Dataset Summary Statistics*. Freddie Mac. Available online: http://www.freddiemac.com/fmac-resources/research/pdf/summary_statistics.pdf (accessed on 3 February 2019).
- Frye, Jon. 2000. Depressing Recoveries. *Risk* 13: 108–11.
- Frye, Jon. 2005. The Effects of Systematic Credit Risk: A False Sense of Security. In *Recovery Risk*. Edited by Edward Altman, Andrea Resti and Andrea Sironi. London: Risk Books, pp. 187–200.
- Galavotti, Maria Carla. 2001. Subjectivism, objectivism and objectivity in bruno de finetti's bayesianism. In *Foundations of Bayesianism*. Edited by David Corfield and Jon Williamson. Dordrecht: Springer, pp. 161–74.
- Genest, Christian, and Louis-Paul Rivest. 1993. Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* 88: 1034–43. [CrossRef]
- Geske, Robert. 1977. The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis* 12: 541–52. [CrossRef]
- Giudici, Paolo. 2001. Bayesian Data Mining, with Application to Credit Scoring and Benchmarking. *Applied Stochastic Models in Business and Industry* 17: 69–81. [CrossRef]
- Giudici, Paolo, Pietro Muliere, and Maura Mezzetti. 2003. Mixtures of Dirichlet Process Priors for Variable Selection in Survival Analysis. *Journal of Statistical Planning and Inference* 17: 867–78.
- Hamerle, Alfred, Michael Knapp, and Nicole Wildenauer. 2011. Modelling Loss Given Default: A “Point in Time”-Approach. In *The Basel II Risk Parameters: Estimation, Validation, Stress Testing—With Applications to Loan Risk Management*. Edited by Bernd Engelmann and Robert Rauhmeier. Berlin: Springer, pp. 137–50.
- Hjort, Nils Lid, Chris Holmes, Peter Mueller, and Stephen G. Walker. 2010. *Bayesian Nonparametrics*. Cambridge: Cambridge University Press.
- Hull, John C. 2015. *Risk Management and Financial Institutions*, 4th ed. New York: Wiley.
- Ivashina, Victoria, and David Scharfstein. 2010. Bank lending during the financial crisis of 2008. *Journal of Financial Economics* 97: 319–38. [CrossRef]
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. New York: Wiley.
- Jones, Philip, Scott P. Mason, and Eric Rosenfeld. 1984. Contingent Claims Analysis of Corporate Capital Structures: An Empirical Investigation. *The Journal of Finance* 39: 611–25. [CrossRef]
- JP Morgan. 1997. Creditmetrics—Technical Document. Available online: http://www.defaultrisk.com/_pdf6j4/creditmetrics_techdoc.pdf (accessed on 3 February 2019).
- Kim, In Joon, Krishna Ramaswamy, and Suresh Sundaresan. 1993. Does Default Risk in Coupons Affect the Valuation of Corporate Bonds?: A Contingent Claims Model. *Financial Management* 22: 117–31. [CrossRef]
- Knight, Will. 2017. The Dark Secret at the Heart of AI. *Technology Review* 120: 54–63.
- Lando, David. 1998. On Cox Processes and Credit Risky Securities. *Review of Derivatives Research* 2: 99–120. [CrossRef]
- Lindley, Dennis. 1991. *Making Decisions*, 2nd ed. New York: Wiley.
- Longstaff, Francis A., and Eduardo S. Schwartz. 1995. A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *The Journal of Finance* 50: 789–819. [CrossRef]
- Mahmoud, Hosam. 2008. *Polya Urn Models*. Boca Raton: CRC Press.
- Maio, Vittorio. 2017. Modelling the Dependence between PD and LGD. A New Regulatory Capital Calculation with Empirical Analysis from the US Mortgage Market. Master's thesis, Politecnico di Milano, Milano, Italy. Available online: <https://www.politesi.polimi.it/handle/10589/137281> (accessed on 10 March 2019).
- McNeil, Alexander J., and Jonathan P. Wendin. 2007. Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance* 14: 131–49. [CrossRef]

- McNeil, Alexander J., Ruediger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management*. Princeton: Princeton University Press.
- Meng, Xiao-Li. 1994. Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science* 9: 538–58. [CrossRef]
- Merton, Robert C. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance* 29: 449–70.
- Miu, Peter, and Bogie Ozdemir. 2006. Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD Correlations. *Journal of Credit Risk* 2: 43–68.
- Muliere, Pietro, Piercesare Secchi, and Stephen G Walker. 2000. Urn Schemes and Reinforced Random Walks. *Stochastic Processes and their Applications* 88: 59–78. [CrossRef]
- Muliere, Pietro, Piercesare Secchi, and Stephen G Walker. 2003. Reinforced Random Processes in Continuous Time. *Stochastic Processes and Their Applications* 104: 117–30. [CrossRef]
- Murphy, Kevin P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge: The MIT Press.
- Narain, Bhavana. 1992. Survival Analysis and the Credit Granting Decision. In *Credit Scoring and Credit Control*. Edited by Lyn C. Thomas, David B. Edelman and Jonathan N. Crook. Oxford: Oxford University Press, pp. 109–21.
- Nelsen, Roger B. 2006. *An Introduction to Copulas*. New York: Springer.
- Nielsen, Lars Tyge, Jesus Saà-Requejo, and Pedro Santa-Clara. 2001. *Default Risk and Interest Rate Risk: The Term Structure of Default Spreads*. Paris: INSEAD.
- Peluso, Stefano, Antonietta Mira, and Pietro Muliere. 2015. Reinforced Urn Processes for Credit Risk Models. *Journal of Econometrics* 184: 1–12. [CrossRef]
- Resti, Andrea, and Andrea Sironi. 2007. *Risk Management and Shareholders' Value in Banking*. New York: Wiley.
- Rueschendorf, Ludger. 2009. On the distributional transform, sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference* 139: 3921–27. [CrossRef]
- Shackle, George Lennox Sharman. 1955. *Uncertainty in Economics and Other Reflections*. Cambridge: Cambridge University Press.
- Taleb, Nassim Nicholas. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.
- Turlakov, Mihail. 2013. Wrong-way risk, credit and funding. *Risk* 26: 69–71.
- Vasicek, Oldrich A. 1984. Credit Valuation. Available online: [http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/c181fb77ee99d464c125757a00505078/\\$FILE/Credit_Valuation.pdf](http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/c181fb77ee99d464c125757a00505078/$FILE/Credit_Valuation.pdf) (accessed on 10 March 2019).
- Walker, Stephen, and Pietro Muliere. 1997. Beta-Stacy Processes and a Generalization of the Pólya-Urn Scheme. *The Annals of Statistics* 25: 1762–80. [CrossRef]
- Wilde, Tom. 1997. *CreditRisk+: A Credit Risk Management Framework*. Technical Report. New York: Credit Suisse First, Boston.
- Wilson, Thomas C. 1998. Portfolio credit risk. *Economic Policy Review* 4: 71–82. [CrossRef]
- Witzany, Jiří. 2011. A Two Factor Model for PD and LGD Correlation. *Bulletin of the Czech Econometric Society* 18. [CrossRef]
- Yao, Xiao, Jonathan Crook, and Galina Andreeva. 2017. Is It Obligor or Instrument That Explains Recovery Rate: Evidence from US Corporate Bond. *Journal of Financial Stability* 28: 1–15. [CrossRef]
- Zhang, Jie, and Lyn C. Thomas. 2012. Comparisons of Linear Regression and Survival Analysis Using Single and Mixture Distributions Approaches in Modelling LGD. *International Journal of Forecasting* 28: 204–15. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Credit Risk Assessment Model for Small and Micro-Enterprises: The Case of Lithuania

Rasa Kanapickiene ^{1,*} and Renatas Spicas ²

¹ Department of Finance, Faculty of Economics and Business Administration, Vilnius University, 10222 Vilnius, Lithuania

² Kaunas Region Credit Union, 44249 Kaunas, Lithuania; renatas.spicas@gmail.com

* Correspondence: rasa.kanapickiene@evaf.vu.lt

Received: 7 April 2019; Accepted: 5 June 2019; Published: 13 June 2019

Abstract: In this research, trade credit is analysed from a seller (supplier) perspective. Trade credit allows the supplier to increase sales and profits but creates the risk that the customer will not pay, and at the same time increases the risk of the supplier's insolvency. If the supplier is a small or micro-enterprise (SMiE), it is usually an issue of human and technical resources. Therefore, when dealing with these issues, the supplier needs a high accuracy but simple and highly interpretable trade credit risk assessment model that allows for assessing the risk of insolvency of buyers (who are usually SMiE). The aim of the research is to create a statistical enterprise trade credit risk assessment (ETCRA) model for Lithuanian small and micro-enterprises (SMiE). In the empirical analysis, the financial and non-financial data of 734 small and micro-sized enterprises in the period of 2010–2012 were chosen as the samples. Based on the logistic regression, the ETCRA model was developed using financial and non-financial variables. In the ETCRA model, the enterprise's financial performance is assessed from different perspectives: profitability, liquidity, solvency, and activity. Varied model variants have been created using (i) only financial ratios and (ii) financial ratios and non-financial variables. Moreover, the inclusion of non-financial variables in the model does not substantially improve the characteristics of the model. This means that the models that use only financial ratios can be used in practice, and the models that include non-financial variables can also be used. The designed models can be used by suppliers when making decisions of granting a trade credit for small or micro-enterprises.

Keywords: trade credit; small and micro-enterprises; financial non-financial variables; risk assessment; logistic regression

1. Introduction

Sellers use a variety of promotional tools to increase sales. One of these is the trade credit (Afrifa and Gyapong 2017; Wang et al. 2018), which benefits both the seller and the buyer. Trade credit is “debt arising from credit sales and recorded as an account receivable by the seller and as an account payable by the buyer” (Brigham and Houston 2004). On the one hand, according to Wang et al. (2018), “trade credit has grown rapidly and become an effective tool to incentivize suppliers to increase sales and profits in supply chain management”. Fabbri and Klapper (2016) state that “the traditional explanation for the existence of trade credit is that trade credit plays a nonfinancial role”. On the other hand, researchers (Lin and Chou 2015; Tsuruta 2015; Fabbri and Klapper 2016; Hill et al. 2017; Shenoy and Williams 2017; Tsao 2018) argue that customers use trade credit as one of the main sources of short-term external financing.

The peculiarity of trade credit is that it involves a joint commodity-financial transaction (Pike and Cheng 2001). The supplier sells the product or service and at the same time gives the customer credit for the purchase. Most authors, in analyzing the concept of trade credit (Cunat 2007; Garcia-Teruel and

Martinez-Solano 2010a, 2010b; Tsao 2010; Mahata and De 2016; Tsuruta 2015), emphasize the financial aspect as a key feature—the trade credit is a deferred payment for goods for a certain period of time. In the literature (Martinez-Sola et al. 2017), the deferred payment is treated as a loan to the customer, which can be compared to the loan granted by a bank. On the other hand, it is emphasized (Paul and Boden 2011) that trade credit is provided by non-financial institutions whose primary purpose is to sell their goods to customers. Having explained the complexity of the concept of trade credit, we now discuss the economic aspects of trade credit from the perspective of both the customer and the supplier.

From the customer's perspective, research (Garcia-Teruel and Martinez-Solano 2010a; Martinez-Sola et al. 2017) shows that a "trade credit is very important for firms that have more difficulty funding themselves through credit institutions as is the case for small- and medium-sized enterprises (SMEs), (i) whose access to the capital markets is very limited, and (ii) who use less external finance, especially bank finance, and (iii) rely more on short-term debt finance" (Garcia-Teruel and Martinez-Solano 2010a). It should be noted that this problem has intensified during the period of the financial crisis. As McGuinness and Hogan (2016) claim, "in the aftermath of the financial crisis in 2008, bank lending to SMEs declined significantly in most developed economies". Consequently, trade credit is particularly important for these companies.

From the supplier's perspective, according to Wang et al. (2018), by providing trade credits, the supplier encourages customers to buy more and thus increases his/her profits. However, on the other hand, the granting of trade credits increases the default risk of the suppliers themselves.

Prior studies show that a trade credit debtor (customer) in a bankruptcy process will almost certainly fail to meet the claims of its trade creditors (suppliers), thereby causing credit losses (Jacobson and von Schedvin 2015; Wang et al. 2018). Consistent with this argument, the researchers conclude that such credit losses could, in turn, influence trading creditors to become insolvent and subsequently go bankrupt. The conclusion is confirmed by Jacobson and von Schedvin's (2015) performed research, wherein scientists emphasize that corporate failures in the corporate sector result in significantly higher credit losses than in the banking sector. The findings are based on an empirical study (Jacobson and von Schedvin 2015), which states that the aggregate annual credit loss incurred by Swedish trade creditors is around 50 percent higher than Swedish bank lending to non-financial corporations. This indicates that trade creditors (suppliers) need a tool to help identify the credit risk of trade credit debtors (customers).

This issue becomes even more complicated when the trade credit debtor (customer) is an SME. The reason for this is the lack of data (Yoshino and Taghizadeh-Hesary 2014; Martinez-Sola et al. 2017) and difficulties in accessing the authentic SME Database (Yoshino and Taghizadeh-Hesary 2014), resulting in lower quality data and lower reliability of information (Tascon et al. 2018). For an example, as Ciampi and Gordini (2013) point out, unaudited financial statements and shares yield no market price (as they are not traded on exchanges), and no credit rating is evaluated, which can be attributed to the objective reasons for the lower quality data. Therefore, as claimed by Ciampi and Gordini (2013) and Tascon et al. (2018), recent research generally recognizes that SMEs require special tools for risk management, taking into account their specific characteristics. Nevertheless, the difficulty in accessing financial data and other information from SMEs is precisely the reason, according to Abdullah et al. (2016), why little research has been done on small- and medium-sized enterprises (SMEs).

Kosmidis and Stavropoulos (2014) have analyzed the literature studies conducted so far and have concluded that there are many studies carried out on the subject of financial distress signaling and corporate failure prediction. However, credit risk assessment and default prediction models are usually developed, as is common or for large enterprises (Altman 1968; Chesser 1974; Taffler and Tisshaw 1977; Zavgren 1985; Bekhet and Eletter 2014; Niklis et al. 2014; Verbraken et al. 2014; Florez-Lopez and Ramon-Jeronimo 2015; Lessmann et al. 2015; Manab et al. 2015; Fernandes and Artes 2016; Sohn et al. 2016; Sousa et al. 2016; Xiao et al. 2016; Petropoulos et al. 2016).

It is recognized that SMEs are an important part of the economy and their specificity in managing credit risk must be taken into account. However, there are significantly fewer models of credit risk

management (Tascon et al. 2018; Zhu et al. 2016; Yoshino and Taghizadeh-Hesary 2014; Ciampi and Gordini 2013; Behr and Guettler 2007; Altman and Sabato 2007). Knowing that SMEs are generally riskier than large corporations (Yoshino and Taghizadeh-Hesary 2014; Terdpaopong and Mihret 2011), researchers (Terdpaopong and Mihret 2011) suggest using financial distress prediction models applied for SMEs.

Two areas of research can be distinguished in Lithuania. First, there are already tested credit risk and bankruptcy prediction models. The most common tested models are the models of linear discriminant analysis (Altman, Springate, Taffler and Tisshaw) and logistic regression (Chesser, Zavgren). The results are contradictory. Some scientists have expressed doubts that bankruptcy prediction models developed in different countries are suitable for predicting the bankruptcy of Lithuanian enterprises, as enterprises operate under different conditions (e.g., Purvinis et al. 2005). Secondly, new models are being developed based on enterprises operating in Lithuania. One of the first Lithuanian researchers to analyze credit risk assessment methods in the context of bankruptcy forecasting was Grigaravicius (2003), who applied a logistic regression model for corporate bankruptcy prediction. Credit risk assessment models are being developed for credit institutions—banks (Valvoniš 2008; Dzidzevičiute 2013) or credit unions (Spicas 2017). If the models are applied for SMEs, the samples of model development are small (Butkus et al. 2014). Thus, standard credit models are developed by financial institutions for large companies. They rely on a large amount of data. Therefore, one cannot directly transpose these models to SMEs and special tools need to be designed.

The aim of this research is to create a statistical enterprise trade credit risk assessment model for Lithuanian small and micro-enterprises (SMiE). In this research, small and micro-enterprises are understood as they are defined in 2010–2012 by the Republic of Lithuania Law on Small- and Medium-Size Business Development, i.e., enterprises that have less than 50 employees and whose financial data meet any of the selected conditions: (i) the annual revenue of the enterprise does not exceed EUR 7 million; (ii) the value of assets indicated in the statement of financial position does not exceed EUR 5 million.

The statistical enterprise trade credit risk assessment model (hereinafter referred to as the ETCRA model) is a technical tool used during the trade creditors' (suppliers') decision-making process that helps to estimate the probability of default (PD) of potential trade credit debtors (customers) by using his/her initial data and making the decision regarding the granting of the extension of payment of goods maturity.

The development of the model is based on the following provision: the model must be practically applicable; it must be designed according to the model users, i.e., the trade creditors (suppliers). As the majority of suppliers are SME enterprises, it is necessary to evaluate the fact that these enterprises face difficulties due to limited human, financial, and technological resources. Regarding these features, the model developed must exhibit special characteristics, such as simplicity, clarity of model results, and the inclusion of accessible external information (in the form of independent variables) into modelling, including financial and non-financial information. Adapting the results of the Spicas et al. (2018) study, the following requirements apply to the model being developed:

1. The accuracy of a model, i.e., the model must hold high discriminatory power.
2. The high interpretability of the results of the model.
3. The simplicity of the model. The application of the model, a periodical review, calibration, and interpretation of the model results should not require particular knowledge in the areas of statistics and information technology.
4. A probabilistic model. The result of the model must be PD (probability of default), i.e., the probability that the debtor will become insolvent within a specified period of time.

The research focuses on three aspects of the ETCRA model formation. Firstly, when the trade creditor (supplier) (hereinafter referred to as creditor) chooses target trade credit debtors (customers) (hereinafter referred to as debtors), a segment of the modelled dependent variable must be defined.

It should be noted that the definition of a dependent variable is usually related to a loss event, i.e., the definition of bad debt. Therefore, the research focuses on how to define the bad trade credit debtor (customer) (hereinafter referred to as a bad debtor). An in-depth discussion of these aspects will be provided further in the methodology of the research. Secondly, for the ETCRA model development, it is important to choose suitable independent variables, i.e., financial and non-financial variables of enterprises. Thirdly, issues regarding the methodology of model development are discussed and the opportunities of logistic regression application are analyzed.

The most important contributions of this research are that the developed ETCRA model is characterized by high interpretability of results, accuracy, and simplicity. It can also be stated that the model features flexibility, i.e., the trade creditors (suppliers) can choose to assess the trade credit debtors using only financial ratios or, additionally, non-financial variables. As these two alternatives have been analyzed in the research, it has led to the conclusion that the inclusion of non-financial variables in the model does not substantially improve the characteristics of the model.

The research analyzes the credit risk of small and micro enterprises separately. There is also the issue of, as since 2016, the information of micro-enterprises in financial statements being significantly reduced, dependent on whether the characteristics of the ETCRA model deteriorate when compared with small enterprises. An important result of the research is that having properly selected financial ratios, the characteristics of the ETCRA model of micro-enterprises do not deteriorate.

The rest of the paper is structured as follows: The subsequent section describes the limitations on the information provided by financial statements of small and micro-enterprises in Lithuania. It is followed by a section describing the methodology of the research. The next section provides the empirical results and discussion. In the last section, the concluding remarks are presented.

2. Limitations on Information Provided by Financial Statements of Small and Micro-Enterprises

Since the entry into force of the Business Accounting Standards in 2004 (which are drafted on the basis of European Union law and International Accounting Standards), the financial accounting regulation has fundamentally changed in Lithuania. Accounting has become much more complicated, requiring a very high level of accountant qualifications to prepare financial statements correctly. This change has also created additional difficulties for such a complex business system, as well as for entrepreneurial development in Lithuania (especially for small- and medium-sized businesses). The small and medium business accounting system was based on the notion that small- and medium-sized enterprises do not differ significantly from large companies and should follow similar accounting requirements. It should be noted that, at that time, it was understood that small businesses could not meet all the requirements for large companies and were allowed to draw up financial statements that contained less information than large companies.

The assumption, prevailing at the end XX century, currently has changed, and it has been recognized that the needs of small- and medium-sized enterprises differ. Studies conducted by the European Commission ([Ramboll Management 2005](#)) found that the smaller the company, the more costly it is to bear accounting, financial reporting, and its audit requirements. Research has confirmed that small businesses are not able to hire employees who are capable of managing their accounts properly or in preparing financial statements, and therefore need to use these services from other enterprises.

As small companies are very important for socio-economic development in Lithuania, it was decided that the accounting and financial reporting of these companies should be simplified in order to reduce the administrative burden. The structure of financial statements has changed in Lithuania again since 2016, and the amount of information included in statements of micro-enterprises has decreased markedly. Therefore, it becomes important to assess whether the amount of the available financial information is sufficient for the enterprise trade credit risk assessment of small and micro-enterprises.

Lower requirements for financial statements of small and micro-enterprises are established in the Law on Financial Statements of Entities of the Republic of Lithuania and in National Accounting Standards (more specifically, in Business Accounting Standards (BASs)).

Article 20 of this Law provides that the set of financial statements of medium-sized and large enterprises shall consist of the following financial statements: (i) statement of financial position; (ii) statement of profit or loss; (iii) statement of cash flows; (iv) statement of changes in equity; and (v) notes¹. However, the level of detail of financial statements of small and micro-enterprises is lower, meaning that (1) a set of financial statements of small enterprises shall consist of the following financial statements: (i) statement of financial position or abridged statement of financial position; (ii) statement of profit or loss; and (iii) notes. (2) Micro-enterprises can refuse drawing up notes and the set of their financial statements shall consist of the following financial statements: (i) short statement of financial position; and (ii) short statement of profit or loss.

Users of financial statements want to get as much information as possible from the financial statements. However, financial statements of a lower level of detail do not allow calculating all the desired financial ratios of the enterprise. When analyzing the small enterprises, these limitations of financial analysis appear.

Firstly, small enterprises are not entitled to generate cash flow statements. Therefore, it is impossible for these enterprises to calculate cash flow ratios. These enterprises are not entitled to generate the statement of changes in equity as well. In this way, the information about cash flows and changes in equity necessary for analysis is lost.

Secondly, small enterprises can make an abridged statement of financial position. Hence, it is impossible to estimate financial ratios that require more thorough financial data. Some examples are given. *Intangible assets*, *tangible assets* and *financial assets* are not grouped into smaller components and, therefore, the structure of these assets cannot be evaluated. *Inventories* are not grouped into smaller components, i.e., *raw materials*, *production in progress*, and *finished goods*, and thus cannot be evaluated. *Amounts receivable within one year* are not specified, thus *trade debtors* cannot be accurately evaluated. All liabilities are only grouped into current liabilities (i.e., *amounts payable within one year and other short-term liabilities*) and non-current liabilities (i.e., *amounts payable after one year and other long-term liabilities*). Therefore, the components of smaller liabilities, such as *amounts owed to credit institutions or trade creditors* cannot be evaluated (see Appendix A, Table A1).

Third, in Lithuania, both large and small enterprises that follow the Business Accounting Standards (BASs) generate profit (loss) statements of the same level of detail, i.e., equally detail all the enterprise's income and expenses of the reporting period: (1) Income is grouped into smaller components. In the statement of profit or loss, there are also presented three items of income from financing and investing activities, i.e., *income from investments in the shares of parent, subsidiaries and associated entities*; *income from other long-term investments and loans*; and *other interest and similar income*. (2) According to International Financial Reporting Standards (IFRSs), "expenses are sub-classified to highlight components of financial performance that may differ in terms of frequency, the potential for gain or loss and predictability" ([International Financial Reporting Standards \(IFRSs\)](#)). In these enterprises, the expenses are provided in the "function of expense" (also called "cost of sales") method, i.e., the expenses are classified according to their function. In the statement of profit or loss, there are presented the following expenses: *cost of sales*; *selling expenses*; *general and administrative expenses*; *the impairment of the financial assets and short-term investments*; and *interest and other similar expenses*. The result of other operating—profit or loss—is provided and it is not possible to estimate the income and expenses earned in this activity separately. Thus, according to [Spicas et al. \(2015\)](#), "it is not possible to calculate the financial ratios of expenditure level because employment costs and depreciation expenses are not known. Since the depreciation is unknown, EBITDA cannot be calculated either".

When analyzing the micro-enterprises, financial analysis is more limited than in the case of small enterprises.

¹ The titles of the financial statements and financial items are used in accordance with International Financial Reporting Standards (IFRSs).

Firstly, micro-enterprises may not generate cash flow statements and statements of changes in equity as well. Therefore, it is impossible for enterprises to calculate cash flow ratios. Changes in equity cannot be evaluated. These enterprises may refuse to draw up an explanatory note that explains the amounts presented in other financial statements and discloses the additional material information that is not presented in them. Therefore, the amount of information analyzed is reduced.

Secondly, micro-enterprises can make a short statement of financial position. Compared to small enterprises, even less information is available in the statement of financial position of micro-enterprises. Three points can be distinguished: (i) Equity structure is not provided. Therefore, items such as *capital* or *retained profit (loss)* and related ratios cannot be evaluated during the analysis. (ii) Only two components of current assets—*inventories* and *other current assets*—are distinguished. Thus, there is no information on cash and cash equivalents, trade debtors, or short-term investments. (iii) Only two components of fixed assets are distinguished—*tangible assets* and *other fixed assets*. Here, a methodological problem of assets grouping is faced. The *other fixed assets* item presented in the statement of financial position of small enterprises includes (i) *assets of the deferred tax on profit*, (ii) *biological assets*, and (iii) *other assets* (not included in preceding items of other fixed assets). On the other hand, the *other fixed assets* item presented in micro-enterprises statement of financial position includes (i) *intangible assets*, (ii) *financial assets*, and (iii) *other fixed assets* (to the extent that this item is presented in the statement of financial position of small enterprises) items (see Appendix A, Table A1). It suggests that the nature of these items—*other fixed assets*—differs. Hence, the structure of the assets of micro-enterprises becomes incomparable with that of other (small, medium, or large) enterprises.

Third, as discussed above, in small enterprises, the expenses are provided in the “function of expense” (also called “cost of sales”) method, i.e., the expenses are classified according to their function, whereas in micro-enterprises, the expenses are provided in the “nature of expense” method. Enterprises aggregate expenses within profit or loss according to their nature (for example, *acquired stocks and used resources*; *expenses related to employment relations*; *depreciation*), and do not reallocate them among functions within the entity. The IFRSs highlight that “this method may be simple to apply because no allocations of expenses to functional classifications are necessary”.

However, the problem of analyzing enterprises is that the statements of profit or loss of small and micro-enterprises are based on different methods, which makes the comparison between them relatively difficult (see Appendix A, Table A2). For example, (i) small enterprises can calculate earnings before interest and taxes (EBIT) (as interest are presented separately)—this is difficult for micro-enterprises as the interest expenses are unknown; (ii) since the expenses of micro-enterprises are not grouped by function, it is not calculated at *cost of sales*, *selling expenses*, and *general and administrative expenses*. Therefore, no gross profit or operating profit can be calculate, and it is also not possible to calculate corresponding ratios (gross profit margin, operating profit margin).

To sum up, it can be assumed that the decline in the amount of information presented in the financial statements may reduce the accuracy of the credit risk models being created.

3. Methodology Formation of Development of Statistical Enterprise Trade Credit Risk Assessment Model for Small and Micro-Enterprises

The credit risk model was developed to focus on small and micro-enterprises in Lithuania. The questions regarding data collection are discussed first.

3.1. Data Collection

First of all, in order to organize the research, the population, sampling, the period considered, the control period, and the sample size were determined.

3.1.1. The Population

As the present research is concerned with small and micro-enterprises, it is necessary to analyze the definition of these enterprises.

Small and micro-enterprises are understood as they are defined by the Republic of Lithuania Law on Small and Medium-size Business Development (2017), i.e., enterprises that have less than 50 employees and whose financial data meets any one of these conditions were selected: (i) annual revenue of the enterprise does not exceed EUR 10 million; (ii) value of assets indicated in the statement of financial position does not exceed EUR 10 million.

As we see, the size of the enterprises is defined by three parameters: the number of employees, annual revenue, and assets. Meanwhile, the Department of Statistics of the Republic of Lithuania details enterprises only by the number of employees, i.e., an enterprise is considered small as with less than 50 employees. There is also a definition of a working enterprise, i.e., an operating enterprise—an enterprise operating with a number of employees and (or) annual revenue. According to the data of the Department of Statistics of the Republic of Lithuania, the number of enterprises in operation at the beginning of 2010 in Lithuania with less than 50 employees was 60,516 (94.8% of all the enterprises in operation in Lithuania), and at the beginning of 2013—63,075 (95.9% of all the enterprises in operation in Lithuania). It should also be noted that the definition of a bad debtor is associated with the concept of bankruptcy; therefore, it is necessary to know the information about bankrupt enterprises. According to the data of the Department of Statistics of the Republic of Lithuania, in 2010–2012, 4311 enterprises went bankrupt or became bankrupt, i.e., 4311 bankruptcy processes had been started. Therefore, a sample of small and micro-enterprises was formed from this population for the development of the model.

3.1.2. Sampling

For this research, the sample of Lithuanian small and micro enterprises was provided by a credit bureau, *Creditinfo Lietuva*. From the population research, the sample was formed by way of random selection—the sample comprised financial and non-financial data from small and micro-enterprises over the period of 2010–2012, i.e., over the period considered.

3.1.3. The Period Considered

The period considered is the period over which the financial and non-financial data of the enterprises were analyzed. In this research, the period considered was 2010–2012.

At this point, it is worth explaining why, in this research, the period considered was 2010–2012. In 2010–2012, the financial statements of both small and micro-enterprises were of the same scope. As mentioned above, the information of a micro-enterprise in the financial statements has decreased significantly since 2016. Therefore, when we examine the financial statements of micro-enterprises for 2010–2012, we can study them in two ways, i.e., it is possible to analyze the amount of information used in the trade credit risk assessment (i) if the statements were not shortened; (ii) if, at that time, micro-enterprises would have provided only as much information as required after 2016.

3.1.4. The Control Period

At the stage of development of the ETCRA model for small and micro-enterprises (SMiE), it was necessary to properly distinguish the bad debtor from the good debtor.

Firstly, if a traditional definition of a bad debtor is used, then the bad debtor is an enterprise that went bankrupt during the period considered or delayed in paying a debt to the creditor for more than 90 days. An enterprise that does not meet these requirements is considered a good debtor. However, [Spicas et al. \(2018\)](#) proved that such a definition is insufficient to form a good debtor sample. Therefore, if we follow the definition of a bad debtor formulated by [Spicas et al. \(2018\)](#), that “an enterprise is considered a bad debtor if during the 24-month monitoring period it went bankrupt or delayed payment for more than 90 calendar days to any creditor (according to the available data)” ([Spicas et al. 2018](#)), it is necessary to assess the state of an enterprise in 24 months. Thus, it is not sufficient to analyze an enterprise only in the period considered. Therefore, the control period is distinguished.

Secondly, because of the specificity of activities of the small and micro-enterprises observed, the cessation of activities without announcing bankruptcy in this target segment can occur more frequently than in the segments of large and medium enterprises. Therefore, the business continuity opportunities of enterprises should be assessed in detail—additional assessment is needed to determine whether an enterprise that is not considered as a bad debtor, continued activities, and had no significant activity failure during the *control period*.

According to Spicas et al. (2018), the enterprises that meet at least one of the following criteria are removed from the good debtor sample; in order to form it in a way that it is made only of enterprises performing real economic and commercial activities: (1) an operating but indebted enterprise (an enterprise that during the *control period* continued its activities but had significant outstanding debts to the State Social Insurance Fund); (2) an actual non-operational enterprise: an enterprise that during the *control period* (i) according to the data from the State Social Insurance Fund, an enterprise that employs two or fewer employees; (ii) annual income of the enterprise does not exceed EUR 10,000; (iii) value of assets indicated in the statement of financial position does not exceed EUR 5000. The assessment of business continuity opportunities was carried out in the control period. In this research, the control period was 2015–2016.

3.1.5. Sample Size

The sample was formed from 734 enterprises of which there were

- (1) Three-hundred and nine bad debtors. During the period considered (2010–2012), these enterprises either (i) went bankrupt or started bankruptcy processes, or (ii) had significant debts. Thus, these enterprises fit the *definition of a bad debtor*.
- (2) Four-hundred and twenty-five good debtors. The enterprises that follow the good debtor sample are those which (i) did not go bankrupt or started bankruptcy processes and did not have significant debts during the period considered (2010–2012) and (ii) during the *control period* continued its activities (2015–2016) and had no indications of activity failure.

3.2. Research Methodology

The logistic regression enterprise trade credit risk assessment (ETCRA) model for small and micro-enterprises (SMiE) was formed in this research. The research consisted of two stages: independent variables were selected at first and, secondly, the logistic regression model was compiled.

3.2.1. Selection of Independent Variables

The most frequently analysed *financial variables* are relative financial ratios that are calculated from the financial statements of enterprises. Since financial reports are standardized, it was appropriate to make a collection of possible financial ratios by analyzing the ratios used in the scientific and professional literature. It was important that the final collection of financial ratios of the model contained ratios that analyzed different areas of the enterprise's activities and would differ in the way of the economic information content. Furthermore, when the possible financial ratios were selected, the information from financial statements available for analysis needed to be taken into account.

The research was a continuation of previous authors' research. In prior empirical research (Spicas et al. 2015; Spicas et al. 2018), a total of 101 different models of bankruptcy prediction and credit risk models were analyzed. The following groups of financial ratios were found to be used in the development of these models: (i) ratios of profitability, liquidity, and expenditure level groups were most commonly used; (ii) the ratios of the groups of asset structure, solvency, and asset turnover were used less frequently; and (iii) the ratios of tangible asset and capital market groups were used least. After the analysis of the scientific literature, 168 financial ratios were distinguished. However, because of the abridgement of the information provided by the aforementioned financial statements of

Lithuanian small enterprises, only 53 financial ratios could be used in the development of a risk model (Spicas et al. 2015; Spicas et al. 2018).

From 2016, regarding the changes in the financial statement requirements for small and micro-enterprises, the amount of information in the financial statements of these enterprises decreased. Therefore, the problem of developing credit risk models is once again relevant and needs to be reconsidered.

The hypothesis H1 was formulated as follows: compared to small enterprises, the characteristics of the ETCRA model for micro-enterprises by providing smaller amounts of information in the financial statements has deteriorated.

Non-financial variables are ratios that show the solvency opportunities of an enterprise by the way of analysis of non-financial sources. For example, credit histories, indicators of litigation, temporary measures applied, behavior in social networks, and other significant information. Usage of non-financial ratios is directly related to outside factors, i.e., information infrastructure; therefore, the opportunities for using these ratios vary significantly in different countries. Because of this, non-financial ratios were selected according to the data available for model development and external information infrastructure.

In this research, the following non-financial ratios collected by the credit bureau, *Creditinfo Lietuva*, were used: number of outstanding debts; sum of outstanding debts; average term of delay of outstanding debts; number of delayed debts; sum of delayed debts; average term of delay of delayed debts; number of requests in credit bureau; and number of valid arrests.

Hypothesis H2 was formulated as follows: the inclusion of non-financial variables in the model substantially improves the characteristics of the model.

Furthermore, in the scientific literature, there is a unanimous agreement regarding the positive influence of the application of variable selection methods to the quality of models (Yang and Duan 2008; Mileris 2012; Cateni et al. 2012).

In this empirical stage of research, the aim was to test the financial and non-financial variables which were used in the credit risk and bankrupt determination. Sixty-one variables were used in the model development: eight non-financial and 53 financial variables (Appendix B).

Control variables. As Li (2016) indicates, three major sources of endogeneity problems are “omitted variables, simultaneity, and measurement error”. To solve the endogeneity problem, it is suggested by Li (2016) to “include as many important and time-variant control variables as possible”.

In the research, different enterprise characteristics variables are used as control variables. For example, in order to analyze the relationship between CSR (corporate social responsibility) and firm value, researchers (Li et al. 2016; Cho et al. 2019) choose firm size, leverage, and research and development expenses as the control variables. Capital expenditure ratio, ROA (return on assets), advertising expense (Li et al. 2016), book-to-market ratio, cash flows from operating activities, and the standard deviation of daily stock returns (Cho et al. 2019) are being chosen too. It is noted that a size of an enterprise is a very important control variable, which, according to Dang et al. (2018), might be measured by total assets, total sales, and market capitalization. In addition, Herrador-Alcaide and Hernandez-Solis (2019) suggest that the size of the enterprise can be measured by the number of employees.

In this research, financial ratios, such as leverage, ROA, the logarithm of total assets, and the logarithm of total sales were selected as financial variables. Financial data, such as research and development expenses and cash flows from operating activities, were not published and were hardly collected in the SMiE financial statements. Since most SMiE enterprises do not trade shares on the stock exchange, market ratios could not be calculated. Therefore, the size of the enterprise selected as a control variable was measured by the number of employees.

Selection of independent variables was performed as follows:

Firstly, the analysis of missing values was performed. According to Spicas et al. (2018), the number of missing values of independent variables are to be analyzed as well as the reasons behind the failure to calculate the ratio. In cases when the ratio could not be calculated, since in financial statements of a significant part of enterprises there were no data required to calculate the variable in question, the question of removal of such a ratio was addressed.

Secondly, an analysis of the discriminatory power of independent variables was performed. In a two-way analysis, the evaluation of individual discriminatory qualities of ratios and an information value method is usually applied (Yap et al. 2011; Nikolic et al. 2013):

$$IV = \sum_{i=1}^n \left(\left(\frac{G_i}{\sum_{i=1}^n G_i} - \frac{B_i}{\sum_{i=1}^n B_i} \right) \cdot WOE_i \right), \quad (1)$$

where IV denotes information value indicator, B is the number of bad debtors, G is the number of good debtors, n is the number of assessed debtors, i is the the index of the attribute being evaluated, and WOE is the weight of evidence, calculated according to the second formula (Yap et al. 2011; Nikolic et al. 2013):

$$WOE_i = \ln \left(\left(\frac{G_i}{\sum_{i=1}^n G_i} \right) / \left(\frac{B_i}{\sum_{i=1}^n B_i} \right) \right). \quad (2)$$

The analysis was performed for every variable by calculating the information value (IV) indicator according to the first formula. According to Siddiqi (2006), estimated IV should be interpreted as follows: (i) almost no predictive power is $IV < 0.05$; (ii) poor predictive power is $0.05 < IV < 0.1$; (iii) medium predictive power is $0.1 < IV < 0.25$; (iv) high predictive power is $IV > 0.25$. Typically, any characteristic with an information value of over 0.1 will be considered for inclusion into the model (Hand and Henley 1997).

Finally, the expediency of the variables included in the model was substantiated. Financial ratios were grouped in the following groups: profitability, liquidity, solvency, activity, and structure ratio groups. It was analyzed in terms of what ratios should be included in the model from each group. The possibility of removing some strongly correlated ratios from the model was considered. The remaining variables were used to compile a logistic regression model.

3.2.2. Development of the Logistic Regression Model

The on-going statistical ETCRA model for small and micro-enterprises (SMiE) shall evaluate the probability that a debtor “will show some undesirable behaviour in the future” (Lessmann et al. 2015). It evaluates how likely an applicant is to default (Lessmann et al. 2015).

Logistic regression is a standard statistical model which is particularly appropriate here. Indeed, scientists agree that the statistical models based on logistic regression are widely used in credit risk analysis for their efficiency (Zhu et al. 2016), reliability (Figini et al. 2017), and ease of interpretation (Han et al. 2018; Figini et al. 2017). Crone and Finlay (2012) state that “the main reason for continuing usage of the logistic regression over other methods of estimation is that it provides a suitable balance of accuracy, efficiency and interpretability of the results”. In addition, this method is simple and universal (Han et al. 2018). Finally, it is indicated that logistic regression is the industry standard (Lessmann et al. 2015). The logistic regression method is used in credit risk models for small- and medium-sized enterprises (Figini et al. 2017; Zhu et al. 2016; Altman and Sabato 2007; Behr and Guettler 2007). In this research, logistic regression was selected for its popularity for default prediction.

The logistic regression ETCRA model was formed by calculating the coefficients of the selected independent variables. The logistic regression model is written as

$$PD = P(Y = 1) = \frac{1}{1 + e^{-z}}, \quad z = \beta_0 + \beta_1 X_1 \dots \beta_k X_k, \quad (3)$$

where the dependent variable in logistic regression (PD = probability of default) is a dummy variable. The PD takes the value 1 ($Y = 1$) if a debtor defaulted in the observation period and 0 ($Y = 0$) otherwise (Behr and Guettler 2007); β_i represents the particular coefficient in a linear combination of k independent variables ($i = 1, \dots, k$). Independent variables X_i are all potentially relevant parameters that may drive credit risk (Behr and Guettler 2007). In this research, independent variables X_i are financial variables, non-financial variables, and the control variable.

In the research, the forward method was applied. Initially, the constant was concluded, and the independent variables, i.e., financial ratios that have a strong correlation with the dependent variable, were gradually applied in the model. Furthermore, the expert method was considered, i.e., each group of financial ratios was analyzed, the relationships between the ratios were analyzed, and only then insignificant ratios were eliminated. The model was considered to be appropriate when the following requirements were complied with: (i) chi-square criterion p -value was less than 0.05; (ii) Cox and Snell R Square and Nagelkerke R Square were not less than 0.2; (iii) statistically significant variables should be included in the model, i.e., p -values of Wald's p -value were less than 0.05.

For the measure of the predictive accuracy of the models, the receiver operating characteristic (ROC) curve was applied. This analysis was widely used (Behr and Guettler 2007; Chang et al. 2018; Figini et al. 2017). According to Crook et al. (2007), "the ROC curve is a plot of the proportion of bads classified as bad against the proportion of goods classified as bad" at all cut-off points. The former is called *sensitivity* and the latter is $1 - \text{specificity}$, since *specificity* is the proportion of goods predicted to be good". Han et al. (2018) refer to the area under the curve (AUC) which can be used as an indicator to quantify "a quantitative performance measure: the area will range from 0.5, for a worthless model, to 1, for a perfect classifier". According to Zhu et al. (2016), if $0.8 < AUC < 0.9$, the model has excellent discriminatory power; if $AUC > 0.9$, the model has outstanding discriminatory power.

Statistical analyses were performed using IBM SPSS Statistics version 25.

4. Research Results and Findings

As stated above, in this research the model development started with 61 variables: eight non-financial and 53 financial variables (Appendix B). One control variable, i.e., the number of employees that describes the size of the enterprise, was selected.

4.1. Selection of Independent Variables

During the formation of a logistic regression ETCRA model, the selection of independent variables was carried out in the order of priority described in the methodology:

Firstly, data sets with missing values and implausible information were eliminated from the sample. During the analysis of the missing values, six variables were removed (see Appendix B): inventories/sales (INV/S), accounts receivable/sales (AR/S), sales/fixed assets (S/FA), working capital/operating expenses (WC/OE), EBIT/interest expenses (EBIT/IE), and accounts receivable/inventories (AR/INV).

Secondly, during an analysis of the discriminatory power of independent variables, it was indicated that the calculated information value (IV) indicator showed that two variables by medium and 16 (eight financial ratios and eight non-financial ratios) by weak discriminatory power.

Financial variables with (i) weak discriminatory power and (ii) with medium discriminatory power were removed from further study (see Appendix B). Even though non-financial variable IV < 0.03, i.e., the variables had weak discriminatory power, it was decided to remove only one financial variable from the further process, i.e., a number of requests in credit bureau (NRCB). In total, during this stage of analysis, 11 indicators were removed from further study, and 43 variables were chosen for further analysis: seven non-financial variables and 36 financial variables (see Appendix B).

Thirdly, financial ratios were grouped. As mentioned earlier, not all financial ratios are calculated by micro-enterprises, since at present, normative documents allow these companies to provide less information in their financial statements (see Appendix A). The ratios analyzed are as follows (see Appendix B): (1) profitability ratios (profitability ratios were divided into two groups: *return of*

sales and return of investment; in this group, three ratios were not calculated for micro-enterprises); (2) liquidity ratios (in this group, seven ratios were not calculated for micro-enterprises); (3) solvency ratios (in this group, one ratio was not calculated for micro-enterprises); (4) activity ratios (in this group, one ratio was not calculated for micro-enterprises); (5) structure ratios (these ratios were divided into two groups: *total assets structure ratios* and *equity and liabilities structure ratios*; in this group, three ratios were not calculated for micro-enterprises).

These groups of financial ratios are important for trade creditors (suppliers). Indubitably, the creditor is primarily interested in the ability of the trade credit debtor (customer) to pay its current liabilities; therefore, the creditor is interested in the liquidity ratios. However, not only the liquidity of the enterprise is important, but also a long-term perspective, i.e., whether the company will be able to pay all its liabilities. In this way, the solvency ratios become important to the creditor. On the other hand, if the enterprise does not yield enough profit, it will face the problem of repayment of debts in the future—the profitability ratios of the debtor become important for the creditor. It is also important for the creditor whether the debtor activity is efficient, i.e., whether the asset is optimally used and the expenses are well managed, therefore, the creditor is interested in the activity ratios. If the debtor funds operate in the process of production or service unreasonably, the fast cycle of the process is not ensured, which results in slowing down the income from the activity, i.e., the likelihood to overdue the payment for the creditors increases. The aim of the ETCRA model implementation was for the model to include ratios from all groups of the financial ratios.

4.2. Logistic Regression Model

The forward (LR) method was applied. Moreover, the expert method was applied, i.e., each group of financial ratios were analyzed, and insignificant ratios were removed only after the analysis of ratio relations.

Separate ETCRA models for small and micro-enterprises were developed during the research.

First of all, different ETCRA models were created for each group of enterprises (i) for small companies (Models SE) and (ii) for micro-enterprises (Models MiE). In addition, each group of enterprises had developed three models that differ according to the use of financial ratios and non-financial variables. Varied model variants were created using (1) only financial ratios (Model SE1 and Model MiE1), (2) financial ratios and non-financial variables. Model SE2 and Model MiE2 use financial ratios and one non-financial variable (number of valid arrests (NVA)); Model SE3 and Model MiE3 use financial ratios and two non-financial variables (number of valid arrests (NVA) and average term of delay of outstanding debts (ATDOD)).

Logistic regression models were formed by calculating the coefficients of the selected independent variables (see Table 1).

Financial ratios selected for the models encompassed the main financial areas of activity analysis of enterprises: profitability, liquidity, solvency, and activity. It was found that the probability of default of small enterprises was dependent on three relative financial ratios: EBIT/sales (EBIT/S), equity/total liabilities (Eq/TL), and sales/total assets (S/TA). Moreover, the probability of default of micro-enterprises was dependent on four relative financial ratios: net profit/sales (NP/S), current assets/current liabilities (CA/CL), Eq/TL, and S/TA (see Table 1).

EBIT/sales is one of the most important profitability ratios of the company, as it shows the profitability of the enterprise during its operating and investment cycles, however, without the evaluation of the financing policy. Since this ratio cannot be calculated by the micro-enterprises (MiE), the net profit/sales ratio, which describes the profitability of all the enterprise's activities, was chosen. This ratio proved significant as it shows whether MiE's income covers all expenses; according to it, we can decide how much funds can be directed to the development of the enterprise.

Table 1. Logistic regression enterprise trade credit risk assessment (ETCRA) models for small (SE) and micro-enterprises (MiE).

Independent Variables	Model SE1	Model SE2	Model SE3	Model MiE1	Model MiE2	Model MiE3
Constant	−0.675 ** (25.751)	0.042 (0.050)	−0.526 ** (5.337)	−0.578 ** (14.091)	0.093 (0.210)	−0.495 ** (4.210)
Financial ratios						
<i>1. Profitability ratios (Return of sales)</i>						
EBIT/S	−1.765 ** (30.145)	−1.663 ** (27.597)	−1.764 ** (29.358)	-	-	-
NP/S	-	-	-	−1.861 ** (33.904)	−1.733 ** (30.149)	−1.832 ** (31.700)
<i>2. Liquidity ratios</i>						
CA/CL	-	-	-	−0.174 ** (4.927)	−0.154 ** (3.896)	−0.167 ** (4.416)
<i>3. Solvency ratios</i>						
Eq/TL	−0.577 ** (30.496)	−0.520 ** (25.470)	−0.564 ** (26.964)	−0.438 ** (14.930)	−0.401 ** (12.650)	−0.439 ** (13.979)
<i>4. Activity ratios</i>						
S/TA	0.088 ** (11.777)	0.107 ** (14.681)	0.105 ** (13.425)	0.094 ** (12.648)	0.113 ** (15.516)	0.110 ** (14.051)
Non-financial variables						
NVA(1)	-	−1.076 ** (29.714)	−1.250 ** (36.090)	-	−1.026 ** (26.459)	−1.204 ** (32.701)
ATDOD	-	-	(22.793)	-	-	(24.300)
ATDOD(1)	-	-	0.971 ** (20.571)	-	-	1.034 ** (22.566)
ATDOD(2)	-	-	1.063 ** (11.461)	-	-	1.053 ** (11.143)
The percentage of the model's correctly classified good debtor cases	83.5	80.2	84.7	84.2	82.1	85.2
The percentage of the model's correctly classified bad debtor cases	57.3	61.5	63.1	58.6	62.5	63.8
The total percentage of the model's correctly classified cases	72.5	72.3	75.6	73.4	73.8	76.2
Chi-square p-value	0.000	0.000	0.000	0.000	0.000	0.000
Cox and Snell R Square	0.238	0.269	0.293	0.253	0.280	0.306
Nagelkerke R Square	0.320	0.362	0.394	0.341	0.377	0.411
df beta (s)	<1	<1	<1	<1	<1	<1

Note: The Wald's chi-square value is reported in parentheses; ** denotes statistical significance at the 5% level; ^a EBIT/S = EBIT/sales; NP/S = net profit/sales; CA/CL = current assets/current liabilities; Eq/TL = equity/total liabilities; S/TA = sales/total assets; NVA = number of valid arrests; ATDOD = average term of delay of delayed debts; source: author's calculations.

Liquidity was characterized by the current assets/current liabilities ratio, i.e., the ability of the company to meet its current liabilities using its current assets was assessed. If the enterprise already had difficulties with the payment of the current liabilities in the past year, it is likely that the current trade debts will not be covered. The solvency ratios equity/total liabilities show how many times equity is higher than total liabilities; the lower the ratio, the higher the financial risk of the enterprise, i.e., big commitments that will have to be covered in the future. The higher value of the activity ratios sales/total assets indicates a higher degree of efficiency in the overall asset management.

Compared to the research already carried out, it can be stated that the financial ratios used in the ETCRA model are widely used in the development of credit risk models. Though these ratios are widely used in global practice, they are rarely used in models created by Lithuanian enterprises. Sales/total Assets was used by Danenas et al. (2011) and Grigaravicius (2003); current assets/current liabilities—by Grigaravicius (2003), Danenas et al. (2011) and Butkus et al. (2014); equity/total liabilities—by Butkus et al. (2014).

Non-financial variables of the model analyzed credit history and other important risk factors. The probability of default was dependent on two non-financial ratios: the average term of delay of outstanding debts and number of valid arrests.

When analyzing the control variable (i.e., the size of the enterprise measured by the number of employees), it was found to be statistically insignificant, and the value of its coefficient β_i was very small, therefore, the variable was disregarded in the model. It is worth mentioning that the logarithm

of total assets and the logarithm of total sales which were analyzed as financial variables, and may also be an enterprise size control variable, were removed from further study for weak discriminatory power. This can be explained by the fact that the model did not include enterprises of different sizes, but SMiE.

After analyzing independent variable coefficients and the weight of evidence of the developed model it can be seen that they conform to the economic logic. For example, the more effective and profitable the enterprise, the smaller the PD; a longer delay concerning repayment of debts means increased PD. If there were instances of valid seizure of assets, the PD was larger. This means that if the coefficient is positive and the independent variable increases, the probability of the event increases ($Y = 1$, in this case, that an enterprise becomes insolvent) and the opposite is true as well, if the coefficient is negative as the independent variable value increases, the higher the probability that $Y = 0$, i.e., that the enterprise will remain solvent. It can be stated that the developed models meet the high interpretability of results of a model.

According to the models formed, the enterprise probability of default (PD) is calculated by Equation (3), where the enterprises with $PD > 50\%$ were classified into the group of bad debtors, and the enterprises with $PD < 50\%$ were classified into the group of good debtors. According to the models formed, the calculation of coefficient z is shown in Appendix C.

These models are suitable, as they comply with the following requirements: (i) chi-square criterion p -value is less than 0.05; (ii) Cox and Snell R Square and Nagelkerke R Square is not less than 0.2; (iii) p values of Wald's p -value is less than 0.05, i.e., the statistically significant variables are included in the model; (iv) the total percentage of the model's correctly classified cases is 72.3–76.2% (Table 1). Furthermore, Figure 1 shows ROC curves, and the area under the curves (AUC) is presented in Table 2.

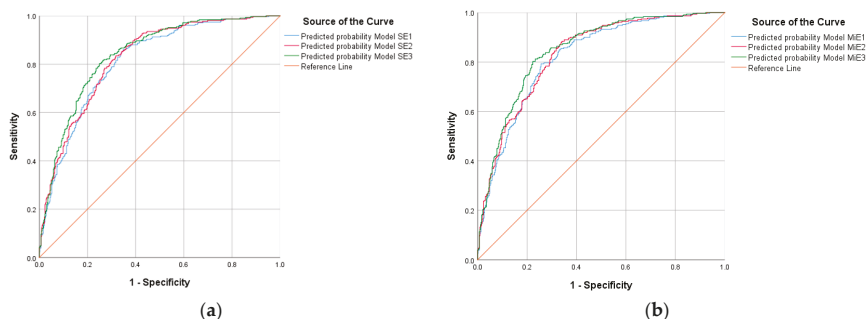


Figure 1. Receiver operating characteristic (ROC) curve: (a) for SE models; (b) for MiE models.

Table 2. Area under the curve (for SE and MiE models).

Test Result Variable(s): Predicted Probability for	Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Model SE1	0.814	0.016	0.000	0.783	0.844
Model SE2	0.825	0.015	0.000	0.795	0.854
Model SE3	0.840	0.015	0.000	0.811	0.868
Model MiE1	0.821	0.015	0.000	0.791	0.851
Model MiE2	0.831	0.015	0.000	0.802	0.860
Model MiE3	0.847	0.014	0.000	0.819	0.875

Note: ^a Under the nonparametric assumption; ^b Null hypothesis: true area = 0.5.

The area under the curves was 0.814 with a 95% confidence interval (0.783, 0.844) for Model SE1; 0.825 and (0.795, 0.854) for Model SE2; 0.840 and (0.811, 0.868) for Model SE3; 0.821 and (0.791, 0.851) for Model MiE1; 0.831 and (0.802, 0.860) for Model MiE2; 0.847 and (0.819, 0.875) for Model MiE3,

respectively (see Figure 1 and Table 2). Also, the area under the curve was significantly different from 0.5, since the p -value was 0.000, meaning that the logistic regression classified the group significantly better than by chance. The formed models met the accuracy requirement: models were characterized by high discriminatory power and the AUC values were over 80.

Finally, the models met the simplicity of a model, i.e., the application of the model and periodic review. The interpretation of the results of the model does not require particular knowledge in the field of statistics and information technologies. Therefore, trade creditors (suppliers) can use the developed ETCRA models to assess the trade credit risk of the trade credit debtors (customers). It should be noted that the data collection required to apply the models was not complicated. At the end of the financial year, enterprises are required to submit financial statements to the State Enterprise *Center of Registers*, and as well as this, credit and financial data of the enterprises are collected by the credit bureaus.

5. Conclusions

After performing the theoretical and empirical research, the following conclusions can be drawn:

- Financial and non-financial variables should be included in the statistical ETCRA models. Relative financial ratios were proposed to be used as financial variables. Non-financial variables were selected according to the data available for model development and external information infrastructure.
- Separate ETCRA models for small and micro-enterprises were developed during the research. Varied model variants have been created using (i) only financial ratios and (ii) financial ratios and non-financial variables.
- In the ETCRA model, the enterprise's financial performance is assessed from different perspectives: profitability, liquidity, solvency and activity. Profitability is expressed using the EBIT/S ratio in SE companies, and the NP/S ratio is used by MiE companies. The *liquidity ratio* CA/CL is used only for MiE. The *solvency ratio* Eq/TL and *activity ratio* S/TA are used in both groups of models for SE and MiE. To assess the probability of default for the trade credit debtor (customer), two non-financial variables are suggested to be used as the non-financial variables: the number of valid arrests and the average term of delay of delayed debts.
- Both hypotheses were not confirmed during the research. Thus, (i) the inclusion of non-financial variables in the model do not substantially improve the characteristics of the model. This means that models that use only financial ratios can be used in practice, and models that include non-financial variables can also be used. (ii) Compared to small enterprises, the characteristics of the ETCRA model of micro-enterprises by providing smaller amounts of information in the financial statements do not deteriorate. ETCRA models developed during the research are suitable for both small and micro-enterprises to predict the probability of the default.
- The designed ETCRA model is probabilistic and satisfies the requirements of the results of high interpretability, accuracy, and simplicity. Therefore, this model can be used by trade creditors (suppliers) when making decisions regarding the granting a trade credit for small or micro-enterprises. In addition, the opportunity to choose is given—trade creditors (suppliers) can evaluate the trade credit debtors (customers) using only the financial ratios or, additionally, non-financial variables.

Author Contributions: Conceptualization, R.K. and R.S.; methodology, R.K. and R.S.; formal analysis, R.K. and R.S.; data curation, R.K. and R.S.; writing—original draft preparation, R.K.; writing—review and editing, R.K.; supervision, R.K.

Funding: This research received no external funding.

Acknowledgments: The authors express their gratitude to *Creditinfo Lietuva* for data provided for the development of statistic ETCRA model.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Comparison of financial statements of enterprises of different sizes—statement of financial position.

Medium-Sized and Large Enterprises	Statements Generated by: Small Enterprises	Micro-Enterprises
Statement of Financial Position	Form of a Statement: Abridged Statement of Financial Position	Short Statement of Financial Position
<i>ASSETS:</i>	<i>ASSETS:</i>	<i>ASSETS:</i>
FIXED ASSETS	FIXED ASSETS	FIXED ASSETS
Intangible assets ¹	Intangible assets	-
Tangible assets ²	Tangible assets	Tangible assets
Financial assets ³	Financial assets	-
Other fixed assets ⁴	Other fixed assets	Other fixed assets
CURRENT ASSETS	CURRENT ASSETS	CURRENT ASSETS
Inventories ⁵	Inventories	Inventories
Amounts receivable within one year ⁶	Amounts receivable within one year	-
Short-term investments ⁷	Short-term investments	Other current assets
Cash and cash equivalents	Cash and cash equivalents	-
PREPAYMENTS AND ACCRUED INCOME	PREPAYMENTS AND ACCRUED INCOME	PREPAYMENTS AND ACCRUED INCOME
<i>EQUITY AND LIABILITIES:</i>	<i>EQUITY AND LIABILITIES:</i>	<i>EQUITY AND LIABILITIES:</i>
<i>EQUITY</i>	<i>EQUITY</i>	<i>EQUITY</i>
Capital ⁸	Capital	-
Share premium account	Share premium account	-
Revaluation reserve	Revaluation reserve	-
Reserves ⁹	Reserves	-
Retained profit (loss) ¹⁰	Retained profit (loss)	-
GRANTS, SUBSIDIES	GRANTS, SUBSIDIES	GRANTS, SUBSIDIES
PROVISIONS ¹¹	PROVISIONS	PROVISIONS
AMOUNTS PAYABLE AND OTHER LIABILITIES	AMOUNTS PAYABLE AND OTHER LIABILITIES	AMOUNTS PAYABLE AND OTHER LIABILITIES
Amounts payable after one year and other long-term liabilities ¹²	Amounts payable after one year and other long-term liabilities	Amounts payable after one year and other long-term liabilities
Amounts payable within one year and other short-term liabilities ¹³	Amounts payable within one year and other short-term liabilities	Amounts payable within one year and other short-term liabilities
ACCRUALS AND DEFERRED INCOME	ACCRUALS AND DEFERRED INCOME	ACCRUALS AND DEFERRED INCOME

The relevant items are detailed as follows: ¹ Assets arising from development; goodwill; software; concessions, patents, licenses, trademarks, and similar rights; other intangible assets; advance payments; ² Land; buildings and structures; machinery and plants; vehicles; other equipment, fittings, and tools; investment property (land; buildings); advance payments and tangible assets under construction (production); ³ Shares in entities of the entities group; loans to entities of the entities group; amounts receivable from entities of the entities group; shares in associated entities; loans to associated entities; amounts receivable from the associated entities; long-term investments; amounts receivable after one year; other financial assets; ⁴ Assets of the deferred tax on profit; biological assets; other assets; ⁵ Raw materials, materials, and consumables; production and work in progress; finished goods; goods for resale; biological assets; fixed tangible assets held for sale; advance payments; ⁶ Trade debtors; amounts owed by entities of the entities group; amounts owed by associates entities; other debtors; ⁷ Shares in entities of the entities group; other investments; ⁸ Authorized (subscribed) or primary capital; subscribed capital unpaid (-); own shares (-); ⁹ Compulsory reserve or emergency (reserve) capital; reserve for acquiring own shares; other reserves; ¹⁰ Profit (loss) for the reporting year; profit (loss) brought forward; ¹¹ Provisions for pensions and similar obligations; provisions for taxation; other provisions; ¹² Debenture loans; amounts owed to credit institutions; payments received on account; trade creditors; amounts payable under the bills and checks; amounts payable to the entities of the entities group; amounts payable to the associated entities; other amounts payable and long-term liabilities; ¹³ Debenture loans; amounts owed to credit institutions; payments received on account; trade creditors; amounts payable under the bills and checks; amounts payable to the entities of the entities group; amounts payable to the associated entities; liabilities of tax on profit; liabilities related to employment relations; other amounts payable and short-term liabilities.

Table A2. Comparison of financial statements of enterprises of different sizes—statement of profit or loss.

Statements Generated by:	
Large, Medium-Sized, and Small Enterprises	Micro-Enterprises
Form of a Statement:	
Statement of Profit or Loss	Short Statement of Profit or Loss
Revenue	Revenue
Cost of sales	Acquired stocks and used resources; The change in stocks value
Fair value adjustments of the biological assets	-
GROSS PROFIT (LOSS)	-
Selling expenses	-
General and administrative expenses	Expenses related to employment relations; Depreciation, amortization and impairment of assets
Other operating results	Other income Other expenses
Income from investments in the shares of parent, subsidiaries, and associated entities	-
Income from other long-term investments and loans	-
Other interest and similar income	-
The impairment of the financial assets and short-term investments	-
Interest and other similar expenses	-
PROFIT (LOSS) BEFORE INCOME TAXES	-
Income taxes	Income taxes
NET PROFIT (LOSS)	NET PROFIT (LOSS)

Appendix B

Table A3. Inclusion of variables into logistic regression model for SE.

Variables	Assigned name	Stage 1 *	Stage 2		Stage 3	Stage 4
			Stage 2.1	Stage 2.2		
Non-financial variable		Selected (S)/Not Selected (NS) for Further Analysis				
Number of outstanding debts	NOD	S	0.0920	Weak	S	S
Sum of outstanding debts	SOD	S	0.0430	Weak	S	S
Average term of delay of outstanding debts	ATDOD	S	0.0540	Weak	S	S
Number of delayed debts	NDD	S	0.0940	Weak	S	S
Sum of delayed debts	SDD	S	0.0860	Weak	S	S
Average term of delay of delayed debts	ATDDD	S	0.0870	Weak	S	S
Number of requests in credit bureau	NRCB	S	0.0340	Weak	NS	NS
Number of valid arrests	NVA	S	0.3740	Strong	S	S

Table A3. Cont.

Variables	Calculation formula	Stage 1 *	Stage 2		Stage 3	Stage 4
			Stage 2.1	Stage 2.2		
Financial variable (Financial ratio)						
<i>1a. Profitability ratios (return of sales)</i>						
Gross profit/sales	GP/S	S	0.0730	Weak	NS	NS
EBIT/sales	EBIT/S	S	0.6950	Very strong	S	S
EBT/sales	EBT/S	S	0.788	Very strong	S	NS
Net profit/sales	NP/S	S	0.7660	Very strong	S	NS
<i>1b. Profitability ratios (return of investment)</i>						
Gross profit/total assets	GP/TA	S	0.2150	Strong	S	S
EBIT/total assets	EBIT/TA	S	0.841	Very strong	S	NS
EBIT/current liabilities	EBIT/CL	S	0.6230	Very strong	S	NS
EBT/total assets	EBT/TA	S	0.9760	Very strong	S	NS
EBT/equity	EBT/Eq	S	0.3580	Strong	S	S
EBT/(equity – current liabilities)	EBT/(Eq-CL)	S	0.2150	Strong	S	NS
Net profit/total assets	ROA	S	0.9780	Very strong	S	NS
Net profit/equity	ROE	S	0.3850	Strong	S	NS
<i>2. Liquidity ratios</i>						
Current assets/current liabilities	CA/CL	S	0.7940	Very strong	S	NS
(Current assets – inventories)/current liabilities	(CA-INV)/CL	S	0.6700	Very strong	S	S
Inventories/current liabilities	INV/CL	S	0	Weak	NS	NS
Accounts receivable/total liabilities	AR/TL	S	0.3260	Strong	S	NS
Accounts receivable/(total liabilities – cash)	AR/(TL-Cash)	S	0.2630	Strong	S	NS
Cash/current liabilities	Cash/CL	S	0.9460	Very strong	S	NS
(Cash – inventories)/current liabilities	(Cash-INV)/CL	S	0.7030	Very strong	S	S
Cash/total liabilities	Cash/TL	S	0.8380	Very strong	S	NS
Cash/equity	Cash/Eq	S	1.0440	Suspiciously strong	S	NS
Working capital/total assets	WC/TA	S	0.8060	Very strong	S	S
Working capital/equity	WC/Eq	S	0.4100	Strong	S	NS
(Current liabilities – cash)/total assets	(CL-Cash)/TA	S	0.9660	Very strong	S	S
<i>3. Solvency ratios</i>						
Total liabilities/total assets	TL/TA	S	1.1870	Suspiciously strong	S	S
Equity/total assets	Eq/TA	S	1.1450	Suspiciously strong	S	NS
Equity/(equity + long term liabilities)	EQ/(EQ-LTTL)	S	0.4070	Strong	S	NS
Equity/total liabilities	Eq/TL	S	1.1830	Suspiciously strong	S	S
Fixed assets/equity	FA/Eq	S	0.8480	Very strong	S	NS
Current assets/total liabilities	CA/TL	S	0.6060	Very strong	S	NS
Current assets/(total liabilities – cash)	CA/(TL-Cash)	S	0.5940	Very strong	S	NS
<i>4. Activity ratios</i>						
Inventories/sales	INV/S	NS	-	-	NS	NS
Accounts receivable/sales	AR/S	NS	-	-	NS	NS
Sales/fixed assets	S/FA	NS	-	-	NS	NS
Sales/current assets	S/CA	S	0.0820	Weak	NS	NS
Sales/total assets	S/TA	S	0.2290	Strong	S	S
Sales/cash	S/Cash	S	0.0670	Weak	NS	NS
Equity/sales	S/Eq	S	0.9640	Very strong	S	S
Cost of sales/sales	CS/S	S	0.0780	Weak	NS	NS
Current liabilities/sales	CL/S	S	0.5370	Very strong	S	S
Working capital/sales	WC/S	S	0.6470	Very strong	S	NS
Working capital/operating expenses	WC/OE	NS	-	-	NS	NS
EBIT/interest expenses	EBIT/IE	NS	-	-	NS	NS
<i>5a. Structure ratios (total assets structure ratios)</i>						
Current assets/total assets	CA/TA	S	0.1680	Medium	NS	NS
Accounts receivable/inventories	AR/INV	NS	-	-	NS	NS
Inventories/total assets	INV/TA	S	0.0930	Weak	NS	NS
Cash/total assets	Cash/TA	S	0.3700	Strong	S	S
<i>5b. Structure ratios (equity and liabilities structure ratios)</i>						
Retained earnings/total assets	RE/TA	S	0.8580	Very strong	S	S
Current liabilities/(total liabilities – cash)	CL/(TL-Cash)	S	0.5840	Very strong	S	NS
<i>6. Other ratios</i>						
Logarithm of total assets	LogTA	S	0.141	Medium	NS	NS
Logarithm of total sales	LogS	S	0	Weak	NS	NS
Sales/capital stock	S/CS	S	0.074	Weak	NS	NS

* Stage 1—an analysis of missing values; Stage 2—discriminatory power analysis (2.1—IV value; 2.2—discriminatory power); Stage 3—correlation analysis; Stage 4—forward (LR) regression.

Table A4. Inclusion of variables into logistic regression model for MiE.

Variables	Assigned name	Stage 1	Stage 2		Stage 3		Stage 4
			Stage 2.1	Stage 2.2	Stage 3.1	Stage 3.2	
Non-financial variable			Selected (S)/Not Selected (NS) for Further Analysis				
Number of outstanding debts	NOD	S	0.0920	Weak	S	S	S
Sum of outstanding debts	SOD	S	0.0430	Weak	S	S	S
Average term of delay of outstanding debts	ATDOD	S	0.0540	Weak	S	S	S
Number of delayed debts	NDD	S	0.0940	Weak	S	S	S
Sum of delayed debts	SDD	S	0.0860	Weak	S	S	S
Average term of delay of delayed debts	ATDDD	S	0.0870	Weak	S	S	S
Number of requests in credit bureau	NRCB	S	0.0340	Weak	NS	NS	NS
Number of valid arrests	NVA	S	0.3740	Strong	S	S	S
Financial variable (Financial ratio)			Calculation formula				
<i>1a. Profitability ratios (return of sales)</i>							
Gross profit/sales	GP/S	S	0.0730	Weak	NS	NS	NS
EBIT/sales	EBIT/S	S	0.6950	Very strong	S	S	NS
EBT/sales	EBT/S	S	0.788	Very strong	S	S	NS
Net profit/sales	NP/S	S	0.7660	Very strong	S	S	S
<i>1b. Profitability ratios (return of investment)</i>							
Gross profit/total assets	GP/TA	S	0.2150	Strong	S	NS	NS
EBIT/total assets	EBIT/TA	S	0.841	Very strong	S	NS	NS
EBIT/current liabilities	EBIT/CL	S	0.6230	Very strong	S	NS	NS
EBT/total assets	EBT/TA	S	0.9760	Very strong	S	S	NS
EBT/equity	EBT/Eq	S	0.3580	Strong	S	S	S
EBT/(equity – current liabilities)	EBT/(Eq-CL)	S	0.2150	Strong	S	S	NS
Net profit/total assets	ROA	S	0.9780	Very strong	S	S	NS
Net profit/equity	ROE	S	0.3850	Strong	S	S	NS
<i>2. Liquidity ratios</i>							
Current assets/current liabilities	CA/CL	S	0.7940	Very strong	S	S	S
(Current assets – inventories)/current liabilities	(CA-INV)/CL	S	0.6700	Very strong	S	S	S
Inventories/current liabilities	INV/CL	S	0	Weak	NS	NS	NS
Accounts receivable/total liabilities	AR/TL	S	0.3260	Strong	S	NS	NS
Accounts receivable/(total liabilities – cash)	AR/(TL-Cash)	S	0.2630	Strong	S	NS	NS
Cash/current liabilities	Cash/CL	S	0.9460	Very strong	S	NS	NS
(cash – inventories)/current liabilities	(Cash-INV)/CL	S	0.7030	Very strong	S	NS	NS
Cash/total liabilities	Cash/TL	S	0.8380	Very strong	S	NS	NS
Cash/equity	Cash/Eq	S	1.0440	Suspiciously strong	S	NS	NS
Working capital/total assets	WC/TA	S	0.8060	Very strong	S	S	S
Working capital/equity	WC/Eq	S	0.4100	Strong	S	S	NS
(Current liabilities – cash)/total assets	(CL-Cash)/TA	S	0.9660	Very strong	S	NS	NS
<i>3. Solvency ratios</i>							
Total liabilities/total assets	TL/TA	S	1.1870	Suspiciously strong	S	S	S
Equity/total assets	Eq/TA	S	1.1450	Suspiciously strong	S	S	NS
Equity/(equity + long term liabilities)	EQ/(EQ-LTL)	S	0.4070	Strong	S	S	NS
Equity/total liabilities	Eq/TL	S	1.1830	Suspiciously strong	S	S	S
Fixed assets/equity	FA/Eq	S	0.8480	Very strong	S	S	NS
Current assets/total liabilities	CA/TL	S	0.6060	Very strong	S	S	NS
Current assets/(total liabilities – cash)	CA/(TL-Cash)	S	0.5940	Very strong	S	NS	NS
<i>4. Activity ratios</i>							
Inventories/sales	INV/S	NS	-	-	NS	NS	NS
accounts receivable/sales	AR/S	NS	-	-	NS	NS	NS
Sales/fixed assets	S/FA	NS	-	-	NS	NS	NS
Sales/current assets	S/CA	S	0.0820	Weak	NS	NS	NS
Sales/total assets	S/TA	S	0.2290	Strong	S	S	S
Sales/cash	S/Cash	S	0.0670	Weak	NS	NS	NS
Equity/sales	S/Eq	S	0.9640	Very strong	S	S	S
Cost of sales/sales	CS/S	S	0.0780	Weak	NS	NS	NS
Current liabilities/sales	CL/S	S	0.5370	Very strong	S	S	S
Working capital/sales	WC/S	S	0.6470	Very strong	S	S	NS
Working capital/operating expenses	WC/OE	NS	-	-	NS	NS	NS
EBIT/interest expenses	EBIT/IE	NS	-	-	NS	NS	NS
<i>5a. Structure ratios (total assets structure ratios)</i>							
Current assets/total assets	CA/TA	S	0.1680	Medium	NS	NS	NS
Accounts receivable/inventories	AR/INV	NS	-	-	NS	NS	NS
Inventories/total assets	INV/TA	S	0.0930	Weak	NS	NS	NS
Cash/total assets	Cash/TA	S	0.3700	Strong	S	NS	NS
<i>5b. Structure ratios (equity and liabilities structure ratios)</i>							
Retained earnings/total assets	RE/TA	S	0.8580	Very strong	S	S	NS
Current liabilities/(total liabilities – cash)	CL/(TL-Cash)	S	0.5840	Very strong	S	S	NS
<i>6. Other ratios</i>							
Logarithm of total assets	LogTA	S	0.141	Medium	NS	NS	NS
Logarithm of total sales	LogS	S	0	Weak	NS	NS	NS
Sales/capital stock	S/CS	S	0.074	Weak	NS	NS	NS

Note: * Stage 1—An analysis of missing values; Stage 2—Discriminatory power analysis (2.1—IV value; 2.2—Discriminatory power); Stage 3.1—The ratios are not calculated for micro-enterprises; Stage 3.2—Correlation analysis; Stage 4—Forward (LR) regression.

Appendix C

Table A5. The calculation of coefficient z for the logistic regression enterprise trade credit risk assessment (ETCRA) models for small (SE) and micro-enterprises (MiE).

Varied Model Variants Have been Created Using:	Models for SE	Models for MiE
(1) only financial ratios	$z_{SE1} = -0.675 - 1.765 \times \text{EBIT}/S - 0.577 \times \text{Eq}/\text{TL} + 0.088 \times S/\text{TA}$	$z_{MiE1} = -0.578 - 1.861 \times \text{NP}/S - 0.174 \times \text{CA}/\text{CL} - 0.438 \times \text{Eq}/\text{TL} + 0.094 \times S/\text{TA}$
(2) financial ratios and non-financial variables:		
(i) model uses financial ratios and one non-financial variable (NVA)	$z_{SE2} = 0.042 - 1.663 \times \text{EBIT}/S - 0.520 \times \text{Eq}/\text{TL} + 0.107 \times S/\text{TA} + k_{SE2}^*$	$z_{MiE2} = 0.093 - 1.733 \times \text{NP}/S - 0.154 \times \text{CA}/\text{CL} - 0.401 \times \text{Eq}/\text{TL} + 0.113 \times S/\text{TA} + k_{MiE2}^*$
(ii) model uses financial ratios and two non-financial variables (NVA and ATDOD)	$z_{SE3} = -0.526 - 1.764 \times \text{EBIT}/S - 0.564 \times \text{Eq}/\text{TL} + 0.105 \times S/\text{TA} + k_{SE3}^* + n_{SE}^{**}$	$z_{MiE3} = -0.495 - 1.832 \times \text{NP}/S - 0.167 \times \text{CA}/\text{CL} - 0.439 \times \text{Eq}/\text{TL} + 0.110 \times S/\text{TA} + k_{MiE3}^* + n_{MiE}^{**}$
where:		
* k denotes number of valid arrests	Model SE2: if NVA > 0: $k_{SE2} = -1.076$, if not—0; Model SE3: if NVA > 0: $k_{SE3} = -1.250$, if not—0	Model MiE2: if NVA > 0: $k_{MiE2} = -1.026$, if not—0; Model MiE3: if NVA > 0: $k_{MiE3} = -1.204$, if not—0
** n describes the average term of delay of outstanding debts	when delayed payment for more than 90 calendar days, $n = 0$; when an enterprise does not have delayed payment, $n_{SE} = 0.971$; in over case $n_{SE} = 1.063$	when delayed payment for more than 90 calendar days, $n = 0$; when an enterprise does not have delayed payment, $n_{MiE} = 1.034$; in over case $n_{MiE} = 1.053$

References

- Abdullah, Nur Adiana Hiau, Muhammad M. Ma'aji, and Hwei Khaw. 2016. The value of governance variables in predicting financial distress among small and medium-sized enterprises in Malaysia. *Asian Academy of Management Journal of Accounting and Finance* 12: 77–91. [\[CrossRef\]](#)
- Afrifa, Godfred Adjapong, and Ernest Gyapong. 2017. Net trade credit: What are the determinants? *International Journal of Managerial Finance* 13: 246–66. [\[CrossRef\]](#)
- Altman, Edward. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 23: 589–609. [\[CrossRef\]](#)
- Altman, Edward, and Gabriele Sabato. 2007. Modelling credit risk for SMEs: Evidence from the US market. *ABACUS-A Journal of Accounting Finance and Business Studies* 43: 332–57. [\[CrossRef\]](#)
- Behr, Patrick, and Andre Guettler. 2007. Credit risk assessment and relationship lending: An empirical analysis of German small and medium-sized enterprises. *Journal of Small Business Management* 45: 194–213. [\[CrossRef\]](#)
- Bekhet, Hussain Ali, and Shorouq Fathi Kamel Eleter. 2014. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance* 4: 20–28. [\[CrossRef\]](#)
- Brigham, Eugene, and Joel Houston. 2004. *Fundamentals of Financial Management*, 10th ed. Mason: Thomson/South-Western.
- Butkus, Mindaugas, Sigita Žakarė, and Diana Cibulskienė. 2014. Bankruptcy diagnostic model and its application to predict company's bankrupt likelihood in Lithuania. *Applied Economics: Systematic Research* 8: 111–32.
- Cateni, Silvia, Marco Vannucci, Marco Vannucci, and Valentina Colla. 2012. Variable Selection and Feature Extraction through Artificial Intelligence Techniques. In *Multivariate Analysis in Management, Engineering and Science*. Edited by Valim de Freitas and A. P. Barbosa Rodrigues de Freitas. London: INTECH, vol. 6, pp. 103–17. [\[CrossRef\]](#)
- Chang, Yung-Chia, Kuei-Hu Chang, and Guan-Jhih Wu. 2018. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing* 73: 914–20. [\[CrossRef\]](#)
- Chesser, Delton. 1974. Predicting loan noncompliance. *Journal of Commercial Bank Lending* 58: 28–38.
- Cho, Sang Jun, Chune Young Chung, and Jason Young. 2019. Study on the Relationship between CSR and Financial Performance. *Sustainability* 11: 343. [\[CrossRef\]](#)

- Ciampi, Francesco, and Niccolo Gordini. 2013. Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. *Journal of Small Business Management* 51: 23–45. [CrossRef]
- Crone, Sven F., and Steven Finlay. 2012. Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting* 28: 224–38. [CrossRef]
- Crook, Jonathan, David Edelman, and Lyn Thomas. 2007. Recent developments in consumer credit risk assessment. *European Journal of Operational Research* 183: 1447–65. [CrossRef]
- Cunat, Vicente. 2007. Trade Credit: Suppliers as Debt Collectors and Insurance Providers. *The Review of Financial Studies* 20: 491–527. [CrossRef]
- Danenas, Paulius, Gintautas Garsva, and Rimvydas Simutis. 2011. Development of Discriminant Analysis and Majority-Voting Based Credit Risk Assessment Classifier. In Paper presented at Proceedings of the 2011 international Conference on Artificial Intelligence, ICAI, Las Vegas, NV, USA, July 18–21; vol. 1, pp. 204–9.
- Dang, Chongyu, Zhichuan (Frank) Li, and Chen Yang. 2018. Measuring Firm Size in Empirical Corporate Finance. *Journal of Banking & Finance* 86: 159–76.
- Dzidzeviciute, Laima. 2013. Possibilities of the Statistical Scoring Models' Application at Lithuanian Banks. Ph.D. dissertation, Vilnius University, Vilnius, Lithuania; 237p.
- Fabbri, Daniela, and Leora F. Klapper. 2016. Bargaining power and trade credit. *Journal of Corporate Finance* 41: 66–80. [CrossRef]
- Fernandes, Guilherme, and Rinaldo Artes. 2016. Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research* 249: 517–24. [CrossRef]
- Figini, Silvia, Federico Bonelli, and Emanuele Giovannini. 2017. Solvency prediction for small and medium enterprises in banking. *Decision Support Systems* 102: 91–97. [CrossRef]
- Florez-Lopez, Raquel, and Juan Manuel Ramon-Jeronimo. 2015. Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal. *Expert Systems with Applications* 42: 5737–53. [CrossRef]
- Garcia-Teruel, Pedro Juan, and Pedro Martinez-Solano. 2010a. Determinants of trade credit: A comparative study of European SMEs. *International Small Business Journal* 28: 215–33. [CrossRef]
- Garcia-Teruel, Pedro Juan, and Pedro Martinez-Solano. 2010b. A Dynamic Approach to Accounts Receivable: A Study of Spanish SMEs. *European Financial Management* 16: 400–21. [CrossRef]
- Grigaravicius, Saulius. 2003. Corporate Failure Diagnosis. Reliability and Practice. *Management of Organizations: Systematic Research* 28: 29–42.
- Han, Jun-Tae, Jae-Seok Choi, Myeon-Jung Kim, and Jina Jeong. 2018. Developing a Risk Group Predictive Model for Korean Students Falling into Bad Debt. *Asian Economic Journal* 32: 3–14. [CrossRef]
- Hand, David J., and William E. Henley. 1997. Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society Series A—Statistics in Society* 160: 523–41. [CrossRef]
- Herrador-Alcaide, Teresa, and Montserrat Hernandez-Solis. 2019. Empirical Study Regarding Non-Financial Disclosure for Social Conscious Consumption in the Spanish E-Credit Market. *Sustainability* 11: 866. [CrossRef]
- Hill, Matthew D., Gary W. Kelly, Lorenzo A. Preve, and Virginia Sarria-Allende. 2017. Trade Credit or Financial Credit? An International Study of the Choice and Its Influences. *Emerging Markets Finance and Trade* 53: 2318–32. [CrossRef]
- International Financial Reporting Standards (IFRSs). 2008. Commission Regulation (EC) No. 1126/2008 of 3 November 2008 Adopting Certain International Accounting Standards in Accordance with Regulation (EC) No. 1606/2002 of the European Parliament and of the Council. Available online: <http://eur-lex.europa.eu> (accessed on 26 February 2019).
- Jacobson, Tor, and Erik von Schedvin. 2015. Trade credit and the propagation of corporate failure: An empirical analysis. *Econometrica* 83: 1315–71. [CrossRef]
- Kosmidis, Kosmas, and Antonios Stavropoulos. 2014. Corporate failure diagnosis in SMEs: A longitudinal analysis based on alternative prediction models. *International Journal of Accounting and Information Management* 22: 49–67. [CrossRef]
- Lessmann, Stefan, Bart Baesen, Hsin-Vonn Seow, and Lyn C. Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research* 247: 124–36. [CrossRef]

- Li, Frank. 2016. Endogeneity in CEO power: A survey and experiment. *Investment Analysts Journal* 45: 149–62. [CrossRef]
- Li, Frank, Tao Li, and Dylan Minor. 2016. CEO power, corporate social responsibility, and firm value: A test of agency theory. *International Journal of Managerial Finance* 12: 611–28. [CrossRef]
- Lin, Tsung-Te, and Jian-Hsin Chou. 2015. Trade credit and bank loan: Evidence from Chinese firms. *International Review of Economics & Finance* 36: 17–29.
- Mahata, Gour Chandra, and Sujit Kumar De. 2016. An EOQ inventory system of ameliorating items for price dependent demand rate under retailer partial trade credit policy. *Operations Research & Decision Theory OPSEARCH* 53: 889–916.
- Manab, Norlida Abdul, Ng Yen Theng, and Rohani Md-Rus. 2015. The Determinants of Credit Risk in Malaysia. *Procedia-Social and Behavioral Sciences* 172: 301–8. [CrossRef]
- Martinez-Sola, Cristina, Pedro J. Garcia-Teruel, and Pedro Martinez-Solano. 2017. SMEs access to finance and the value of supplier financing. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad* 46: 455–83. [CrossRef]
- McGuinness, Gerard, and Teresa Hogan. 2016. Bank credit and trade credit: Evidence from SMEs over the financial crisis. *International Small Business Journal* 34: 412–45. [CrossRef]
- Mileris, Ricardas. 2012. Assessment of enterprise default probability by credit rating model—[monių finansinių įsipareigojimų nevykdymo tikimybės vertinimas nustatant kredito reitingus. *Applied Economics: Systematic Research* 6: 127–43.
- Niklis, Dimitrios, Michael Doumpos, and Constantin Zopounidis. 2014. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. *Applied Mathematics and Computation* 234: 69–81. [CrossRef]
- Nikolic, Nebojsa, Nevenka Zarkic-Joksimovic, Djordje Stojanovski, and Iva Joksimovic. 2013. The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications* 40: 5932–44. [CrossRef]
- Paul, Salima Y., and Rebecca Boden. 2011. Size matters: The late payment problem. *Journal of Small Business and Enterprise Development* 18: 732–47. [CrossRef]
- Petropoulos, Anastasios, Sotirio Chartzis, and Stylianos Xanthopoulos. 2016. A novel corporate credit rating system based on Student's-t hidden Markov models. *Expert Systems with Applications* 53: 87–105. [CrossRef]
- Pike, Richard H., and Nam Sang Cheng. 2001. Credit management: An examination of policy choices, practices and late payment in UK companies. *Journal of Business Finance and Accounting* 28: 1013–41. [CrossRef]
- Purvinis, Ojaras, Povilas Sukys, and Ruta Virbickaite. 2005. Research of the Possibility of Bankruptcy Diagnostics Applying Neural Network. *Engineering Economics* 41: 16–22.
- Ramboll Management. 2005. *Report on Impacts of Raised Thresholds Defining SMEs*; Stockholm. Available online: <https://docplayer.net/storage/24/3408074/1560288073/UtH7rKTuqd84H219mPWwGQ/3408074.pdf> (accessed on 26 February 2019).
- Shenoy, Jaideep, and Ryan Williams. 2017. Trade credit and the joint effects of supplier and customer financial characteristics. *Journal of Financial Intermediation* 27: 68–80. [CrossRef]
- Siddiqi, Naeem. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. Hoboken: John Wiley & Sons.
- Sohn, So Young, Dong Ha Kim, and Jin Hee Yoon. 2016. Technology credit scoring model with fuzzy logistic regression. *Applied Soft Computing* 43: 150–58. [CrossRef]
- Sousa, Maria Rocha, João Gama, and Elísio Brandão. 2016. A new dynamic modeling framework for credit risk assessment. *Expert Systems with Application* 45: 341–51.
- Spicas, Renatas. 2017. Statistical Credit Risk Assessment Model of Small and Very Small Enterprises for Lithuanian Credit Unions. Ph.D. dissertation, Vilnius University, Vilnius, Lithuania; 236p.
- Spicas, Renatas, Rasa Kanapickiene, and Monika Ivaskeviciute. 2015. Filter Methods of Variable Selection for Enterprise Credit Risk Prediction. Paper presented at 15th International Scientific Conference on Perspectives of Business and Entrepreneurship Development—Economic, Management, Finance and System Engineering from the Academic and Practitioners Views, Brno Univ Technol, Fac Business and Management, Brno, Czech Republic, May 28–29.

- Spicas, Renatas, Rasa Kanapickiene, Mindaugas Vijunas, and Robertas Kirka. 2018. Development of Enterprise Credit Risk Assessment Model for Lithuanian Credit Unions. *Transformations in Business & Economics* 17: 152–77.
- Taffler, Richard, and Howard Tisshaw. 1977. Going, Going, Gone—Four Factors Which Predict. *Accountancy* 88: 50–54.
- Tascon, Maria, Francisco Castano, and Paula Castro. 2018. A new tool for failure analysis in small firms: Frontiers of financial ratios based on percentile differences (PDFR). *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad* 47: 433–63. [CrossRef]
- Terdpaong, Kanitsorn, and Dessalegn Getie Mihret. 2011. Modelling SME credit risk: Thai Empirical evidence. *Small Enterprise Research* 18: 63–79. [CrossRef]
- Tsao, Yu-Chung. 2010. Two-phase pricing and inventory management for deteriorating and fashion goods under trade credit. *Mathematical Methods of Operations Research* 72: 107. [CrossRef]
- Tsao, Yu-Chung. 2018. Trade credit and replenishment decisions considering default risk. *Computers & Industrial Engineering* 117: 41–46.
- Tsuruta, Daisuke. 2015. Bank loan availability and trade credit for small businesses during the financial crisis. *Quarterly Review of Economics and Finance* 55: 40–52. [CrossRef]
- Valvoniš, Vytautas. 2008. Credit Risk Assessment and Management Model: Practice and Perspectives of Lithuanian Banks. Ph.D. dissertation, Vilnius University, Vilnius, Lithuania; 204p.
- Verbraken, Thomas, Cristian Bravo, Richard Weber, and Bart Baesens. 2014. Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research* 238: 505–13. [CrossRef]
- Wang, Kai, Ruiqing Zhao, and Jin Peng. 2018. Trade credit contracting under asymmetric credit default risk: Screening, checking or insurance. *European Journal of Operational Research* 266: 554–68. [CrossRef]
- Xiao, Hongshan, Zhi Xiao, and Yu Wang. 2016. Ensemble classification based on supervised clustering for credit scoring. *Applied Soft Computing* 43: 73–86. [CrossRef]
- Yang, Chenguang, and Xiao-Bo Duan. 2008. Credit Risk Assessment in Commercial Banks Based on SVM Using PCA. *Machine Learning and Cybernetics* 2: 1207–11.
- Yap, Bee Wah, Seng Huat Ong, and Nor Huselina Mohamed Husain. 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications* 38: 13274–83. [CrossRef]
- Yoshino, Naoyuki, and Farhad Taghizadeh-Hesary. 2014. Analytical framework on credit risks for financing small and medium-sized enterprises in Asia. *Asia-Pacific Development Journal* 21: 1–21. [CrossRef]
- Zavgren, Christine. 1985. Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis. *Journal of Business Finance & Accounting* 12: 19–45.
- Zhu, You, Chi Xie, Bo Sun, Gang-Jin Wang, and Xin-Guo Yan. 2016. Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models. *Sustainability* 8: 433. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Credit Valuation Adjustment Compression by Genetic Optimization

Marc Chataigner and Stéphane Crépey *

LaMME, Univ Evry, CNRS, Université Paris-Saclay, CEDEX, 91037 Évry, France; marc.chataigner@univ-evry.fr

* Correspondence: stephane.crepey@univ-evry.fr

Received: 4 August 2019; Accepted: 23 September 2019; Published: 29 September 2019

Abstract: Since the 2008–2009 financial crisis, banks have introduced a family of X-valuation adjustments (XVAs) to quantify the cost of counterparty risk and of its capital and funding implications. XVAs represent a switch of paradigm in derivative management, from hedging to balance sheet optimization. They reflect market inefficiencies that should be compressed as much as possible. In this work, we present a genetic algorithm applied to the compression of credit valuation adjustment (CVA), the expected cost of client defaults to a bank. The design of the algorithm is fine-tuned to the hybrid structure, both discrete and continuous parameter, of the corresponding high-dimensional and nonconvex optimization problem. To make intensive trade incremental XVA computations practical in real-time as required for XVA compression purposes, we propose an approach that circumvents portfolio revaluation at the cost of disk memory, storing the portfolio exposure of the night so that the exposure of the portfolio augmented by a new deal can be obtained at the cost of computing the exposure of the new deal only. This is illustrated by a CVA compression case study on real swap portfolios.

Keywords: counterparty risk; credit valuation adjustment (CVA); XVA (X-valuation adjustments) compression; genetic algorithm

MSC: 91B30; 91G20; 91G40; 91G60

JEL Classification: D52; D53; G01; G13; G24

1. Introduction

XVAs, where VA stands for valuation adjustment and X is a catch-all letter to be replaced by C for credit, F for funding, M for margin, or K for capital, have been implemented by banks in reaction to the regulatory changes aroused by 2008 financial turmoils. They monetize counterparty risk and its funding and capital consequences by add-ons to derivative entry prices sourced from clients. According to the cost-of-capital XVA approach of [Albanese and Crépey \(2019\)](#), accounting for the impossibility for a bank to replicate the jump-to-default related cash flows, the final, all-inclusive XVA formula reads

$$CVA + FVA + MVA + KVA, \quad (1)$$

to be sourced by the bank from clients on an incremental run-off basis at every new deal.

As stated by the [Basel Committee on Banking Supervision \(2015\)](#), major counterparty credit losses on OTC (Over-the-counter) derivative portfolios in 2008 arose from XVA accounting losses rather than from actual client defaults. In particular, a bank incurs a CVA loss when the market perceives a deterioration of the credit risk of a client. This has motivated the creation of XVA desks for dealing with these risks.

In this paper, we deal with CVA compression, i.e., the minimization of the CVA of a client portfolio by the introduction of an incremental trade, subject to the constraint of not altering too much the market risk of the portfolio. In the financial derivative industry, the term compression term is generally applied in the context of “trade compression”, i.e., the reduction of the gross notional of positions in the market. Trade compression aims notably at reducing certain capital requirements, the number of transactions, and their amount (see Section 5.3 of (Gregory 2015)). As reflected by the proliferation of related industry presentations¹, this kind of balance sheet optimization is very active in top tier banks at the moment.

XVAs reflect market inefficiencies that should be compressed as much as possible. Here, we focus on CVA compression for concreteness and simplicity, but the developed XVA compression methodology is generic. It could and should be applied to further XVA metrics, as soon as these are available with sufficient speed, for computation at the portfolio level, and accuracy, for numerical significance of the results at the incremental trade level (see Section 5 in (Albanese et al. 2018), which emphasizes the XVA compression perspective on the pricing and risk management of financial derivatives in the post-2008–2009 global financial crisis era, and cf. (Konratyev and Giorgidze 2017), which uses a genetic algorithm for determining an optimal trade-off between MVA compression and transaction costs).

The complexity of XVA compression problems stems, in particular, from the hybrid nature of the state space of the corresponding optimization problems. Indeed, a new trade (financial derivative) is described by a combination of continuous and discrete parameters. This rules out the use of standard convex optimization schemes for such problems. Instead, we are lead to the use of metaheuristic algorithms. In this paper, we show how a genetic algorithm with penalization can efficiently find a CVA offsetting trade, while limiting the impact of the trade on the market exposure profile. The latter is necessary for staying in line with the separation of mandates between the XVA desks, in charge of managing counterparty risk, and the other, dubbed “clean”, trading desks of the bank, in charge of hedging the market risk of the bank positions.

The other XVA compression challenge is execution time, with intensive valuation of the involved XVA metrics as a bottleneck. The XVA metrics are primarily defined at the portfolio level: Time-0 XVAs can be formulated as expectations of nonlinear functionals of the bank derivative portfolio exposure, i.e., “clean” valuation (or “mark-to-market” (MtM) ignoring counterparty risk) of the bank portfolio, assessed at randomly sampled times and scenarios. Each new deal gives rise to XVA add-ons computed as the corresponding trade incremental XVA amounts, i.e., the differences between the XVAs of the portfolios including and excluding the new deal. To make intensive trade incremental XVA computations practical in real-time as required for XVA compression purposes, our proposed MtM store-and-reuse approach circumvents clean revaluation at the cost of disk memory, storing the portfolio exposure of the night so that the exposure of the portfolio augmented by a new deal can be obtained at the cost of computing the exposure of the new deal only.

Outline and Contributions

The paper is outlined as follows. Section 2 formulates the penalized CVA compression problem and introduces the related genetic optimization algorithm. Section 3 is about two key acceleration techniques in this regard. Section 4 presents a numerical case study on real swap portfolios. Section 5 concludes.

The main contributions of the paper are the design of a parallelized genetic algorithm for the CVA compression task, the MtM store-and-reuse acceleration technique for trade incremental XVA computations, and the numerical CVA compression case study on real swap portfolios.

¹ See also, e.g., David Bachelier’s panel discussion Capital and Margin Optimisation, Quantminds International 2018 conference, Lisbon, 16 May 2018.

More broadly, this paper enriches the literature on the use of genetic (also called evolutionary) optimization algorithms in finance. [Cont and Hamida \(2005\)](#) applied evolutionary algorithms to investigate a set of co-calibrated model parameterizations in order to assess the associated model risk. [Kroha and Friedrich \(2014\)](#) compared different genetic algorithms for automatic trading. [Jin et al. \(2019\)](#) applied evolutionary algorithms to optimal investment and consumption stochastic control problems. For wider reviews of genetic algorithms in finance, we refer the readers to ([Chen 2012](#); [Drake and Marks 2002](#)).

We refer the reader to the end of the paper for a list of the main abbreviations.

2. CVA Compression Modeling

2.1. Credit Valuation Adjustment

We consider a complete stochastic basis $(\Omega, \mathbb{F}, \mathbb{P})$, for a reference market filtration (ignoring the default of the bank itself) $\mathbb{F} = (\mathfrak{F}_t)_{t \in \mathbb{R}_+}$, satisfying the usual conditions, and a risk-neutral pricing measure \mathbb{P} , calibrated to market quotes of fully collateralized transactions. All the processes of interest are \mathbb{F} adapted and all the random times of interest are \mathbb{F} stopping times. This holds at least after so-called reduction of all the data to \mathbb{F} , starting from a larger filtration \mathbb{G} including the default of the bank itself as a stopping time, assuming immersion from \mathbb{F} into \mathbb{G} for simplicity (see ([Albanese and Crépey 2019](#)) for details). The \mathbb{P} expectation and $(\mathfrak{F}_t, \mathbb{P})$ conditional expectation are denoted by \mathbb{E} and \mathbb{E}_t , respectively.

In developed markets, the overnight indexed swap (OIS) rate is together the reference remuneration rate for posted collateral and the best market proxy for a risk-free rate. We denote by $r = (r_t)_{t \in \mathbb{R}_+}$ an \mathbb{F} progressive OIS rate process and we write $\beta = e^{-\int_0^\cdot r_s ds}$ for the corresponding risk-neutral discount factor.

By clean valuation or mark-to-market of a contract (or portfolio), we mean the (trade additive) risk-neutral conditional expectation of its OIS discounted future promised cash flows, ignoring counterparty risk and its capital and funding implications.

We consider a bank engaged in bilateral trading with a single corporate counterparty (client) with default time and recovery rate τ_c and R_c , respectively. This setup, which is chosen for simplicity, is consistent with a common situation where credit risk budget is assigned at each counterparty level within the bank. We denote by MtM the corresponding mark-to-market process of the client portfolio to the bank.

The (time 0) CVA of the bank is its expected discounted loss in case of client default, i.e.,

$$CVA = \mathbb{E}[\mathbb{1}_{\{\tau_c \leq T\}} \beta_t^{-1} \beta_{\tau_c} (1 - R_c) \text{MtM}_{\tau_c}^+]. \tag{2}$$

Assuming deterministic interest rates, this can be rewritten as

$$CVA = (1 - R_c) \int_0^T \beta_t \text{EPE}(t) \mathbb{P}(\tau_c \in dt), \tag{3}$$

where the expected positive exposure (EPE) is defined as

$$\text{EPE}(t) = \mathbb{E}(\text{MtM}_s^+ | s = \tau_c)_{\tau_c=t}. \tag{4}$$

Equation (3) is popular with practitioners because it allows obtaining the CVA as the integral of the EPE against the client CDS (Credit default swap) curve. However, it is only really practical in simplistic models where the market and credit sides of the problem are independent, so that $\text{EPE}(t) = \mathbb{E}(\text{MtM}_t^+)$. However, a key CVA modeling issue is wrong-way risk, i.e., the risk of adverse dependence between market and credit (see ([Brigo and Vrins 2018](#); [Crépey and Song 2016 2017](#); [Glasserman and Yang 2018](#); [Hull and White 2012](#); [Iben Taarit 2018](#); [Li and Mercurio 2015](#); [Pykhtin 2012](#))).

Assuming the client default time endowed with an intensity γ^c , a more flexible formula is

$$CVA = (1 - R_c) \mathbb{E} \int_0^T \beta_s e^{-\int_0^s \gamma_u^c du} \gamma_s^c \text{MtM}_s^+ ds. \tag{5}$$

Under a credit support agreement (CSA), MtM should be replaced by $(\text{MtM} - C)$ in all equations above, where C is the collateral posted by the counterparty. Obviously, collateral can mitigate the EPE and the CVA considerably. In the data of our case study there is no CSA, i.e., $C = 0$.

Non-linearity of MtM^+ with respect to the portfolio payoff components imposes CVA calculations at the counterparty portfolio (netting set) level.

Similar approaches apply to FVA computations, with analogous comments, whereas the MVA can be computed based on quantile regression for the embedded dynamic initial margin calculations (see (Crépey et al. 2019)). In any case, the numerical bottleneck of XVA computations lies in intensive MtM calculations.

2.2. Fitness Criterion

By the augmented and initial portfolios, respectively, we mean the portfolio of the bank inclusive and exclusive of a newly considered deal with the client. The aim of an XVA compression problem is to find a new trade that minimizes the corresponding XVA metric of the augmented portfolio. This is equivalent to minimize the incremental CVA, which we denote by

$$\begin{aligned} \Delta CVA &= (1 - R_c) \mathbb{E} \int_0^T \beta_s e^{-\int_0^s \gamma_u^c du} \gamma_s^c (\text{MtM}_s^{augm})^+ ds - (1 - R_c) \mathbb{E} \int_0^T \beta_s e^{-\int_0^s \gamma_u^c du} \gamma_s^c (\text{MtM}_s^{init})^+ ds \\ &= (1 - R_c) \mathbb{E} \int_0^T \beta_s e^{-\int_0^s \gamma_u^c du} \gamma_s^c ((\text{MtM}_s^{augm})^+ - (\text{MtM}_s^{init})^+) ds, \end{aligned} \tag{6}$$

where the indices *init* and *augm* refer to the initial portfolio and augmented portfolio, respectively. We emphasize that trade incremental CVA computations require two portfolio-wide calculations: one without the new trade and another one including it.

Minimizing an XVA metric is most easily obtained through a significant deformation of the portfolio exposure process (especially in the context of this work of a portfolio with a single counterparty). However, an XVA compression procedure should not affect too much the market risk of the portfolio, because market risk is the mandate of the clean desks of the bank, who, in particular, are subject to trading limits.

This motivates the addition of a penalization to the incremental XVA criterion. In our case study, the incremental deal will consist of an interest rate swap. As such product is mostly sensitive to interest rate moves, a natural penalization is then in terms of its DV01 (dollar value of an 01), i.e., the variation of its mark-to-market (at time 0) under a parallel shift of the yield curve by one basis point ($=10^{-4}$).

More precisely, an interest rate swap exchanges one leg indexed on a floating interest rate against one leg paying a fixed interest rate, called swap rate. The swap is said to be payer (respectively, receiver) for the party that pays (respectively, receives) the floating payments. A monocurrency swap exchanges both legs in the same currency. It is mainly sensitive to the fluctuations of the corresponding floating interest rate term structure. DV01 measures the associated risk as the difference between the prices of the swap under the baseline (actual market data observed in the real market) and for a bumped yield curve defined as the concatenation of the money market rates, forward rates, and swap rates, on the relevant (successive) time segments. Bumping the yield curve typically means adding 10^{-4} to each tenor of this curve and updating the other reference curves (zero coupon rates, forward rates, etc.) accordingly.

Focusing on the CVA metric in this paper, we obtain the following fitness minimization problem:

$$\underset{x \in \mathcal{A}}{\text{minimize}} \quad f(x) = \Delta CVA(x) + \alpha |DV01(x)|, \tag{7}$$

where x parameterizes a new deal (swap) to be found in a suitable search space \mathcal{A} (see Section 4.1), $\Delta\text{CVA}(x)$ is its incremental CVA (cf. Equation (6)), $\text{DV01}(x)$ is its DV01, and α is a penalization parameter controlling the trade-off between CVA reduction and market risk profile preservation. By solving Equation (7), we aim at identifying a new deal which, if added to the current client portfolio of the bank, would diminish its counterparty risk without impacting too much its market risk. Note that, for scaling reasons (with, in particular, market penalization), we address the XVA compression problem in terms of trade incremental (as opposed to augmented portfolio) XVA numbers.

A new deal is determined by a combination of quantitative (e.g., notional, maturity, etc.) and qualitative (e.g., currency, long or short direction, etc.) parameters, so that no gradient or Hessian is available for the fitness function f in Equation (7). Moreover, one is interested in exploring a variety of local minima of f , to see different trading opportunities emerge from the optimization procedure. Furthermore, we can guess that some (crucial) parameters need be learned first, such as currency or maturity; other parameters, such as notional, can be refined in a second stage. All these features lead us to addressing Equation (7) by means of a genetic optimization algorithm.

2.3. Genetic Optimization Algorithm

Genetic optimization algorithms belong to the class of derivative-free optimizers, which is surveyed and benchmarked numerically in (Rios and Sahinidis 2013) (including the CMA-ES and DAKOTA/EA genetic algorithms).

The idea of genetic (or evolutionary) optimization algorithms is to evolve a population of individuals through cycles of modification (mutation) and selection in order to improve the performance of its individuals, as measured by a given fitness function. In addition, so-called crossover is used to enhance the search in parameter space. To the best of our knowledge, evolutionary algorithms were first explicitly introduced in (Turing 1950, chp. 7 *Learning Machines*, p. 456). See the classical monographs by the authors of (Back 1996; Goldberg 1989; Holland 1975). They then experienced the general artificial intelligence disgrace and comeback before and after the 2000s. However, they always stayed an active field of research, seen from different perspectives, such as particle filtering, MCMC, or sequential monte carlo methods (see (Del Moral 2004)). Beyond its financial applications reviewed at the end of Section 1, genetic optimization has been used in many different fields, such as mechanics (Verma and Lakshminarayanan 2006), calibration of neural networks hyperparameters (Young et al. 2015), or operational research (Larranaga et al. 1999).

Genetic optimization offers no theoretically guaranteed rate of convergence, but it is often found the most efficient approach in practice for dealing with hybrid (partly continuous, partly discrete/combinatorial, hence without well defined gradient and Hessian), nonconvex (in the sense of one local minimum, at least, for each set of values of the discrete parameters), and high-dimensional optimization problems such as Equation (7).

At each iteration, the fitness $f(x)$ is computed for each individual (also named chromosome) x of an initial population (a set of chromosomes). The values returned by the objective function are used for selecting chromosomes from the population. Among numerous selection methods (see (Blickle and Thiele 1995)), we can quote fitness proportionate selection, ranking proportionate selection (in order to avoid the overrepresentation of the chromosomes with the highest fitness values), and tournament selection (selection of the best among randomly drawn chromosomes). The common intention of these selection methods is to sample in priority individuals with the best fitness values. A genetic algorithm is dubbed elitist if the selection operator always keeps the chromosome with the best fitness value. Otherwise (as in our case), there is no guarantee that the best visited chromosome is contained in the population corresponding to the final iteration.

The mutation stage is intended to maintain some diversity inside the population, in order to avoid the algorithm being trapped by local minima. A mutation randomly changes one gene, i.e., one component (e.g., in our case, the notional of a new swap) of a chromosome.

Selection and mutation play opposite roles: a focus on fitness leads to a quicker convergence toward a local minimum; conversely, a too heavily mutated population results in a slow random research.

In addition, a crossover operator plays the role of a reproduction inside the algorithm. The principle of crossover is to build two children chromosomes from parent chromosomes. A distribution (often the same as the one used for selection) is chosen for picking chromosomes from a population of the previous iteration and for recombining pairs of selected chromosomes. Children share gene values of their parents but a gene value from one parent cannot be inherited by both children. A crossover mask decides for each gene in which parent a child can copy the gene version. One of the most popular crossover masks is single point crossover (see Appendix A).

The role of the crossover operator is paradoxical, as crossover can be seen as a combination of mutations, which increase the genetic diversity, while crossover also promotes chromosomes with higher fitnesses. Crossover aims at benefiting of a presupposed proximity of best solutions.

The above operators are applied iteratively until a suitable stopping condition is satisfied. The most basic one is a fixed number of iteration, but customized criteria may also be used to limit further the number of iterations. For instance, the algorithm can be interrupted when the minimum (or sometimes even the maximum) fitness value within the population at the beginning of an iteration is below a predefined threshold.

See Algorithm 1 and Figure 1 for the algorithm in pseudo-code and skeleton forms, denoting by r_m the mutation rate, i.e., the percentage of individuals in a population affected by a mutation, and by r_c the crossover rate, i.e., the percentage of individuals affected by crossover recombination.

The behavior of a genetic optimization algorithm is essentially determined by the choice of the selection operator, the number of solutions affected by a mutation, and the number of chromosomes affected by a crossover. See the work of Tabassum and Kuruvilla (2014) for a user guide to the main genetic algorithm ingredients and Carvalho et al. (2011) for applications of genetic optimization algorithms to benchmark functions.

Algorithm 1: Pseudo-code of an optimization genetic algorithm

Data: An initial population \mathcal{P}_{init} of size P and the associated fitness values for each chromosome, a crossover rate r_c , and a mutation rate r_m .

Initialization; **while** a stopping condition is not satisfied **do**

Save $\lfloor (1 - r_c) * P \rfloor$ chromosomes, chosen by an appropriate selection method from \mathcal{P}_{init} , in $\mathcal{P}_{selected}$; Save $\lfloor r_c * P \rfloor$ chromosomes, chosen by an appropriate selection method from \mathcal{P}_{init} , in $\mathcal{P}_{crossover}$; Recombine, uniformly without replacement, $\lfloor \frac{r_c * P}{2} \rfloor$ pairs from $\mathcal{P}_{crossover}$; Merge $\mathcal{P}_{crossover}$ and $\mathcal{P}_{selected}$ in $\mathcal{P}_{mutated}$; Mutate randomly $\lfloor r_m * P \rfloor$ in $\mathcal{P}_{mutated}$;

for Each chromosome c in $\mathcal{P}_{mutated}$ **do**

Compute the fitness value of c ;

end

end

Result: A new population and the associated fitness values.

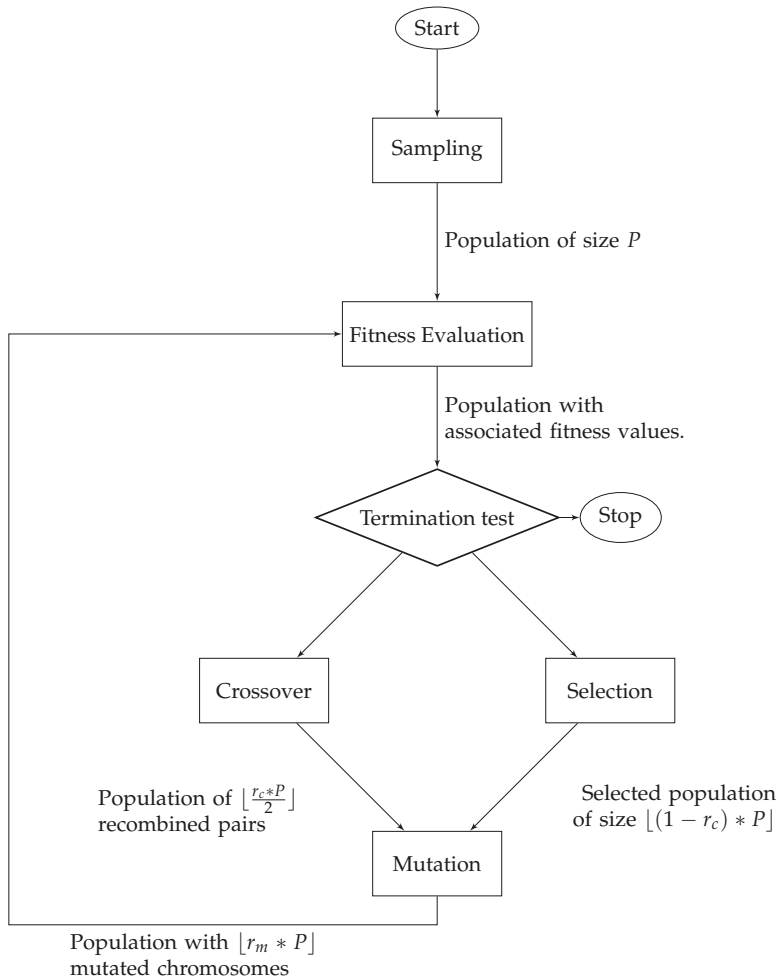


Figure 1. Skeleton of an optimization genetic algorithm.

3. Acceleration Techniques

Without suitable acceleration techniques, the above CVA compression approach is not workable in real time on realistic banking portfolios: in the examples of Section 4, a naive (desktop) implementation requires about 20 h of computations. This becomes even more problematic for hyperparameters tuning (such as α , crossover rate r_c , etc.). Hyperparameters are generally chosen with grid search, random search (see (Bergstra et al. 2011)), Bayesian optimization (see (Snoek et al. 2012)) or even evolutionary algorithms again (see (Young et al. 2015)). In any case, their calibration is greedy in terms of overall genetic algorithm execution.

In this section, we deal with the two following acceleration techniques, which may be used simultaneously:

- a MtM store-and-reuse approach for trade incremental XVA computations, speeding up the unitary evaluation of the fitness function; and
- a parallelization of the genetic algorithm accelerating the fitness evaluation at the level of the population.

3.1. MtM Store-and-Reuse Approach for Trade Incremental XVA Computations

Most of the time in portfolio-wide XVA calculations is spent in clean valuation (i.e., mark-to-market MtM) computations: by comparison, simulation of the risk factors or of the collateral is typically negligible.

Our case study is based on the CVA metric. As observed after Equation (6), by lack of trade-additivity of the (portfolio-wide) CVA, trade incremental XVA computations require two portfolio-wide calculations: one without the new trade and another one including it. However, it is possible to store the (including MtM) paths simulated for the initial portfolio and reuse them each time we want to compute a new trade incremental XVA. Then, each trade incremental XVA computation only requires the forward simulation of the mark-to-market process of the new deal.

The corresponding MtM store-and-reuse approach to trade incremental XVA computations circumvents repeated valuations at the cost of disk memory. It exploits the trade additivity of clean valuation by recording the MtM paths of the initial portfolio on a disk. For every new deal, the augmented portfolio exposure is obtained by adding, along the paths of the risk factors, the mark-to-market of the initial portfolio and of the new deal. This augmented portfolio exposure is then plugged into the XVA engine.

An optimally implemented MtM store-and-reuse approach brings down trade incremental XVA computations to the time of generating the clean price process of the trade itself, instead of the one of the augmented portfolio as a whole. Another advantage of this approach is its compliance with desk segregation: As far as clean valuation is concerned, the XVA desks just use the pricers of the clean desks. Hence, the MtM process plugged into the XVA computations is consistent with the one used for producing the market risk hedging sensitivities.

However, such an approach comes at the costs of memory disk (obviously), but also data slippage as, for consistency, it requires anchoring all the trade incremental XVA computations at the market data and parameters corresponding to the generation of the initial portfolio exposure. In practice, an MtM process at the overall portfolio level can only be generated during night runs, between two market sessions.

Moreover, we have to distinguish between first-order (or first-generation) XVAs, which are options on the MtM process, and higher-order (or second-generation) XVAs (see (Crépey et al. 2019)), which can be viewed as compound options of order two or more on the MtM process. Second-generation XVAs may also involve conditional risk measures, e.g., conditional value-at-risk for the dynamic initial margin calculations that are required for MVA² computations, as opposed to conditional expectations only in the case of first generation XVAs.

A Monte Carlo simulation diffuses risk factors X (such as interest rates, credit spreads, etc.) along drivers Z (such as Brownian motions, Poisson processes, etc.), according to a model formulated as a Markovian system of stochastic differential equations, starting from some given initial condition X_0 for all risk factors, suitably discretized in time and space. Modulo calibration, X_0 can be identified with the time 0 market data. We denote by \widehat{Y} a suitable estimate of a process Y at all (outer) nodes of a Monte Carlo XVA engine. In particular, $\widehat{\text{MtM}}$ is the fully discrete counterpart of the MtM process of the initial portfolio, namely the clean value of the portfolio at future exposure dates in a time grid and for different scenario paths.

At first sight, an MtM store-and-reuse approach is unsuitable for second-order XVAs, such as the MVA and the KVA (but also the CVA in the case of a CSA where the bank receives so-called initial margin). Indeed, in their case, the principle of swapping computations against storage would require to store not one portfolio exposure $\widehat{\text{MtM}}$, but a whole family of resimulated, future conditional portfolio exposures, (at least, over a certain time horizon), which seems hardly feasible in practice. However, even in the case of second-order XVA metrics, an MtM store and reuse approach can be

² For details regarding the initial margin and the MVA, see (Crépey et al. 2019, sets. 5.2 and 6.4).

implemented with the help of appropriate regression techniques (at the cost of an additional regression error; see (Crépey et al. 2019)).

Formalizing the above discussion, the conditions for a straightforward and satisfactory application of the MtM store-and-reuse approach to a given XVA metric are as follows, referring by indices *init*, *incr*, and *augm* to the initial portfolio, the new deal, and the augmented portfolio:

1. (No nested resimulation of the portfolio exposure required) The formula for the corresponding (portfolio-wide, time-0) XVA metric should be estimatable without nested resimulation, only based on the portfolio exposure rooted at $(0, X_0)$. A priori, additional simulation level makes impractical the MtM store-and-reuse idea of swapping execution time against storage.
2. (Common random numbers) \widehat{MtM}^{incr} should be based on the same paths of the drivers as \widehat{MtM}^{init} . Otherwise, numerical noise (or variance) would arise during \widehat{MtM} aggregation.
3. (Lagged market data) \widehat{MtM}^{incr} should be based on the same time, say 0, and initial condition X_0 (including, modulo calibration, market data), as \widehat{MtM}^{init} . This condition ensures a consistent aggregation of \widehat{MtM}^{init} and \widehat{MtM}^{incr} into \widehat{MtM}^{augm} .

These conditions have the following implications:

1. The first seems to ban second-order generation XVAs, such as CVA in presence of initial margin, but these can in fact be included with the help of appropriate regression techniques.
2. The second implies storing the driver paths that were simulated for the purpose of obtaining \widehat{MtM}^{init} ; it also puts a bound on the accuracy of the estimation of \widehat{MtM}^{incr} , since the number of Monte Carlo paths is imposed by the initial run. Furthermore, the XVA desks may want to account for some wrong way risk dependency between the portfolio exposure and counterparty credit risk (see Section 2.1); approaches based on correlating the default intensity and the market exposure in Equation (5) are readily doable in the present framework, provided the trajectories of the drivers and/or risk factors are shared between the clean and XVA desks.
3. The third induces a lag between the market data (of the preceding night) that are used in the computation of \widehat{MtM}^{incr} and the exact \widehat{MtM}^{incr} process; when the lag on market data becomes unacceptably high (because of time flow and/or volatility on the market), a full reevaluation of the portfolio exposure is required.

Figure 2 depicts the embedding of an MtM store-and-reuse approach into the trade incremental XVA engine of a bank.

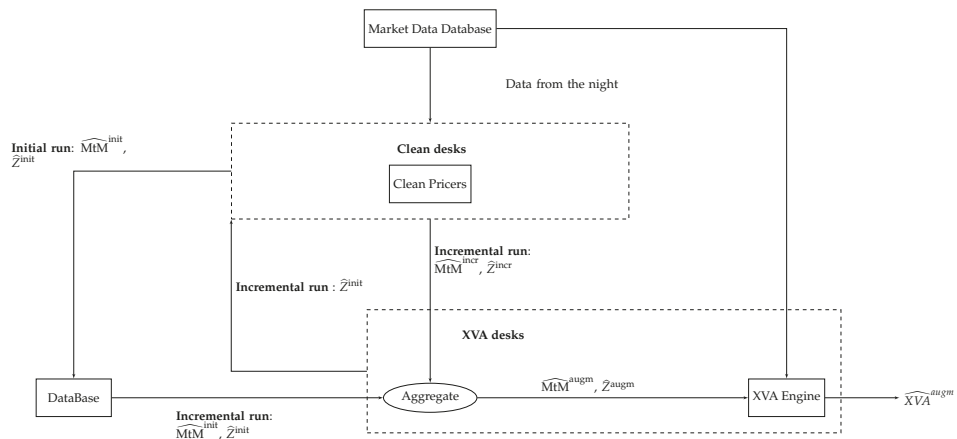


Figure 2. MtM store-and-reuse implementation of a trade incremental XVA engine with drivers Z.

3.2. Parallelization of the Genetic Algorithm

Most of the XVA compression computational time is spent in the evaluation of the incremental XVA metric involved in the fitness criterion visible in Equation (7). The MtM store-and-reuse approach allows reducing the complexity of such trade incremental XVA computations to trade (as opposed to portfolio) size. However, to achieve XVA compression in real time, this is not enough; another key step is the parallelization of the genetic algorithm that is used for solving Equation (7).

The genetic algorithm is a population based method, which implies to maintain a population of individuals (tentative new deals) through each iteration of the algorithm. The calculation of the objective function, for a given individual, does not depend on the fitness value of the other individuals. Therefore we can vectorize the computation of the fitness values within the population. Provided a suitable parallel architecture is available, a perfectly distributed genetic algorithm makes the execution time independent of the population size P (see Algorithm 1 and Figure 1).

This makes an important difference with other metaheuristic optimization algorithm, such as simulated annealing or stochastic hill climbing, which only evaluate one or very few solutions per iteration, but need much more iterations to converge toward a good minimum (see (Adler 1993; Janaki Ram et al. 1996)). As discussed in (Pardalos et al. 1995), the above parallelization of the fitness function evaluation, for a given population, should not be confused with a parallel genetic algorithm in the sense of an independent evolution of several smaller populations.

In our context, where individuals only represent incremental trades, a parallelization of population fitness evaluation is compatible with an MtM store-and-reuse approach for the trade incremental XVA computations. Combining the two techniques results in an XVA compression time independent of the sizes of the initial portfolio of the bank and of the population of the genetic algorithm used for the optimization, which represents an XVA compression computation time gain factor of the order of

$$\text{Number of trades in the initial portfolio} \times \text{population size.}$$

4. Case Study

In the remainder of the paper, we present CVA compression results on real swap portfolios³, using an additional swap for the CVA compression. We aim at addressing questions such as:

- Which type of swap is suitable for achieving the compression of the CVA, in the context of a given initial portfolio?
- How does the compression distort the portfolio exposure, with or without penalization?

To ease the implementation of the MtM store-and-reuse approach, we assume no CSA (cf. Section 3.1).

4.1. New Deal Parameterization

A swap is parameterized by its notional, its maturity, its direction, and its currency. The quantitative parameters are encoded through grids of values:

- Notional: From 10^5 to 10^7 by step of 10^5 dollars.
- Maturity: From 1 to 20 years by step of 1 year, 30 years and 50 years.

The qualitative parameters are encoded as enumerations of values:

- Currency: Euro, US dollar, GBP or Yen.
- Direction: A binary variable for payer or receiver.

³ The underlying interest rate and FX models are proprietary and cannot be disclosed in the paper. We use a deterministic credit spread model for the counterparty, calibrated to the CDS term structure of the latter.

Moreover, we impose the additional swap to be at par so that it can be entered at no cost, which is equally desirable from the bank and the client perspectives.

The above parameterization defines a discrete search space \mathcal{A} with $100 \times 22 \times 4 \times 2 = 1.76 \times 10^4$ elements.

4.2. Design of the Genetic Algorithm

We address the optimization problem in Equation (7) by a genetic algorithm as per Section 2.3. The new deal space \mathcal{A} in Equation (7) is viewed as a space of chromosomes x , the genes (deal parameters) of which evolve randomly along the iterations of the algorithm as detailed in Section 2.3.

In the theoretical literature on genetic algorithms, an individual is represented as a bit string. In practice, however, bit string representation of parameters does not give enough control on the mutation distribution. Namely, in bit string representation, mutations affect all bits uniformly, whereas we might want to mutate some parameters more frequently (the quantitative parameters, in particular, as the algorithm tends to quickly identify the relevant values of the qualitative parameters). Hence, we rather model our individuals x by a variable string, a choice also made in (Kondratyev and Giorgidze 2017).

We choose rank proportionate selection to avoid fitness scaling issues. More precisely, if we have a population $\mathcal{P} = \{1, \dots, P\}$ of P individuals and the associated fitnesses $(f_i)_{i \in \mathcal{P}}$, then the probability to select chromosome i is

$$p_i = \frac{2rank(f_i)}{P(P+1)},$$

where *rank* is a function that ranks chromosomes according to their fitness value (returning one for the highest value, in the context of a minimization problem).

Regarding the crossover operator, we use a uniform crossover mask, i.e., the choice of gene inherited from one parent or another is drawn with a uniform probability.

Regarding mutations, the probability to mutate a gene is proportional to the number of alleles (values) that it can take. The mutation operator then selects uniformly a new gene allele (value).

In our experiments, the notional of the new swap can take 100 different values, its currency 4 values, its position 2 values, and its maturity 22 values. Hence, when a chromosome is selected for mutation, the probability to mutate each of its genes is equal to $\frac{100}{128}$ for the notional, $\frac{22}{128}$ for the maturity, $\frac{4}{128}$ for the currency, and $\frac{2}{128}$ for the position. Indeed, a more frequent mutation of notional and maturity parameters are desirable. Diversity for currency and position is ensured at the initialization of the algorithm (with a large population) and maintained across the iterations thanks to the crossover operator. A prerequisite for a successful implementation is a reasonable specification of the search space \mathcal{A} .

Hyperparameters strongly impact the behavior of the algorithm. In the case of XVA compression, which is time-consuming, searching good values for the hyperparameters by a grid search method would be too demanding computationally. In our numerics, the mutation rate r_m is set to 20% and the crossover rate r_c to 50%. In the genetic algorithms literature, the crossover rate is often close to one, but for problems with few genes (i.e., components of x , or parameters, only four in our case), it is recommended to select a smaller value.

With parallelization in mind (see Section 3.2), we prefer to decrease the number of iterations even if it implies exploring more solutions. We set the genetic algorithm population size to $P = 100$ individuals and we limit the number of iterations to 5. Hence, we value the fitness function on 600 tentative new swaps x .

4.3. Results in the Case of Payer Portfolio Without Penalization

First, we consider a portfolio only composed of payer swaps. The expected exposure (EE) and the expected positive and negative exposures (EPE and ENE), i.e., $\mathbb{E}MtM_t^\pm$, are shown as a function of time t in Figure 3, which illustrates the asymmetric market risk profile of the portfolio.

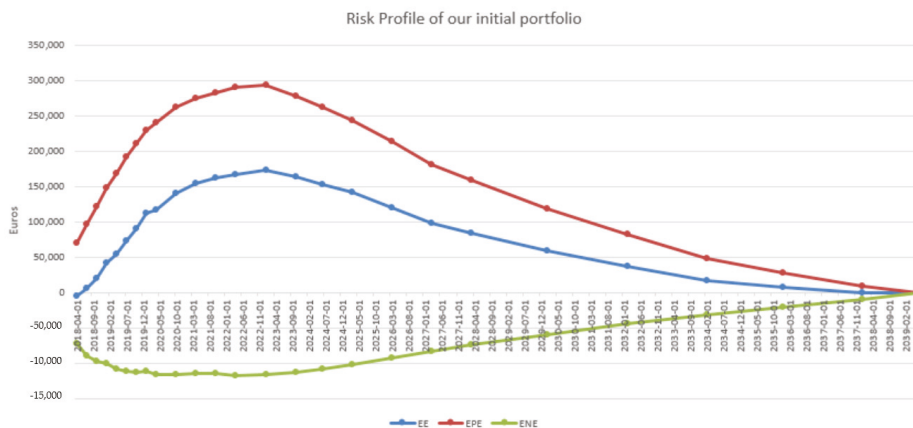


Figure 3. Market risk profile of the portfolio (payer portfolio without penalization).

Our first point is to verify that the algorithm without penalization, i.e., for $\alpha = 0$ in Equation (7), will select a receiver swap with a maturity comparable to those of the swap of the initial portfolio.

Table 1 reports after each iteration the three best solutions (from top to bottom) ever found since the beginning of the algorithm (in terms of the fitness criterion in Equation (7) with $\alpha = 0$, i.e., ΔCVA). A negative incremental CVA means that the new swap decreases the counterparty risk of the bank. The initial portfolio CVA amounts to 34,929€. We also report the $|DV01|$ s of the augmented portfolios in order to be able to assess the impact of the penalization in our next experiment.

Table 1. Evolution of optimal solutions after each iteration (payer portfolio without penalization).

Iter.	Mat. (yrs)	Not. (K€)	Rate (%)	Curr.	Pos.	ΔCVA (€)	$\frac{-\Delta CVA}{CVA}$ (in %)	$ DV01 $ (€)
0	10	4,800,000	1.6471	GBP	Receive	-8019	23.0	4484
	10	4,700,000	1.6471	GBP	Receive	-7948	22.8	4390
	10	4,600,000	1.6471	GBP	Receive	-7872	22.5	4297
1	17	5,600,000	1.4623	EUR	Receive	-17,249	49.4	8648
	12	5,400,000	1.7036	GBP	Receive	-9163	26.2	5957
	16	3,900,000	0.6377	JPY	Receive	-8760	25.1	6137
2	14	6,600,000	1.3416	EUR	Receive	-21,680	62.1	8626
	17	5,100,000	1.4623	EUR	Receive	-19,729	56.5	7875
	17	5,600,000	1.4623	EUR	Receive	-17,249	49.4	8648
3	14	6,600,000	1.3416	EUR	Receive	-21,680	62.1	8626
	17	5,100,000	1.4623	EUR	Receive	-19,729	56.5	7875
	17	5,600,000	1.4623	EUR	Receive	-17,249	49.4	8648
4	17	3,300,000	1.4623	EUR	Receive	-27,300	78.2	5096
	12	6,100,000	1.2203	EUR	Receive	-25,382	72.7	6959
	11	5,600,000	1.147	EUR	Receive	-23,009	65.9	5908
5	17	3,300,000	1.4623	EUR	Receive	-27,300	78.2	5096
	12	6,100,000	1.2203	EUR	Receive	-25,382	72.7	6959
	12	5,100,000	1.2203	EUR	Receive	-25,264	72.3	5818

As shown in Figure 4, a stabilization of the algorithm is observed after four iterations, on a new swap leading to a CVA gain of about 27,300€, i.e., about 78% of the initial portfolio CVA. The maturity and the notional are found the two most sensitive genes in the optimization. The maturity of the swap

is chosen by the algorithm so as to reduce the exposure peak: The decrease of the exposure in the first eight years of the portfolio is visible in terms of EPE profile in Figure 5 and of CVA profile⁴ in Figure 6.

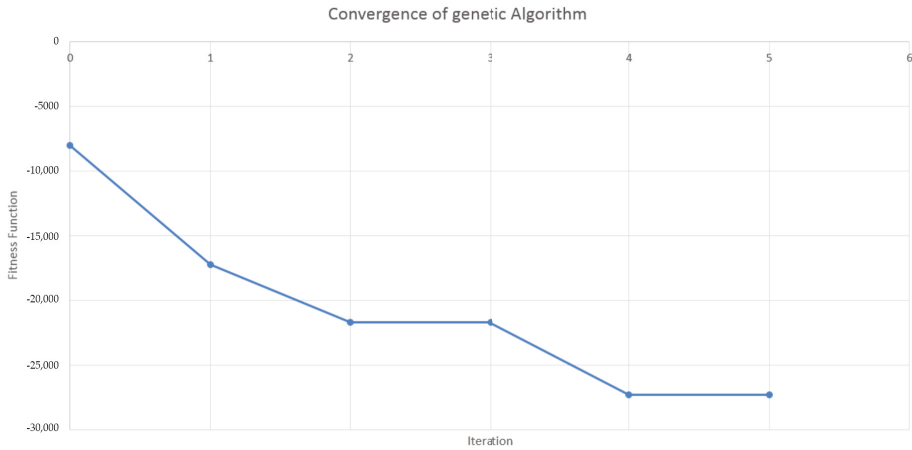


Figure 4. Fitness value as a function of iteration number (payer portfolio without penalization).

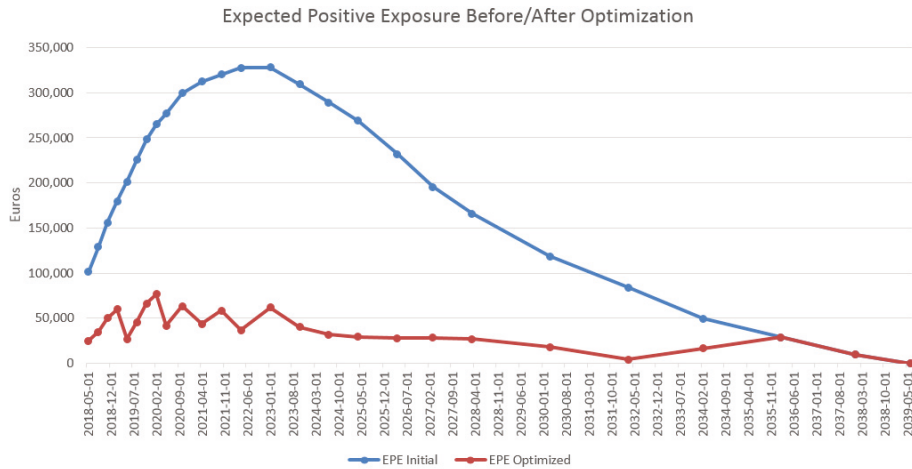


Figure 5. Market risk profile of the portfolio before and after optimization (payer portfolio without penalization).

⁴ Term structure obtained by integrating the EPE profile against the CDS curve of the counterparty from time 0 to an increasing upper bound $t \leq T$ (cf. Equation (3)).

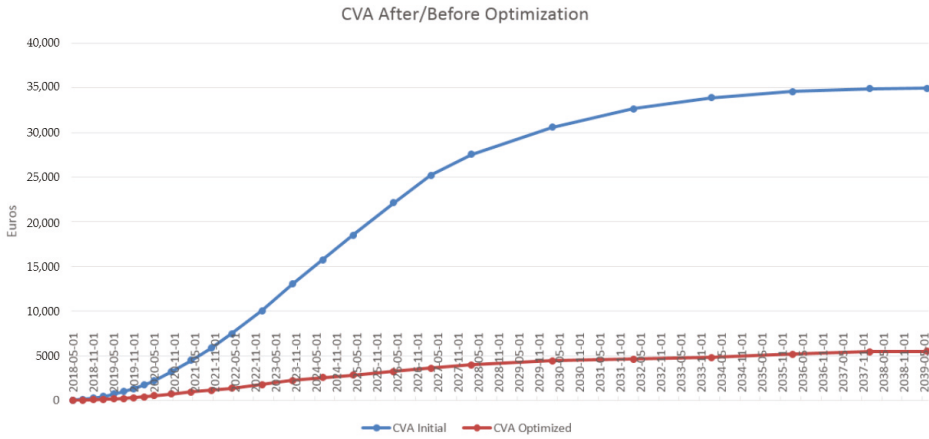


Figure 6. CVA profile before and after optimization (payer portfolio without penalization).

4.4. Results in the Case of Payer Portfolio With Penalization

We keep the same initial portfolio but we now penalize our objective function by the $|DV01|$ of the new swap, setting the regularization parameter α to one in Equation (7). As shown below, this choice achieves a good balance between the two terms ΔCVA and $\alpha DV01$ in Equation (7).

In the present context of a payer portfolio, $|DV01|$ control and CVA gain are two antagonistic targets. This may explain why the algorithm seems to struggle in finding a stable solution: indeed, the last iteration still decreases the fitness significantly (see Figure 7).

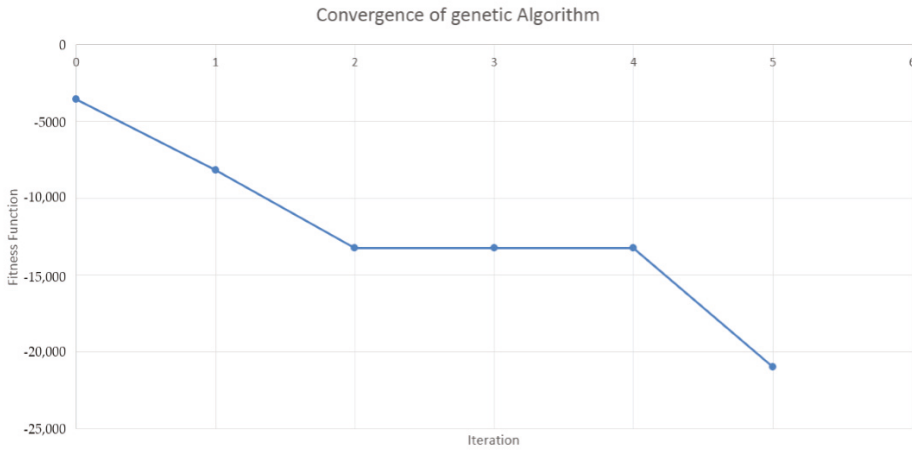


Figure 7. Fitness as a function of iteration number (payer portfolio with penalization).

During the execution (see Table 2), the algorithm first optimizes the CVA and then (in iteration 5) reduces the $|DV01|$. This is due to the difference of order of magnitude between ΔCVA and $|DV01|$ (recalling $\alpha = 1$): ΔCVA is more important, hence the algorithm only takes care of the penalization once ΔCVA has been compressed.

Table 2. Evolution of optimal solutions after each iteration (payer portfolio with penalization).

Iter.	Mat. (yrs)	Not. (K€)	Rate (%)	Curr.	Pos.	ΔCVA (€)	$\frac{-\Delta CVA}{CVA}$ (in %)	DV01 (€)
0	10	4,500,000	1.6471	GBP	Receive	-7790	22.3	4218
	10	4,600,000	1.6471	GBP	Receive	-7871	22.5	4311
	10	4,700,000	1.6471	GBP	Receive	-7947	22.8	4405
1	17	5,600,000	1.4731	EUR	Receive	-16,892	48.4	8706
	10	4,500,000	1.6471	GBP	Receive	-7790	22.3	4217
	10	4,600,000	1.6471	GBP	Receive	-7871	22.5	4311
2	14	6,600,000	1.3336	EUR	Receive	-21,888	62.7	8654
	17	5,600,000	1.4731	EUR	Receive	-16,892	48.4	8706
	17	6,100,000	1.4531	EUR	Receive	-15,038	43.1	9466
3	14	6,600,000	1.3336	EUR	Receive	-21,888	62.7	8654
	17	5,600,000	1.4731	EUR	Receive	-16,892	48.4	8706
	9	4,500,000	0.9584	EUR	Receive	-10,454	29.9	3945
4	10	6,600,000	1.3336	EUR	Receive	-21,888	62.7	8654
	11	6,600,000	1.3999	EUR	Receive	-18,825	53.9	9207
	17	5,600,000	1.4731	EUR	Receive	-16,892	48.4	8706
5	11	2,900,000	1.3811	EUR	Receive	-25,059	71.7	4039
	18	1,500,000	1.48	EUR	Receive	-18,258	52.3	2442
	17	1,500,000	1.4531	EUR	Receive	-16,553	47.4	2327

In the end, the gains in CVA are of the same order of magnitude as in the case without penalization (92% of the CVA gain without penalization), but for about 20% of |DV01| less than before. The second and third best solutions also achieve a great CVA gain, while diminishing the |DV01| by a factor three with respect to the nonpenalized case. By comparison with the unpenalized case (cf. Tables 1 and 2), the trades identified by the algorithm have a lower maturity or a smaller notional, hence a smaller |DV01|.

Figures 8 and 9 show the corresponding market risk and CVA profiles before and after the optimization.

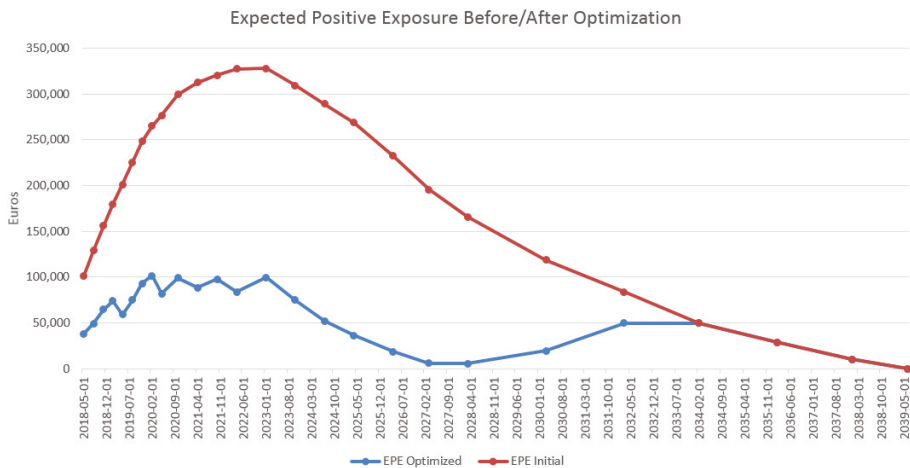


Figure 8. Market risk profile of portfolio before and after optimization (payer portfolio with penalization).

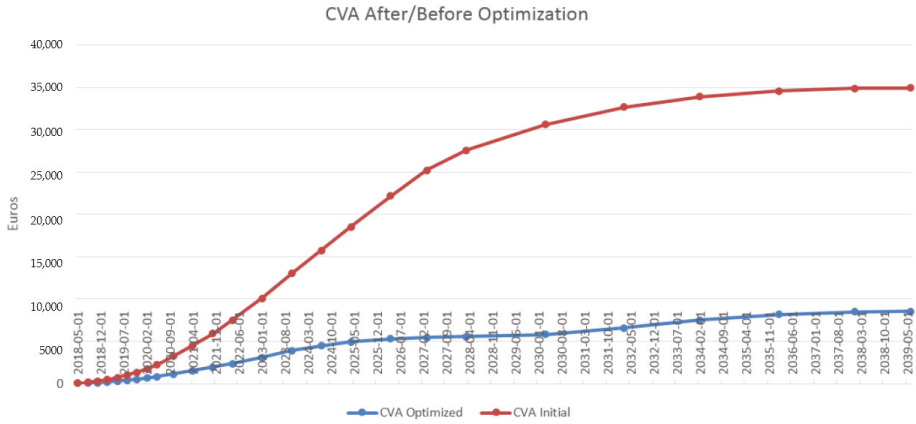


Figure 9. CVA profile before and after optimization (payer portfolio with penalization).

4.5. Results in the Case of a Hybrid Portfolio With Penalization

Next, we challenge our algorithm with a more balanced initial portfolio, as shown in Figure 10 (to be compared with Figure 3). The initial CVA is now 6410€. We set the regularization parameter α in Equation (7) to 0.3, as opposed to 1 in the previous case, in view of the lower CVA of the initial portfolio.

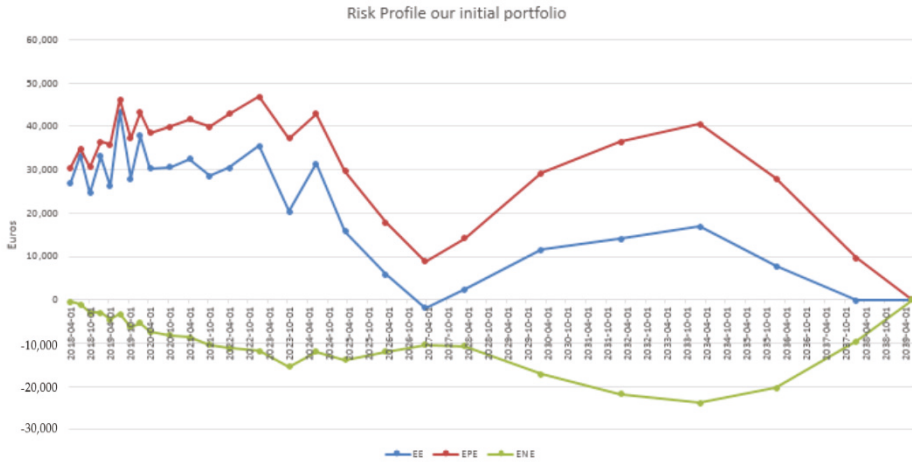


Figure 10. Market risk profile of the portfolio (hybrid portfolio with penalization).

As visible in Table 3 and Figure 11, the stabilization of the algorithm occurs after three iterations, showing that, for the hybrid portfolio, $|DV01|$ penalization and ΔCVA play less antagonistic roles. This is obtained by a relatively small notional and a maturity limited to 9 years, versus 11 years in the previous case of a payer portfolio with penalization. The corresponding market risk and CVA profiles, before and after the optimization, are displayed in Figures 12 and 13. Figure 12 explains the choices operated by the algorithm: as we restrict our incremental strategy to one swap, the algorithm limits the EPE until the first positive peak before 2026. A better strategy, but one outside our search space \mathcal{A} , would be to add a second swap with entry date in 2028 and end date in 2037.

Table 3. Evolution of optimal solutions after each iteration (hybrid portfolio with penalization).

Iter.	Mat. (yrs)	Not. (K€)	Rate (%)	Curr.	Pos.	ΔCVA (€)	$\frac{-\Delta CVA}{CVA}$ (in %)	$ DV01 $ (€)
0	1	6,000,000	0.025	JPY	Receive	14	-0.2	609
	1	6,100,000	0.025	JPY	Receive	14	-0.2	619
	1	6,300,000	0.025	JPY	Receive	14	-0.2	640
1	8	1,500,000	0.8565	EUR	Receive	-1905	29.7	1177
	6	2,300,000	0.586	EUR	Receive	-1166	18.2	1370
	9	700,000	1.608	GBP	Receive	-820	12.8	595
2	8	1,500,000	0.8565	EUR	Receive	-1905	29.7	1177
	6	2,300,000	0.586	EUR	Receive	-1166	18.2	1370
	9	700,000	1.608	GBP	Receive	-82	12.8	595
3	9	1,900,000	0.9584	EUR	Receive	-2284	35.6	1665
	8	1,500,000	0.8565	EUR	Receive	-1905	29.7	1177
	7	2,700,000	0.7225	EUR	Receive	-1628	25.4	1865
4	9	1,900,000	0.9584	EUR	Receive	-2284	35.6	1665
	8	1,500,000	0.8565	EUR	Receive	-1905	29.7	1177
	7	2,700,000	0.7225	EUR	Receive	-1628	25.4	1865
5	9	1,900,000	0.9584	EUR	Receive	-2284	35.6	1665
	8	1,500,000	0.8565	EUR	Receive	-1905	29.7	1177
	9	2,500,000	0.9584	EUR	Receive	-1942	30.3	2192

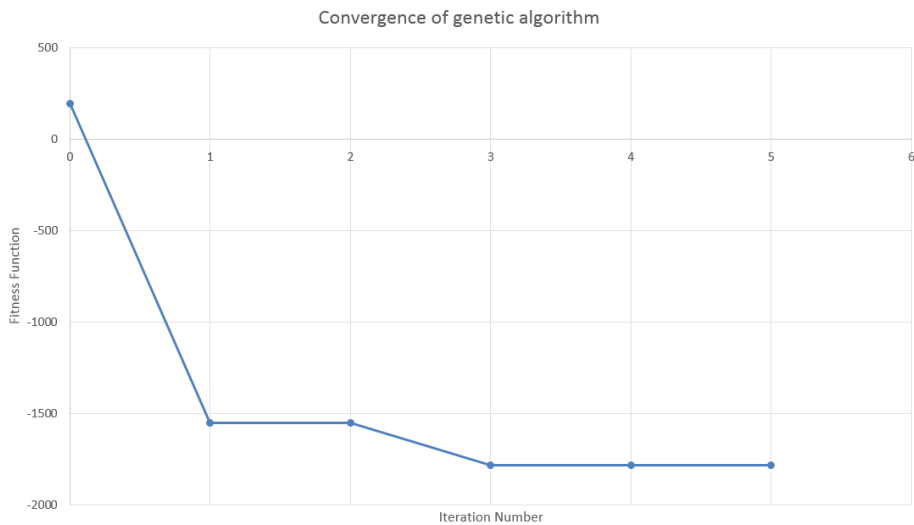


Figure 11. Fitness value as a function of iteration number (hybrid portfolio with penalization).

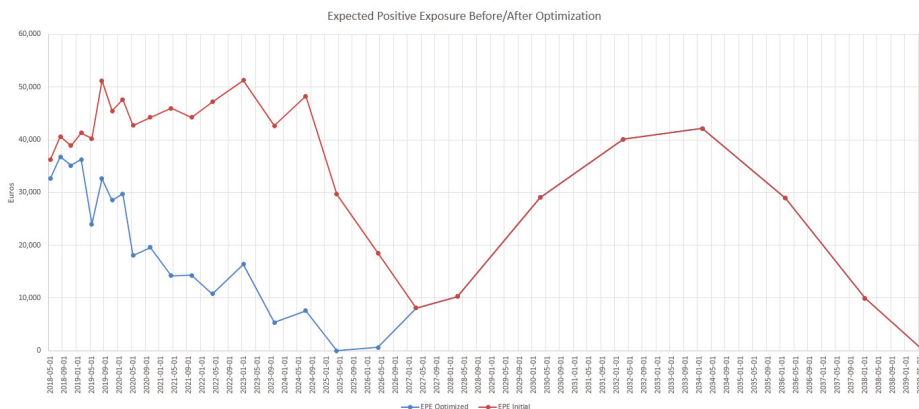


Figure 12. Market risk profile of portfolio before and after optimization (hybrid portfolio with penalization).

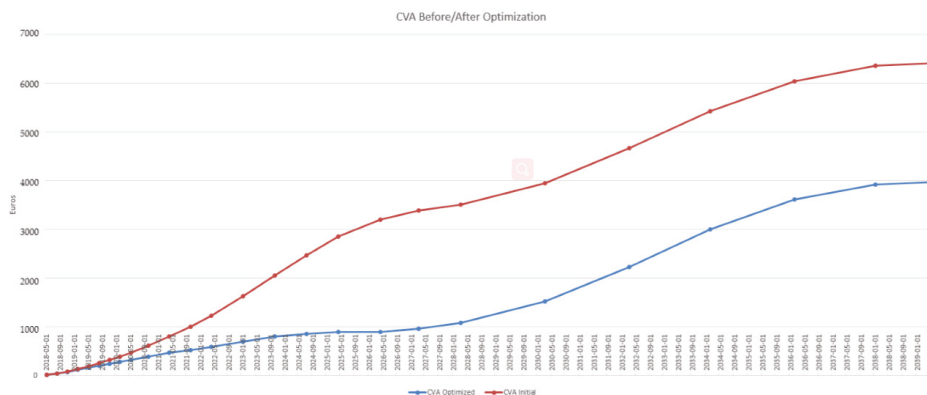


Figure 13. CVA profile before and after optimization (hybrid portfolio with penalization).

5. Conclusions

There exists a trade-off between CVA compression and DV01 penalization, which have antagonistic influences on the incremental exposure. Provided the search space for incremental trades is adequately chosen and parameterized, genetic optimization can result in significant CVA gains and, under DV01 penalization, this can be achieved without too much impact on the market risk of the bank position.

On the portfolios considered in our case studies, with 10–20 trades, a basic XVA compression run on a standard PC without the acceleration techniques of Section 3 takes about 20 h. The time gain resulting from an MtM store-and-reuse implementation of the trade incremental XVA computations as per Section 3.1 primarily depends on the size of the initial portfolio, but also on the maturity, and complexity more generally (vanilla vs. callable or path-dependent, etc.), of the constituting trades. Likewise, the time gain resulting from a parallel implementation of the genetic algorithm as per Section 3.2 primarily depends on the population size P , but it can be deteriorated by grid latency, hardware limitation, or data flow management features. In our simulations, an MtM store-and-reuse implementation of the trade incremental XVA computations reduces the XVA compression time to about 7 h. A further parallel implementation of the genetic optimization algorithm lowers the execution time to about 1 h.

The case study of this paper is only a first step toward more complex optimizations. One could thus enlarge the search space with, e.g., crosscurrency swaps. In this case, the market risk penalization

should be revisited to penalize other risk factors, beyond interest rate risk that is already accounted for by $|DV01|$. The penalization could also be refined with a focus on forward mark-to-market, i.e., market risk in the future (our current $|DV01|$ penalization only controls spot market risk).

CVA compression strategies involving several additional trades could be implemented. A first step toward such a multi-variate, multi-trade, compression would be an iterated application of single-trade XVA compressions, whereby, after each compression, the optimally augmented portfolio becomes the initial portfolio for the next compression. The benefit of such an iterative approach would be the ability to work with a search space \mathcal{A} (or a sequence of them) of constant size, as opposed to a global search space \mathcal{A} that would need to grow exponentially with the number of new trades in the case of a single multi-trade compression cycle.

Additional XVA metrics, and ultimately the all-inclusive XVA add-on in Equation (1), should be included in the compression (which, in particular, would allow one to identify possible XVA cuts across different netting sets).

Author Contributions: Conceptualization, M.C. and S.C.; methodology, M.C.; software, M.C.; validation, M.C.; formal analysis, M.C. investigation, M.C. and S.C.; writing—original draft preparation, S.C.; writing—review and editing, S.C.; supervision, S.C.; project administration, S.C.; and funding acquisition, S.C.

Funding: This research was conducted with the support of the Research Initiative “Modélisation des marchés actions, obligations et dérivés” financed by HSBC France under the aegis of the Europlace Institute of Finance. The views and opinions expressed in this presentation are those of the author alone and do not necessarily reflect the views or policies of HSBC Investment Bank, its subsidiaries or affiliates. The PhD thesis of Marc Chataigner is co-funded by a public grant as part of investissement d’avenir project, reference ANR-11-LABX-0056-LLH LabEx LMH.

Acknowledgments: The authors would like to thank Hugo Lebrun, Rémi Liguereau, and Guillaume Macey (HSBC France) for their assistance in the conduct of this study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CDS	Credit default swap
CVA	Credit valuation adjustment
DV01	Dollar value of an 01
EE	Expected exposure
EPE	Expected positive exposure
ENE	Expected negative exposure
FVA	Funding valuation adjustment
KVA	Capital valuation adjustment
MtM	Mark-to-market
MVA	Margin valuation adjustment
OIS	Overnight indexed swap
OTC	Over-the-counter
XVA	Generic “X” valuation adjustment

Appendix A. Single Point Crossover

Let (p_1, p_2) be a pair of chromosomes chosen as parents and let (c_1, c_2) denote the children. We assume that each chromosome has four genes A, B, C, D , that p_1 has gene versions $\{A_1, B_1, C_1, D_1\}$ and p_2 has gene versions $\{A_2, B_2, C_2, D_2\}$. For a single point crossover, we draw uniformly an integer i such the first i genes for c_1 are inherited from p_1 and the remaining genes are transferred from p_2 to c_1 , and symmetrically so for c_2 . For instance, if we draw $i = 2$, then c_1 has gene versions $\{A_1, B_1, C_2, D_2\}$, and c_2 has gene values $\{A_2, B_2, C_1, D_1\}$.

References

- Adler, Dan. 1993. Genetic algorithms and simulated annealing: A marriage proposal. Paper presented at the IEEE International Conference on Neural Networks, San Francisco, CA, USA, March 28–April 1, pp. 1104–9.
- Albanese, Claudio, and Stéphane Crépey. 2019. XVA Analysis from the Balance Sheet. Available online: <https://math.maths.univ-evry.fr/crepey> (accessed on 3 August 2019).
- Albanese, Claudio, Marc Chataigner, and Stéphane Crépey. 2018. Wealth transfers, indifference pricing, and XVA compression schemes. In *From Probability to Finance—Lecture Note of BICMR Summer School on Financial Mathematics*. Mathematical Lectures from Peking University Series. Edited by Y. Jiao. Berlin: Springer.
- Back, Thomas. 1996. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford: Oxford University Press.
- Basel Committee on Banking Supervision. 2015. *Review of the Credit Valuation Adjustment Risk Framework*. Consultative Document. Basel: Basel Committee on Banking Supervision.
- Bergstra, James, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. 2011. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*. Cambridge: The MIT Press, pp. 2546–54.
- Blickle, Tobias, and Lothar Thiele. 1995. *A Comparison of Selection Schemes Used in Genetic Algorithms*. TIK-Report. Zurich: Computer Engineering and Networks Laboratory (TIK).
- Brigo, Damiano, and Frédéric Vrins. 2018. Disentangling wrong-way risk: Pricing credit valuation adjustment via change of measures. *European Journal of Operational Research* 269: 1154–64. [CrossRef]
- Carvalho, Delmar, João Bittencourt, and Thiago Maia. 2011. The simple genetic algorithm performance: A comparative study on the operators combination. Paper presented at the First International Conference on Advanced Communications and Computation, Barcelona, Spain, October 23–29.
- Chen, Shu-Heng. 2012. *Genetic Algorithms and Genetic Programming in Computational Finance*. Berlin: Springer Science & Business Media.
- Cont, Rama, and Sana Ben Hamida. 2005. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance* 8: 43–76. [CrossRef]
- Crépey, Stéphane, and Shiqi Song. 2016. Counterparty risk and funding: Immersion and beyond. *Finance and Stochastics* 20: 901–30. [CrossRef]
- Crépey, Stéphane, and Shiqi Song. 2017. Invariance Properties in the Dynamic Gaussian Copula Model. *ESAIM: Proceedings and Surveys* 56: 22–41.
- Crépey, Stéphane, Rodney Hoskinson, and Bouazza Saadeddine. 2019. Balance Sheet XVA by Deep Learning and GPU. Available online: <https://math.maths.univ-evry.fr/crepey> (accessed on 3 August 2019).
- Del Moral, Pierre. 2004. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems With Applications*. Berlin: Springer.
- Drake, Adrian E., and Robert E. Marks. 2002. Genetic algorithms in economics and finance: Forecasting stock market prices and foreign exchange—A review. In *Genetic Algorithms and Genetic Programming in Computational Finance*. Berlin: Springer, pp. 29–54.
- Glasserman, Paul, and Linan Yang. 2018. Bounding wrong-way risk in CVA calculation. *Mathematical Finance* 28: 268–305. [CrossRef]
- Goldberg, David. 1989. *Genetic Algorithms in Search Optimization and Machine Learning*. Boston: Addison-Wesley.
- Gregory, Jon. 2015. *The xVA Challenge: Counterparty Credit Risk, Funding, Collateral and Capital*. Hoboken: Wiley.
- Holland, John H. 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Ann Arbor: University of Michigan Press.
- Hull, John, and Alan White. 2012. CVA and wrong way risk. *Financial Analyst Journal* 68: 58–69. [CrossRef]
- Iben Taarit, M. 2018. Pricing of XVA Adjustments: From Expected Exposures to Wrong-Way risks. Ph.D. thesis, Université Paris-Est, Marne-la-Vallée, France. Available online: <https://pastel.archives-ouvertes.fr/tel-01939269/document> (accessed on 3 August 2019).
- Janaki Ram, D., T. Sreenivas, and K. Ganapathy Subramaniam. 1996. Parallel simulated annealing algorithms. *Journal of Parallel and Distributed Computing* 37: 207–12.
- Jin, Zhuo, Zhixin Yang, and Quan Yuan. 2019. A genetic algorithm for investment-consumption optimization with value-at-risk constraint and information-processing cost. *Risks* 7: 32. [CrossRef]

- Kondratyev, Alexei, and George Giorgidze. 2017. Evolutionary Algos for Optimising MVA. *Risk Magazine*, December. Available online: <https://www.risk.net/cutting-edge/banking/5374321/evolutionary-algos-for-optimising-mva> (accessed on 3 August 2019).
- Kroha, Petr, and Matthias Friedrich. 2014. Comparison of genetic algorithms for trading strategies. Paper presented at International Conference on Current Trends in Theory and Practice of Informatics, High Tatras, Slovakia, January 25–30. Berlin: Springer, pp. 383–94.
- Larranaga, P., C. Kuijpers, R. Murga, I. Inza, and S. Dizdarevic. 1999. Genetic algorithms for the travelling salesman problem: A review of representations and operators. *Artificial Intelligence Review* 13: 129–70. [CrossRef]
- Li, Minqiang, and Fabio Mercurio. 2015. Jumping with Default: Wrong-Way Risk Modelling for CVA. *Risk Magazine*, November. Available online: <https://www.risk.net/risk-management/credit-risk/2433221/jumping-default-wrong-way-risk-modelling-cva> (accessed on 3 August 2019).
- Pardalos, Panos, Leonidas Pitsoulis, T. Mavridou, and M. Resende. 1995. Parallel search for combinatorial optimization: Genetic algorithms, simulated annealing, tabu search and GRASP. Paper presented at International Workshop on Parallel Algorithms for Irregularly Structured Problems, Lyon, France, September 4–6. Berlin: Springer, pp. 317–31.
- Pykhtin, Michael. 2012. General wrong-way risk and stress calibration of exposure. *Journal of Risk Management in Financial Institutions* 5: 234–51.
- Rios, Luis, and Nikolaos Sahinidis. 2013. Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization* 56: 1247–93. [CrossRef]
- Snoek, Jasper, Hugo Larochelle, and Ryan Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*. Cambridge: The MIT Press, pp. 2951–59.
- Tabassum, Mujahid, and Mathew Kuruvilla. 2014. A genetic algorithm analysis towards optimization solutions. *International Journal of Digital Information and Wireless Communications (IJDIIWC)* 4: 124–42. [CrossRef]
- Turing, Alan M. 1950. Computing machinery and intelligence. *Mind* 59: 433–60. [CrossRef]
- Verma, Rajeev, and P. Lakshminiarayanan. 2006. A case study on the application of a genetic algorithm for optimization of engine parameters. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 220: 471–79. [CrossRef]
- Young, Steven, Derek Rose, Thomas Karnowski, Seung-Hwan Lim, and Robert Patton. 2015. Optimizing deep learning hyper-parameters through an evolutionary algorithm. Paper presented at Workshop on Machine Learning in High-Performance Computing Environments, Austin, TX, USA, November 15, p. 4.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Risk Factor Evolution for Counterparty Credit Risk under a Hidden Markov Model

Ioannis Anagnostou ^{1,2,*} and Drona Kandhai ^{1,2}

¹ Computational Science Lab, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands; b.d.kandhai@uva.nl

² Quantitative Analytics, ING Bank, Foppingadreef 7, 1102BD Amsterdam, The Netherlands

* Correspondence: i.anagnostou@uva.nl; Tel.: +31-20-525-6789

Received: 31 March 2019; Accepted: 5 June 2019; Published: 12 June 2019

Abstract: One of the key components of counterparty credit risk (CCR) measurement is generating scenarios for the evolution of the underlying risk factors, such as interest and exchange rates, equity and commodity prices, and credit spreads. Geometric Brownian Motion (GBM) is a widely used method for modeling the evolution of exchange rates. An important limitation of GBM is that, due to the assumption of constant drift and volatility, stylized facts of financial time-series, such as volatility clustering and heavy-tailedness in the returns distribution, cannot be captured. We propose a model where volatility and drift are able to switch between regimes; more specifically, they are governed by an unobservable Markov chain. Hence, we model exchange rates with a hidden Markov model (HMM) and generate scenarios for counterparty exposure using this approach. A numerical study is carried out and backtesting results for a number of exchange rates are presented. The impact of using a regime-switching model on counterparty exposure is found to be profound for derivatives with non-linear payoffs.

Keywords: Counterparty Credit Risk; Hidden Markov Model; Risk Factor Evolution; Backtesting; FX rate; Geometric Brownian Motion

1. Introduction

One of the main factors that amplified the financial crisis of 2007–2008 was the failure to capture major risks associated with over-the-counter (OTC) derivative-related exposures ([Basel Committee on Banking Supervision 2010a](#)). Counterparty exposure, at any future time, is the amount that would be lost in the event that a counterparty to a derivative transaction would default, assuming zero recovery at that time. Banks are required to hold regulatory capital against their current and future exposures to all counterparties in OTC derivative transactions.

A key component of the counterparty exposure framework is modeling the evolution of underlying risk factors, such as interest and exchange rates, equity and commodity prices, and credit spreads. Risk Factor Evolution (RFE) models are, arguably, the most important part of counterparty exposure modeling, since small changes in the underlying risk factors may have a profound impact on the exposure and, as a result, on the regulatory and economic capital buffers. It is, therefore, crucial for financial institutions to put significant effort in the design and calibration of RFE models and, in addition, have a sound framework in place in order to assess the forecasting capability of the model.

Although the Basel Committee on Banking Supervision has stressed the importance of the ongoing validation of internal models method (IMM) for counterparty exposure ([Basel Committee on Banking Supervision 2010b](#)), there are no strict guidelines on the specifics of this validation process. As a result, there is some degree of ambiguity regarding the regulatory requirements that financial institutions are expected to meet. In an attempt to reduce this ambiguity, [Anfuso et al. \(2014\)](#) introduced a complete framework for counterparty credit risk (CCR) model backtesting which is compliant with Basel III and

the new Capital Requirements Directives (CRD IV). A detailed backtesting framework for CCR models was also introduced by Ruiz (2014), who expanded the corresponding framework for Value-at-Risk (VaR) models by the Basel Committee (Basel Committee on Banking Supervision 1996).

The most ubiquitous model for the evolution of exchange rates is Geometric Brownian Motion (GBM). Under GBM, the exchange rate dynamics are assumed to follow a continuous-time stochastic process, in which the returns are log-normally distributed. Although simplicity and tractability render GBM a particularly popular modeling choice, it is generally accepted that it cannot adequately describe the empirical facts exhibited by real exchange rate returns (Booth and Glassman 1987). More specifically, exchange rate returns can be leptokurtic, exhibiting tails that exceed those of the normal distribution. As a result, a scenario-generation framework based on GBM may assign unrealistically low probabilities to extreme scenarios, leading to the under-estimation of counterparty exposure and, consequently, regulatory and economic capital buffers.

The main reason for the inability of GBM to produce return distributions with realistically heavy tails is the assumption of constant drift and volatility parameters. In this paper, we present a way to address this limitation without entirely departing from the convenient GBM framework. We propose a model where the GBM parameters are allowed to switch between different states, governed by an unobservable Markov process. Thus, we model exchange rates with a hidden Markov model (HMM) and generate scenarios for counterparty exposure using this approach.

A HMM is a mathematical model in which the system being modeled is assumed to follow a Markov chain whose states are hidden from the observer. HMMs have a broad range of applications, in speech recognition (Juang and Rabiner 1991), computational biology (Krogh et al. 1994), gesture recognition (Wilson and Bobick 1999), and in other areas of artificial intelligence and pattern recognition (Ghahramani 2001). HMMs have gained significant popularity in the mathematical and computational finance fields. The application of HMMs in financial and economic time-series was pioneered by Hamilton in Hamilton (1988; 1989). Since then, a significant amount of literature has been published, focusing on the ability of HMMs to reproduce stylized facts of asset returns (Bulla et al. 2011; Nystrup et al. 2015; Rydén et al. 1998), asset allocation (Ang and Bekaert 2004; Guidolin and Timmermann 2007; Nystrup et al. 2015), and option pricing (Bollen 1998; Guo 2001; Naik 1993).

Our paper expands the counterparty exposure literature by introducing a hidden Markov model for the evolution of exchange rates. We provide a detailed description of HMMs and their estimation process. In our numerical experiments, we use GBM and HMM to generate scenarios for the Euro against two major and two emerging currencies. We perform a thorough backtesting exercise, based on the framework proposed by Ruiz (2014), and find similar performances for GBM and a two-state HMM. Finally, we use the generated scenarios to calculate credit exposure for foreign exchange (FX) options, and find significant differences between the two models, which are even more pronounced for deep out-of-the-money instruments.

The remainder of the paper is organized as follows. Section 2 provides the fundamentals of HMMs, along with the algorithms for determining their parameters from data. Section 3 gives background information on modeling the evolution of exchange rates. Section 4 outlines the framework for performance evaluation of RFE models. A numerical study is presented in Section 5. Finally, in Section 6, we draw conclusions and discuss future research directions.

2. An Introduction to Hidden Markov Models

The hidden Markov model (HMM) is a statistical model in which a sequence of observations is generated by a sequence of unobserved states. The hidden state transitions are assumed to follow a first-order Markov chain. The theory of hidden Markov models (HMMs) originates from the work of Baum et al. in the late 1960s (Baum and Petrie (1966), Baum and Eagon (1967)). In the rest of this section, we introduce the theory of hidden Markov models (HMMs), following Rabiner (1990).

2.1. Formal Definition of a HMM

In order to formally define a hidden Markov model (HMM), the following elements are required:

1. N , the number of hidden states. Even though the states are not directly observed, in many practical applications they have some physical interpretation. For instance, in financial time-series, hidden states may correspond to different phases of the business cycle, such as prosperity and depression. We denote the states by $X = \{X_1, X_2, \dots, X_N\}$, and the state at time t by q_t .
2. M , the number of distinct observation symbols per state. These symbols represent the physical output of the system being modeled. The individual symbols are denoted by $V = \{v_1, v_2, \dots, v_M\}$.
3. The transition probability distribution between hidden states, $A = \{a_{ij}\}$, where

$$a_{ij} = P [q_{t+i} = X_j | q_t = X_i], \quad 1 \leq i, j \leq N. \tag{1}$$

4. The observation symbol probability distribution in state j , $B = \{b_j(k)\}$, where

$$b_j(k) = P [v_k \text{ at } t | q_t = X_j], \quad 1 \leq j \leq N, 1 \leq k \leq M. \tag{2}$$

5. The initial distribution of the hidden states, $\pi = \{\pi_i\}$, where

$$\pi_i = P [q_1 = X_i], \quad 1 \leq i \leq N. \tag{3}$$

The parameter set of the model is denoted by $\lambda = (A, B, \pi)$. A graphical representation of a hidden Markov model with two states and three discrete observations is given by Figure 1.

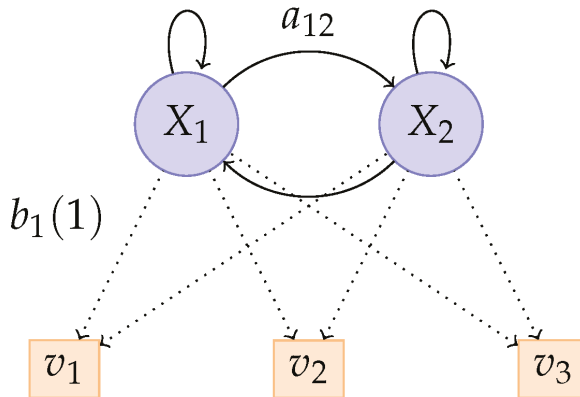


Figure 1. A hidden Markov model (HMM) with two states and three discrete observations, where a_{ij} is the probability of transition from state X_i to state X_j and $b_j(k)$ is the emission probability for symbol v_k in state X_j .

In the case where there are an infinite amount of symbols for each hidden state, v_k is omitted and the observation probability $b_j(k)$, conditional on the hidden state X_j , can be replaced by

$$b_j(O_t) = P(O_t | q_t = X_j).$$

If the observation symbol probability distributions are Gaussian, then $b_j(O_t) = \phi(O_t | u_j, \sigma_j)$, where $\phi(\cdot)$ is the Gaussian probability density function, and u_j and σ_j are the mean and standard deviation of the corresponding state X_j , respectively. In that case, the parameter set of the model is $\lambda = (A, u, \sigma, \pi)$, where u and σ are vectors of means and standard deviations, respectively.

2.2. The Three Basic Problems for HMMs

The idea that HMMs should be characterized by three fundamental problems originates from the seminal paper of Rabiner (1990). These three problems are the following:

Problem 1 (Likelihood). Given the observation sequence $O = O_1O_2 \dots O_T$ and a model $\lambda = (A, B, \pi)$, how do we compute the conditional probability $P(O|\lambda)$ in an efficient manner?

Problem 2 (Decoding). Given the observation sequence $O = O_1O_2 \dots O_T$ and a model λ , how do we determine the state sequence $Q = q_1q_2 \dots q_T$ which optimally explains the observations?

Problem 3 (Learning). How do we select model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$?

2.3. Solutions to the Three Basic Problems

2.3.1. Likelihood

Our objective is to calculate the likelihood of a particular observation sequence, $O = O_1O_2 \dots O_T$, given the model λ . The most intuitive way of doing this is by summing the joint probability of O and Q for all possible state sequences Q of length T :

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda) \cdot P(Q|\lambda). \tag{4}$$

The probability of a particular observation sequence O , given a state sequence $Q = q_1q_2 \dots q_T$, is

$$\begin{aligned} P(O|Q, \lambda) &= \prod_{t=1}^T P(O_t|q_t, \lambda) \\ &= b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T), \end{aligned} \tag{5}$$

as we have assumed that the observations are independent. The probability of a state sequence Q can be written as

$$P(Q|\lambda) = \pi_{q_1} a_{q_1q_2} a_{q_2q_3} \dots a_{q_{T-1}q_T}. \tag{6}$$

The joint probability of O and Q is the product of the above two terms; that is,

$$P(O, Q|\lambda) = P(O|Q, \lambda) \cdot P(Q|\lambda). \tag{7}$$

Although the calculation of $P(O|\lambda)$ using the above definition is rather straightforward, the associated computational cost is huge.

Thankfully, a dynamic programming approach, called the Forward Algorithm, can be used instead.

Consider the forward variable $\alpha_i(t)$, defined as

$$\alpha_i(i) = P(O_1O_2 \dots O_i, q_i = X_i|\lambda). \tag{8}$$

We can solve for $\alpha_i(i)$ inductively using Algorithm 1.

Algorithm 1 The Forward Algorithm.

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \tag{9}$$

2. Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad \begin{matrix} 1 \leq t \leq T-1 \\ 1 \leq j \leq N. \end{matrix} \tag{10}$$

3. Termination:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \tag{11}$$

Correspondingly, we can define a backward variable $\beta_t(i)$ as

$$\beta_t(i) = P(O_{t+1}O_{t+2} \cdots O_T | q_t = X_i, \lambda). \tag{12}$$

Again, we can solve for $\beta_t(i)$ inductively using Algorithm 2.

Algorithm 2 The Backward Algorithm.

1. Initialization:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N. \tag{13}$$

2. Induction:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad \begin{matrix} t = T-1, T-2, \dots, 1 \\ 1 \leq i \leq N. \end{matrix} \tag{14}$$

2.3.2. Decoding

In order to identify the best sequence $Q = \{q_1 q_2 \cdots q_T\}$ for the given observation sequence $O = \{O_1 O_2 \cdots O_T\}$, we need to define the quantity

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1 q_2 \cdots q_t = i, O_1 O_2 \cdots O_t | \lambda). \tag{15}$$

By induction, we have

$$\delta_{t+1}(j) = \left[\max_i \delta_t(i) a_{ij} \right] \cdot b_j(O_{t+1}). \tag{16}$$

To actually retrieve the state sequence, it is necessary to keep track of the argument which maximizes Equation (16), for each t and j . We do so via the array $\psi_t(j)$. The complete procedure for finding the best state sequence is presented in Algorithm 3.

Algorithm 3 Viterbi algorithm.

1. Initialization:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \tag{17}$$

$$\psi_1(i) = 0. \tag{18}$$

2. Recursion:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T \tag{19}$$

$$1 \leq j \leq N$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T \tag{20}$$

$$1 \leq j \leq N.$$

3. Termination:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \tag{21}$$

$$q_T^* = \arg \min_{1 \leq i \leq N} [\delta_T(i)].$$

4. Sequence back-tracking:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1. \tag{22}$$

2.3.3. Learning

The model which maximizes the probability of an observation sequence O , given a model $\lambda = (A, B, \pi)$, cannot be determined analytically. However, a local maximum can be found using an iterative algorithm, such as the Baum-Welch method or the expectation-maximization (EM) method (Dempster et al. 1977). In order to describe the iterative procedure of obtaining the HMM parameters, we need to define $\xi_t(i, j)$, the probability of being at the state X_i at time t , and the state X_j at time $t + 1$, given the model and observation sequence; that is,

$$\xi_t(i, j) = P(q_t = X_i, q_{t+1} = X_j | O, \lambda). \tag{23}$$

Using the earlier defined forward and backward variables, $\xi_t(i, j)$ can be rewritten as

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)}. \tag{24}$$

We define

$$\gamma_t(i) = \sum_{j=i}^N \xi_t(i, j) \tag{25}$$

as the probability of being in state X_i at time t . It is clear that

$$\sum_{t=i}^{T-1} \gamma_t(i) = \text{expected number of transitions from } X_i, \text{ and} \tag{26}$$

$$\sum_{t=i}^{T-1} \xi_t(i, j) = \text{expected number of transitions from } X_i \text{ to } X_j. \tag{27}$$

Using these formulas, the parameters of a HMM can be estimated, in an iterative manner, as follows:

$$\hat{\pi}_i = \gamma_1(i) = \text{expected number of times in state } X_i \text{ at time } t = 1; \tag{28}$$

$$\begin{aligned} \hat{a}_{ij} &= \frac{\sum_{t=i}^{T-1} \zeta_t(i, j)}{\sum_{t=i}^{T-1} \gamma_t(i)} \\ &= \frac{\text{expected number of transitions from } X_i \text{ to } X_j}{\text{expected number of transitions from } X_i}; \end{aligned} \tag{29}$$

$$\begin{aligned} \hat{b}_j(k) &= \frac{\sum_{t=1}^T \mathbf{1}_{\{O_t=v_k\}} \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \\ &= \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j}. \end{aligned} \tag{30}$$

If $\lambda = (A, B, \pi)$ is the current model and $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ is the re-estimated one, then it has been shown, by Baum and Eagon (1967); Baum and Petrie (1966), that $P(O|\hat{\lambda}) \geq P(O|\lambda)$.

In case the observation probabilities are Gaussian, the following formulas are used to update the model parameters u and σ :

$$\hat{u}_j = \frac{\sum_{t=1}^T \gamma_t(j) O_t}{\sum_{t=1}^T \gamma_t(j)}, \tag{31}$$

$$\hat{\sigma}_j = \sqrt{\frac{\sum_{t=1}^T \gamma_t(j) (O_t - u_j)^2}{\sum_{t=1}^T \gamma_t(j)}}. \tag{32}$$

3. Modelling the Evolution of Exchange Rates

As discussed in the introduction, the first step in calculating the future distribution of counterparty exposure is the generation of scenarios using the models that represent the evolution of the underlying market factors. These factors typically include interest and exchange rates, equity and commodity prices, and credit spreads. This article is concerned with the modeling of exchange rates.

3.1. Geometric Brownian Motion

In mathematical finance, the Geometric Brownian Motion (GBM) model is the stochastic process which is usually assumed for the evolution of stock prices (Hull 2009). Due to its simplicity and tractability, GBM is also a widely used model for the evolution of exchange rates.

A stochastic process, S_t , is said to follow a GBM if it satisfies the following stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{33}$$

where W_t is a Wiener process, and μ and σ are constants representing the drift and volatility, respectively.

The analytical solution of Equation (33) is given by:

$$S_t = S_0 \exp \left(\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right). \tag{34}$$

With this expression in hand, and knowing that $W_t \sim N(0, t)$, one can generate scenarios simply by generating standard normal random numbers.

3.2. A Hidden Markov Model for Drift and Volatility

One of the main shortcomings of the GBM model is that, due to the assumption of constant drift and volatility, some important characteristics of financial time-series, such as volatility clustering and heavy-tailedness in the return distribution, cannot be captured. To address these limitations, we consider a model with an additional stochastic process. The observations of the exchange rates are assumed to be generated by a discretised GBM, in which both the drift and volatility parameters are able to switch, according to the state of an unobservable process which satisfies the Markov property. In other words, the conditional probability distribution of future states depends solely upon the current state, not on the sequence of states that preceded it. The observations also satisfy a Markov property with respect to the states (i.e., given the current state, they are independent of the history).

Thus, we consider a hidden Markov model with Gaussian emissions $\lambda = (A, u, \sigma, \pi)$, as was presented in Section 2.1. We denote the hidden states by $X = \{X_1, X_2, \dots, X_N\}$, and the state at time t as q_t . The unobservable Markov process governs the distribution of the log-return process $Y = \{Y_2, \dots, Y_T\}$, where $Y_t = \log \frac{S_t}{S_{t-1}}, t = 2, \dots, T$. The dynamics of Y are then as follows:

$$Y_t = u(q_t) + \sigma(q_t)Z_t, \tag{35}$$

where $u(q_t) = \left(\mu(q_t) - \frac{\sigma^2(q_t)}{2} \right)$ and $Z_t \sim N(0,1)$ are independent standard normal random numbers.

The transition probabilities of the hidden process, as well as the drift and volatility of the GBM, can be estimated from a series of observations, using the algorithms presented in Section 2. The number of hidden states has to be specified in advance. In many practical applications, the number of hidden states can be determined based on intuition. For example, stock markets are often characterized as “bull” or “bear”, based on whether they are appreciating or depreciating in value. A bull market occurs when returns are positive and volatility is low. On the other hand, a bear market occurs when returns are negative and volatility is high. It would, therefore, be in line with intuition to assume that stock market observations are driven by a two-state process. The number of states can also be determined empirically; for example, using the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). Once the model parameters have been estimated, scenarios can be generated by generating the hidden Markov chain and sampling the log-returns from the corresponding distributions.

4. RFE Model Performance Evaluation

4.1. Backtesting

In this sub-section, we give a brief overview of a framework for the backtesting of RFE models. For a more detailed description, the reader is referred to Ruiz (2014). Backtesting is the process of comparing the distributions given by the RFE models with the realized history of the corresponding risk factors. In accordance with regulatory requirements, RFE models have to be backtested at

multiple forecasting horizons, making use of various distributional tests (Basel Committee on Banking Supervision 2010b).

To test whether a set of realizations can reasonably be modeled as arising from a specific distribution, we use the Probability Integral Transform (PIT) (see (Diebold et al. 1997)), defined as

$$F(x_n) = \int_{-\infty}^{x_n} f(u)du, \tag{36}$$

where x_n is the realization of a random variable and $f(\cdot)$ is its predicted density. Note that, if one applies PIT using the true density of x_n to construct a set of values, it follows that the distribution of the constructed set will simply be $\mathcal{U}(0, 1)$. As a result, one is able to evaluate the quality of the model $f(\cdot)$ for x_n , simply by measuring the distance between the distribution of the constructed set and $\mathcal{U}(0, 1)$.

For a given set of realizations x_{t_i} of the risk factor to be tested, we set a starting point t_{start} and an ending point t_{end} . The size of the backtesting window is then $T_b = t_{end} - t_{start}$. If the time horizon over which we want to test our model is Δ , we proceed as follows:

1. We set $t_1 = t_{start}$. We, then, calculate the PIT $F(x_{t_1+\Delta})$ of the realized value at $t_1 + \Delta$ using the model risk factor distribution for that point. If an analytical expression is not available, the distribution can be approximated numerically. This yields a value F_1 .
2. We, then, move forward to $t_2 = t_1 + \Delta$. We calculate $F(x_{t_2+\Delta})$ using the model risk factor distribution at $t_2 + \Delta$ and obtain a value F_2 .
3. We repeat the above, until $t_i + \Delta = t_{end}$.

This exercise yields a set $\{F_i\}_{i=1}^K$, where K is the number of steps taken. As mentioned previously, if the empirical distribution of the realizations is the same as the predicted distribution, then the constructed set $\{F_i\}_{i=1}^K$ will be uniformly distributed.

In order to measure the distance d between the distribution of the constructed set and $\mathcal{U}(0, 1)$ we can use a number of metrics, such as:

The Anderson–Darling metric:

$$d_{AD} = \int_{-\infty}^{\infty} (F_e(x) - F(x))^2 \omega(F(x))dF(x)$$

$$\omega(x) = \frac{1}{x(1-x)}, \tag{37}$$

the Cramer–von Mises metric:

$$d_{CVM} = \int_{-\infty}^{\infty} (F_e(x) - F(x))^2 \omega(F(x))dF(x)$$

$$\omega(x) = 1, \text{ or} \tag{38}$$

the Kolmogorov–Smirnov metric:

$$d_{KS} = \sup_x |F_e(x) - F(x)|, \tag{39}$$

where F_e is the empirical and F is the theoretical cumulative distribution function. Note that each of these metrics provides a single distance value \bar{d} between the distribution of the realized set and $\mathcal{U}(0, 1)$. To obtain an understanding of whether this distance is acceptable, we simulate time-series using the model being tested. Although the simulated time-series will follow the model by definition, there will still be some distance, d , due to numerical noise. By repeating this experiment a sufficiently large number of times (say, M), we can obtain a set $\{d_i\}_{i=1}^M$ and approximate, numerically, its cumulative distribution function $\psi(d)$. With $\psi(d)$ in hand, we can assess the distance \bar{d} , as follows: If \bar{d} falls in a range with high probability with respect to $\psi(d)$, then the probability of the model being accurate is

high. By defining d_y and d_r as the 95th and the 99.99th percentiles, respectively, we can obtain three color bands for model performance:

- Green band: $\vec{d} \in [0, d_y)$;
- Yellow band: $\vec{d} \in [d_y, d_r)$; and
- Red band: $\vec{d} \in [d_r, \infty)$.

An example of the three-color scoring scheme is shown in Figure 2. The backtesting process can be carried out for a set of time horizons, and for every horizon a single result can be produced, in terms of a probability $\psi(\vec{d})$ and a color band.

4.2. Long-Term Percentiles of Distribution Cones

Backtesting provides a statistical judgement of the performance of the model for relatively short-term forecast horizons. Assessing the distribution cones of the risk factor evolution provides insight into the behavior of the model for long forecast horizons. The high and low percentiles of the distribution cones need to be compared to the long-term percentiles of the observed risk factor data. Please note that assessing the long-term percentiles needs expert judgement to some extent, as it is difficult to statistically unambiguously state what the long-term percentile of a distribution cone should be.

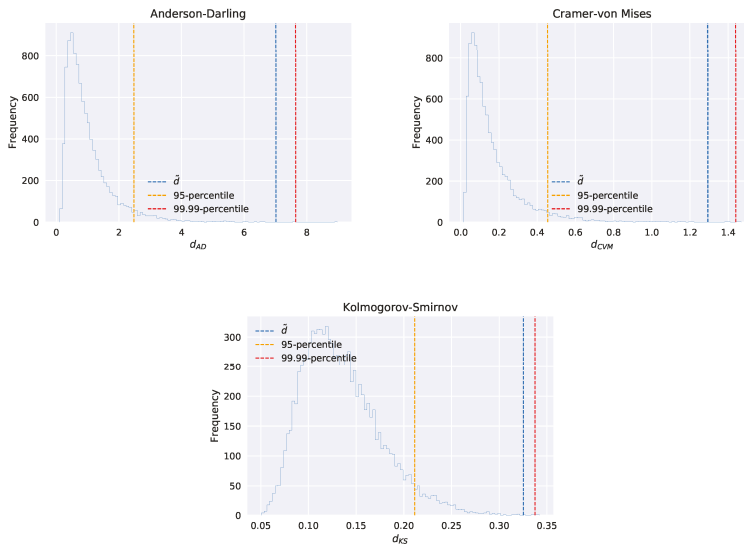


Figure 2. Examples of the three-color scoring scheme. The model in the example receives yellow scores for all three metrics, as $d_y < \vec{d} < d_r$.

5. Numerical Experiments

5.1. Overview of Data Selections

In order to evaluate the performance of the HMM approach, we used the foreign exchange rates of the Euro against two G10 and two emerging-market currencies. We used daily observations, between 1 January 2004 and 31 December 2016, for the following FX rates:

- USD/EUR,
- GBP/EUR,
- RUB/EUR, and
- MXN/EUR.

5.2. Selection of the Number of Hidden States

Choosing the appropriate number of hidden states for a HMM is not a trivial task. Two commonly used criteria for model comparison are the Akaike information criterion (AIC):

$$AIC = -2 \log L + 2p, \tag{40}$$

and the Bayesian information criterion (BIC):

$$BIC = -2 \log L + p \log T, \tag{41}$$

where L is the likelihood of the fitted model, p is the number of free parameters in the model, and T denotes the number of observations (Zucchini et al. 2016). The number of free parameters in a HMM with a Gaussian distribution for each hidden state is:

$$p = N^2 + 2N - 1, \tag{42}$$

where N is the number of hidden states. Thus, in both criteria, the second term is a penalty term which increases with increasing N . Compared to the AIC, the penalty term of the BIC has more weight when $T > e^2$ and, therefore, the BIC often favors models with fewer parameters than the AIC does.

A bank that uses internal models to measure exposure for capital purposes must use at least three years of historical data for calibration, where the parameters have to be updated quarterly or more frequently, if market conditions warrant. During the course of backtesting, re-calibration of the RFE model parameters needs to be done at the same frequency as for production to make the re-calibration effects visible (Basel Committee on Banking Supervision 2010b). Consequently, in the backtesting exercise that follows (in Section 5.3), we use calibration blocks of three years and move the block forward by one quarter every time.

To choose the appropriate number of hidden states, we calibrate HMMs with 2–5 states for each of the three-year blocks and calculate the AIC and BIC. The results are shown in Figures 3–6. Based on the AIC results, the performance of HMMs with 2, 3, 4, or 5 states is almost the same for the emerging-market currencies. For USD/EUR, models with higher number of hidden states seem to perform better while, for GBP/EUR, the two-state model is preferable. However, based on the BIC, the HMM with two states is the best candidate for all four currency pairs. Therefore, we focus on the HMM with two states for the rest of our numerical experiments.

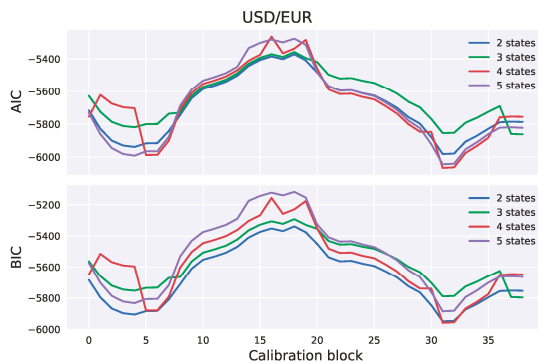


Figure 3. Akaike information criterion (AIC) and Bayesian information criterion (BIC) for HMMs calibrated using USD/EUR time-series.

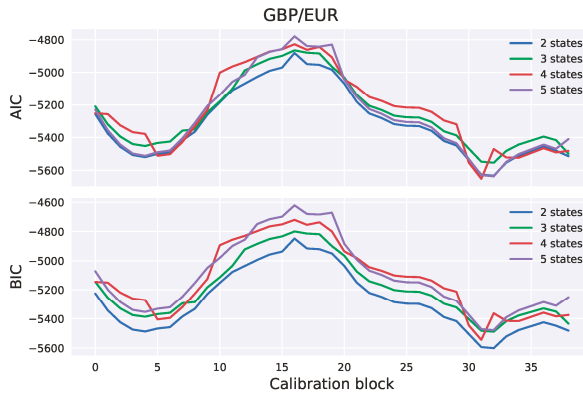


Figure 4. AIC and BIC for HMMs calibrated using GBP/EUR time-series.

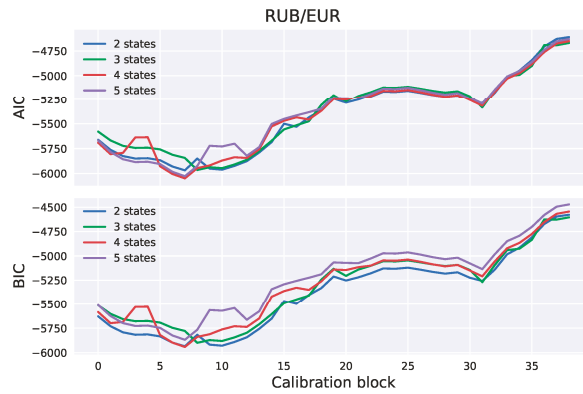


Figure 5. AIC and BIC for HMMs calibrated using RUB/EUR time-series.

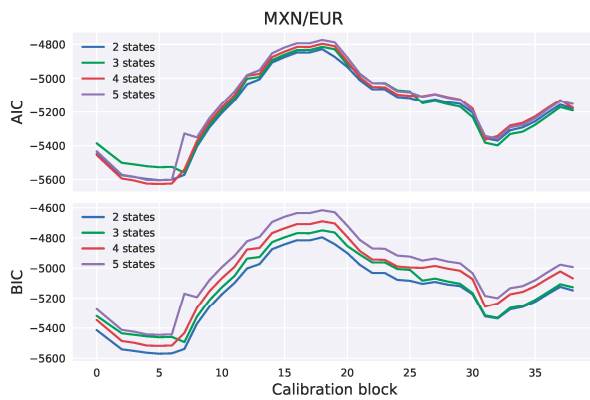


Figure 6. AIC and BIC for HMMs calibrated using MXN/EUR time-series.

5.3. Model Backtesting

We applied the backtesting algorithm (presented in Section 4) using observations between 1 January 2004 and 31 December 2016 for the selected FX rates. We used a calibration window T_c of three years with quarterly re-calibration ($\delta_c = 3$ months). The length of the backtesting window was $T_b = 10$ years and we tested model performance for time horizons Δ of length 1 week, 2 weeks, 1 month, and 3 months.

In order to generate scenarios of length Δ , the following steps were taken. At every time point t with $1 \leq t \leq \Delta$ and, given the current hidden state $q_t = X_i$, the next hidden state $q_{t+1} = X_j$ was chosen using the transition probability matrix A . The observation O_t was then generated, according to the corresponding emission probability distribution b_j . The initial hidden state q_0 was assumed to be the most probable state at the end of the learning procedure.

It is important to note that the backtesting procedure provides a statistical assessment of the model performance for relatively short-term forecast horizons. For instance, a backtesting window T_b of 10 years and a time horizon $\Delta = 1$ year would translate to only 10 independent points. As a result, the statistical relevance of the backtesting exercise would be limited. In order to gain an insight into model behavior for longer forecast horizons, we consider the distribution cones of the risk factor evolution. The high and the low percentiles of the distribution cones are compared to observed risk factor data.

In the following, we discuss the results obtained for each of the FX rates.

5.3.1. USD/EUR

Table 1 summarizes the results of the backtesting exercise for USD/EUR, in terms of probabilities as well as color bands. When the forecasting horizon was 1 week, both the GBM and the two-state HMM scored yellow under the Anderson–Darling and the Cramerthree–von Mises metrics, and green under the Kolmogorov–Smirnov metric. For the two-week forecasting horizon, both models obtained a yellow score under all three metrics. Finally, for the longer horizons (1 and 3 months), both models performed significantly better, with green scores under all metrics. The backtesting results do not indicate any notable difference in performance between the HMM and the GBM.

Table 1. Backtesting results for USD/EUR with calibration window $T_c = 3$ years, frequency of re-calibration $\delta_c = 3$ months, and backtesting window $T_b = 10$ years. GBM, Geometric Brownian Motion.

Time Horizon	GBM			HMM2		
	AD	CVM	KS	AD	CVM	KS
1W	0.9783	0.9709	0.9192	0.9884	0.9846	0.9189
2W	0.9553	0.9507	0.9831	0.9744	0.9662	0.9890
1M	0.4849	0.5829	0.5620	0.6486	0.6957	0.6023
3M	0.3476	0.2130	0.1156	0.6145	0.4968	0.3397

In order to gain an insight into the performance of the models for longer time horizons, we present, in Figure 7, the 5th and 95th percentiles of the forecast distributions for a horizon of 7 years, between 2011 and end of 2016. It can be seen that the HMM gave slightly more conservative forecasts, compared to the GBM, but the realized time-series fell within the 90% probability region under both models, at the end of the 7 year period.

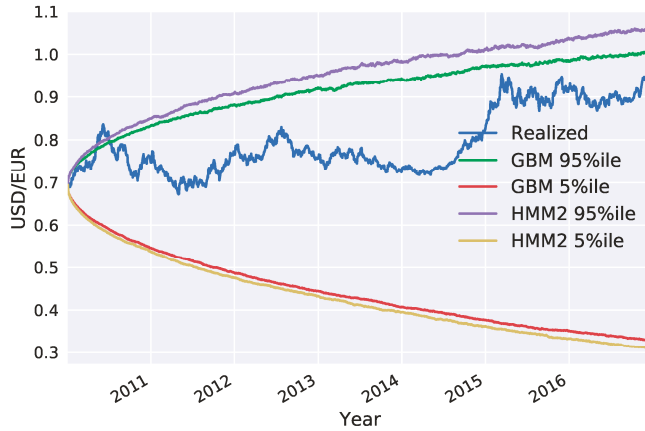


Figure 7. Percentiles of long-term distribution cones for USD/EUR under GBM and HMM with two states.

5.3.2. GBP/EUR

The backtesting results for GBP/EUR are summarized in Table 2. The two models achieved similar performance when the time horizon was 2 weeks or longer. When the forecasting horizon was 2 weeks, both models scored yellow. In the 1-month horizon, both models had green scores under the Anderson–Darling and Cramer–von Mises metrics, and a yellow score under the Kolmogorov–Smirnov metric. The scores were green for both models under all metrics when the time horizon was 3 months. The greatest difference between the two models was observed for the 1-week forecasting horizon, where the two-state model performed significantly better, scoring green under all three metrics, while the corresponding scores for GBM were yellow.

Table 2. Backtesting results for GBP/EUR with calibration window $T_c = 3$ years, frequency of re-calibration $\delta_c = 3$ months, and backtesting window $T_b = 10$ years.

Time Horizon	GBM			HMM2		
	AD	CVM	KS	AD	CVM	KS
1W	0.9861	0.9816	0.9820	0.8698	0.9280	0.6391
2W	0.9936	0.9919	0.9964	0.9934	0.9920	0.9952
1M	0.9085	0.8965	0.9594	0.9299	0.9140	0.9726
3M	0.8702	0.8273	0.8716	0.8573	0.8014	0.8535

Figure 8 shows the 5th and 95th percentiles of the forecast distributions between 2011 and end of 2016. Similarly to the results for USD/EUR, HMM gave slightly more conservative forecasts and the realized time-series fell within the 90% probability region under both models, at the end of the 7 year period. However, in 2016, the realized time-series fell outside the 95th percentile of the GBM distribution, while it was still within this bound for the HMM.

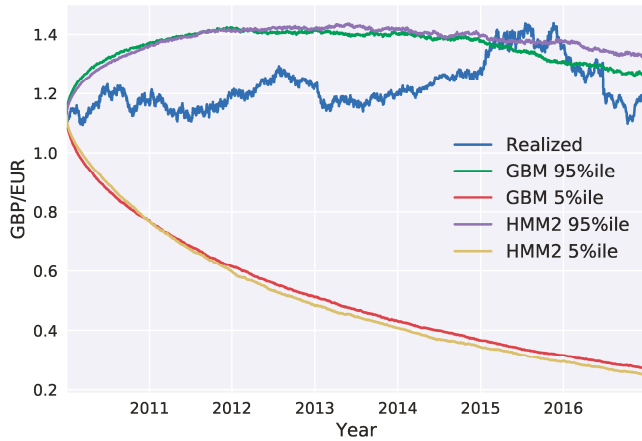


Figure 8. Percentiles of long-term distribution cones for GBP/EUR under GBM and HMM with two states.

5.3.3. RUB/EUR

Table 3 presents the results of the backtesting exercise for RUB/EUR. It can be seen that both GBM and HMM did not perform very well when the forecasting horizon was 1 week, with HMM having yellow scores under every metric. The results were similar for the 2 week forecasting horizon. In the longer time horizons, however, both models performed better. HMM outperformed the one-state model GBM, achieving green scores in the 1-month horizon. The scores were green for both models when the forecasting horizon was 3 months.

Table 3. Backtesting results for RUB/EUR with calibration window $T_c = 3$ years, frequency of re-calibration $\delta_c = 3$ months, and backtesting window $T_b = 10$ years.

Time Horizon	GBM			HMM2		
	AD	CVM	KS	AD	CVM	KS
1W	0.9991	0.9988	0.9988	0.9997	0.9996	0.9996
2W	0.9992	0.9989	0.9992	0.9996	0.9996	0.9988
1M	0.9830	0.9809	0.9446	0.9485	0.9457	0.8898
3M	0.5526	0.1406	0.0651	0.4399	0.3394	0.1624

Figure 9 shows the percentiles of the long-term distribution cones for RUB/EUR. It is clear that the difference between GBM and HMM was more pronounced, with the HMM yielding significantly more conservative forecasts. The realized time-series was close to the 95th percentile of the GBM distribution until mid-2014, exceeding it on a number of occasions in 2011 and in 2013. Despite a sharp decline in 2015, the realized time-series remained above the 5th percentile for both models throughout the 7 year period.

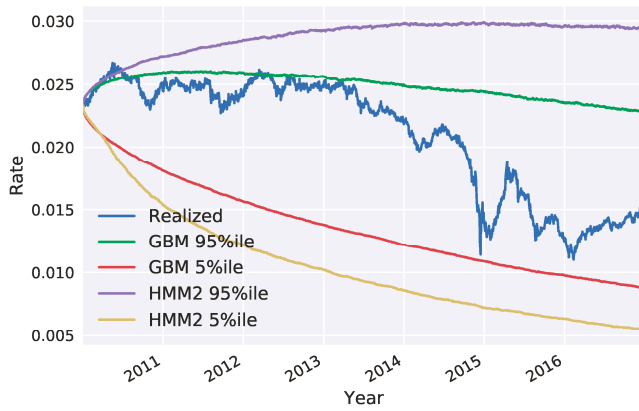


Figure 9. Percentiles of long-term distribution cones for RUB/EUR under GBM and HMM with two states.

5.3.4. MXN/EUR

Table 4 summarizes the results of the backtesting exercise for MXN/EUR, in terms of scores as well as color bands. Both HMM and GBM had yellow scores for the shorter time horizons (1 and 2 weeks), under all metrics. The models performed better for the longer time horizons (1 and 3 months), achieving green scores. Figure 10 shows the long-term distribution cones. Similar to the the GBP/EUR case, we do not observe a clear difference in performance between GBM and HMM with two states.

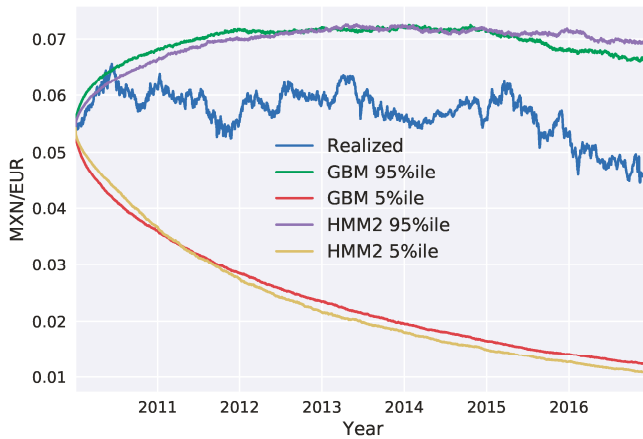


Figure 10. Percentiles of long-term distribution cones for MXN/EUR under GBM and HMM with two states.

Table 4. Backtesting results for MXN/EUR, with a 3-year calibration window, quarterly re-calibration, and a 10-year backtesting window.

Time Horizon	GBM			HMM2		
	AD	CVM	KS	AD	CVM	KS
1W	0.9967	0.9955	0.995	0.9967	0.9956	0.9938
2W	0.9895	0.9864	0.9677	0.9841	0.9768	0.9742
1M	0.5185	0.5963	0.7136	0.5501	0.6071	0.6225
3M	0.7124	0.6643	0.4422	0.7045	0.7373	0.6563

5.4. Impact on Credit Exposure: A Case Study for FX Options

5.4.1. Exposure at Default (EAD)

Prior to presenting the case study on FX options, we provide a brief introduction to credit exposure calculation. For a more detailed description, the reader is referred to [Zhu and Pykhtin \(2007\)](#) and [Gregory \(2012\)](#).

When a financial institution is permitted to use the IMM to calculate credit exposure, the following steps need to be taken:

1. *Scenario Generation.* Market scenarios are simulated for a fixed set of exposure dates $\{t_k\}_{k=1}^N$ in the future, using the RFE models.
2. *Instrument Valuation.* Instrument valuation is performed for each exposure date and for each simulated scenario.

The outcome of this process is a set of realizations of credit exposure at each exposure date in the future. One can then estimate the expected exposure EE_k as the average exposure at future date t_k , where the average is taken across all simulated scenarios of the relevant risk factors.

The Expected Positive Exposure (EPE) is defined as the weighted average of the EE over the first year

$$EPE = \sum_{k=1}^{\min(1 \text{ year, maturity})} EE_k \times \Delta t_k, \tag{43}$$

where the weights $\Delta t_k = t_k - t_{k-1}$ are the proportion that an individual expected exposure represents over the entire one-year time horizon.

In order to account for potential non-conservative aging effects, a modification is necessary. First, an Effective EE profile is obtained from the EE profile by adding the non-decreasing constraint for maturities below one year. Effective EE can be calculated, recursively, as follows:

$$\text{Effective } EE_k = \max \{ \text{Effective } EE_{k-1} - EE_k \}, \tag{44}$$

where the current date is denoted as t_0 and EE_0 equals the current exposure.

Effective EPE can, then, be calculated from the Effective EE profile, in the same way that EPE is calculated from the EE profile:

$$\text{Effective EPE} = \sum_{k=1}^{\min(1 \text{ year, maturity})} \text{Effective } EE_k \times \Delta t_k. \tag{45}$$

Finally, the Exposure at Default (EAD) is the product of a multiplier α and the Effective EPE

$$EAD = \alpha \times \text{Effective EPE}. \tag{46}$$

The multiplier α , introduced by [Picoult \(2002\)](#), is a correction coefficient that accounts for wrong-way risk. Under the IMM, α is fixed at a rather conservative level of 1.4. However, banks using

the IMM have an option to use their own estimate of α , with the prior approval of the supervisor and a floor of 1.2.

5.4.2. Results

In order to study the impact of using a two-state HMM, instead of a GBM, on regulatory and economic capital, we consider the case of FX call options on the RUB/EUR rate. The rationale behind this choice was that the Russian currency suffered a crisis in 2014, which will be included in our calibration data set.

Our starting date was 2 January 2016. We estimated the parameters of a GBM and a two-state HMM, using three years of data (between January 2013 and December 2015). Following the methodology presented in Section 5.4.1, we generated market scenarios for the following set of future exposure dates:

$$\{t_k\}_{k=1}^9 = \{1 \text{ week, 2 weeks, 3 weeks, 4 weeks, 2 months, 3 months, 6 months, 9 months, 1 year}\}. \quad (47)$$

For each generated scenario and each exposure date, option valuation was performed using the Garman–Kohlhagen model (Garman and Kohlhagen (1983)).

The value of a call option at time t is given by the analytical formula

$$C_t = S_t e^{-r_f(T-t)} N(x + \sigma\sqrt{T-t}) - K e^{-r_d(T-t)} N(x), \quad (48)$$

where

$$x \equiv \frac{\ln(S_t/K) + (r_d - r_f - (\sigma^2/2))(T-t)}{\sigma\sqrt{T-t}},$$

S_t is the spot price of the deliverable currency at time t (domestic units per foreign unit),

K is the strike price of the option (domestic units per foreign unit),

$T - t$ is the time to maturity,

r_d is the domestic risk-free interest rate,

r_f is the foreign risk-free interest rate,

σ is the volatility of the spot currency price, and

$N(\cdot)$ is the cumulative normal distribution function.

Note that, in the formula, both spot and strike price are quoted in units of domestic currency per unit of foreign currency. As a result, the option price will be in the same units, as well. In order to obtain the market value of a position in such an option, it is necessary to multiply by a notional amount Λ in the foreign currency.

In our example, the foreign and domestic currencies are RUB and EUR, respectively. In order to achieve a candid comparison of the two RFE models for the exchange rate, we do not consider interest rate and volatility as risk factors for FX options. Instead, we make the simplistic assumptions of $r_d = r_f = 0$ and constant volatility $\sigma = 0.15$ (equal to the supervisory volatility for foreign exchange options in the standardised approach, see Basel Committee on Banking Supervision (2014)). The notional amount Λ is set to RUB 100,000,000. The spot RUB/EUR exchange rate on 2 January 2016 was $S_0 = 0.01263$.

The credit exposure values for out-of-the-money (OTM) call options on the RUB/EUR exchange rate, for a range of strike prices, are illustrated in Figure 11a. The impact of using a two-state HMM, instead of a GBM, is shown in Figure 11b. These results are summarized in Table 5. It is clear that exposure values under HMM exceeded the exposure values under GBM markedly for deep-out-the-money options. This difference would have a direct impact on how these positions would be capitalized against counterparty default, with a difference that could exceed 400% for the strike price $K = 0.023$. It is also important to note that, given the exchange rate movements over recent years, it is not unrealistic for the moneyness of such options to change dramatically, leading to large

unexpected losses. For in-the-money call options, the two models produced identical exposure values. Thus, these results are omitted from this paper.

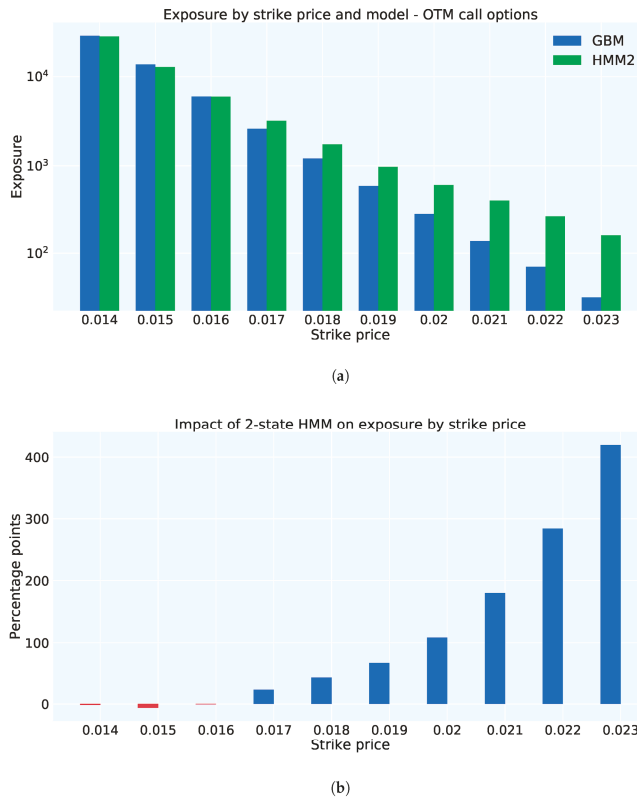


Figure 11. Credit exposure values for out-of-the-money (OTM) call options on the RUB/EUR exchange rate (a) and the impact of using a two-state HMM, instead of a GBM (b).

Table 5. Credit exposure values for out-of-the-money (OTM) options on the RUB/EUR exchange rate.

Strike K	Credit Exposure		Impact (%)
	GBM	HMM2	
0.014	29,507.54	29,013.22	-1.68
0.015	13,684.44	12,838.95	-6.18
0.016	5,981.10	5,939.89	-0.69
0.017	2,598.64	3,199.17	23.11
0.018	1,207.04	1,740.47	44.19
0.019	580.34	973.64	67.77
0.020	285.70	595.99	108.61
0.021	143.08	401.04	180.29
0.022	70.20	269.96	284.60
0.023	31.87	165.56	419.47

6. Conclusions

In this paper, we presented a hidden Markov model for the evolution of exchange rates with regards to counterparty exposure. In the proposed model, the observations of the exchange rates were assumed to be generated by a discretized GBM, in which both the drift and volatility parameters are

able to switch, according to the state of a hidden Markov process. The main motivation of using such a model is the fact that GBM can assign unrealistically low probabilities to extreme scenarios, leading to the under-estimation of counterparty exposure and the corresponding capital buffers. The proposed model is able to produce distributions with heavier tails and capture extreme movements in exchange rates without entirely departing from the convenient GBM framework.

We generated exchange rate scenarios for four currency pairs: USD/EUR, GBP/EUR, RUB/EUR, and MXN/EUR. A risk factor evolution model backtesting exercise was performed, in line with Basel III requirements, and the percentiles of the long-term distribution cones were obtained. The performances of the one-state and two-state models (GBM and the two-state HMM, respectively) were found to be very similar, with the two-state model HMM being slightly more conservative. However, when the generated scenarios were used to calculate exposure profiles for options on the RUB/EUR exchange rate, we found significant differences between the results of the two models. These differences were even more pronounced for deep out-of-the-money options.

Our study highlights some of the limitations of backtesting as a tool for comparing the performance of RFE models. Backtesting can be a useful way to objectively assess model performance. However, it can only be performed over short time horizons; with our available data, we could perform a statistically sound test of modeling assumptions for a time horizon of maximum length three months. It is, therefore, important to put effort into the interpretation of backtesting results, before they are translated into conclusions about model performance. Our results show how two models with similar performances in a backtesting exercise can result in very different exposure values and, consequently, in very different regulatory and economic capital buffers. This can lead to regulatory arbitrage and potentially weaken financial stability and, further, turn into a systemic risk.

The research presented in this paper can be extended in a number of ways, such as considering the evolution of risk factors other than exchange rates. Another topic worthy of investigation is the enhancement of the backtesting framework presented by Ruiz (2014), by considering statistical tests similar to the ones presented by Berkowitz (2001) and Amisano and Giacomini (2007). Finally, an interesting research direction is the development of an agent-based simulation model with heterogeneous modeling approaches, with regards to the RFE models. This model could potentially give valuable insights into the impact of heterogeneous models in financial stability.

Author Contributions: Both authors conceived and planned the research. Ioannis Anagnostou performed the numerical experiments. Both authors discussed the results and contributed to the final version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement no. 675044 (<http://bigdatafinance.eu/>), Training for Big Data in Financial Research and Risk Management.

Acknowledgments: The authors are grateful to Jori Hoencamp, Steven van Haren, Marcel Boersma, and the reviewers for the useful comments.

Conflicts of Interest: The opinions expressed in this article are solely those of the authors and do not represent in any way those of their current and past employers. The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- Amisano, Gianni, and Raffaella Giacomini. 2007. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25: 177–90.
- Anfuso, Fabrizio, Dimitrios Karyampas, and Andreas Nawroth. 2014. Credit exposure models backtesting for Basel III. *Risk*. September. Available online: <https://www.risk.net/2362332> (accessed on 1 March 2019).
- Ang, Andrew, and Geert Bekaert. 2004. How regimes affect asset allocation. *Financial Analysts Journal* 60: 86–99. [[CrossRef](#)]
- Basel Committee on Banking Supervision. 1996. *Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*. Basel: Basel Committee on Banking Supervision.
- Basel Committee on Banking Supervision. 2010a. *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Basel: Bank for International Settlements.

- Basel Committee on Banking Supervision. 2010b. *Sound Practices for Backtesting Counterparty Credit Risk Models*. Basel: Bank for International Settlements.
- Basel Committee on Banking Supervision. 2014. *The Standardised Approach for Measuring Counterparty Credit Risk Exposures*. Basel: Bank for International Settlements.
- Baum, Leonard E., and John Alonzo Eagon. 1967. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society* 73: 360–63. [\[CrossRef\]](#)
- Baum, Leonard E., and Ted Petrie. 1966. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics* 37: 1554–63. [\[CrossRef\]](#)
- Berkowitz, Jeremy. 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19: 465–74.
- Bollen, Nicolas P. B. 1998. Valuing options in regime-switching models. *The Journal of Derivatives* 6: 38–49. [\[CrossRef\]](#)
- Boothe, Paul, and Debra Glassman. 1987. The statistical distribution of exchange rates: Empirical evidence and economic implications. *Journal of International Economics* 22: 297–319. [\[CrossRef\]](#)
- Bulla, Jan, Sascha Mergner, Ingo Bulla, André Sesboué, and Christophe Chesneau. 2011. Markov-switching asset allocation: Do profitable strategies exist? *Journal of Asset Management* 12: 310–21. [\[CrossRef\]](#)
- Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39: 1–38. [\[CrossRef\]](#)
- Diebold, Francis X., Todd A. Gunther, and Anthony Tay. 1997. Evaluating density forecasts. *International Economic Review* 39: 863–83. [\[CrossRef\]](#)
- Garman, Mark B., and Steven W. Kohlhagen. 1983. Foreign currency option values. *Journal of International Money and Finance* 2: 231–37. [\[CrossRef\]](#)
- Ghahramani, Zoubin. 2001. An introduction to hidden Markov models and bayesian networks. In *Hidden Markov Models: Applications in Computer Vision*. Singapore: World Scientific, pp. 9–41.
- Gregory, Jon. 2012. *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*. Hoboken: John Wiley & Sons.
- Guidolin, Massimo, and Allan Timmermann. 2007. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31: 3503–44. [\[CrossRef\]](#)
- Guo, Xin. 2001. An explicit solution to an optimal stopping problem with regime switching. *Journal of Applied Probability* 38: 464–81. [\[CrossRef\]](#)
- Hamilton, James D. 1988. Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* 12: 385–423. [\[CrossRef\]](#)
- Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society* 57: 357–84. [\[CrossRef\]](#)
- Hull, John C. 2009. *Options, Futures, and Other Derivatives*, 7th ed. New York: Pearson Prentice Hall.
- Juang, Biing Hwang, and Laurence R. Rabiner. 1991. Hidden Markov models for speech recognition. *Technometrics* 33: 251–72. [\[CrossRef\]](#)
- Krogh, Anders, Michael Brown, I. Saira Mian, Kimmen Sjölander, and David Haussler. 1994. Hidden Markov models in computational biology: Applications to protein modeling. *Journal of Molecular Biology* 235: 1501–31. [\[CrossRef\]](#)
- Naik, Vasanttilak. 1993. Option valuation and hedging strategies with jumps in the volatility of asset returns. *The Journal of Finance* 48: 1969–84. [\[CrossRef\]](#)
- Nystrup, Peter, Bo William Hansen, Henrik Madsen, and Erik Lindström. 2015. Regime-based versus static asset allocation: Letting the data speak. *The Journal of Portfolio Management* 42: 103–9. [\[CrossRef\]](#)
- Nystrup, Peter, Henrik Madsen, and Erik Lindström. 2015. Stylised facts of financial time series and hidden Markov models in continuous time. *Quantitative Finance* 15: 1531–41. [\[CrossRef\]](#)
- Picoult, Evan. 2002. *Quantifying the Risks of Trading*. Cambridge: Cambridge University Press.
- Rabiner, Lawrence R. 1990. A tutorial on hidden Markov models and selected applications in speech recognition. In *Readings in Speech Recognition*. Amsterdam: Elsevier, pp. 267–96.
- Ruiz, Ignacio. 2014. Backtesting counterparty risk: How good is your model? *The Journal of Credit Risk* 10: 87. [\[CrossRef\]](#)

- Rydén, Tobias, Timo Teräsvirta, and Stefan Åsbrink. 1998. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* 13: 217–44. [CrossRef]
- Wilson, Andrew D., and Aaron F Bobick. 1999. Parametric hidden Markov models for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21: 884–900. [CrossRef]
- Zhu, Steven H., and Michael Pykhtin. 2007. A guide to modeling counterparty credit risk. *GARP Risk Review* 37. July/August. Available online: <https://ssrn.com/abstract=1032522> (accessed on 11 June 2019).
- Zucchini, Walter, Iain L. MacDonald, and Roland Langrock. 2016. *Hidden Markov Models for Time Series: An Introduction Using R*. London: Chapman and Hall/CRC.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Default Ambiguity

Tolulope Fadina ¹ and Thorsten Schmidt ^{1,2,3,*}

¹ Department of Mathematical Stochastics, University of Freiburg, Ernst-Zermelo Str.1, 79104 Freiburg im Breisgau, Germany; tolulope.fadina@stochastik.uni-freiburg.de

² Freiburg Research Institute of Advanced Studies (FRIAS), 79104 Freiburg im Breisgau, Germany

³ University of Strasbourg Institute for Advanced Study (USIAS), 67081 Strasbourg, France

* Correspondence: thorsten.schmidt@stochastik.uni-freiburg.de

Received: 29 March 2019; Accepted: 17 May 2019; Published: 10 June 2019

Abstract: This paper discusses ambiguity in the context of single-name credit risk. We focus on uncertainty in the default intensity but also discuss uncertainty in the recovery in a fractional recovery of the market value. This approach is a first step towards integrating uncertainty in credit-risky term structure models and can profit from its simplicity. We derive drift conditions in a Heath–Jarrow–Morton forward rate setting in the case of ambiguous default intensity in combination with zero recovery, and in the case of ambiguous fractional recovery of the market value.

Keywords: model ambiguity; default time; credit risk; no-arbitrage; reduced-form HJM models; recovery process

1. Introduction

Recently, an increasing amount of literature focuses on uncertainty as it relates to financial markets. The problem is that, the probability distribution of randomness in these markets is unknown. Typically, the unknown distribution is either estimated by statistical methods or calibrated to given market data by means of a model for the financial market. For example, in credit risk, the default probability is not observed, hence, have to be estimated from observable data. These methods introduce a large *model risk*.

Already, Knight (1921) pointed towards a formulation of risk which is able to treat such challenges in a systematic way. He was followed by Ellsberg (1961), who called random variables with known probability distribution *certain*, and those where the probability distribution is not known *uncertain*.

In this paper, we address these problems by constructing a model such that the parameters are characterised by uncertainty. Then, a single probability measure in a classical model is replaced by a family of probability measures, that is, a full class of models.

Following the modern literature in the area, we will call the feature that the probability distribution is not entirely fixed or, cannot be modelled by a single probability measure, *ambiguity*. This area has recently renewed the attention of researchers in mathematical finance to fundamental subjects such as arbitrage conditions, pricing mechanisms, and super-hedging. In equity markets, volatility uncertainty plays a crucial role, and has been extensively investigated, see for example, Avellaneda et al. (1995); Denis and Martini (2006); Lyons (1995); Vorbrink (2014). A major difficulty in this setting is that volatility uncertainty is characterized (at least in continuous time) by probabilities measures being mutually singular. Thus, the classical fundamental theorem of asset pricing fails to justify the no-arbitrage conditions, and new techniques are demanded, see Bayraktar and Zhang (2013); Bouchard and Nutz (2015); Burzoni et al. (2017), and Biagini et al. (2017).

In this paper, we introduce the concept of ambiguity to *defaultable term structure models*. The starting point for term structure models are typically bond prices of the form

$$P(t, T) = e^{-\int_t^T f(t,u)du}, \quad 0 \leq t \leq T, \tag{1}$$

where $(f(t, T))_{0 \leq t \leq T}$ is the instantaneous forward rate and T is the maturity time. This follows the seminal approach proposed in Heath et al. (1992). The presence of credit risk¹ in the model introduces an additional factor known as the default time. In this setting, bond prices are assumed to be absolutely continuous with respect to the maturity of the bond. This assumption is typically justified by the argument that, in practice, only a finite number of bonds are liquidly traded and the full term structure is obtained by interpolation, thus is smooth. There are two classical approaches to model market default risk: the *structural approach* Merton (1974) and the *reduced-form approach* (see for example, Artzner and Delbaen (1995); Duffie et al. (1996); Lando (1994) for some of the first works in this direction).

Structural models of credit risk describe the modelling of credit events specific to a particular corporate firm. Here, the underlying state is the value of a firm’s assets which is observable. Default time is define as the first time the firm’s asset value falls below a certain barrier level (for example, its liabilities). Hence, default is not a surprise. The approach links the default events to the firm’s economic fundamentals, consequently, default time is endogenously within the model. However, the assumption that the firm’s asset value is observable is often too strong in real applications, see Duffie and Lando (2001) and the survey article Frey and Schmidt (2011). Structural models in credit risk have been studied under many different viewpoints, see Black and Cox (1976); Frey and Schmidt (2009 2012); Gehmlich and Schmidt (2018); Geske (1977); Kim et al. (1993); Leland and Toft (1996); Merton (1974).

In comparison to the structural approach, reduced form approaches take a less stringent viewpoint regarding the mechanisms leading to default and model default events via an additional random structure. This additional degree of freedom together with their high tractability led to a tremendous success of this model class in credit risk modelling. For more details on the reduced form approach, we refer to Bielecki and Rutkowski (2002), and the references therein. Structural models can be embedded into (generalized) reduced form models as pointed out in Bélanger et al. (2004) and Fontana and Schmidt (2018).

Reduced-form models typically postulate that default time is totally inaccessible and, consequently, bond prices are absolutely continuous with respect to the maturity. Under the assumption of zero *recovery*², this implies that credit risky bond prices $P(t, T)$ are given by

$$P(t, T) = \mathbb{1}_{\{\tau > t\}} e^{-\int_t^T f(t,u)du} \tag{2}$$

with τ denoting the *random default time*. Since τ is totally inaccessible, it has an intensity λ . For example, if the intensity is constant, the default time is the first jump of a Poisson process with constant arrival rate λ . More generally, λ_t may be viewed as the conditional rate of occurrence of default at time t , given information up to that time. In a situation where the owner of a defaultable claim recovers part of its initial investment upon default, the associated survival process $\mathbb{1}_{\{\tau > t\}}$ in (2), is replaced by a semimartingale. The quantity of the investment recovered is the so-called *recovery*.

Under ambiguity, we suggest that there is some prior information at hand which gives a upper and lower bounds on the *intensity*. The implicit assumption that the probability distribution of default is known is quite restrictive. Thus, we analyse our problem in a multiple priors model which describe

¹ The risk that an agent fails to fulfil contractual obligations. Example of an instrument bearing credit risk is a corporate bond.

² The amount that the owner of a defaulted claim receives upon default.

uncertainty about the “true probability distribution”. By means of the Girsanov theorem, we construct the set of priors from the reference measure. The assumption is that all priors are equivalent.

In view of our framework, it is only important to acknowledge that a rating class provides an estimate of the one-year default probability in terms of a confidence interval. Also estimates for 3-, and 5-year default probabilities can be obtained from the rating migration matrix. Thus, leading to a certain amount of model risk.

The aim of this paper is to incorporate ambiguity into the context of single-name credit risk. We focus on ambiguity on the default intensity, and also discuss ambiguity on the recovery.

The main results are as follows: we obtain a necessary and sufficient condition for a reference probability measure to be a local martingale measure for credit risky bond markets under default ambiguity, thereby ensuring the absence of arbitrage in a sense to be precisely specified below. Furthermore, we consider the case where we have partial information on the amount that the owner of a defaulted claim receives upon default.

The next section of this paper introduces homogeneous ambiguity, and its example. Section 3 introduces the fundamental theorem of asset pricing (FTAP) under homogeneous ambiguity. In Sections 4 and 6, we derive the no-arbitrage conditions for defaultable term structure models with zero-recovery, and fractional recovery of the market value, in our framework. We conclude in section 7.

2. Intensity-Based Models

Intensity-based models are the most used model class for modeling credit risk (see (Bielecki and Rutkowski 2002, Chapter 8) for an overview of relevant literature). The default intensity, however, is difficult to estimate and therefore naturally carries a lot of uncertainty. This has led to the emergence of rating agencies which, since the early 20th century, estimate bond’s credit worthiness³.

Modeling of credit risk has up to now incorporated uncertainty in the default intensity in a systematic way. On the other side, a number of Bayesian approaches exist, utilizing filtering technologies (see, for example, Duffie and Lando (2001); Frey and Schmidt (2009), among many others).

Here, we introduce an alternative treatment of the lack of precise knowledge of the default intensity based on the concept of *ambiguity* following the seminal ideas from Frank Knight in Knight (1921).

Uncertainty in our setting will be captured through a family of probability measures \mathcal{P} replacing the single probability measure \mathbb{P} in classical approaches. Intuitively, each \mathbb{P} represents a model and the family \mathcal{P} collects models which we consider equally likely.

In this spirit, working with a single \mathbb{P} , or with a set $\mathcal{P} = \{\mathbb{P}\}$ which contains only one element, is in a one-to-one correspondence to assuming that the parameters of the underlying processes are exactly known. In financial markets, this is certainly not the case and ambiguity helps to incorporate this uncertainty into the used models.

We consider throughout a fixed finite time horizon $T^* > 0$. In light of our discussion above, let (Ω, \mathcal{F}) be a measurable space and \mathcal{P} be a set of probability measures on the measurable space (Ω, \mathcal{F}) . In particular, there is no fixed and known measure \mathbb{P} (except in the special case where \mathcal{P} contains only one element which we treat en passant).

Intensity-based default models correspond to the case where the ambiguity is *homogeneous*, i.e., there is a measure \mathbb{P}' such that $\mathbb{P} \sim \mathbb{P}'$ for all $\mathbb{P} \in \mathcal{P}$. Here, $\mathbb{P} \sim \mathbb{P}'$ means that \mathbb{P} and \mathbb{P}' are *equivalent*, that is, they have the same nullsets. The reference measure \mathbb{P}' only has the role of fixing events of measure zero for all probability measures under consideration. Intuitively, this means there is no ambiguity on these events of measure zero. In the following, E' is the expectation with respect to the reference measure \mathbb{P}' .

³ For a historical account, see Sylla (2002): John Moody founded the first rating agency in 1909, in the United States.

Remark 1. As a consequence of the equivalence of all probability measures in \mathcal{P} , all equalities and inequalities will hold almost-surely with respect to any probability measure $\mathbb{P} \in \mathcal{P}$, or, respectively, to \mathbb{P}' .

Ambiguity in Intensity-Based Models

In this section, we introduce ambiguity in intensity-based models. Our goal is not the most general approach in this setting: we rather focus on simpler, but still practically highly relevant cases. For a more general treatment, we refer the reader to Biagini and Zhang (2017). The main mathematical tool we use here is enlargement of filtrations and we refer the reader to Aksamit and Jeanblanc (2017) for further details and a guide to the literature.

Assume that under \mathbb{P}' we have a d -dimensional Brownian motion W with canonical and augmented⁴ filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}$ and a standard exponential random variable τ , independent of \mathcal{F}_{T^*} , that is, $\mathbb{P}'(t < \tau | \mathcal{F}_t) = \exp(-t)$, $0 \leq t \leq T^*$. The Brownian motion W has the role of modelling market movements and general information, excluding default information. We therefore call \mathbb{F} the *market filtration* in the following. The filtration $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T^*}$ includes default information and is obtained by a progressive enlargement of \mathbb{F} with τ , i.e.,

$$\mathcal{G}_t = \bigcap_{\epsilon > 0} \sigma(\mathbb{1}_{\{t \geq \tau\}}, W_s : 0 \leq s \leq t + \epsilon), \quad 0 \leq t \leq T^*.$$

To finalize our setup, we assume that $\mathcal{F} = \mathcal{G}_{T^*}$.

Note that up to now, everything has been specified under the reference measure \mathbb{P}' and nothing was said about the concrete models we are interested in (except about the nullsets). These models will now be introduced using the Girsanov theorem, i.e., by changing from \mathbb{P}' to the measures we are interested in.

Consequently, the next step is to construct measures \mathbb{P}^λ with appropriate processes λ – under \mathbb{P}^λ , the default time τ will have the intensity λ . More precisely, assume that λ is some positive process which is predictable with respect to the market filtration, \mathbb{F} . Define the density process Z^λ by

$$Z_t^\lambda := \begin{cases} \exp\left(\int_0^t (1 - \lambda_s) ds\right), & t < \tau \\ \lambda_\tau \exp\left(\int_0^\tau (1 - \lambda_s) ds\right) & t \geq \tau. \end{cases} \tag{3}$$

Note that Z^λ is a \mathbb{G} -local martingale and corresponds to a Girsanov-type change of measure (see Theorem VI.2.2 in Brémaud (1981)). If $E'[Z_{T^*}^\lambda] = 1$ we obtain an equivalent measure $\mathbb{P}^\lambda \sim \mathbb{P}'$ via

$$\mathbb{P}^\lambda(A) := E'[\mathbb{1}_A Z_{T^*}^\lambda] \quad \forall A \in \mathcal{F}. \tag{4}$$

Under the measure \mathbb{P}^λ , τ has intensity λ : more precisely, this means that the process

$$M_t^\lambda := \mathbb{1}_{\{t \leq \tau\}} - \int_0^{t \wedge \tau} \lambda_s ds, \quad 0 \leq t \leq T^*, \tag{5}$$

is a \mathbb{P}^λ -martingale.

Now we introduce a precise definition of ambiguity on the default intensity which is very much in spirit of the G -Brownian motion in Peng (2010): we consider an interval $[\underline{\lambda}, \bar{\lambda}] \subset (0, \infty)$ where $\underline{\lambda}$ and $\bar{\lambda}$ denote lower (upper) bounds in the default intensity. Intuitively, we include all possible

⁴ Augmentation can be done in a standard fashion with respect to \mathbb{P}' .

intensities lying in these bounds in our family of models \mathcal{P} . More precisely, we define the set of density generators H by

$$H := \{\lambda : \lambda \text{ is } \mathbb{F}\text{-predictable and } \mathbb{P}'(\underline{\lambda} \leq \lambda_t \leq \bar{\lambda}, t \in [0, T^*]) = 1\}.$$

Ambiguity on the default intensity is now covered by considering the concrete family of probability measures

$$\mathcal{P} := \{\mathbb{P}^\lambda : \lambda \in H\}. \tag{6}$$

In the following, we will always consider this \mathcal{P} . First, we observe that this set is convex.

Lemma 1. \mathcal{P} is a convex set.

Proof. Consider $\mathbb{P}^{\lambda'}, \mathbb{P}^{\lambda''} \in \mathcal{P}$ and $\alpha \in (0, 1)$. Then,

$$\alpha \mathbb{P}^{\lambda'}(A) + (1 - \alpha) \mathbb{P}^{\lambda''}(A) = E'[\mathbb{1}_A(\alpha Z_{T^*}^{\lambda'} + (1 - \alpha) Z_{T^*}^{\lambda''})].$$

Now consider the (well-defined) intensity λ , given by

$$\int_0^t \lambda_s ds := t - \log \left[\alpha e^{\int_0^t (1 - \lambda'_s) ds} + (1 - \alpha) e^{\int_0^t (1 - \lambda''_s) ds} \right],$$

$0 \leq t \leq T^*$. Then,

$$\alpha Z_{T^*}^{\lambda'} + (1 - \alpha) Z_{T^*}^{\lambda''} = Z_{T^*}^\lambda$$

such that by (4), $\mathbb{P}^\lambda \sim \mathbb{P}'$ refers to an equivalent change of measure. Finally, we have to check that $\lambda \in H$, which means that $\lambda \in [\underline{\lambda}, \bar{\lambda}]$, $0 \leq t \leq T^*$: note that

$$\begin{aligned} t - \log \left[\alpha e^{\int_0^t (1 - \lambda'_s) ds} + (1 - \alpha) e^{\int_0^t (1 - \lambda''_s) ds} \right] \\ \leq t - \log \left[\alpha e^{\int_0^t (1 - \bar{\lambda}) ds} + (1 - \alpha) e^{\int_0^t (1 - \bar{\lambda}) ds} \right] \\ \leq t - t(1 - \bar{\lambda}) = \bar{\lambda}t, \end{aligned}$$

and $\lambda_t \leq \bar{\lambda}$ follows. Similarly, $\underline{\lambda} \leq \lambda_t$ and the claim follows since t was arbitrary. \square

Remark 2. Intuitively, the requirement $\underline{\lambda} > 0$ states that there is always a positive risk of experiencing a default, which is economically reasonable. Technically it has the appealing consequence that all considered measures in \mathcal{P} are equivalent.

It turns out that the set of possible densities will play an important role in connection with measure changes. In this regard, we define *admissible measure changes* with respect to \mathcal{P} by

$$\mathcal{A} := \{\lambda^* : \lambda^* \text{ is positive, } \mathbb{F}\text{-predictable and } E'[Z_{T^*}^{\lambda^*}] = 1\}.$$

The associated Radon-Nikodym derivatives $Z_{T^*}^{\lambda^*}$ for $\lambda^* \in \mathcal{A}$ are the possible Radon-Nikodym derivatives for equivalent measure changes.

Remark 3. It is of course possible to consider an ambiguity setting more general than the specific one in (6). One possibility is to consider only a subset of \mathcal{P} . Another possibility is to allow the bounds $\underline{\lambda}$ and $\bar{\lambda}$ to depend on time, or even on the state of the process – this latter case is important for considering affine processes under uncertainty and we refer to [Fadina et al. \(2019\)](#) for further details. In Section 5, we consider indeed such a more general setting.

3. Absence of arbitrage under ambiguity

Absence of arbitrage and the respective generalizations, *no free lunch* (NFL), and *no free lunch with vanishing risk* (NFLVR), are well established concepts when the underlying probability measure \mathbb{P} is known and fixed. Here, we give a small set of sufficient conditions for absence of arbitrage extended to the setting with ambiguity. In this regard, consider, a fixed set \mathcal{P} of probability measures on the measurable space (Ω, \mathcal{F}) . In addition, let $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T^*}$ be a right-continuous filtration.

Discounted price processes of the traded assets are given by a finite dimensional \mathbb{G} -semimartingale $X = (X_t)_{0 \leq t \leq T^*}$. The semimartingale property holds equivalently in any of the filtration \mathbb{G}_+ or the augmentation of \mathbb{G}_+ , see (Neufeld and Nutz 2014, Proposition 2.2). It is well known that then X is a semimartingale for all $\mathbb{P} \in \mathcal{P}$.

A self-financing trading strategy is a predictable and X -integrable process Φ and the associated discounted gains process is given by the stochastic integral of Φ with respect to X ,

$$(\Phi \cdot X)_t = \int_0^t \Phi_u dX_u, \quad 0 \leq t \leq T^*.$$

Intuitively, an *arbitrage* is an admissible self-financing trading strategy which starts from zero initial wealth, has non-negative pay-off under all possible future scenarios, hence for all $P \in \mathcal{P}$, there is at least one $P \in \mathcal{P}$, such that the pay-off is positive. This is formalized in the following definition, compare for example Vorbrink (2014). As usual a trading strategy is called a -admissible, if $(\Phi \cdot X)_t \geq -a$ for all $0 \leq t \leq T^*$.

Definition 1. A self-financing trading strategy Φ is called \mathcal{P} -arbitrage if it is a -admissible for some $a > 0$ and

- (i) for every $\mathbb{P} \in \mathcal{P}$ we have that $(\Phi \cdot X)_{T^*} \geq 0$, \mathbb{P} -almost surely, and
- (ii) for at least one $\mathbb{P} \in \mathcal{P}$ it holds that $\mathbb{P}((\Phi \cdot X)_{T^*} > 0) > 0$.

Since all probability measures $\mathbb{P} \in \mathcal{P}$ are considered as possible, a \mathcal{P} -arbitrage is a riskless trading strategy for all possible models (i.e., for all $\mathbb{P} \in \mathcal{P}$) while it is a profitable strategy for at least one scenario (i.e., for at least one $\mathbb{P} \in \mathcal{P}$).

The main tool for classifying arbitrage free markets will be local martingale measures, even in the setting with ambiguity. In this regard, we call a probability measure \mathbb{Q} a *local martingale measure* if X is a \mathbb{Q} -local martingale.

It is well-known that *no arbitrage* or, more precisely, *no free lunch with vanishing risk* (NFLVR) in a market where discounted price processes are locally bounded semimartingales is equivalent to the existence of an equivalent local martingale measure (ELMM), see Delbaen and Schachermayer (1994, 1998). The technically difficult part of this result is to show that a precise criterion of absence of arbitrage implies the existence of an ELMM. In the following, we will not aim at such a deep result under ambiguity, but utilize the easy direction, namely that existence of an ELMM implies the absence of arbitrage as formulated below.

From the classical *fundamental theorem of asset pricing* (FTAP), the following result follows easily.

Theorem 1. If, for every $\mathbb{P} \in \mathcal{P}$ there exists an equivalent local martingale measure $\mathbb{Q}(\mathbb{P})$, then there is no arbitrage in the sense of Definition 1.

Proof. Indeed, assume on the contrary that there is an arbitrage Φ with respect to some measure $\mathbb{P} \in \mathcal{P}$ which we fix for the remainder of the proof. If there exists an ELMM $\mathbb{Q}(\mathbb{P})$ then Φ would be an arbitrage strategy together with an ELMM, a contradiction to the classical FTAP. \square

This (sufficient) condition directly corresponds to the existing results in the literature (see, for example, Biagini et al. (2017)) where arbitrages of the first kind are studied under the additional assumption of continuity for the traded assets.

4. Ambiguity on the default intensity

Our aim is to discuss dynamic term structure models under default risk with ambiguity on the default intensity. The relevance of this issue has, for example, already been reported in Riedel (2015). Here, we take this as motivation to propose a precise framework taking ambiguity on the default intensity into account.

4.1. Dynamic defaultable term structures

We specialize the considerations of absence of arbitrage in section 3 to defaultable bond markets. Recall that we have a filtration \mathbb{G} at hand and that τ is the \mathbb{G} -stopping time at which the company defaults. We define the *default indicator process* H by

$$H_t = \mathbb{1}_{\{t \geq \tau\}}, \quad 0 \leq t \leq T^*.$$

The associated *survival process* is $1 - H$. A credit risky bond with a maturity time $T \leq T^*$ is a contingent claim promising to pay one unit of currency at T . We denote the price of such a bond at time $t \leq T$ by $P(t, T)$. If no default occurs prior to T , $P(t, T) = 1$. We will first consider *zero recovery*, i.e., assume that the bond loses its total value at default. Then $P(t, T) = 0$ on $\{t \geq \tau\}$.

Besides zero recovery, we only make the weak assumption that bond-prices prior to default are positive and absolutely continuous with respect to maturity T . This follows the well-established approach by Heath et al. (1992). More formally, we assume that

$$P(t, T) = \mathbb{1}_{\{\tau > t\}} \exp\left(-\int_t^T f(t, u) du\right) \quad 0 \leq t \leq T. \tag{7}$$

The initial forward curve $T \mapsto f(0, T)$ is then assumed to be sufficiently integrable and the *forward rate processes* $f(\cdot, T)$ are assumed to follow Itô processes satisfying

$$f(t, T) = f(0, T) + \int_0^t a(s, T) ds + \int_0^t b(s, T) dW_s, \tag{8}$$

for $0 \leq t \leq T$. Recall that W was chosen to be a Brownian motion. We denote by \mathcal{O} the optional σ -algebra and by \mathcal{B} the Borel σ -algebra.

Assumption 1. We require the following technical assumptions:

- (i) the initial forward curve is measurable, and integrable on $[0, T^*]$:

$$\int_0^{T^*} |f(0, u)| du < \infty,$$

- (ii) the drift parameter $a(\omega, s, t)$ is \mathbb{R} -valued $\mathcal{O} \otimes \mathcal{B}$ -measurable and integrable on $[0, T^*]$:

$$\int_0^{T^*} \int_0^{T^*} |a(s, t)| ds dt < \infty,$$

- (iii) the volatility parameter $b(\omega, s, t)$ is \mathbb{R}^d -valued, $\mathcal{O} \otimes \mathcal{B}$ -measurable, and

$$\sup_{s, t \leq T^*} \|b(s, t)\| < \infty.$$

(iv) Let r_t be the short rate process at time t , for $0 \leq t \leq T^*$. With probability one, it holds that

$$0 < f(t, t) - r_t, \quad 0 \leq t \leq T^*.$$

Set for $0 \leq t \leq T \leq T^*$,

$$\begin{aligned} \bar{a}(t, T) &= \int_t^T a(t, u) du, \\ \bar{b}(t, T) &= \int_t^T b(t, u) du. \end{aligned}$$

Lemma 2. Under Assumption 1 it holds that,

$$\int_t^T f(t, u) du = \int_0^T f(0, u) du + \int_0^t \bar{a}(\cdot, u) du + \int_0^t \bar{b}(\cdot, u) dW_u - \int_0^t f(u, u) du$$

for $0 \leq t \leq T \leq T^*$, almost surely.

This follows as in Heath et al. (1992): for the case W is a Brownian motion, this is Lemma 6.1 in Filipović (2009). This result could also be generalized where W is replaced by a semimartingale with absolutely continuous characteristics, see Proposition 5.2 in Björk et al. (1997). Note that the strong condition (iii) of uniform boundedness of b in Assumption 1 is needed for the application of the stochastic Fubini theorem.

4.2. Absence of arbitrage without ambiguity on the default intensity

The first step towards the study of term structure models with default ambiguity is the study of absence of arbitrage in (classical) intensity-based dynamic term structure models. Consider $\lambda = (\lambda_t)_{0 \leq t \leq T^*} \in \mathcal{A}$ and the probability measure \mathbb{P}^λ . Then, the dual predictable projection H^p of H is given by $H_t^p = \int_0^{t \wedge \tau} \lambda_s ds$ (under \mathbb{P}^λ). Moreover, the Doob-Meyer decomposition yields that

$$M^\lambda := H - \int_0^{\cdot \wedge \tau} \lambda_s ds$$

is \mathbb{P}^λ -martingale, compare equation (5).

For discounting, we use the bank account. Its value is given by a stochastic process starting with 1 which is then upcounted by the short rate r , i.e., the value process of the bank account is $\gamma(t) = \exp(\int_0^t r_s ds)$ with an \mathbb{G} -predictable process r .

In the bond market context considered here, a measure \mathbb{Q} is called local martingale measure if, for any maturity $T \in (0, T^*]$, the discounted bond price process for the bond with maturity T is a \mathbb{Q} -local martingale. Then, we obtain the following result.

Proposition 1. Assume that Assumption 1 holds. Consider a measure \mathbb{Q} on (Ω, \mathcal{F}) , such that M^λ is a \mathbb{Q} -martingale, that W is a \mathbb{Q} -Brownian motion and that $\mathbb{Q}(\int_0^{T^*} |r_s| ds < \infty) = 1$. Then \mathbb{Q} is a local martingale measure if and only if

- (i) $f(t, t) = r_t + \lambda_t$,
- (ii) the drift condition

$$\bar{a}(t, T) = \frac{1}{2} \|\bar{b}(t, T)\|^2,$$

holds $dt \otimes d\mathbb{Q}$ -almost surely for $0 \leq t \leq T \leq T^*$ on $\{\tau > t\}$.

Proof. We set $E = 1 - H$ and $F(t, T) = \exp\left(-\int_t^T f(t, u)du\right)$. Then (7) can be written as $P(t, T) = E(t)F(t, T)$. Integrating by part yields

$$dP(t, T) = F(t-, T)dE(t) + E(t-)dF(t, T) + d[E, F(\cdot, T)]_t.$$

For $\{t < \tau\}$,

$$dP(t, T) = P(t-, T) \left(-\lambda_t dt + \left(f(t, t) + \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{a}(t, T) \right) dt \right) - P(t-, T) \left(dM^\lambda + \bar{b}(t, T)dW_t \right). \tag{9}$$

The discounted bond price process is a local martingale if and only if the predictable part in the semimartingale decomposition vanishes, i.e.,

$$f(t, t) - r_t - \lambda_t - \bar{a}(t, T) + \frac{1}{2} \|\bar{b}(t, T)\|^2 = 0. \tag{10}$$

Letting $T = t$ we obtain (i) and (ii) and the result follows. \square

4.3. Absence of arbitrage with ambiguity on the default intensity

Next, we derive the no-arbitrage conditions for the forward rate in term of the intensity and the short rate, and also the conditions for the drift and volatility parameters, under ambiguity on the default intensity. In this regard, we require a bit more structure: we assume that the setting detailed in section 2 holds, in particular, we consider the family of probability measures \mathcal{P} constructed in equation (6). Recall that for all $\mathbb{P}^\lambda \in \mathcal{P}$, W is a Brownian motion and that $\mathbb{P}^\lambda \sim \mathbb{P}'$. We may, for the moment, safely assume that the market filtration \mathbb{F} satisfies the usual conditions under \mathbb{P}' .

For a generic real-valued, \mathbb{F} -progressive process $\theta = (\theta_t)_{t \geq 0}$, let the process $z^\theta = (z_t^\theta)_{0 \leq t \leq T^*}$ be given as the unique strong solution of

$$dz_t^\theta = \theta_t z_t^\theta dW_t, \quad z_0^\theta = 1. \tag{11}$$

Then, z^θ is a continuous local martingale. If $E'[z_{T^*}^\theta] = 1$, we can define a probability measure $\tilde{\mathbb{P}}^\theta$ by letting

$$\tilde{\mathbb{P}}^\theta(A) := E'[\mathbb{1}_A z_{T^*}^\theta], \quad \forall A \in \mathcal{F}, \tag{12}$$

just as in equation (4). Under $\tilde{\mathbb{P}}^\theta$ the process $\tilde{W} = W - \int_0^\cdot \theta_s ds$ is a Brownian motion, i.e., W itself became a Brownian motion with drift θ , see Theorem 5.1 in Chapter 3 of Karatzas and Shreve (1998).

Moreover, set $\tilde{\lambda}_t := (f(t, t) - r_t) \cdot \lambda_t^{-1}$, $t \in [0, T^*]$. Note that under Assumption 1, $\tilde{\lambda}$ is positive (which is necessary for an equivalent change of measure). The associated density is abbreviated by

$$Z_{T^*}^* = Z_{T^*}^{\tilde{\lambda}}.$$

Theorem 2. Consider $\mathbb{P}^\lambda \in \mathcal{P}$. Under Assumption 1, there exists an ELMM to \mathbb{P}^λ , if there exists an \mathbb{F} -progressive process θ^* such that

- (i) $E^{\mathbb{P}^\lambda} [Z_{T^*}^* z_{T^*}^{\theta^*}] = 1$,
- (ii) the drift condition

$$\bar{a}(t, T) = \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{b}(t, T)\theta_t^*, \quad 0 \leq t \leq T \leq T^*$$

holds $dt \otimes d\mathbb{P}^\lambda$ -almost surely on $\{t < \tau\}$.

Intuitively, the theorem states that for the probability measure \mathbb{P}^λ , we find an ELMM if we are able to perform an equivalent change of measure (condition (i)) in such a way that under the new measure the drift condition holds for the Brownian motion with drift θ^* (condition (ii)).

Proof. We start from some $\mathbb{P}^\lambda \in \mathcal{P}$ and fix this measure in the following. This means that, under \mathbb{P}^λ , W is a Brownian motion and τ has intensity λ . In the search for an ELMM we are looking for an equivalent measure \mathbb{P}^* which satisfies the conditions of Proposition 1.

In this regard, note the following: by its definition, (11), together with condition (i), z^{θ^*} is a density process for a change of measure via the Girsanov theorem for Itô processes, see Theorem 5.1 in Chapter 3 of Karatzas and Shreve (1998). Moreover, by (3) together with (i), $Z^* = Z^{\bar{\lambda}}$ is the density for the change in intensity from λ under \mathbb{P}^λ to the intensity given by $\lambda_t^* := (f(t, t) - r_t)_{0 \leq t \leq T^*}$, see (Brémaud 1981, Theorem VI.2.T2). We set

$$d\mathbb{P}^* := Z_{T^*}^* z_{T^*}^{\theta^*} d\mathbb{P}^\lambda.$$

According to Theorem 3.40 in Chapter III of Jacod and Shiryaev (2003), this refers to a Girsanov-type (and equivalent) change of measure. Moreover, $W_t^* = W_t - \int_0^t \theta_s^* ds$, $0 \leq t \leq T^*$, is a \mathbb{P}^* -Brownian motion and M^{λ^*} is a \mathbb{P}^* -martingale.

We now show that \mathbb{P}^* is also a local martingale measure. Recall from (9) that, under \mathbb{P}^λ ,

$$\frac{dP(t, T)}{P(t-, T)} = \left(f(t, t) + \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{a}(t, T) \right) dt - dH_t - \bar{b}(t, T) dW_t,$$

for $\{t < \tau\}$; here $H_t = \mathbb{1}_{\{t \geq \tau\}}$ is the default indicator. We introduce the martingales W^* and M^{λ^*} into this representation: note that

$$\begin{aligned} \frac{dP(t, T)}{P(t-, T)} &= \left(-\lambda^*(t) + f(t, t) + \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{a}(t, T) - \bar{b}(t, T)\theta_t^* \right) dt \\ &\quad - dM_t^{\lambda^*} - \bar{b}(t, T) dW_t^*, \end{aligned}$$

again on $\{t < \tau\}$. Since, by assumption, $-\lambda_t^* + f(t, t) = r_t$, together with the drift condition (ii), we obtain for the discounted bond price process $\tilde{P}(t, T) = P(t, T) / \gamma_t$,

$$d\tilde{P}(t, T) = \tilde{P}(t-, T) \cdot \left(-dM_t^{\lambda^*} - \bar{b}(t, T) dW_t^* \right),$$

which is a \mathbb{P}^* -local martingale and the proof is finished. \square

5. Examples

The setting proposed above can, in dimension one, be directly linked to a special case of the non-linear affine processes introduced in Fadina et al. (2019). Indeed, note that for a progressive process λ , the integral

$$X_t := x + \int_0^t \lambda_s ds, \quad 0 \leq t \leq T^*$$

is a special semimartingale. Moreover, there are affine bounds on drift and volatility (the bound of the volatility is simply zero) since

$$\underline{\lambda} \leq \lambda_t \leq \bar{\lambda},$$

such that X is a non-linear affine process.

The major advantage of this setting is that numerical methods via non-linear PDE come into reach. More precisely, Theorem 4.1 in Fadina et al. (2019) shows that whenever ψ is Lipschitz, the non-linear expectation

$$\mathcal{E}[\psi(X_{T^*})] := \sup_{\mathbb{P} \in \mathcal{P}} E^{\mathbb{P}}[\psi(X_{T^*})] \tag{13}$$

can be expressed as viscosity solution of the fully non-linear PDE

$$\begin{cases} -\partial_t v(t, x) - G(x, \partial_x v(t, x)) = 0 & \text{on } [0, T^*) \times [\underline{\lambda}, \bar{\lambda}], \\ v(T, x) = \psi(x) & x \in [\underline{\lambda}, \bar{\lambda}], \end{cases} \tag{14}$$

where G is defined by

$$G(x, p) := \sup_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} \{\lambda p\} \tag{15}$$

and $v(0, x) = \mathcal{E}[\psi(X_{T^*})]$ (the dependency on x arises through $X_0 = x$).

Clearly, when p is either strictly positive (hence $\partial_x v(t, x)$) or negative, then the supremum in (15) is immediate and the PDE (14) can be solved using standard methods. This means that the solution to the non-linear expectation is obtained simply by the upper bound $\bar{\lambda}$ (or the lower bound $\underline{\lambda}$, respectively). Such a condition holds if ψ is monotone. The more general case has to be solved using numerical methods and we provide a simple example now.

Example 1. Consider a butterfly on X_{T^*} , i.e., the derivative with the payoff

$$\psi(x) = (x - K_1)^+ - 2(x - K_2)^+ + (x - K_3)^+,$$

where we choose $K_1 = -0.2, K_2 = 0.3,$ and $K_3 = 0.8$. Moreover, let $\underline{\lambda} = 0.1$ and $\bar{\lambda} = 0.5$. Then the upper and lower price bounds for the butterfly are shown in Figure 1 (the upper prices are given by the nonlinear expectation in equation (13), while the lower prices are obtained by replacing the supremum in (13) by an infimum).

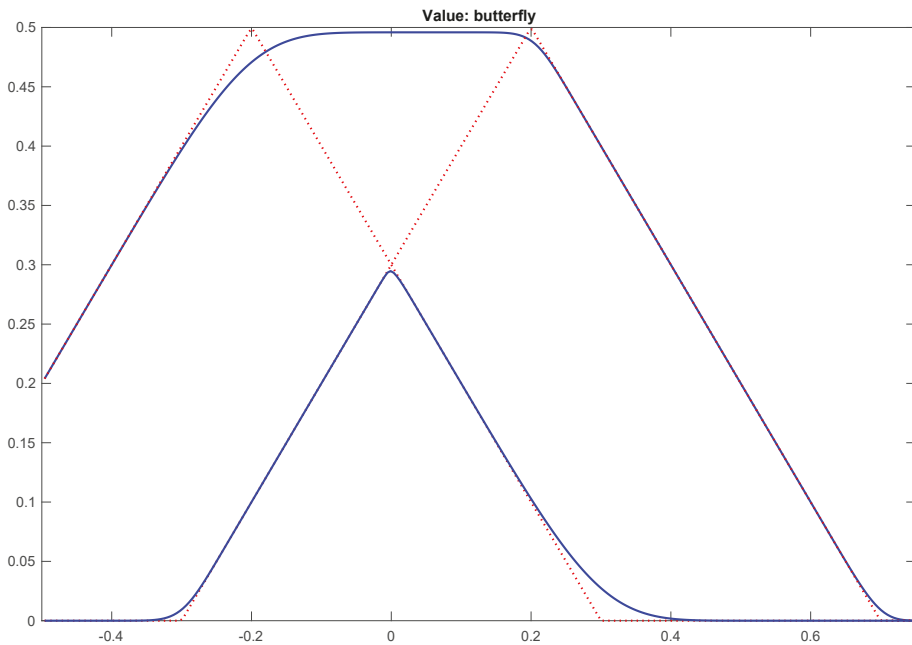


Figure 1. This figure shows the solution of the nonlinear PDE in Equation (14) with boundary condition $\psi(y) = (y - K_1 + x)^+ - 2(y - K_2 + x)^+ + (y - K_3 + x)^+, K_1 = -0.2, K_2 = 0.3, K_3 = 0.8,$ and $x \in [-0.5, 0.7]$ is depicted on the x-axis of the plot. The dashed lines show the solution for the lower bound (upper bound, respectively), i.e., for the constants $\underline{\lambda} = 0.1$ and $\bar{\lambda} = 0.5$. The upper and lower solid lines show the upper and lower price bounds.

6. Ambiguity on the recovery

A detailed study of bond markets beyond zero recovery is often neglected, the high degree of uncertainty about the recovery mechanism being a prime reason for this. This motivates us to take some time for developing a deeper understanding of a suitable recovery model under ambiguity.

We start from the observation that intensity-based models always need certain recovery assumptions, as for example, zero recovery, fractional recovery of treasury, and fractional recovery of par value, see (Bielecki and Rutkowski 2002, Chapter 8). We have so far considered the case where the credit risky bond becomes worthless at default (zero recovery). In the following, we will consider *fractional recovery of market value* where the credit risky bond loses a fraction of its market value upon default. Other recovery models can be treated in a similar fashion.

6.1. Fractional recovery without ambiguity

Fractional recovery of market value (RMV) is specified through a marked point process $(T_n, R_n)_{n \geq 1}$ where the stopping times (T_n) denote the default times and $R_n \in (0, 1]$ denotes the associated fractional recovery. Let

$$R_t = \prod_{T_n \leq t} R_n, \quad 0 \leq t \leq T^*, \tag{16}$$

denote the *recovery process*. Then, R is non-increasing, positive with $R_0 = 1$. The recovery process replaces the default indicator in (7). More precisely, we assume that the family of defaultable bond prices under RMV satisfy

$$P_R(t, T) = R_t \exp\left(-\int_t^T f(t, u) du\right), \quad 0 \leq t \leq T \leq T^*. \tag{17}$$

Remark 4. *If a default occurs at $t = T_n$, the bond loses a random fraction $q_t = 1 - R_{T_n} \in [0, 1)$ of its pre-default value. Thus, the value $(1 - q_t)P(t-, T)$ is immediately available to the bond owner at default. It is still subject to default risk because of the possible future defaults occurring at T_{n+1}, T_{n+2}, \dots*

First, we state a generalization of Proposition 1 to this setting. To this end, we require more structure and continue in the setting of the section 2. Assume that the marked point process $(T_n, R_n)_{n \geq 1}$ is independent from W and standard in the following sense: the random times (T_n) are the jumping times from a Poisson process with intensity one, and the recovery values (R_n) are independent from (T_n) and W , and uniformly distributed in $[\underline{r}, \bar{r}] \subset (0, 1]$.

The filtration $\mathbb{G} = (\mathcal{G}_t)_{0 \leq t \leq T^*}$ is obtained by a progressive enlargement of \mathbb{F} with default and recovery information (given by R), i.e.,

$$\mathcal{G}_t = \bigcap_{\epsilon > 0} \sigma(R_s, W_s : 0 \leq s \leq t + \epsilon), \quad 0 \leq t \leq T^*.$$

We assume again $\mathcal{F} = \mathcal{G}_{T^*}$. As next step, we introduce measure changes for the marked point process. Let

$$\Phi_t = \sum_{T_n \leq t} R_n, \quad 0 \leq t \leq T^*.$$

This implies that the defaultable bond prices under RMV, $P_R(0, t)$, at maturity time t of this bond, pays out Φ_t , the accumulated fractional default losses. Then, Φ is a special semimartingale w.r.t. \mathbb{G} . Let

$$\mu^\Phi(dt, dx) = \sum_{n \geq 1} \delta_{(T_n, R_n)}(dt, dx)$$

denote the associated jump measure and let $\nu^\Phi(dt, dx)$ denote its compensator, see Chapter II.1 in [Jacod and Shiryaev \(2003\)](#) or Chapter VIII.1 in [Brémaud \(1981\)](#). Note that $\nu^\Phi(dt, dx) = \mathbb{1}_{\{x \in [\underline{l}, \bar{r}]\}}(\bar{r} - \underline{l})^{-1} dx dt$.

We introduce the densities

$$L^{\mu,h} = \left(\prod_{T_n \leq T^*} \mu_{T_n} h(T_n, R_n) \right) \cdot e^{\int_0^{T^*} \int_{\underline{l}}^{\bar{r}} (1 - \mu_t h(t,x)) (\bar{r} - \underline{l})^{-1} dx dt}, \tag{18}$$

where the predictable process μ is positive and, for any $x \in [\underline{l}, \bar{r}]$, the \mathbb{G} -predictable process $(h(t, x))_{0 \leq t \leq T^*}$ is also positive. If $E[L_{T^*}] = 1$, we can define the equivalent measure $\mathbb{P}^{\mu,h}$ by

$$d\mathbb{P}^{\mu,h} = L_{T^*} d\mathbb{P}'. \tag{19}$$

By \mathcal{A}^* we denote all pairs (μ, h) which satisfy the above properties. Then, the compensator of the jump measure μ^Φ under $\mathbb{P}^{\mu,h}$ is

$$\mu_t h(t, x) \nu^\Phi(dt, dx) = \mu_t h(t, x) \mathbb{1}_{\{x \in [\underline{l}, \bar{r}]\}} (\bar{r} - \underline{l})^{-1} dx dt =: K_t^{\mu,h}(dx) dt, \tag{20}$$

see T10 in Section VIII.3 of [Brémaud \(1981\)](#). Next, we compute the compensator of R . We obtain from (16) that

$$R_t - \int_0^t \int R_{s-} (x - 1) K_s^{\mu,h}(dx) ds, \quad 0 \leq t \leq T^* \tag{21}$$

is a $\mathbb{P}^{\mu,h}$ -martingale. For a \mathbb{G} -progressive process g , we denote

$$M^g = R + \int_0^\cdot R_{s-} g_s ds.$$

Proposition 2. *Assume that Assumption 1 holds and let g be a positive and \mathbb{G} -predictable process. Consider a measure \mathbb{Q} on (Ω, \mathcal{F}) , such that M^g is a \mathbb{Q} -martingale, W is a \mathbb{Q} -Brownian motion, and $\mathbb{Q}(\int_0^{T^*} |r_s| ds < \infty) = 1$. Then \mathbb{Q} is a local martingale measure if and only if*

- (i) $f(t, t) = r_t + g_t$,
- (ii) the drift condition

$$\bar{a}(t, T) = \frac{1}{2} \left\| \bar{b}(t, T) \right\|^2, \quad 0 \leq t \leq T \leq T^*,$$

holds $dt \otimes d\mathbb{Q}$ -almost surely.

Proof. We generalize the proof of Proposition 1 to the case of RMV. To this end, let $F(t, T) = \exp\left(-\int_t^T f(t, u) du\right)$. Then (17) reads $P_R(t, T) = R(t)F(t, T)$. Integrating by part yields

$$dP_R(t, T) = F(t-, T) dR(t) + R(t-) dF(t, T) + d[R, F(\cdot, T)]_t.$$

Note that, by assumption,

$$M_t^g = R_t + \int_0^t R_{s-} g_s ds, \quad 0 \leq t \leq T^*,$$

is a Q -martingale and that $[R, F(\cdot, T)] = 0$ since R is of finite variation and $F(\cdot, T)$ is continuous. Hence, by Lemma 2,

$$dP_R(t, T) = P_R(t-, T) \left(-g_t + f(t, t) + \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{a}(t, T) \right) dt + P_R(t-, T) \left(r_t dt + dM^S - \bar{b}(t, T) dW_t \right), \tag{22}$$

and we obtain the result as in the proof of Proposition 1. \square

Example 2. A classical example is when the defaults (T_n) arrive at rate $\lambda > 0$, and the recovery values (R_n) are i.i.d. Then, $\int (x - 1) \nu^\Phi(dt, dx) = \lambda E[R_1 - 1] dt$. We obtain that the instantaneous forward rate of the defaultable bond $f(t, t)$ equals $r_t + \lambda E[1 - R_1]$. In the case of zero recovery, we recover $f(t, t) = r_t + \lambda$, and, in the case of full recovery (the case without default risk), $f(t, t) = r_t$.

6.2. Fractional recovery with ambiguity

We introduce ambiguity in this setting by changing from the standardized measure \mathbb{P}' to various appropriate measures via the Girsanov theorem. We also generalize the setting for ambiguity from the quite specific \mathcal{P} to a general set of probability measures \mathcal{P}^* here, see Remark 3. The reason for this is also economic: while bounding the intensity from above and below seems to be quite plausible, an upper/lower bound on the recovery (i.e., on (R_n)) sounds too strong for some applications.

Recall that \mathcal{A}^* was the set of all candidates (μ, h) which induce the measure changes via (19). Ambiguity is introduced by the set \mathcal{P}^* of probability measures satisfying

$$\emptyset \neq \mathcal{P}^* \subset \{\mathbb{P}^{\mu, h} : (\mu, h) \in \mathcal{A}^*\}. \tag{23}$$

If \mathcal{P}^* contains only one probability measure, we are in the classical setting, otherwise there is ambiguity in the market. Measure changes from $\mathbb{P}^{\mu, h}$ to a new measure are done via the density L^{μ^*, h^*} (see (18)) where, as above, $\mu^*, h^*(\cdot, x)$, $x \in [\underline{r}, \bar{r}]$ are positive and progressive. Recall the definition of the density z^θ in (11).

Theorem 3. Let $g_t^* := f(t, t) - r_t$, $t \in [0, T^*]$, and assume that Assumption 1 holds. Then there exists an ELMM for $\mathbb{P}^{\mu, h} \in \mathcal{P}^*$ if

- (i) there exists an \mathbb{F} -progressive θ^* such that $E[z_{T^*}^{\theta^*}] = 1$,
- (ii) there exist μ^* and $h^*(t, x)$, such that $E[L^{\mu^*, h^*}] = 1$ and

$$g_t^* = \int (x - 1) \mu_t^* h^*(t, x) K_t^{\mu, h}(dx), \quad 0 \leq t \leq T^*,$$

$dt \otimes dP'$ -almost surely, and

- (iii) the drift condition

$$\bar{a}(t, T) = \frac{1}{2} \|\bar{b}(t, T)\|^2 - \bar{b}(t, T) \theta_t^*, \quad 0 \leq t \leq T \leq T^*,$$

holds $dt \otimes dP'$ -almost surely.

Absence of arbitrage in this general ambiguity setting can now be classified, thanks to Theorem 1 as follows: if an ELMM exists for each $\mathbb{P}^{\mu, h} \in \mathcal{P}^*$, then the market is free of arbitrage in the sense of Definition 1.

Proof. Fix $\mathbb{P}^{\mu, h} \in \mathcal{P}^*$. We can define an equivalent measure $\mathbb{P}^* \sim \mathbb{P}^{\mu, h}$ by

$$d\mathbb{P}^* := L^{\mu^*, h^*} z_{T^*}^{\theta^*} d\mathbb{P}^{\mu, h},$$

with μ^* and h^* as in (ii). According to Theorem 3.40 in Chapter III of Jacod and Shiryaev (2003), this refers to a Girsanov-type (and equivalent) change of measure. Moreover, $W^* = W - \int_0^t \theta_s^* ds$ is a \mathbb{P}^* -Brownian motion. Next, note that the compensator of the jump measure μ^Φ under \mathbb{P}^* computes, according to T10 in Section VIII.3 in Brémaud (1981), to

$$v^*(dt, dx) := \mu_t^* h^*(t, x) K_t^{\mu, h}(dx) dt$$

with $K_t^{\mu, h}$ from Equation (20). This implies that

$$M_t^{\mathcal{G}^*} = R_{s-} (x - 1) v^*(ds, dx) = R_t - \int_0^t R_{s-} \mathcal{G}_s^* ds \tag{24}$$

is a \mathbb{P}^* -martingale.

Now, we show that \mathbb{P}^* is indeed a martingale measure: from (22) we obtain that

$$\frac{dP_R(t, T)}{P_R(t-, T)} = \left(f(t, t) + \frac{1}{2} \left\| \bar{b}(t, T) \right\|^2 - \bar{a}(t, T) \right) dt + dR_t - \bar{b}(t, T) dW_t.$$

It follows that

$$\begin{aligned} \frac{dP_R(t, T)}{P_R(t-, T)} &= \left(f(t, t) + \frac{1}{2} \left\| \bar{b}(t, T) \right\|^2 - \bar{a}(t, T) - \bar{b}(t, T) \theta_t^* \right) dt + dM_t^{\mathcal{G}^*} - \bar{b}(t, T) dW_t^* \\ &= r_t dt + dM_t^{\mathcal{G}^*} - \bar{b}(t, T) dW_t^*, \end{aligned}$$

by the definition of \mathcal{G}^* and the drift condition (iii). Hence, discounted bond prices are \mathbb{P}^* -local martingales and the proof is finished. \square

Remark 5. We can view zero recovery in the above setting by assuming that $\mathbb{P}^l(R_1 = 0) = 1$ and letting $\tau = T_1$. Note that this case is excluded in RMV setting, under this assumption, at the first default all prices drop to zero and further defaults can not occur.

7. Conclusions

This paper provides a first step towards including ambiguity in intensity based models for credit risk. Many research questions are still open: first, the extension of constant boundaries $\underline{\lambda}$, $\bar{\lambda}$ to time-dependent, or, as in Fadina et al. (2019), state-dependent boundaries. Second, the extension to two or more defaultable assets, where default dependence comes into play. Third, estimation and determination of the bounds by statistical methods or calibration to market data. We hope that our paper provides the foundation for future works in these directions.

Author Contributions: All authors contributed equally to the paper. The author appearance is in alphabetical order.

Funding: Financial support by Carl-Zeiss-Stiftung, and German Research Foundation (DFG) via CRC 1283 is gratefully acknowledged. We also thank the Freiburg Institute of Advanced Studies (FRIAS) for its hospitality and financial support.

Conflicts of Interest: The authors declare no conflict of interest.

References

Aksamit, Anna, and Monique Jeanblanc. 2017. *Enlargements of Filtration with Finance in View*. SpringerBriefs in Quantitative Finance. Berlin: Springer.

Artzner, Philippe, and Freddy Delbaen. 1995. Default risk insurance and incomplete markets. *Mathematical Finance* 5: 187–95. [CrossRef]

Avellaneda, Marco, Arnon Levy, and Antonio Parás. 1995. Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance* 2: 73–88. [CrossRef]

- Bayraktar, Erhan and Yuchong Zhang. 2013. Fundamental theorem of asset pricing under transaction costs and model uncertainty. arXiv:1309.1420v2.
- Bélanger, Alain, Steven E. Shreve, and Dennis Wong. 2004. A general framework for pricing credit risk. *Mathematical Finance* 14: 317–50. [CrossRef]
- Biagini, Francesca, and Yinglin Zhang. 2017. Reduced-form framework under model uncertainty. arXiv:1707.04475.
- Biagini, Sara, Bruno Bouchard, Constantin Kardaras, and Marcel Nutz. 2017. Robust fundamental theorem for continuous processes. *Mathematical Finance* 27: 963–87. [CrossRef]
- Bielecki, Tomasz, and Marek Rutkowski. 2002. *Credit Risk: Modeling, Valuation and Hedging*. Berlin, Heidelberg and New York: Springer.
- Björk, Tomas, Giovanni Di Masi, Yuri Kabanov, and Wolfgang Runggaldier. 1997. Towards a general theory of bond markets. *Finance and Stochastics* 1: 141–74. [CrossRef]
- Black, Fischer, and John C. Cox. 1976. Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance* 31: 351–67. [CrossRef]
- Bouchard, Bruno, and Marcel Nutz. 2015. Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability* 25: 823–59. [CrossRef]
- Brémaud, Pierre. 1981. *Point Processes and Queues*. Berlin, Heidelberg and New York: Springer.
- Burzoni, Matteo, Frank Riedel, and Mete H. Soner. 2017. Viability and arbitrage under knightian uncertainty. arXiv:1707.03335 .
- Delbaen, Freddy, and Walter Schachermayer. 1994. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen* 300: 463–520. [CrossRef]
- Delbaen, Freddy, and Walter Schachermayer. 1998. The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen* 312: 215–50. [CrossRef]
- Denis, Laurent, and Claude Martini. 2006. A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *The Annals of Applied Probability* 16: 827–52. [CrossRef]
- Duffie, Darrell, and David Lando. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica* 69: 633–64. [CrossRef]
- Duffie, Darrell, Mark Schroder, and Costis Skiadas. 1996. Recursive valuation of defaultable securities and the timing of resolution of uncertainty. *The Annals of Applied Probability* 6: 1075–90. [CrossRef]
- Ellsberg, Daniel. 1961. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics* 75: 643–69. [CrossRef]
- Fadina, Tolulope, Ariel Neufeld, and Thorsten Schmidt. 2019. Affine processes under parameter uncertainty. *Probability, Uncertainty and Quantitative Risk* 4: 1–35. [CrossRef]
- Filipović, Damir. 2009. *Term Structure Models: A Graduate Course*. Berlin, Heidelberg and New York: Springer.
- Fontana, Claudio, and Thorsten Schmidt. 2018. General dynamic term structures under default risk. *Stochastic Processes and their Applications* 128: 3353–86. [CrossRef]
- Frey, Rüdiger, and Thorsten Schmidt. 2009. Pricing corporate securities under noisy asset information. *Mathematical Finance* 19: 403–21. [CrossRef]
- Frey, Rüdiger, and Thorsten Schmidt. 2011. Filtering and Incomplete Information. In *Credit Risk Frontiers*. Edited by T. Bielecki, D. Brigo and F. Patras. Hoboken: Wiley.
- Frey, Rüdiger, and Thorsten Schmidt. 2012. Pricing and hedging of credit derivatives via the innovations approach to nonlinear filtering. *Finance and Stochastics* 16: 105–133. [CrossRef]
- Gehrmlich, Frank, and Thorsten Schmidt. 2018. Dynamic defaultable term structure modelling beyond the intensity paradigm. *Mathematical Finance* 28: 211–39. [CrossRef]
- Geske, Robert 1977. The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis* 12: 541–52. [CrossRef]
- Heath, David, Robert Jarrow, and Andrew Morton. 1992. Bond pricing and the term structure of interest rates. *Econometrica* 60: 77–105. [CrossRef]
- Jacod, Jean, and Albert N Shiryaev. 2003. *Limit Theorems for Stochastic Processes*, 2nd ed. Berlin: Springer.
- Karatzas, Ioannis, and Steven E Shreve. 1998. *Methods of Mathematical Finance*. New York: Springer.
- Kim, In Joon, Krishna Ramaswamy, and Suresh Sundaresan. 1993. The valuation of corporate fixed income securities. Working paper, Wharton School, University of Pennsylvania: Philadelphia.
- Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin.
- Lando, David. 1994. Three Essays on Contingent Claim Pricing. Ph.D. thesis, Cornell University, Ithaca, NY, USA.

- Leland, Hayne E., and Klaus Bjerre Toft. 1996. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance* 51: 987–1019. [[CrossRef](#)]
- Lyons, Terry. J. 1995. Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance* 2: 117–133. [[CrossRef](#)]
- Merton, Robert. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 29: 449–70.
- Neufeld, Ariel, and Marcel Nutz. 2014. Measurability of semimartingale characteristics with respect to the probability law. *Stochastic Processes and their Applications* 124: 3819–45. [[CrossRef](#)]
- Peng, Shige. 2010. Nonlinear expectations and stochastic calculus under uncertainty. arXiv:1002.4546.
- Riedel, Frank. 2015. Financial economics without probabilistic prior assumptions. *Decisions in Economics and Finance* 38: 75–91. [[CrossRef](#)]
- Sylla, Richard. 2002. *An Historical Primer on the Business of Credit Rating*. Boston: Springer, pp. 19–40.
- Vorbrink, Jörg. 2014. Financial markets with volatility uncertainty. *Journal of Mathematical Economics* 53: 64–78. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Article

Contingent Convertible Debt: The Impact on Equity Holders

Delphine Boursicot, Geneviève Gauthier * and Farhad Pourkalbassi

Department of Decision Sciences, HEC Montréal and GERAD, 3000 chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7, Canada; delphine.boursicot@hec.ca (D.B.); farhad.esfahani@hec.ca (F.P.)

* Correspondence: genevieve.gauthier@hec.ca

Received: 30 March 2019; Accepted: 16 April 2019; Published: 29 April 2019

Abstract: Contingent Convertible (CoCo) is a hybrid debt issued by banks with a specific feature forcing its conversion to equity in the event of the bank's financial distress. CoCo carries two major risks: the risk of default, which threatens any type of debt instrument, plus the exclusive risk of mandatory conversion. In this paper, we propose a model to value CoCo debt instruments as a function of the debt ratio. Although the CoCo is a more expensive instrument than traditional debt, its presence in the capital structure lowers the cost of ordinary debt and reduces the total cost of debt. For preliminary equity holders, the presence of CoCo in the bank's capital structure increases the shareholder's aggregate value.

Keywords: credit risk; contingent convertible debt; financial modelling; risk management; financial crisis

1. Introduction

The 1988 Basel accord ties bank capitalization to portfolio risk by introducing the Capital Adequacy Requirement (CAR). Subsequently, Basel II obliges banks to maintain sufficient loss-absorbing capital on an annual basis. However, several studies on the 2008 financial crisis, such as [Flannery \(2014\)](#) and [Duffie \(2010\)](#) reveal that, in practice, regulators are unable to force banks to maintain adequate loss-absorbing capital. To alleviate banks under-capitalization, [Flannery \(2005\)](#) introduced Contingent Convertibles, hereafter referred to as CoCo in accordance with most of the main-related studies. CoCo is a hybrid debt with a specific clause stipulating that the issuer would either convert it to equity or write down its face value if the level of loss-absorbing capital falls below a certain threshold. This is supposed to help with a firm's recapitalization under distress, while equity holders would be reluctant to raise capital voluntarily by issuing new shares. Basel III recommends large financial institutions to issue CoCo as a part of their capital adequacy requirements (CAR).

CoCo carries two major risks: the risk of default, which threatens any type of debt instrument, plus the exclusive risk of mandatory conversion. CoCo differs greatly from a traditional convertible bond, which the conversion is optional and generally rewards the bondholder; hence, it is not a risk. Mandatory conversion of CoCo is a punishing mechanism which decreases the value of the bondholder in most scenarios; hence, it is a risk factor.

Most literature employs the structural approach to model CoCo dynamics. Studies generally define a trigger threshold as the barrier, then calculate the conditional probability of hitting the barrier. What makes this group of studies different is the choice of underlying instrument that triggers the conversion and the dynamics of the underlying trigger. The trigger is the book-equity-to-book-asset ratio in [Glasserman and Nouri \(2012\)](#), where they model the book asset process using a geometric Brownian motion (GBM). Conversion occurs if the book-equity-to-book-asset ratio exceeds a predetermined exogenous ratio. [Chen et al. \(2013\)](#) use the same underlying instrument while they model the book asset process using a jump-diffusion. Conversion is triggered if the asset value goes

over a predetermined exogenous threshold. In [Brigo et al. \(2015\)](#), firm asset value is a GBM process and the conversion barrier is a linear function of the asset-to-equity ratio. [De Spiegeleer et al. \(2017\)](#), in a more recent study, come closer to the Basel III accord and define the implied Common Equity Tier 1 (CET1) volatility as the conversion trigger while modelling the CET1 capital ratio as a GBM. Structural modelling based on a GBM makes the conversion time predictable, while being counterfactual. Uncertainty with respect to the conversion is due not only to the use of accounting capital ratio as the conversion trigger, but also to the stipulations in Basel III that allow regulators to choose the conversion time at their discretion. Studies such as [Albul et al. \(2013\)](#) and [Hilscher and Raviv \(2014\)](#) model the CoCo as a contingent claim on the value of the bank's assets. The results are relatively tractable models in which the bank's incentive to issue CoCo voluntarily could be examined. Once again, the main challenge of these approaches is in triggering the conversion by the unobservable market value of assets.

In [Cheridito and Xu \(2015\)](#), the CoCo price is modelled using a pure reduced form approach where the conversion and default events are modelled with a time-changed Poisson process. However, the reduced form approach is less intuitive by ignoring the link between the capital ratio and the trigger event ([Brigo et al. 2015](#)). [Chung and Kwok \(2016\)](#) use a structural approach to model the conversion when the capital ratio falls below a certain threshold and also use a Poisson process to model the potential unexpected conversion imposed by the regulator.

Conversion price is also a matter of debate in the literature. As a basic design, [Flannery \(2005\)](#) proposes that the number of shares received by CoCo holders at conversion is determined by the face value of the CoCo bonds divided by the market price of stock on the day of conversion. However, this basic conversion mechanism gives an opportunity to short sellers to bid down the share price, force conversion and dilute the market by increasing the number of outstanding shares. To avoid share price manipulation, [Duffie \(2010\)](#) argues that the number of shares should be based on a multi-day average of closing prices. Other studies such as [Flannery \(2016\)](#) and [Pennacchi \(2010\)](#) propose converting the CoCo into a predetermined number of shares at a fixed price. However, there is a high risk that CoCo investors will suffer some value losses upon conversion due to a jump in the market price of shares.

Regulators insist that the CoCo conversion trigger should be based on the accounting capital ratio. Indeed, this is determined by the Basel Committee on Banking Supervision (BCBS) at a global level in Basel III, the European Banking Authority through the Capital Requirements Directive IV/Capital Requirements Regulation (CRD IV/CRR) and the Office of the Superintendent of Financial Institutions (OSFI) in Canada through the Capital Adequacy Requirements (CAR) guidelines. The Common Equity Tiers 1 (CET1) should not fall below a certain percent of risk-weighted asset (RWA). Although accounting measures are not forward looking and can be manipulated by managers, they ensure that CoCo conversion occurs when a firm encounters serious financial problems. [Glasserman and Nouri \(2012\)](#) and [Chen et al. \(2013\)](#) choose asset book value as the trigger through the book-equity-to-book-asset ratio, claiming that the book asset value truly approximates the market asset value. Many studies are against employing book value as a basis of conversion because it generates a delayed signal of financial distress. [McDonald \(2013\)](#) and [Bolton and Samama \(2012\)](#) assume that the market price of equity can measure a bank's loss-absorbing capacity. They propose a CoCo design in which the share price functions as the conversion trigger. [Sundaresan and Wang \(2015\)](#) point out two shortcomings of employing the market price of shares as the conversion trigger within their framework, both of which are linked to the fact that CoCo conversion generates a value exchange between CoCo holders and initial equity holders. First, if the value is transferred from shareholders to CoCo investors at conversion, there can be more than one rational expected equilibrium price for both the stocks and CoCos. Second, if the value is transferred to shareholders, the model sometimes lacks an equilibrium share price. [Sundaresan and Wang \(2015\)](#) conclude that a unique competitive equilibrium exists if the conversion does not induce a value transfer. [Glasserman and Nouri \(2012\)](#) maintain that this multiple equilibrium problem is a feature of discrete-time models and can be alleviated in a continuous time framework.

Closer to our paper, [Chen et al. \(2017\)](#) study the design and incentive effects of contingent convertible debt on optimal capital structure. The asset value is modelled as a jump diffusion, default and conversion occur whenever the firm value triggers some critical value. The default threshold is determined by the equity holder in such a way to maximize the equity value. They conclude that equity holders often have a positive incentive of issuing CoCos because the presence of CoCos reduces the bankruptcy costs when the conversion trigger is large enough.

In this paper, we also study the impact of having a contingent debt instrument in the debt structure in the perspective of equity holders. Based on a discrete time dynamic optimization approach, we reach similar conclusions as [Chen et al. \(2017\)](#). Indeed, our framework differs from the latter in various ways. First, the endogenous floating coupon rates of the standard and the CoCo debts account for the indebtedness of the firm. Second, the number of shares received by CoCo debt holders at conversion time is also designed differently. In our study, the optimal dividend stream is solved through a dynamic optimization approach: the equity holders have an incentive to control the firm debt ratio to maximize their share of the firm equity value, avoiding the dilution effect caused by conversion. It also benefits the bondholder as it mitigates the default risk.

A numerical simulation based on realistic data evaluates the benefits and the costs of having CoCos in the bank's debt structure. The parameters are estimated using three banks from three different regions (Europe, Canada and the United States), for three different periods of time (pre-crisis, crisis, post-crisis). The results help to understand how CoCos can help "Too big to fail" banks in different economic conditions. Although there are important differences among these three cases, common behaviours are observed: the presence of CoCos in the debt structure reduces the probability of default, the coupon of the standard debt, the cost of debt and capital. However, the CoCo is a more expensive instrument than the standard debt, mainly because the investor bears more risk. This study contributes to the literature by evaluating the effectiveness of adding CoCos to the financial firm's capital structure. We do not only evaluate the CoCo debt itself, but also examine its impact on the firm's management strategy by optimizing the per-share value of the cumulated dividend stream. Equity holders modify the optimal dividend policy to account for the conversion risk that affect them through the dilution effect (after conversion, there are more equity shares).

This paper is structured as follows. Section 2 presents the model to value CoCo and defines the conversion as well as the default intensity. Section 3 presents the dynamic programming model use to examine how CoCo impact the firm's management strategy. Section 4 is decomposed into three parts. First, the data used to have realistic scenarios are presented. Second, the results are presented for each bank and each year. Third, a sensitivity analysis follows. Section 5 presents conclusions.

2. The Model

The asset value satisfies

$$A_t = E_t + D_t, \quad (1)$$

where E_t and D_t are the equity and debt values, respectively. The debt is decomposed into three main components: the deposit whose time t value is F_t , the coupon-paying bonds and other debt instruments for a value of B_t , and a convertible contingent instrument (CoCo) whose value is C_t :

$$D_t = F_t + B_t + C_t.$$

The debt ratio is defined as

$$X_t = \frac{D_t}{A_t}. \quad (2)$$

The presence of the CoCo debt alleviates the default risk since, in case of financial distress, the CoCo debt is converted into equity, leading to a smaller debt ratio. CoCo debt holders bear not only a default risk, but also a conversion risk; therefore, they require compensation in the form of a coupon payment that differs from that of a standard debt, which is subject only to credit risk. More precisely,

once converted into equity, the CoCo debt has a zero recovery rate in case of default. However, if the firm survives after the conversion, then it is not clear whether the proportion of equity value held by the CoCo debt holders will be more or less profitable than for a standard debt.

The default time is denoted τ_D , and the conversion time satisfies $\tau_C = \min(\tau_\alpha, \tau_D)$, which means that the CoCo debt is always converted before the default event. Because other factors may influence the conversion decision, it does not necessarily occur as soon as the leverage ratio triggers α , but it is very likely to occur. Consequently, the conversion intensity driving τ_α ,

$$G_t = g(X_t) = \left(\frac{X_t}{\theta_C}\right)^{\beta_C},$$

is a positive and convex function of X_t , taking large values whenever X_t is above $\theta_C \in (0, 1]$ and small values otherwise. θ_C should be close to α . The parameter β_C is usually a large positive constant. Therefore,

$$\mathbb{P}_t(\tau_\alpha = t | \tau_\alpha > t - 1, \tau_D > t - 1) = [1 - \exp(-G_t)] \mathbf{1}_{\tau_\alpha > t - 1} \mathbf{1}_{\tau_D > t - 1}.$$

Intuitively, the conversion should occur whenever the leverage ratio is close to the critical level α determined by the regulator. If the CoCo debt still exists at time t , then the superscript t^- denotes the pre-conversion values of the considered variables. The convertible contingent instrument is a coupon-paying bond with the floating coupon rate $c(X_t)$. At conversion, CoCo debt holders receive, in the form of equity shares, an amount equivalent to the debt's face value $C_{\tau_C^-}$ and a fraction ρ_C of the coupon. In other words, at conversion, the convertible debt holders receive

$$N_{\tau_C} = \frac{C_{\tau_C^-} (1 + \rho_C c(X_{\tau_C-1}))}{S_{\tau_C}}$$

equity shares where S_{τ_C} is the post-conversion price per share. Assuming that initially there were N outstanding equity shares, the post-conversion price per share becomes

$$S_{\tau_C} = \frac{E_{\tau_C}}{N + N_{\tau_C}} = \frac{E_{\tau_C^-} + C_{\tau_C^-} (1 + \rho_C c(X_{\tau_C-1}))}{N + \frac{C_{\tau_C^-} (1 + \rho_C c(X_{\tau_C-1}))}{S_{\tau_C^-}}}$$

Multiplying both sides by $N + \frac{C_{\tau_C^-} (1 + \rho_C c(X_{\tau_C-1}))}{S_{\tau_C^-}}$ and isolating S_{τ_C} leads to

$$S_{\tau_C} = \frac{E_{\tau_C^-}}{N} = S_{\tau_C^-},$$

which implies that the price per share is not affected by the conversion. Then, the equity value becomes

$$E_{\tau_C} = E_{\tau_C^-} + C_{\tau_C^-} (1 + \rho_C c(X_{\tau_C-1})).$$

Letting

$$y_t = \frac{C_t}{D_t} \in [0, \gamma] \subseteq [0, 1] \tag{3}$$

be the proportion of convertible debt in the total debt value. Because there are always deposits, then the convertible instrument represents less than 100% of the debt value, which implies that $\gamma < 1$. The number of additional shares issued at conversion satisfies

$$N_{\tau_C} = N (1 + \rho_C c(X_{\tau_C-1})) y_{\tau_C^-} \frac{X_{\tau_C^-}}{1 - X_{\tau_C^-}} \cong N (1 + \rho_C c(\alpha)) y_{\tau_C^-} \left(\frac{\alpha}{1 - \alpha}\right). \tag{4}$$

Proof.

$$\begin{aligned}
 N_{\tau_C} &= \frac{C_{\tau_C^-} (1 + \rho_{CC} (X_{\tau_C-1}))}{S_{\tau_C}} = N \frac{C_{\tau_C^-} (1 + \rho_{CC} (X_{\tau_C-1}))}{E_{\tau_C^-}} = N \frac{C_{\tau_C^-} (1 + \rho_{CC} (X_{\tau_C-1}))}{D_{\tau_C^-}} \frac{D_{\tau_C^-} A_{\tau_C^-}}{A_{\tau_C^-} E_{\tau_C^-}} \\
 &= N (1 + \rho_{CC} (X_{\tau_C-1})) \frac{C_{\tau_C^-} D_{\tau_C^-}}{D_{\tau_C^-} A_{\tau_C^-}} \frac{1}{1 - \frac{D_{\tau_C^-}}{A_{\tau_C^-}}} = N (1 + \rho_{CC} (X_{\tau_C-1})) y_{\tau_C^-} \frac{X_{\tau_C^-}}{1 - X_{\tau_C^-}}.
 \end{aligned}$$

□

The standard debt pays the floating coupon rate of $b(X_t, y_t)$ unless default occurs. Interestingly, the debt coupon is affected by the presence of the CoCo instrument because the latter mitigates both the default and recovery risks. If

$$1 - z_t = \frac{F_t}{F_t + B_t} = \frac{F_t}{D_t - C_t} = \frac{1}{1 - y_t} \frac{F_t}{D_t}$$

denotes the proportion of the non-convertible contingent debt value that consists of deposits, then

$$F_t = (1 - z_t) (1 - y_t) D_t, B_t = z_t (1 - y_t) D_t \text{ and } C_t = y_t D_t. \tag{5}$$

2.1. The Debt Ratio \mathbb{P} -Dynamics

Assume that the firm survives up to time $t : \tau_D > t$. At the beginning of the $t + 1^{\text{th}}$ period, the invested capital yields returns:

$$R_{t+1} = m_{t+1} + \sigma_{t+1} \varepsilon_{t+1}^{\mathbb{P}},$$

where m and σ^2 are predictable processes and the sequence of ε_t is formed with independent standard normal random variables. The information structure is provided by the filtration $\{\mathcal{F}_t\}_{t=1}^{\infty}$, where $\mathcal{F}_t = \sigma\left(\{\varepsilon_u^{\mathbb{P}}\}_{u=1}^t\right)$ is the σ -field modelling the information available at time t . Throughout the period, decisions about the convertible contingent debt conversion and the dividend payment affect the debt and debt ratio values. At time $t + 1$, free cash flow (FCF) is

$$FCF_{t+1} = A_t R_{t+1}.$$

The financial flow (FF) is decomposed into the deposit interest payment $r_{t+1} F_t$, where r_{t+1} is the predictable risk-free rate, the interest payment on standard and convertible debts, $b(X_t, y_t) B_t$ and $c(X_t) C_t$ respectively, the dividend payment $\delta_{t+1} A_{t+1}$, and the debt structure variation,

$$D_{t+1} - D_t = (F_{t+1} - F_t) + (B_{t+1} - B_t) + (C_{t+1} - C_t) \mathbf{1}_{\tau_C > t+1} - C_t \mathbf{1}_{\tau_C = t+1}.$$

The floating interest rates reflect the risk embedded in both standard and CoCo debts. Both types of debts contain credit risk because the firm may default. See Appendix A.

The debt structure variation is expressed as a ratio:

$$\eta_{t+1} = \frac{D_{t+1} - D_t}{D_t}.$$

New debt issuance makes $\eta_{t+1} > 0$, whereas debt reaching maturity or CoCo debt conversion leads to $\eta_{t+1} < 0$.

Assumption 1. Let $y_t = y_0 \mathbf{1}_{\tau_C > t}$, where $y_0 = C_0 / D_0$. In other words, whenever the contingent convertible debt exists, its proportion of the total debt value remains constant.

Similarly, assuming that the proportion of deposits with respect to the non-convertible debt value remains constant over time, $(1 - z_t) = (1 - z_0) = F_0 / (F_0 + B_0)$ and $z_0 = B_0 / (F_0 + B_0)$.

Assumption 2. At conversion time, if there is no variation in the other types of debt, then $\eta_{\tau_C} = -y_{t_C-1}$. In the numerical implementation, we assume that $\eta_{t+1} = -y_t \mathbf{1}_{\tau_C=t+1}$. In other words, the debt value at conversion is modified only by the conversion of the CoCo debt to equity.

Thus, the weighted average interest rate is

$$\begin{aligned} \mu_{t+1} &= \mu_{t+1} (X_t, z_t, y_t, \mathbf{1}_{\tau_C>t+1}) \\ &= (r_{t+1}F_t + b(X_t, y_t) B_t + c(X_t) C_t \mathbf{1}_{\tau_C \geq t+1}) / D_t \\ &= r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t) + c(X_t) y_t \mathbf{1}_{\tau_C>t+1}. \end{aligned} \tag{6}$$

Since

$$D_{t+1} = D_t (1 + \eta_{t+1}), \tag{7}$$

the financial flow satisfies

$$\begin{aligned} FF_{t+1} &= \mu_{t+1} D_t + D_t - D_{t+1} + \delta_{t+1} A_{t+1} \\ &= (\mu_{t+1} - \eta_{t+1}) D_t + \delta_{t+1} A_{t+1}. \end{aligned}$$

The bank's profit is the difference $FCF - FF$ between the free cash flow and the financial flow. Therefore,

$$\begin{aligned} A_{t+1} &= A_t + FCF_{t+1} - FF_{t+1} \\ &= (1 + R_{t+1}) A_t - (\mu_{t+1} - \eta_{t+1}) D_t - \delta_{t+1} A_{t+1}, \end{aligned}$$

which is equivalent to

$$A_{t+1} = \frac{(1 + R_{t+1}) A_t - (\mu_{t+1} - \eta_{t+1}) D_t}{1 + \delta_{t+1}}. \tag{8}$$

Dividing both sides of Equation (8) by A_t ,

$$\frac{A_{t+1}}{A_t} = \frac{(1 + R_{t+1}) - (\mu_{t+1} - \eta_{t+1}) \frac{D_t}{A_t}}{1 + \delta_{t+1}}.$$

Since

$$\frac{A_{t+1}}{A_t} = \frac{A_{t+1}}{D_{t+1}} \frac{D_{t+1}}{D_t} \frac{D_t}{A_t} = \frac{1}{X_{t+1}} (1 + \eta_{t+1}) X_t,$$

comparing both equations and isolating X_{t+1} implies that the (post-dividend) debt ratio must satisfy

$$X_{t+1} = \frac{(1 + \delta_{t+1}) (1 + \eta_{t+1}) X_t}{(1 + R_{t+1}) - (\mu_{t+1} - \eta_{t+1}) X_t}. \tag{9}$$

Note that

$$X_{t+1} = (1 + \delta_{t+1}) X_{t+1}^0, \tag{10}$$

where X_{t+1}^0 denotes the pre-dividend debt ratio

$$X_{t+1}^0 = X_{t+1} |_{\delta_{t+1}=0} = \frac{(1 + \eta_{t+1}) X_t}{(1 + R_{t+1}) - (\mu_{t+1} - \eta_{t+1}) X_t}.$$

2.2. Conversion

The conversion decision is taken under the assumption that $\delta_{t+1} = 0$, $\eta_{t+1} = 0$ (no debt issuing or refunding) and that the full interest payment includes the CoCo debt coupon:

$$\mu_{t+1}^C = r_{t+1} (1 - z_t) (1 - y_t) + b (X_t, y_t) z_t (1 - y_t) + c (X_t) y_t. \tag{11}$$

The conversion intensity

$$G_{t+1} = g \left(X_{t+1}^C \right) = \left(\frac{X_{t+1}^C}{\theta_C} \right)^{\beta_C} \tag{12}$$

is a positive, increasing, and convex function of the debt ratio

$$X_{t+1}^C = \frac{X_t}{(1 + R_{t+1}) - \mu_{t+1}^C X_t}, \tag{13}$$

which is a particular case of Equation (9). The parameter θ_C is the critical level from which the conversion probability grows fast beyond this threshold. Because of the convex relation between the conversion intensity and the conversion probability, θ_C is not exactly equal to the regulator critical level α , but it is in its neighborhood. The parameter β_C captures the growth speed. In the numerical implementation, both parameters are obtained through a calibration approach based on the regulator critical level α . For this reason, τ_α denotes the conversion time characterized by the conversion intensity α . The conditional conversion probability at time $t + 1$ triggered by the critical level α (letting $\beta_C \rightarrow \infty$, we recover the case where conversion occurs as soon as $X_{t-} > \alpha$) is

$$\mathbb{P}_{t+1} (\tau_\alpha = t + 1 | \tau_C > t) = [1 - \exp (-G_{t+1})] \mathbf{1}_{\tau_C > t}.$$

Because immediate conversion may also arise as a consequence of an unpredicted default, the survival conversion probability is

$$\begin{aligned} \mathbb{P}_{t+1} (\tau_C > t + 1 | \tau_C > t) &= \mathbb{P}_{t+1} (\tau_\alpha > t + 1, \tau_D > t + 1 | \tau_C > t) \\ &= \mathbb{P}_{t+1} (\tau_\alpha > t + 1 | \tau_C > t) \mathbb{P}_{t+1} (\tau_D > t + 1 | \tau_C > t) \\ &= \exp (-H_{t+1} \mathbf{1}_{\tau_C > t} - G_{t+1}). \end{aligned}$$

Then, the conditional conversion probability arising from both the critical level and the potential default is

$$\mathbb{P}_{t+1} (\tau_C = t + 1 | \tau_C > t) = [1 - \exp (-H_{t+1} \mathbf{1}_{\tau_C > t} - G_{t+1})] \mathbf{1}_{\tau_C > t}, \tag{14}$$

where the default intensity H_{t+1} is defined at Equation (17).

2.3. Default

Since default occurs after conversion, the interest payment does not include the CoCo debt coupon. Indeed, a fraction of the CoCo coupon is paid back to CoCo debt holders in the form of equity shares. This means that

$$\mu_{t+1}^D = r_{t+1} (1 - z_t) (1 - y_t) + b (X_t, y_t) z_t (1 - y_t). \tag{15}$$

The default intensity is based on the pre-dividend debt ratio. More precisely, assuming that there are no dividends, $\delta_{t+1} = 0$, and that the only debt variation considered is the one arising from an immediate conversion, $\eta_{t+1} = -y_t \mathbf{1}_{\tau_C = t+1}$, the pre-dividend debt ratio is

$$X_{t+1}^D = \frac{(1 - y_t) X_t}{(1 + R_{t+1}) - (\mu_{t+1}^D + y_t) X_t} \mathbf{1}_{\tau_\alpha > t} + \frac{X_t}{(1 + R_{t+1}) - \mu_{t+1}^D X_t} \mathbf{1}_{\tau_\alpha \leq t}. \tag{16}$$

Since a debt ratio augmentation has more impact on the default probability when the debt ratio is already high, the default intensity H_{t+1} is a positive, increasing, and convex function of the debt ratio. More precisely,

$$H_{t+1} = h\left(X_{t+1}^D, y_t\right) = \lambda_D + \left(\frac{X_{t+1}^D}{\theta_{D,t}}\right)^{\beta_D}, \tag{17}$$

where $\theta_{D,t} = \theta_D + y_t$, $\lambda_D \geq 0$, $\theta_D > \alpha$, $\beta_D > 1$. Indeed, $\theta_{D,t}$ represents the critical debt ratio from which the increasing behaviour of the default probability (seen as a function of the debt ratio) accelerates. Because the CoCo instrument provides the standard debt holders with an additional protection against default risk, the critical debt ratio is $\theta_{D,t}$ slightly higher whenever the CoCo debt is present in the debt structure. Consequently, the conditional default probabilities are

$$\mathbb{P}_{t+1}(\tau_D = t + 1 | \tau_C > t, \tau_D > t) = [1 - \exp(-H_{t+1} \mathbf{1}_{\tau_C > t})] \mathbf{1}_{\tau_D > t} \tag{18}$$

and

$$\mathbb{P}_{t+1}(\tau_D = t + 1 | \tau_C \leq t, \tau_D > t) = [1 - \exp(-H_{t+1} \mathbf{1}_{\tau_C \leq t})] \mathbf{1}_{\tau_D > t}. \tag{19}$$

3. Stochastic Optimum Control Problem

The period $]t, t + s]$ discount factor is

$$DF_{t,t+s} = \prod_{u=t}^{t+s-1} \frac{1}{1 + w(X_u)},$$

where the cost of capital is a weighted average of the cost of equity, r^E , and the cost of debt:

$$w(X_u) = (1 - X_u) r_u^E + X_u \mu_u. \tag{20}$$

The current equity holders want to maximize their share of dividends. More precisely, given a stream of dividend rates $\delta_{1:\infty} = \{\delta_i\}_{i=1}^\infty$, the expected value of the discounted dividends at time t is

$$\begin{aligned} V\left(t, X_t^0, \delta_{1:\infty}\right) \mathbf{1}_{\tau_D > t} &= \sum_{u=t}^\infty \mathbb{E}_t^\mathbb{P} \left[DF_{t,u} \delta_u A_u \left(\mathbf{1}_{\tau_C > u} + \frac{N}{N + N_{\tau_C}} \mathbf{1}_{\tau_C \leq u} \right) \mathbf{1}_{\tau_D > u} \right] \mathbf{1}_{\tau_D > t} \\ &\cong \sum_{u=t}^\infty \mathbb{E}_t^\mathbb{P} \left[DF_{t,u} \delta_u A_u \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq u} \right) \mathbf{1}_{\tau_D > u} \right] \mathbf{1}_{\tau_D > t}, \end{aligned}$$

since Assumption 1 and Equation (4) imply that

$$\mathbf{1}_{\tau_C > u} + \frac{N}{N + N_{\tau_C}} \mathbf{1}_{\tau_C \leq u} = 1 - \frac{N_{\tau_C}}{N + N_{\tau_C}} \mathbf{1}_{\tau_C \leq u} \cong 1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq u}.$$

Since V allows for the decomposition

$$\begin{aligned} V\left(t, X_t^0, \delta_{1:\infty}\right) \mathbf{1}_{\tau_D > t} &\cong \delta_t A_t \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) \mathbf{1}_{\tau_D > t} \\ &\quad + \mathbb{E}_t^\mathbb{P} \left[DF_{t,t+1} V\left(t + 1, X_{t+1}^0, \delta_{t+1:\infty}\right) \mathbf{1}_{\tau_D > t+1} \right] \mathbf{1}_{\tau_D > t}, \end{aligned} \tag{21}$$

the optimal dividend rate sequence $\delta_{1:\infty}^{\text{opt}}$ can be constructed recursively using backward recursion over time:

$$\delta_t^{\text{opt}} = \arg \max_{\delta_t \in [0, \delta_t^{\text{max}}]} \left\{ A_t \delta_t \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) + E_t^{\mathbb{P}} \left[\frac{V(t+1, X_{t+1}^0, \delta_{t+1:\infty}^{\text{opt}})}{1 + w(X_t)} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}. \tag{22}$$

See Appendix D. Because $\frac{A_{t+1}}{A_t} = \frac{D_t}{A_t} \frac{D_{t+1}}{D_t} \frac{A_{t+1}}{D_{t+1}} = \frac{X_t}{X_{t+1}} (1 + \eta_t)$,

$$\begin{aligned} v(t, X_t^0, \delta_{t:\infty}) \mathbf{1}_{\tau_D > t} &= \frac{V(t, X_t^0, \delta_{t:\infty})}{A_t} \mathbf{1}_{\tau_D > t} \\ &\cong \delta_t \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) \mathbf{1}_{\tau_D > t} \\ &\quad + E_t^{\mathbb{P}} \left[DF_{t,t+1} \frac{X_t}{X_{t+1}} (1 + \eta_t) v(t+1, X_{t+1}^0, \delta_{t+1:\infty}) \mathbf{1}_{\tau_D > t+1} \right] \mathbf{1}_{\tau_D > t} \end{aligned} \tag{23}$$

and

$$\delta_t^{\text{opt}} = \arg \max_{\delta_t \in [0, \delta_t^{\text{max}}]} \left\{ \delta_t \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) + E_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} (1 + \eta_t) \frac{v(t+1, X_{t+1}^0, \delta_{t+1:\infty}^{\text{opt}})}{1 + w(X_t)} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}.$$

Therefore, we work with a standardized version of the primary equity holders' share of cumulated discounted dividends. Indeed, under this form, the dynamic optimization does not require the modelling of the dynamics of A .

The dividend rate is bounded above. Indeed, if the dividend payment is too large, the equity value will drop below its current level. More precisely, noting that $X_t = (1 + \delta_t) X_t^0$, the expected variation of the pre-dividend equity is (see Appendix B.3)

$$E_t^{\mathbb{P}} \left[E_{t+1} |_{\delta_{t+1}=0} - E_t \right] \leq (m_{t+1} - (1 + \delta_t) X_t^0 \mu_{t+1} \left((1 + \delta_t) X_t^0 \right)) A_t.$$

We restrict δ_t , making the right-hand side bound positive. Now, let x_0 be the solution of $m_{t+1} - x_0 \mu_{t+1}(x_0) = 0$. Indeed, since $x \mu(x)$ is an increasing function of x starting at 0, a unique solution exists. Since equity holders do not reduce the expected equity value deliberately, it follows that

$$\begin{aligned} m_{t+1} - (1 + \delta_t) X_t^0 \mu_{t+1} \left((1 + \delta_t) X_t^0 \right) &\geq 0 \\ \iff (1 + \delta_t) X_t^0 \mu_{t+1} \left((1 + \delta_t) X_t^0 \right) &\leq m_{t+1} \\ \iff (1 + \delta_t) X_t^0 &\leq x_0 \\ \iff \delta_t &\leq \frac{x_0}{X_t^0} - 1. \end{aligned}$$

In addition, since $\frac{x_0}{X_t^0} - 1 \rightarrow \infty$ as $X_t^0 \rightarrow 0$, this upper dividend rate bound is not active for small debt ratio values. However, the dividend rate is generally lower than the expected asset returns, which leads to $\delta_t \in [0, \delta_t^{\text{max}}]$, where

$$\delta_t^{\text{max}} = \min \left(\max \left(\frac{x_0 - X_t^0}{X_t^0}, 0 \right), m_{t+1} \right). \tag{24}$$

Dynamic programming optimization allows using a recursive method starting from T with backward induction. At each time period, the value function is the sum of the immediate dividend and the expected future dividend gain. To initialize the recursion, a long time horizon T is chosen for which some simplifications are made. Because the model is Markovian, after some iterations, the terminal conditions vanish and, for that reason, the following assumption has no impact on our numerical results. The algorithm stops when the variations in the optimal dividend become very small. Then, for time T , the following simplifications are to be assumed.

Assumption 3. For any $t > T$,

1. The asset returns are no longer uncertain, that is, $R_t = m_T, \sigma_t^2 = 0$;
2. There is no more possibility of conversion, that is, the CoCo debt becomes a standard debt;
3. The dividends are the remaining part of the returns once the interest rate payment on the debts is deducted:

$$\delta_t A_t = \max(m_T A_t - \mu_t D_{t-1}, 0); \tag{25}$$

4. If the dividend payment $m_T A_t - \mu_t D_{t-1}$ is positive, then there is no variation of the debt value, that is, $D_t = D_{t-1}$ or, equivalently, $\eta_t = 0$;
5. The risk-free rate r_t is constant and equal to r .

Since a potential conversion is no longer possible (Assumption 3-2), for all $t > T$, the coupon on the CoCo debt is the same as the one on the ordinary debt, that is,

$$\begin{aligned} \mu_t &= r(1 - z_{t-1})(1 - y_T) + b(X_{t-1}, 0)z_{t-1}(1 - y_T) + b(X_{t-1}, 0)y_T \\ &= r(1 - z_{t-1})(1 - y_T) + b(X_{t-1}, 0)(1 - (1 - z_{t-1})(1 - y_T)). \end{aligned}$$

Assume for a moment that $m_T A_t - \mu_t D_{t-1} \geq 0$. As a consequence of Assumption 3-3, $\forall t > T$,

$$FCF_t - FF_t = m_T A_{t-1} - (\mu_t - \eta_t) D_{t-1} - \delta_t A_t = 0,$$

which implies that $A_t = A_{t-1}$. Therefore,

$$\delta_t = \frac{1}{A_t} \max(m_T A_t - \mu_t D_{t-1}, 0) = \max\left(m_T - \mu_t \frac{A_{t-1} D_{t-1}}{A_t A_{t-1}}, 0\right) = \max(m_T - \mu_t X_{t-1}, 0).$$

Moreover, from Assumption 3-4, $D_t = D_{t-1}$, it follows that $X_t = X_{t-1}$ and $\mu_{t+1} = \mu_t$. In addition, a recursive argument leads to the following conclusion: if the dividend rate is positive, that is

$$0 \leq \delta_{T+1} = \frac{m_T A_{T+1} - \mu_{T+1} D_T}{A_{T+1}} = m_T - \mu_{T+1} \frac{A_T}{A_{T+1}} X_T = m_T - \mu_{T+1} X_T,$$

then for all $t > T$, $X_t = X_T$, $A_t = A_T$, $\mu_t = \mu_{T+1}$ and $\delta_t = \delta_{T+1} = m_T - \mu_{T+1} X_T$. The discount factor then becomes

$$DF_{t,t+s} = \prod_{u=t}^{t+s-1} (1 + w(X_T))^{-1} = (1 + w(X_T))^{-s}.$$

If $m_T A_t - \mu_t D_{t-1} < 0$, then not only is there no dividend payment, but also the firm needs to issue more debt to cover the interest expenses: $D_t = D_{t-1} + \mu_t D_{t-1} - m_T A_t$. In that case,

$$\begin{aligned} FF_t &= \mu_t D_{t-1} + D_{t-1} - D_t \\ &= \mu_t D_{t-1} + D_{t-1} - D_{t-1} - \mu_t D_{t-1} + m_T A_t \\ &= m_T A_t \\ &= FCF_t \end{aligned}$$

and $A_t = A_{t-1} + FCF_t - FF_t = A_{t-1}$. Therefore, the debt is growing and the asset value is stable, which implies that the debt ratio will increase until default.

Lemma 1. Under Assumption 3, the expected value of the discounted dividends at time T satisfies

$$V(T, X_T^0, \delta_T) \mathbf{1}_{\tau_D > T} \tag{26}$$

$$\cong A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) \left\{ \delta_T + \delta_{T+1} \left(\frac{\exp(-h(X_T))}{1 + w(X_T) - \exp(-h(X_T))} \right) \right\} \mathbf{1}_{\tau_D > T},$$

where $\delta_{T+1} = m_T - \mu_{T+1} X_T$, $h(X_T) = \lambda_D + \left(\frac{1}{\theta_D} \max(X_T; 0) \right)^{\beta_D}$ and $X_T = (1 + \delta_T) X_T^0$.

See proof in Appendix B.4.

4. Numerical Results

To generate realistic scenarios, the parameters correspond to the financial ratio of three banks in three different countries (Europe, Canada and the United States). The aim is to analyze how “too big to fail” banks can react in the case where there is or not CoCo in the debt structure. We focus the analysis on three different periods corresponding to the pre-crisis (2006), the crisis (2008) and the post-crisis (2015).

4.1. Data

The sample is composed of three banks listed as Global-Systemically Important Banks (G-SIBs): Société Générale for Europe, Royal Bank of Canada for Canada and Bank of America in the United States. For each bank and for each year, the debt ratio is calculated from Equation (2). The total amount of asset (A_t), deposit (F_t), long and short-term debt (B_t) and CoCo (C_t) are obtained using Bloomberg (Bloomberg Financial Analysis and Bloomberg Contingent Convertibles Search).

Figure 1 shows the evolution of the debt ratio over time, according to the banks. The French bank has the highest debt ratio during the financial crisis of 2008. Nowadays, the French and the Canadian banks appear to have the same level of debt in their financial structure. The decomposition of the debt structure, as presented in Table 1, bring out that the French bank finance these activities with fewer deposit and more bonds than the Canadian and U.S. Bank.

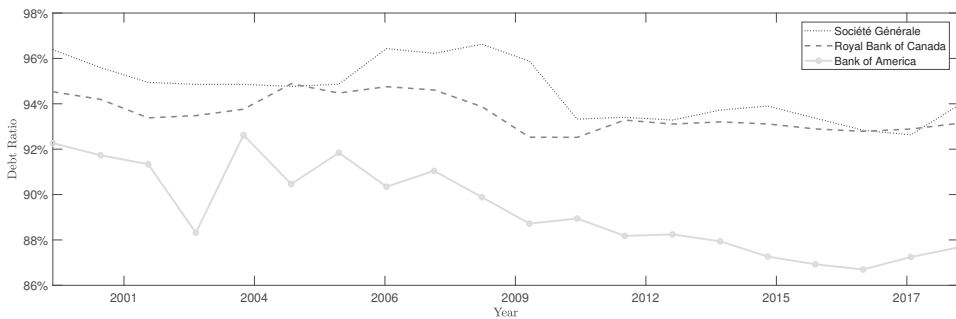


Figure 1. Evolution of the debt ratio over time.

Table 1. Source of assets.

	Société Générale	Royal Bank of Canada	Bank of America
Equity/Asset	5.41%	6.43%	10.62%
Deposit/Asset	40.60%	77.37%	58.86%
Bonds (including CoCo)/Asset	53.99%	16.20%	32.52%

The table presents the average of three financial ratios over 19 years, from 1999 to 2018. Asset is computed using Equation (1). Deposit, Equity and Bonds (including CoCo) are based on the financial statement available on Bloomberg.

CoCo was first launched in Europe at the end of 2009, after the financial crisis, to fulfill a need in terms of risk management. CoCos appear as a potential solution to absorb losses when the capital of banks fall below a certain level. CoCos issuance has started to rapidly increase since 2013/2014: under Basel III, certain specific CoCos are categorized as additional Tier 1 Capital (AT1). Nowadays, CoCos throughout the world are mainly issued by European and Asian companies. Canadian banks started to issue CoCos in 2014 while American banks do not.

To initialize the dynamic optimization, the percentage of CoCos in the debt structure is assumed to be 1%, since it is quite representative of the European and Canadian banks. Indeed, based on the debt value and the amount of CoCos issued by banks, obtained from Bloomberg, the percentage of CoCos in the debt structure is around 1.15% for European banks and around 0.64% for Canadian banks. For the American bank, assuming 1% of CoCos could shed light on how it could help the bank in case of financial distress.

Table 2 presents the parameter values obtained from the financial ratio of three banks, for three different periods corresponding to the pre-crisis (2006), the crisis (2008) and the post-crisis (2015). The critical debt ratio parameters (θ_C and θ_D) and the convexity parameters (β_C and β_D) characterizing the conversion and default intensities were obtained by calibrating the conversion and default probabilities (more details in Appendix C). The results obtained for the critical debt level of the conversion intensity θ_C is near 95%. This is due to the fact that, under Basel III, CoCos that have a minimum trigger level of 5.125% in terms of Common Equity Tier 1 by risk-weighted assets and having perpetual coupons are qualified as additional tiers 1. Previously, there had been a trend towards issuing CoCos with such characteristics. This corresponds to a trigger level imposed by the regulator (α) in terms of debt ratio close to 1–5.125% of risk-weighted assets. The critical debt ratio of the default intensity obtained, θ_D , is slightly greater than 100%. When a firm has more liabilities than assets, there is a high risk of insolvent and excess leverage. The convexity conversion (default) parameter, β_C (β_D), is designed to increase the conversion probability (default probability) rapidly around the critical debt ratio θ_C (θ_D). This leads to high convexity parameters.

The credit risk management literature suggests that the recovery rate is around 40%, based mainly on observed data (Altman and Kishore 1996; Duffie and Lando 2001). The risk-free rate corresponds to the one-year zero coupon curve computed by the European Central Banking, Bank of Canada and the Federal Reserve Bank of St. Louis. r_t^E corresponds to the Return on Equity (ROE). m is the expected capital return expected by investors and is therefore not directly observable. We have chosen to proxy m_t by the average ROE over the last five years and the average risk-free rate, such that $m_t = (1 - x)\bar{r}_t^E + x\bar{r}_t$.

Table 2. Parameter values of the dynamic optimization per bank per year.

	Société Générale			Royal Bank of Canada			Bank of America		
	2006	2008	2015	2006	2008	2015	2006	2008	2015
Returns									
r_t	3.76%	1.99%	−0.2%	4.07%	0.89%	0.51%	4.94%	0.28%	0.73%
r_t^E	20.04%	7.02%	6.23%	23.21%	17.64%	18.42%	18.07%	1.81%	6.27%
m	3.07%	3.43%	1.26%	3.5%	4.35%	2.25%	3.84%	4.66%	0.36%
σ^2		2%			2%			2%	
Initial debt structure									
F_0	31.41%	29.71%	44.35%	79.58%	86.39%	83.45%	54.78%	56.13%	70.29%
y_0		1%			1%			1%	
Conversion risk									
α	98.47%	98.43%	98.63%	97.86%	98.03%	98.03%	96.30%	96.28%	96.65%
β_C		74.48			66.55			39.85	
θ_C		97.42%			96.82%			94.69%	
ρ_C		90%			90%			90%	
Default risk									
θ_D		107.7%			104.11%			107.29%	
β_D		47.2			69.9			40.46	
λ_D		0			0			0	
ρ_D		0.4			0.4			0.4	
Numerical scheme									
T		30			30			30	
Δ_x		0.002			0.002			0.002	
Δ_t		1			1			1	

r_t stands for risk-free rate and correspond to the one-year zero coupon curve computed by each central bank of each region. r_t^E stands for the Return on Equity. m stands for the average of the expected return on capital and σ^2 its variance. F_0 means the percentage of deposit at time 0. $y_0 = C_0/D_0$ is the initial proportion of CoCo debt. The conversion and default intensities functions are respectively Equations (12) and (17). The debt ratio mesh parameter Δ_x represents the distance between two consecutive debt ratios. The time discretization is $\Delta_t = 1$ year.

4.2. Empirical Results

Using the parameters presented in Table 2, Equations (14), (18) and (19) are used to compute the one-year conversion probability and the one-year default probabilities as a function of the actual debt ratio. Figure 2 shows that, for each bank and for each year, the conversion probability increases with the debt ratio. Around the debt ratio critical level, the conversion probability tends to climb to 100% reflecting the mandatory conversion. As expected, the presence of CoCo debt instruments reduces the default probabilities. When there is no more CoCo in the debt structure, the default probability increases more rapidly. It confirms that having CoCos in the debt structure acts like a safety buffer. The conversion and default probabilities are not only driven by the intensity parameters, but also by the actual market conditions, that is, the risk-free rate, the expected return on the capital, and the debt structure. Figure 2 shows that the probabilities of default were greater in 2006, with more indebted banks.

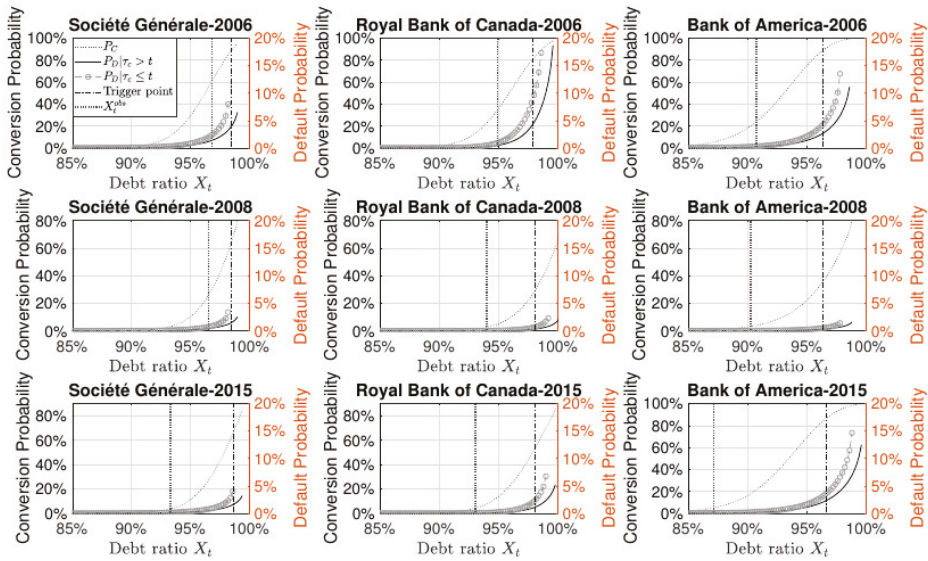


Figure 2. One-year conversion and default probabilities. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The dark grey dotted line represents the conversion probability. The black line represents the one-year default probability in the presence of CoCos in the debt structure. The light gray circle dash line represents the one-year default probability without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

Equations (A4)–(A6) give the coupon rate on the standard debt (with and without CoCos) and the CoCo debt. CoCo debt holders face two risks: the risk of default (like the standard debt holders), plus the risk of mandatory conversion. However, since at conversion, CoCo debt holders receive in terms of equity the value of their investment, there is no conversion recovery risk. Figure 3 shows that when the contingent debt is present in the debt structure, the CoCo debt is more expensive than the standard debt. Moreover, as the contingent debt mitigates the default risk, the coupon rate of the ordinary debt is notably lower in the presence of CoCos. Comparing the result for each bank and each year, Figure 3 shows that the coupons are lower over the years due to lower interest rate. The risk-free rates were higher before the financial crisis and stayed at low levels after the crisis.

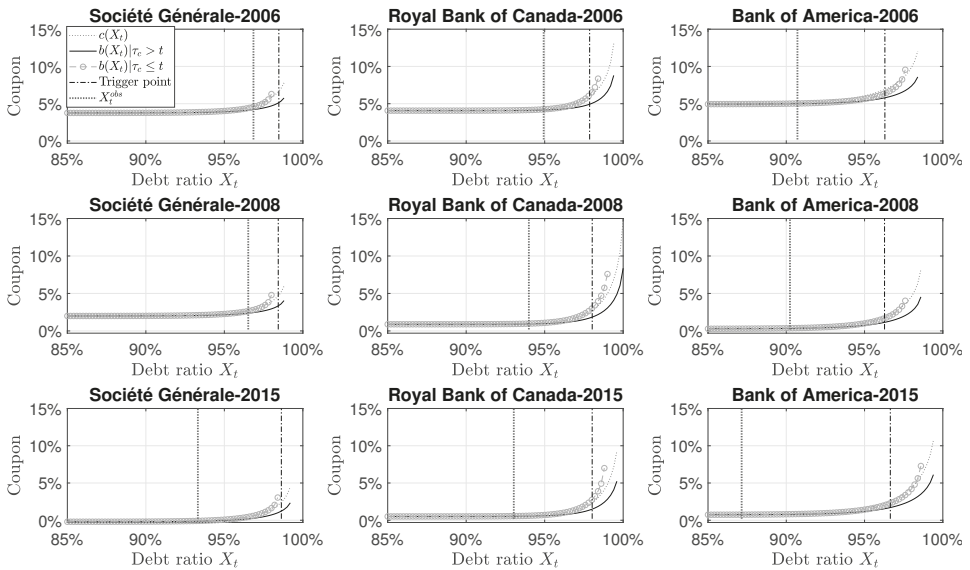


Figure 3. CoCos and standard debts coupons. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The dark grey dotted line represents the coupon on the CoCo debt. The black line represents the coupon on the standard debt in the presence of CoCos in the debt structure. The light gray circle dash line represents the coupon on the standard debt without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

The cost of debt (Equation (6)) is a weighted average of each component (deposit, standard debt and CoCo debt). Figure 4 shows that the cost of debt increases for large debt ratio which is a direct consequence of the coupon curve behavior (Figure 3). Even if the CoCo coupon rate is large relatively to the standard debt coupon rate, for high-debt ratio, the presence of CoCo in the debt structure reduces the cost of debt. Over the years, the cost of debt decreases, mainly due to the lower risk-free rate. Because Royal Bank of Canada uses mainly deposits over bonds, the cost of debt is not so much affected by the increase in the coupon value for high-debt ratio.

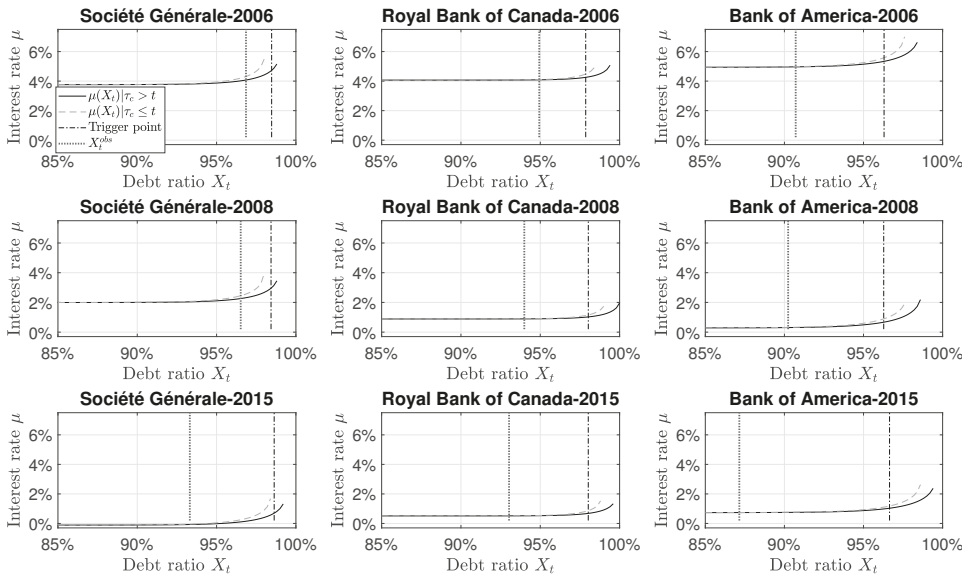


Figure 4. Cost of the debt. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the cost of debt when there is CoCos in the debt structure. The light gray dashed line represents the cost of debt without CoCos in the debt structure. The vertical black dashed-dotted line corresponds to the trigger level (α) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

The cost of debt directly affects the cost of capital (Equation (20)). Indeed, the lower the cost of capital, the more likely the bank is creating value. The cost of capital indicates the minimum rate of return before generating profit. When the debt ratio is close to 0%, the bank is financed by equity: the cost of capital corresponds to the cost of equity. Between the two extremes, the cost of capital tends to decrease: the cost of debt is generally a cheaper source of financing than equity, except when the firm is in financial distress. As shown in Figure 5, the equity return declines substantially during the 2008 financial crisis, especially for the American and European banks and stays at a low level after the crisis. The low cost of capital for Société Générale and Bank of America in 2008 and 2010 is in line with the low Return on Equity and risk-free rate observed for these periods. The effects of the 2008 financial crisis were less important for the Canadian banks. In Europe, the 2011 debt crisis is also a reason for these low interest rates and returns on equity. In all studied cases, the cost of capital is at its lowest just before the debt ratio conversion trigger, around 95%. In Figure 5, we see that, for all studied cases, the debt ratio is to the left of the debt ratio level that minimizes the cost of capital. In a multi-period setting, banks use a precautionary cushion to stay away from the conversion and default thresholds.

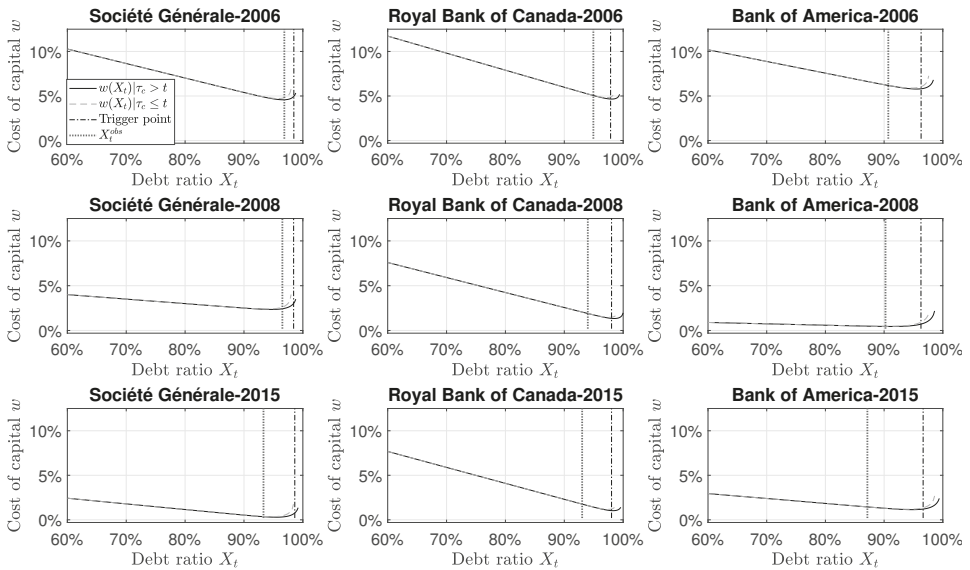


Figure 5. Cost of the capital. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the cost of capital when there is CoCos in the debt structure. The light gray dashed line represents the cost of capital without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

To mitigate the effect of Assumption 3, the dynamic optimization program (21) and (22) is applied recursively using a backward recursion until there are no significant changes in the dividend policy and the value of the discounted aggregated dividend per share. Figure 6 shows the optimal dividend rate in function of the debt ratio at time $t = T - 30$. The optimal dividend rate is more permissive when there is CoCo in the debt structure in 2008 and 2015. The difference between the optimal dividend rate with and without CoCos is the largest for Bank of America during the financial crisis. The U.S. banks were the most affected by the subprime crisis. The presence of CoCos in the debt structure has a significant effect on the optimal dividend rate. The optimal dividend rate in 2015 for Bank of America is very small (0.36%) due to the expected return on capital: we use a weighted average over the mean ROE and risk-free rate over the last five years, and these parameters are very small due to the low interest rate policy and the effect of the crisis.

Figure 7 shows the (normalized) discounted cumulated dividend value of the primary equity holders at time $t = T - 30$. The primary equity holders have larger dividend gains when the CoCo debt is part of the debt structure. This is due to the dilution effect created by the CoCo conversion: CoCo debt holders become equity holders, then, mechanically, decrease the dividend-per-share. The (normalized) discounted cumulated dividend value is at its maximum for a debt ratio around 50%, except for Bank of America in 2015. For all studied cases, the observed debt ratio does not maximize the future dividends. The level of the discounted cumulated dividend for Bank of America in 2008 reflects the sudden financial crisis and our parametrization of the expected return on capital. The Return on Equity was in average equal to 18.82% from 2003 to 2006, 10.77% in 2007 and 1.81% in 2008. The low ROE in 2008 affected the level of the cost of capital and so the discount factor used in the dynamic program. However, the expected return on capital is difficult to estimate and so we use past observations on

the ROE and risk free rate instead of prospective data. This leads to a high expected return on capital, which affects the level of the optimal dividend rate. A sensitivity analysis follows.

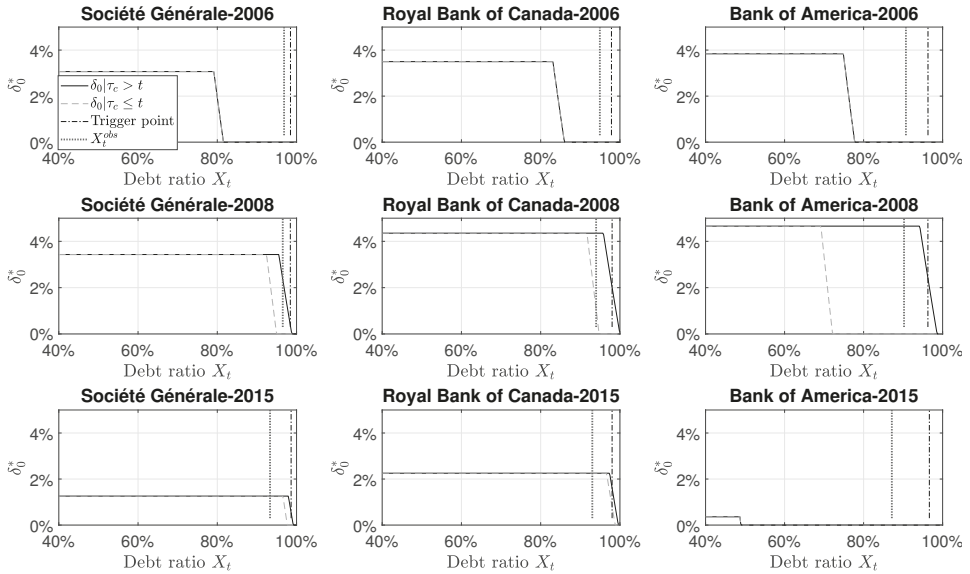


Figure 6. Optimal dividend at time $T = 0$. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the optimal dividend rate when there is CoCos in the debt structure. The light gray dashed line represents the optimal dividend rate without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (α) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

4.3. Sensitivity Analysis

Sensitivity analysis is performed to examine how changes in the parameters affect our conclusions and in particular the discounted cumulated dividend. First, what is the impact of letting the proportion of CoCos in the debt structure increases from 1% to 10%, all other parameters being the same? Obviously, Figure 8 shows that, after conversion, discounted cumulated dividends are much lower because the dilution effect is more important: there is more debt converted to equity. Before the conversion, there is not much impact or a slight decrease due to the modest increase of the conversion risk. For a large debt ratio, the discounted cumulated dividends are slightly smaller.

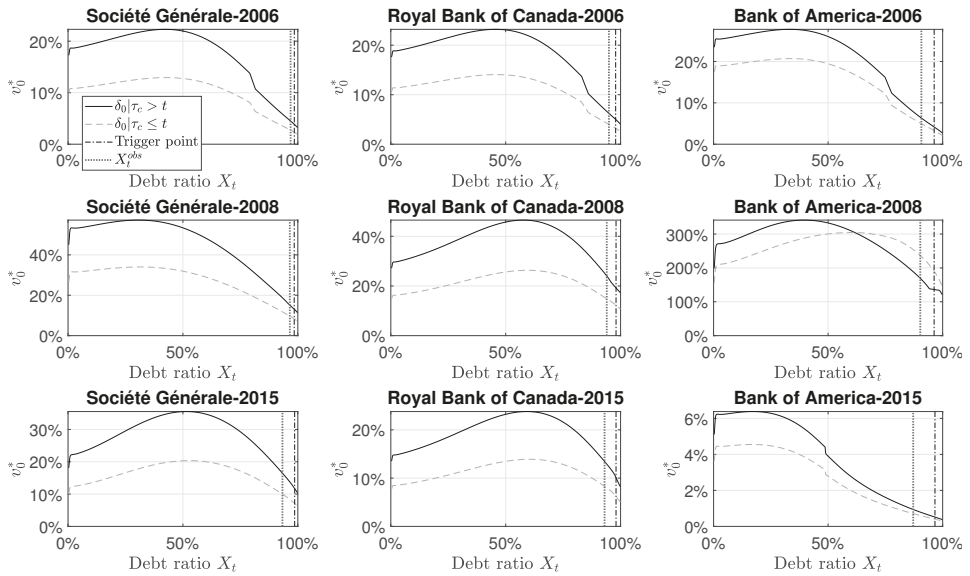


Figure 7. (Standardized) value of discounted cumulated dividends at time $T = 0$. All curves are obtained from a Monte Carlo simulation based on 2×10^6 paths. The parameters are described in Table 2. Each column corresponds to a specific bank and each line corresponds to a specific year. The black line represents the case when there is CoCos in the debt structure. The light gray dashed line represents the case without CoCos in the debt structure. The vertical dashed-dotted line corresponds to the trigger level (a) of 1–5.125% of risk-weighted assets. The dark grey dotted line corresponds to the debt ratio observed for the specified bank at the specified year.

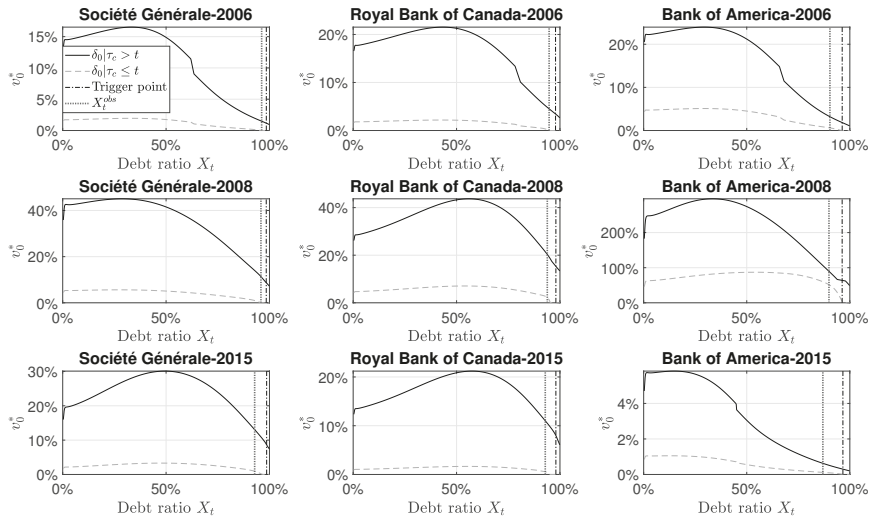


Figure 8. Discounted cumulated dividends at time $T = 0$, with $y = 10\%$

Second, we analyse how good the situation is for the firms, that is, how firms with high Return on Equity affect the dividends. We fix the ROE for each bank equal to the maximum ROE observed from 1999 to 2017, all other parameters being the same as in the base case. Changing the ROE also impacts

the expected return on capital (m). These two parameters impact the cost of capital, the discount factor and thus the discounted cumulated dividends as well as the optimal dividend rates. Indeed, an increase of the expected return on capital leads to higher optimal dividend rates. For example, the optimal dividend rate for Bank of America in 2015 increases at 2.97% and create higher discounted cumulated dividends. In 2008, for Bank of America, increasing the ROE allows to have reasonable values of discounted cumulated dividends, as shown in Figure 9. For Société Générale in 2008 and 2015, increasing the ROE slightly decrease the discounted cumulated dividends. There is a trade-off between having high dividend today and a high dividend in the future, which can create impatience for the equity holder. Indeed, increasing the Return on Equity increase the expected return on capital due to our parametrization. This leads to an increase in the cost of capital that affects the discount factor of the future dividend. It is not clear what is the expectation of the discounted cumulated dividends in that context.

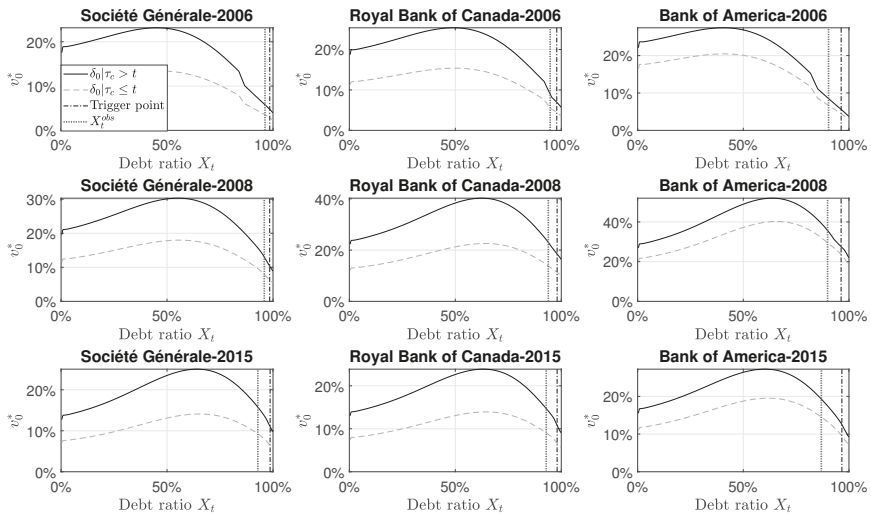


Figure 9. Optimal dividend and discounted cumulated dividends at time $T = 0$, with high Return on Equity.

Finally, we change the default intensity to have higher default probabilities, reflecting bad conditions for banks. We assume a one-year default probability of 8% around the critical debt ratio imposed by the regulator. An increase of the default probability decreases the value of the discounted cumulated dividend faster when the bank is largely indebted, as shown in Figure 10. As expected, when the bank is not so much indebted, there is no or a slight decrease in the level of the discounted cumulated dividend, especially, in the absence of CoCo in the debt structure, as observed for Société Générale in 2006 and 2008.

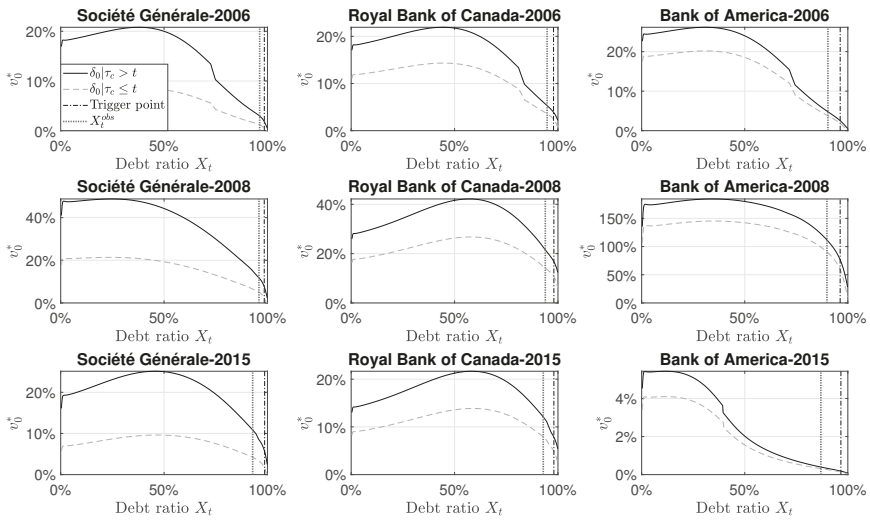


Figure 10. Optimal dividend and discounted cumulated dividends at time $T = 0$, with higher default probabilities

5. Conclusions

We introduced a reduced form approach that also includes information about the financial health of the firm to model the impact of CoCos in the debt structure by taking into account the uncertainty and the time-varying default and conversion risks. To analyze the impact of CoCos in the bank’s capital structure, we set up a hypothetical capital structure consisting of equity, ordinary debt and CoCo debt in order to understand the benefits/costs of convertible contingent debt. Real parameters are used, corresponding to three banks considered as G-SIBs (too big to fail) in three different markets (Europe, Canada and the United States) and for three periods (pre-crisis, crisis and post-crisis) to generate scenarios. Results reveal that CoCos reduce the cost and the risk of standard debt when they are being added to the capital structure. In fact, CoCos acts as a precautionary buffer to prolong the default time and hence to reduce the risk of default for ordinary debt. Meanwhile, CoCo itself is a more expensive instrument compared to the ordinary debt, remembering that it has zero chance of recovery on default. Furthermore, the results show that the presence of CoCos in the debt structure reduce the total cost of debt, knowing that the cost of debt is a weighted average of each debt component. From the point of view of primary equity holders, the presence of CoCos in the bank’s capital structure increases the aggregated value for shareholders. The optimal dividend policy derived from the dynamic optimization suggests paying more dividends when the CoCo is active and leads to higher aggregated dividends, especially in the context of crisis and low return.

Author Contributions: Conceptualization, G.G. and F.P.; Methodology, G.G., F.P. and D.B.; Software, D.B. and F.P.; Validation, D.B., G.G. and F.P.; Formal Analysis, D.B., G.G. and F.P.; Writing D.B., G.G. and F.P.; Supervision, G.G.

Funding: This research was funded by IFSID (Montreal Institute from Structured Finance and Derivatives), GRI (Global Risk Institute) and HEC Montréal.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Floating Rates

Since pricing is performed under the risk-neutral measure, the first step is to determine the risk factor Q -dynamics.

Appendix A.1. The \mathbb{Q} –Dynamics of the Debt Ratio

The model uncertainty is captured by the noise series $\{\varepsilon_t^{\mathbb{P}}\}_{t \in \mathbb{N}}$. The link between the \mathbb{P} probability measure and some risk neutral probability measure \mathbb{Q} is achieved through a Radon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$. The latter is based on the Girsanov theorem,

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_{t+1}} = \exp \left(\gamma \varepsilon_{t+1}^{\mathbb{P}} - \frac{\gamma^2}{2} \right),$$

where the γ parameter is interpreted as the price of risk. The conditional risk-neutral moment generating function (MGF) of $\varepsilon_{t+1}^{\mathbb{P}}$ is

$$M_t^{\mathbb{Q}}(x) = E^{\mathbb{Q}} \left[e^{-x\varepsilon_{t+1}^{\mathbb{P}}} \Big| \mathcal{F}_t \right] = E^{\mathbb{P}} \left[\exp \left(\gamma \varepsilon_{t+1}^{\mathbb{P}} - \frac{\gamma^2}{2} \right) e^{-x\varepsilon_{t+1}^{\mathbb{P}}} \Big| \mathcal{F}_t \right] = \exp \left(\frac{1}{2} x^2 - x\gamma \right),$$

which corresponds to the MGF of a Gaussian random variable with expectation $-\gamma$ and variance 1. Thus, we construct a risk-neutral noise term centred at zero:

$$\varepsilon_{t+1}^{\mathbb{Q}} = \varepsilon_{t+1}^{\mathbb{P}} + \gamma.$$

Under the risk-neutral measure \mathbb{Q} , the asset return satisfies

$$R_{t+1} = m_{t+1} + \sigma_{t+1} \varepsilon_{t+1}^{\mathbb{P}} = m_{t+1} + \sigma_{t+1} \left(\varepsilon_{t+1}^{\mathbb{Q}} - \gamma \right) = \underbrace{m_{t+1} - \gamma \sigma_{t+1}}_{r_{t+1}} + \sigma_{t+1} \varepsilon_{t+1}^{\mathbb{Q}} = R_{t+1}^{\mathbb{Q}}.$$

Therefore, $m_{t+1} = r_{t+1} + \gamma \sigma_{t+1}$ and the risk-neutral debt ratio dynamics are

$$\begin{aligned} X_{t+1}^{\mathbb{Q}} &= \frac{(1 + \delta_{t+1})(1 + \eta_{t+1}) X_t}{(1 + R_{t+1}^{\mathbb{Q}}) - (\mu_{t+1} - \eta_{t+1}) X_t} & (A1) \\ &= \frac{(1 + \delta_{t+1})(1 + \eta_{t+1}) X_t}{(1 + R_{t+1}^{\mathbb{Q}}) - (r_{t+1}(1 - z_t)(1 - y_t) + b(X_t, y_t) z_t(1 - y_t) + c(X_t) y_t) X_t} \mathbf{1}_{\tau_C > t+1} \\ &\quad + \frac{(1 + \delta_{t+1})(1 + \eta_{t+1}) X_t}{(1 + R_{t+1}^{\mathbb{Q}}) - (r_{t+1}(1 - z_t)(1 - y_t) + b(X_t, y_t) z_t(1 - y_t) + y_t) X_t} \mathbf{1}_{\tau_C = t+1} \\ &\quad + \frac{(1 + \delta_{t+1})(1 + \eta_{t+1}) X_t}{(1 + R_{t+1}^{\mathbb{Q}}) - (r_{t+1}(1 - z_t) + b(X_t, 0) z_t) X_t} \mathbf{1}_{\tau_C < t+1}. \end{aligned}$$

The conversion decision depends on the debt ratio, given that there is no dividend payment, $\delta_{t+1} = 0$, no change in debt structure, $\eta_{t+1} = 0$, and with the CoCo interest payment. More precisely,

$$X_{t+1}^{\mathbb{Q},C} = \frac{X_t}{1 + R_{t+1}^{\mathbb{Q}} - (r_{t+1}(1 - z_t)(1 - y_t) + b(X_t, y_t) z_t(1 - y_t) + c(X_t) y_t) X_t}$$

and

$$G_{t+1}^{\mathbb{Q}} = g \left(X_{t+1}^{\mathbb{Q},C} \right). \tag{A2}$$

The default intensity is computed after the conversion (if the latter has not already occurred):

$$X_{t+1}^{Q,D} = \frac{(1 + \eta_{t+1}) X_t}{\left(1 + R_{t+1}^Q\right) - (r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t) + y_t) X_t} \mathbf{1}_{\tau_C > t} + \frac{X_t}{1 + R_{t+1}^Q - (r_{t+1} (1 - z_t) + b(X_t, 0) z_t) X_t} \mathbf{1}_{\tau_C \leq t},$$

and

$$H_{t+1}^Q = h\left(X_{t+1}^{Q,D}\right). \tag{A3}$$

Appendix A.2. Credit Sensitive Debt

Lemma A1. Assume that the time $t + 1$ value of the standard risky debt is

$$B_{(t+1)} \mathbf{1}_{\tau_D > t} = B_t \left((1 + b(X_t, y_t)) \mathbf{1}_{\tau_D > t+1} + \rho_D \mathbf{1}_{\tau_D = t+1} \right) \mathbf{1}_{\tau_D > t}.$$

Then, the interest rate $b(X_t, y_t)$ of the standard debt satisfies

$$b\left(X_t^Q, 0\right) \mathbf{1}_{\tau_C \leq t} = \left[r_t + (1 - \rho_D + r_t) \left(\frac{1}{\mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C \leq t]} - 1 \right) \right] \mathbf{1}_{\tau_C \leq t} \tag{A4}$$

and

$$b\left(X_t^Q, y_t\right) \mathbf{1}_{\tau_C > t} = \left[r_t + (1 - \rho_D + r_t) \left(\frac{1}{\mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C > t]} - 1 \right) \right] \mathbf{1}_{\tau_C > t}, \tag{A5}$$

where r_t is the risk-free rate and ρ_D represents the recovery rate.

See proof in Appendix B.1. The risk-neutral survival probabilities are provided by Lemma A5.

Appendix A.3. Convertible Contingent Debt

Lemma A2. Assume that the time $t + 1$ value of the convertible contingent debt is

$$C_t \left((1 + c(X_t)) \mathbf{1}_{\tau_C > t+1} + (1 + \rho_{CC}(X_t)) \mathbf{1}_{\tau_C = t+1} \right) \mathbf{1}_{\tau_D > t+1}.$$

It follows that the convertible contingent debt interest rate satisfies

$$c(X_t) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} = \frac{(1 + r_t) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} - \mathbb{Q}_t[\tau_D > t + 1 | \tau_C > t, \tau_D > t]}{\mathbb{Q}_t[\tau_C > t + 1 | \tau_C > t, \tau_D > t] + \rho_{CC} \mathbb{Q}_t[\tau_D > t + 1 \text{ and } \tau_C = t + 1 | \tau_C > t, \tau_D > t]} \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t}. \tag{A6}$$

Remark A1. If $\rho_{CC} = 1$, then $c(X_t) = b\left(X_t^Q, y_t\right)$, where $\rho_D = 0$.

See proof in Appendix B.2. The risk-neutral conditional probabilities are detailed in Lemmas A3–A5.

Appendix A.4. Conditional Probabilities

The proofs of the following lemmas are available in the Appendix B.5.

Lemma A3 (No conversion risk-neutral probability).

$$\begin{aligned} \mathbb{Q}_t[\tau_C > t + 1 | \tau_\alpha > t, \tau_D > t] &= \mathbb{Q}_t[\tau_\alpha > t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] \\ &= E_t^Q \left[\exp\left(-H_{t+1}^Q \mathbf{1}_{\tau_C > t} - G_{t+1}^Q\right) | \tau_\alpha > t, \tau_D > t \right], \end{aligned}$$

where G_{t+1}^Q and H_{t+1}^Q are defined by Equations (A2) and (A3).

Lemma A4 (Conversion and survival risk-neutral probability).

$$\begin{aligned} & \mathbb{Q}_t [\tau_\alpha = t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] \\ &= \mathbb{E}_t^Q \left[\exp \left(-H_{t+1}^Q \mathbf{1}_{\tau_C > t} \right) \left(1 - \exp \left(-G_{t+1}^Q \right) \right) \middle| \tau_C > t, \tau_D > t \right], \end{aligned}$$

where G_{t+1}^Q and H_{t+1}^Q are defined at Equations (A2) and (A3).

Lemma A5 (Survival risk-neutral probabilities).

$$\begin{aligned} \mathbb{Q}_t [\tau_D > t + 1 | \tau_D > t, \tau_\alpha > t] &= \mathbb{E}_t^Q \left[\exp \left(-H_{t+1}^Q \mathbf{1}_{\tau_C > t} \right) \middle| \tau_D > t, \tau_\alpha > t \right], \\ \mathbb{Q}_t [\tau_D > t + 1 | \tau_D > t, \tau_\alpha \leq t] &= \mathbb{E}_t^Q \left[\exp \left(-H_{t+1}^Q \mathbf{1}_{\tau_C \leq t} \right) \middle| \tau_D > t, \tau_\alpha \leq t \right], \end{aligned}$$

where H_{t+1}^Q is defined at Equation (A3).

Appendix A.5. Approximations

Appendix A.5.1. Approximation of $b(x, 0)$

From Equation (A1), we note that $X_{t+1}^{Q,D}$ can be viewed as a function of R_{t+1}^Q . To emphasize this relation, we write $X_{t+1}^{Q,D} = X_{t+1}^{Q,D} \left(R_{t+1}^Q \right)$. Since R_{t+1}^Q is centered at r_{t+1} , a Taylor expansion of $\exp \left(-h \left(X_{t+1}^{Q,D} \right) \right)$ around $R_{t+1}^Q = r_{t+1}$ leads to

$$\begin{aligned} & \mathbb{E}_t^Q \left[\exp \left(-h \left(X_{t+1}^{Q,D} \right) \right) \right] \mathbf{1}_{\tau_D > t} \cong \exp \left(-h \left(X_{t+1}^{Q,D} \left(r_{t+1} \right) \right) \right) \\ & - \exp \left(-h \left(X_{t+1}^{Q,D} \left(r_{t+1} \right) \right) \right) h' \left(X_{t+1}^{Q,D} \left(r_{t+1} \right) \right) \frac{\partial X_{t+1}^{Q,D}}{\partial R_{t+1}^Q} \left(r_{t+1} \right) \underbrace{\mathbb{E}_t^Q \left[\left(R_{t+1}^Q - r_{t+1} \right) \right]}_{=0}, \end{aligned}$$

that is,

$$\mathbb{Q}_t [\tau_D > t + 1 | \tau_D > t, \tau_\alpha \leq t] \cong \exp \left(-h \left(\frac{X_t}{1 + r_{t+1} - (r_{t+1} (1 - z_t) + b(X_t, 0) z_t) X_t} \right) \right) \mathbf{1}_{\tau_C \leq t}.$$

Placing it back in Equation (A4),

$$b(x, 0) \cong r_t + (1 - \rho_D + r_t) \left(\exp \left(h \left(\frac{x}{1 + r_{t+1} - (r_{t+1} (1 - z_t) + b(x, 0) z_t) x} \right) \right) - 1 \right).$$

Appendix A.5.2. Approximation of $b(x, y)$ and $c(x)$

Similarly, let

$$\begin{aligned} X_{t+1}^{Q,D} \left(r_{t+1} \right) \mathbf{1}_{\tau_C > t} &= \frac{(1 - y_t) X_t}{(1 + r_{t+1}) - (r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t) + y_t) X_t}, \\ X_{t+1}^{Q,C} \left(r_{t+1} \right) &= \frac{X_t}{1 + r_{t+1} - (r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t) + c(X_t) y_t) X_t}. \end{aligned}$$

From Lemmas A3–A5, we derive the following approximations:

$$\begin{aligned} \mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_\alpha > t] &\cong e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t})}, \\ \mathbb{Q}_t[\tau_C > t + 1 | \tau_C > t, \tau_D > t] &\cong e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t}) - g(X_{t+1}^{Q,C}(r_{t+1}))}, \\ \mathbb{Q}_t[\tau_\alpha = t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] &\cong e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t})} \left(1 - e^{-g(X_{t+1}^{Q,C}(r_{t+1}))}\right). \end{aligned}$$

Placing it back in Equations (A5) and (A6) leads to

$$\begin{aligned} b(X_t, y_t) &\cong r_t + (1 - \rho_D + r_t) \left(e^{h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t})} - 1 \right), \\ c(X_t) &\cong \frac{(1 + r_t) - e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t})}}{e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t}) - g(X_{t+1}^{Q,C}(r_{t+1}))} + \rho_C e^{-h(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t})} \left(1 - e^{-g(X_{t+1}^{Q,C}(r_{t+1}))}\right)} \\ &= \frac{(1 + r_t) \exp\left(h\left(X_{t+1}^{Q,D}(r_{t+1})\mathbf{1}_{\tau_C > t}\right)\right) - 1}{(1 - \rho_C) \exp\left(-g\left(X_{t+1}^{Q,C}(r_{t+1})\right)\right) + \rho_C}. \end{aligned}$$

Appendix B. Proofs

Appendix B.1. Standard Bond Floating Coupon Rate

Proof of Lemma A1. The time t price of a credit-sensitive debt is denoted B_t . At time $t + 1$, there is an interest rate payment of $B_t b(X_t^Q, y_t)$ if no default occurs and there is a recovery of $\rho_D B_{(t+1)^-}$ in case of default. We set the interest rate $b(X_t^Q, y_t)$ such that the next period debt value, before the dividend payment, remains constant, that is, $B_{(t+1)^-} = B_t$. Interestingly, the floating rate is affected by the presence (or absence) of the CoCo debt instrument. We therefore study two cases: $\tau_C > t$ and $\tau_C \leq t$.

If $\tau_C \leq t$, then the standard risky bond pricing corresponds to the classic case:

$$\begin{aligned} B_t \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} &= E_t^Q \left[\frac{\left(B_t + B_t b(X_t^Q, 0) \right) \mathbf{1}_{\tau_D > t+1} + \rho_D B_t \mathbf{1}_{\tau_D = t+1}}{1 + r_t} \right] \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} \\ &= B_t E_t^Q \left[\frac{\left(1 + b(X_t^Q, 0) \right) \mathbf{1}_{\tau_D > t+1} + \rho_D (1 - \mathbf{1}_{\tau_D > t+1})}{1 + r_t} \right] \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} \\ &= B_t \frac{\rho_D + \left(1 - \rho_D + b(X_t^Q, 0) \right) \mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C \leq t]}{1 + r_t} \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t}. \end{aligned}$$

Since

$$\frac{\rho_D + \left(1 - \rho_D + b(X_t^Q, 0) \right) \mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C \leq t]}{1 + r_t} \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} = \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t},$$

$$b(X_t^Q, 0) \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t} = \left[r_t + (1 - \rho_D + r_t) \left(\frac{1}{\mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C \leq t]} - 1 \right) \right] \mathbf{1}_{\tau_C \leq t} \mathbf{1}_{\tau_D > t}.$$

Similarly, if $\tau_C > t$, then the risky debt interest rate is

$$b(X_t^Q, y_t) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} = \left[r_t + (1 - \rho_D + r_t) \left(\frac{1}{\mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_C > t]} - 1 \right) \right] \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t}.$$

□

Appendix B.2. Convertible Bond Floating Coupon Rate

Proof of Lemma A2. The goal is to choose the coupon rate $c(X_t)$ so that the convertible contingent debt value at time $t + 1$ is the same as its value at time t , that is $C_{(t+1)^-} = C_t$.

Assume that $\tau_C > t$ and that $\tau_D > t$, since, otherwise, the convertible instrument is worth 0. In that case, the time t value of the convertible contingent debt value satisfies

$$C_t \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} = E_t^Q \left[\frac{C_t + c(X_t) C_t}{1 + r_t} \mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a > t+1} + \frac{C_t + \rho_C c(X_t) C_t}{1 + r_t} \mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a = t+1} \right] \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t},$$

which is equivalent to

$$\begin{aligned} \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} &= E_t^Q \left[\frac{1 + c(X_t)}{1 + r_t} \mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a > t+1} + \frac{1 + \rho_C c(X_t)}{1 + r_t} \mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a = t+1} \right] \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} \\ &= \frac{1}{1 + r_t} \left((1 + c(X_t)) E_t^Q [\mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a > t+1} | \tau_C > t, \tau_D > t] \right. \\ &\quad \left. + (1 + \rho_C c(X_t)) E_t^Q [\mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a = t+1} | \tau_C > t, \tau_D > t] \right) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t}. \end{aligned}$$

Consequently,

$$\begin{aligned} &c(X_t) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} \\ &= \frac{(1 + r_t) \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t} - E_t^Q [\mathbf{1}_{\tau_D > t+1} | \tau_C > t, \tau_D > t]}{E_t^Q [\mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a > t+1} | \tau_C > t, \tau_D > t] + \rho_C E_t^Q [\mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a = t+1} | \tau_C > t, \tau_D > t]} \mathbf{1}_{\tau_C > t} \mathbf{1}_{\tau_D > t}. \end{aligned}$$

Lastly, note that $\mathbf{1}_{\tau_D > t+1} \mathbf{1}_{\tau_a > t+1} = \mathbf{1}_{\tau_C > t+1}$. \square

Appendix B.3. The Equity Value Variation

From Equations (7) and (8), it follows that

$$\begin{aligned} &E_{t+1} - E_t \\ &= A_{t+1} - D_{t+1} - E_t \\ &= \frac{(1 + R_{t+1}) A_t - (\mu_{t+1} - \eta_{t+1}) D_t}{1 + \delta_{t+1}} - D_t (1 + \eta_{t+1}) - (A_t - D_t) \\ &= \left(\frac{R_{t+1} - \delta_{t+1}}{1 + \delta_{t+1}} - \frac{\mu_{t+1} + \eta_{t+1} \delta_{t+1}}{1 + \delta_{t+1}} X_t \right) A_t. \end{aligned}$$

Therefore,

$$E_t^{\mathbb{P}} [E_{t+1} |_{\delta_{t+1}=0} - E_t] = E_t^{\mathbb{P}} [R_{t+1} - \mu_{t+1} X_t] A_t = (m_{t+1} - E_t^{\mathbb{P}} [\mu_{t+1}] X_t) A_t \mathbf{1}_{\tau_C \leq t},$$

where

$$\begin{aligned} &E_t^{\mathbb{P}} [\mu_{t+1}] \\ &= (r_{t+1} (1 - z_t) + b(X_t, 0) z_t) \mathbf{1}_{\tau_C \leq t} \\ &\quad + (r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t)) \mathbb{P}_t [\tau_C = t + 1 | \tau_C > t] \mathbf{1}_{\tau_C > t} \\ &\quad + (r_{t+1} (1 - z_t) (1 - y_t) + b(X_t, y_t) z_t (1 - y_t) + c(X_t) y_t) \mathbb{P}_t [\tau_C > t + 1 | \tau_C > t] \mathbf{1}_{\tau_C > t}. \end{aligned}$$

Appendix B.4. The Value of Expected Discounted Dividends at T

Proof of Lemma 1. Since $u > T$, we can substitute $X_u = X_T$, $A_u = A_T$, $\mu_u = \mu_{T+1}$, $\delta_u^* = \delta_{T+1}^* = m_T - \mu_{T+1}X_T$, and $DF_{T+1,T+1+u} = (1 + w(X_T))^{-u}$ in Equation (21). Therefore,

$$\begin{aligned} & V(T + 1, X_{T+1}^0) \mathbf{1}_{\tau_D > T+1} \\ \cong & \sum_{u=T+1}^{\infty} E_{T+1}^{\mathbb{P}} \left[DF_{T+1,u} \delta_u A_u \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq u} \right) \mathbf{1}_{\tau_D > u} \right] \mathbf{1}_{\tau_D > T+1} \\ = & \delta_{T+1} A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) \sum_{u=T+1}^{\infty} (1 + w(X_T))^{-(T+1-u)} E_{T+1}^{\mathbb{P}} [\mathbf{1}_{\tau_D > u}] \mathbf{1}_{\tau_D > T+1}. \end{aligned}$$

Because the debt ratio remains constant over time, the default intensity is

$$H_u = H_T = \lambda_D + \left(\frac{1}{\theta_D} \max(X_T; 0) \right)^{\beta_D}.$$

Therefore, the conditional survival probability $E_{T+1}^{\mathbb{P}} [\mathbf{1}_{\tau_D > u}] \mathbf{1}_{\tau_D > T+1} = \exp(- (u - T - 1) H_T)$ and

$$V(T + 1, X_{T+1}^0) \mathbf{1}_{\tau_D > T+1} \cong \delta_{T+1} A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) \sum_{s=0}^{\infty} \left(\frac{\exp(-H_T)}{1 + w(X_T)} \right)^s \mathbf{1}_{\tau_D > T+1}.$$

The final result is obtained using the geometrical series property $\sum_{s=0}^{\infty} a^s = (1 - a)^{-1}$, provided that $|a| < 1$:

$$V(T + 1, X_{T+1}^0) \mathbf{1}_{\tau_D > T+1} \cong \delta_{T+1} A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) \left(1 + \frac{\exp(-H_T)}{1 + w(X_T) - \exp(-H_T)} \right) \mathbf{1}_{\tau_D > T+1}$$

and

$$\begin{aligned} & V(T, X_T^0, \delta_T) \mathbf{1}_{\tau_D > T} \\ \cong & \left\{ \delta_T A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) + E_T^{\mathbb{P}} \left[\frac{V(T + 1, X_{T+1}^0) \mathbf{1}_{\tau_D > T+1}}{1 + w(X_T)} \right] \right\} \mathbf{1}_{\tau_D > T} \\ = & \left\{ \delta_T A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) + E_T^{\mathbb{P}} \left[\frac{V(T + 1, X_{T+1}^0) \mathbf{1}_{\tau_D > T+1}}{1 + w(X_T)} E_{T+1}^{\mathbb{P}} [\mathbf{1}_{\tau_D > T+1}] \right] \right\} \mathbf{1}_{\tau_D > T} \\ = & \left\{ \delta_T A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) + E_T^{\mathbb{P}} \left[\frac{V(T + 1, X_{T+1}^0) \exp(-H_{T+1})}{1 + w(X_T)} \right] \right\} \mathbf{1}_{\tau_D > T} \\ = & \left\{ \delta_T A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) + \frac{V(T + 1, X_T^0) \exp(-H_T)}{1 + w(X_T)} \right\} \mathbf{1}_{\tau_D > T} \\ = & A_T \left(1 - \frac{y_0 \alpha}{1 - \alpha + y_0 \alpha} \mathbf{1}_{\tau_C \leq T} \right) \left\{ \delta_T + \delta_{T+1} \left(\frac{\exp(-H_T)}{1 + w(X_T) - \exp(-H_T)} \right) \right\} \mathbf{1}_{\tau_D > T}. \end{aligned}$$

□

Appendix B.5. Proofs of Lemmas A3–A5

Proof of Lemma A3. Note that $\{\tau_C > t - 1\} = \{\tau_\alpha > t - 1\} \cap \{\tau_D > t - 1\}$. Once we condition on the time t debt ratio, the events $\{\tau_\alpha > t\}$ and $\{\tau_D > t\}$ are independent. Therefore,

$$\begin{aligned} & \mathbb{Q}_t(\tau_\alpha > t \text{ and } \tau_D > t | \tau_\alpha > t - 1, \tau_D > t - 1) \\ &= \mathbb{Q}_t(\tau_D > t | \tau_\alpha > t - 1, \tau_D > t - 1) \mathbb{Q}_t(\tau_\alpha > t | \tau_\alpha > t - 1, \tau_D > t - 1) \\ &= \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C > t - 1}\right) \exp\left(-G_t^{\mathbb{Q}}\right) \mathbf{1}_{\tau_\alpha > t - 1} \mathbf{1}_{\tau_D > t - 1}. \end{aligned}$$

Finally, the law of iterated conditional expectation implies that

$$\begin{aligned} & \mathbb{Q}_t(\tau_\alpha > t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t) \\ &= E_t^{\mathbb{Q}}\left[\mathbb{E}_{t+1}^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha > t+1} \mathbf{1}_{\tau_D > t+1} | \tau_\alpha > t, \tau_D > t\right] | \tau_\alpha > t, \tau_D > t\right] \\ &= E_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) \exp\left(-G_{t+1}^{\mathbb{Q}}\right) | \tau_\alpha > t, \tau_D > t\right]. \end{aligned}$$

□

Proof of Lemma A4. Given the time t debt ratio, the events $\{\tau_\alpha = t\}$ and $\{\tau_D > t\}$ are independent. Therefore,

$$\begin{aligned} & \mathbb{Q}_t(\tau_\alpha = t \text{ and } \tau_D > t | \tau_\alpha > t - 1, \tau_D > t - 1) \\ &= \mathbb{Q}_t(\tau_D > t | \tau_\alpha > t - 1, \tau_D > t - 1) \mathbb{Q}_t(\tau_\alpha = t | \tau_\alpha > t - 1, \tau_D > t - 1) \\ &= \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C > t - 1}\right) \left(1 - \exp\left(-G_t^{\mathbb{Q}}\right)\right) \mathbf{1}_{\tau_\alpha > t - 1} \mathbf{1}_{\tau_D > t - 1} \end{aligned}$$

and

$$\begin{aligned} & \mathbb{Q}_t(\tau_\alpha = t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t) \\ &= E_t^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha = t+1} \mathbf{1}_{\tau_D > t+1} | \tau_\alpha > t, \tau_D > t\right] \\ &= E_t^{\mathbb{Q}}\left[\mathbb{E}_{t+1}^{\mathbb{Q}}\left[\mathbf{1}_{\tau_\alpha = t+1} \mathbf{1}_{\tau_D > t+1} | \tau_\alpha > t, \tau_D > t\right] | \tau_\alpha > t, \tau_D > t\right] \\ &= E_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) \left(1 - \exp\left(-G_{t+1}^{\mathbb{Q}}\right)\right) | \tau_\alpha > t, \tau_D > t\right]. \end{aligned}$$

□

Proof of Lemma A5. First, $\mathbb{Q}_t[\tau_D > t + 1 | \tau_\alpha > t, \tau_D > t]$ is a consequence of Lemmas A3 and A4:

$$\begin{aligned} & \mathbb{Q}_t[\tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] \\ &= \mathbb{Q}_t[\tau_\alpha > t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] \\ &\quad + \mathbb{Q}_t[\tau_\alpha = t + 1 \text{ and } \tau_D > t + 1 | \tau_\alpha > t, \tau_D > t] \\ &= E_t^{\mathbb{Q}}\left[\exp\left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C > t}\right) | \tau_C > t, \tau_D > t\right]. \end{aligned}$$

We now compute $\mathbb{Q}_t[\tau_D > t + 1 | \tau_D > t, \tau_\alpha \leq t]$. Since

$$\mathbb{Q}_t[\tau_D > t | \tau_\alpha \leq t - 1, \tau_D > t - 1] = \exp\left(-H_t^{\mathbb{Q}} \mathbf{1}_{\tau_C \leq t - 1}\right) \mathbf{1}_{\tau_\alpha \leq t - 1} \mathbf{1}_{\tau_D > t - 1},$$

$$\begin{aligned} \mathbb{Q}_t [\tau_D > t + 1 | \tau_D > t, \tau_\alpha \leq t] &= \mathbb{E}_t^{\mathbb{Q}} [\mathbf{1}_{\tau_D > t+1} | \tau_D > t, \tau_\alpha \leq t] \\ &= \mathbb{E}_t^{\mathbb{Q}} \left[\mathbb{E}_{t+1}^{\mathbb{Q}} [\mathbf{1}_{\tau_D > t+1} | \tau_D > t, \tau_\alpha \leq t] | \tau_D > t, \tau_\alpha \leq t \right] \\ &= \mathbb{E}_t^{\mathbb{Q}} \left[\exp \left(-H_{t+1}^{\mathbb{Q}} \mathbf{1}_{\tau_C \leq t} \right) | \tau_D > t, \tau_\alpha \leq t \right]. \end{aligned}$$

□

Appendix C. Calibration of Default and Conversion Probabilities

We use the one-year default probability computed by Bloomberg in order to calibrate the one-year default probability. We also impose a default probability of 1.5% around the critical debt ratio imposed by the regulator. Using a log-linearization of Equation (19), the function H_{t+1} can be expressed such that

$$H_{t+1} = -\log(1 - \mathbb{P}_{t+1}(\tau_D = t + 1 | \tau_C \leq t, \tau_D > t)).$$

Fixing the coefficient β_D , a log-linear regression can be done. We have

$$-\log(1 - \mathbb{P}_{t+1}(\tau_D = t + 1 | \tau_C \leq t, \tau_D > t)) = \beta_0 + \beta_1 (X_{t+1})^{\beta_D},$$

where $\beta_0 = \lambda$ and $\beta_1 = \left(\frac{1}{\theta_D}\right)^{\beta_D}$. A simple transformation of β_1 gives the estimated parameter θ_D :

$$\theta_D = \left(\frac{1}{\beta_1}\right)^{1/\beta_D}.$$

From Equation (14), we can isolate the parameter θ_C , such that

$$\theta_C = \frac{X_{t+1}^C}{(-\log(1 - \mathbb{P}_{t+1}(\tau_C = t + 1 | \tau_C > t)) - H_{t+1} \mathbf{1}_{\tau_C > t+1})^{1/\beta_C}}.$$

The conversion probability is calibrated such that there is a 10% conversion probability over the observed mean debt ratio and a 90% conversion probability at the trigger level imposed by the regulator. We are in the presence of two equations with two unknown parameters: θ_C and β_C . The observed mean debt ratio, denoted by \bar{X}^{obs} , corresponds to the average debt ratio from 2004 to 2017. The trigger level used, denoted by \bar{X}_α^{obs} , is the average trigger level imposed by the regulator from 2004 to 2017. Solving the system of equation, we have

$$\begin{aligned} \theta_C &= \exp \left\{ \frac{\log(f_\alpha) \log(\bar{X}^{obs}) - \log(f_X) \log(\bar{X}_\alpha^{obs})}{\log(f_\alpha) - \log(f_X)} \right\}, \\ \beta_C &= \frac{\log(f_\alpha) \log(f_\alpha) - \log(f_\alpha) \log(f_X)}{\log(f_\alpha) \log(\bar{X}_\alpha^{obs}) - \log(f_\alpha) \log(\bar{X}^{obs})}, \end{aligned}$$

where

$$\begin{aligned} f_\alpha &= -\log(1 - 0.9) - \left(\lambda + \left(\frac{\bar{X}_\alpha^{obs}}{\theta_C}\right)^{1/\beta_D} \right), \\ f_X &= -\log(1 - 0.1) - \left(\lambda + \left(\frac{\bar{X}^{obs}}{\theta_C}\right)^{1/\beta_D} \right). \end{aligned}$$

Appendix D. Finding the Optimal Dividend Rate Sequence

Appendix D.1. Post-Conversion Optimal Dividend Rates

We use the subscript C to indicate that the conversion occurred, that is, to indicate that $\tau_C \leq t$. Let $v_C(t, X_t^0, \delta_{t:\infty}) = A_t^{-1} V_C(t, X_t^0, \delta_{t:\infty})$ and $v_C^*(t, X_t^0) = A_t^{-1} V_C(t, X_t^0, \delta_{t:\infty}^*)$, where $\delta_{t:\infty}^*$ represents the optimal dividend rates from t onward when $\tau_C \leq t$. In that case, since

$$\frac{A_{t+1}}{A_t} = \frac{D_t}{A_t} \frac{D_{t+1}}{D_t} \frac{A_{t+1}}{D_{t+1}} \mathbf{1}_{\tau_C \leq t} = \frac{X_t}{X_{t+1}} (1 + \eta_{t+1}) \mathbf{1}_{\tau_C \leq t} = \frac{X_t}{X_{t+1}},$$

Equation (22) becomes

$$\begin{aligned} \delta_t^* = & \arg \max_{\delta_t \in [0, \delta_t^{\max}]} A_t \left\{ \delta_t \left(1 - \frac{(1 + \rho_{CC}(\alpha)) y_0 \alpha}{1 - \alpha + (1 + \rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) \right. \\ & \left. + \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_C^*(t+1, X_{t+1}^0)}{1 + w(X_t)} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}. \end{aligned}$$

The conditional expectation is evaluated numerically. More precisely, let

$$0 = x_0 < x_1 < \dots < x_n = 1$$

be a discretization of the pre-dividend debt ratio X_{t+1}^0 support. For $i \in \{1, 2, \dots, n\}$, $\xi_i = (x_{i-1} + x_i) / 2$ is the mid-point of each interval. Assuming that $X_t^0 = \xi_i$ and that the dividend rate for that particular state is $\delta_{t,i}$, Equations (10), (9), (6), (16), and (24) become respectively

$$\begin{aligned} X_{t,i} &= (1 + \delta_{t,i}) \xi_i, \\ X_{t+1} &= \frac{(1 + \delta_{t+1}) X_{t,i}}{(1 + R_{t+1}) - \mu_{t+1,i} X_{t,i}}, \\ \mu_{t+1,i} &= r_{t+1} (1 - z_t) + b(X_{t,i}, 0) z_t, \\ X_{t+1}^D &= \frac{X_{t,i}}{(1 + R_{t+1}) - \mu_{t+1,i} X_{t,i}} = X_{t+1} |_{\delta_{t+1}=0}, \text{ and} \\ \delta_{t,i}^{\max} &= \min \left(\max \left(\frac{x_0^* - \xi_i}{\xi_i}, 0 \right), m_{t+1} \right), \end{aligned}$$

where x_0^* is the solution of $m_{t+1} - x_0^* [r_{t+1} (1 - z_t) + b(x_0^*, 0) z_t] = 0$. As shown in the Appendix D.3, the transition probabilities are

$$\begin{aligned} \pi_{ij}(\delta_{t,i}) &= \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0 < x_j, \tau_D > t + 1 \mid \tau_C \leq t, X_t^0 = \xi_i \right] \\ &\cong [\Phi(\varphi_{i,j-1}(\delta_{t,i})) - \Phi(\varphi_{i,j}(\delta_{t,i}))] \exp(-h(\xi_j)), \end{aligned} \tag{A7}$$

where Φ is the cumulative distribution function of a standard normal random variable and

$$\varphi_{i,j}(\delta_{t,i}) = \frac{(x_j^{-1} + \mu_{t+1,i}) X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}}. \tag{A8}$$

Lastly,

$$\begin{aligned}
 & \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_{t,i}}{X_{t+1}} \frac{v_C^*(t+1, X_{t+1}^0)}{1+w(X_{t,i})} \mathbf{1}_{\tau_D > t+1} \middle| X_t |_{\delta_i=0} = \xi_i, \tau_\alpha \leq t, \tau_D > t \right] \\
 & \cong \sum_{j=1}^n \frac{(1+\delta_{t,i}) \xi_i}{(1+\delta_{t+1,j}^*) \xi_j} \frac{v_C^*(t+1, \xi_j)}{1+w((1+\delta_{t,i}) \xi_i)} \pi_{ij}(\delta_{t,i}), \\
 \\
 & \delta_{t,i}^* = \arg \max_{\delta_{t,i} \in [0, \delta_{t,i}^{\max}]} \delta_{t,i} \left(1 - \frac{(1+\rho_{CC}(\alpha)) y_0 \alpha}{1-\alpha + (1+\rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \right) \\
 & \quad + \sum_{j=1}^n \frac{(1+\delta_{t,i}) \xi_i}{(1+\delta_{t+1,j}^*) \xi_j} \frac{v_C^*(t+1, \xi_j)}{1+w((1+\delta_{t,i}) \xi_i)} \pi_{ij}(\delta_{t,i}),
 \end{aligned} \tag{A9}$$

and

$$\begin{aligned}
 v_C^*(t, \xi_i) &= \delta_{t,i}^* \frac{(1+\rho_{CC}(\alpha)) y_0 \alpha}{1-\alpha + (1+\rho_{CC}(\alpha)) y_0 \alpha} \mathbf{1}_{\tau_C \leq t} \\
 & \quad + \sum_{j=1}^n \frac{(1+\delta_{t,i}^*) \xi_i}{(1+\delta_{t+1,j}^*) \xi_j} \frac{v_C^*(t+1, \xi_j)}{1+w((1+\delta_{t,i}^*) \xi_i)} \pi_{ij}(\delta_{t,i}^*).
 \end{aligned} \tag{A10}$$

Appendix D.2. Pre-Conversion Optimal Dividend Rates

We use the subscript NC to indicate that the CoCo debt is not yet converted, that is, to indicate that $\tau_C > t$. Let $v_{NC}(t, X_t^0, \delta_{t:\infty}) = A_t^{-1} V_{NC}(t, X_t^0, \delta_{t:\infty})$ and $v_{NC}^*(t, X_t^0) = A_t^{-1} V_{NC}(t, X_t^0, \delta_{t:\infty}^{\text{opt}})$, where $\delta_{t:\infty}^{\text{opt}}$ is the optimal dividend rate sequence:

$$\delta_{t+1}^{\text{opt}} = \delta_{t+1}^* \mathbf{1}_{\tau_C = t+1} + \delta_{t+1}^{**} \mathbf{1}_{\tau_C > t+1}.$$

In that case, Equation (22) becomes

$$\begin{aligned}
 \delta_t^{**} &= \arg \max_{\delta_t \in [0, \delta_t^{\max}]} A_t \left\{ \delta_t + \mathbb{E}_t^{\mathbb{P}} \left[(1+\eta_{t+1}) \frac{X_t}{X_{t+1}} \frac{A_{t+1}^{-1} V_C(t+1, X_{t+1}, \delta_{t+1:\infty}^*)}{1+w(X_t)} \mathbf{1}_{\tau_C = t+1} \mathbf{1}_{\tau_D > t+1} \right] \right. \\
 & \quad \left. + \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{A_{t+1}^{-1} V_{NC}(t+1, X_{t+1}^0, \delta_{t+1:\infty}^{\text{opt}})}{1+w(X_t)} \mathbf{1}_{\tau_C > t+1} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t} \\
 &= \arg \max_{\delta_t \in [0, \delta_t^{\max}]} A_t \left\{ \delta_t + (1-y_t) \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_C^*(t+1, X_{t+1}^0)}{1+w(X_t)} \mathbf{1}_{\tau_C = t+1} \mathbf{1}_{\tau_D > t+1} \right] \right. \\
 & \quad \left. + \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_{NC}^*(t+1, X_{t+1}^0)}{1+w(X_t)} \mathbf{1}_{\tau_C > t+1} \mathbf{1}_{\tau_D > t+1} \right] \right\} \mathbf{1}_{\tau_D > t}.
 \end{aligned} \tag{A11}$$

Assuming that $X_t^0 = \xi_i$ and that the dividend rate for that particular state is $\delta_{t,i}$, then Equations (10), (9), (6), (13), (16) and (24) become

$$\begin{aligned} X_{t,i} &= (1 + \delta_{t,i}) \xi_i \\ X_{t+1} &= \frac{(1 + \delta_{t+1})(1 - y_t) X_{t,i}}{(1 + R_{t+1}) - (\mu_{t+1,i} - c(X_{t,i}) y_t + y_t) X_{t,i}} \mathbf{1}_{\tau_\alpha=t+1} + \frac{(1 + \delta_{t+1}) X_{t,i}}{(1 + R_{t+1}) - \mu_{t+1,i} X_{t,i}} \mathbf{1}_{\tau_\alpha>t+1}, \\ \mu_{t+1,i} &= r_{t+1} (1 - z_t) (1 - y_t) + b(X_{t,i}, y_t) z_t (1 - y_t) + c(X_{t,i}) y_t, \\ X_{t+1}^C &= X_{t+1}^0 \Big|_{\tau_\alpha>t+1} = \frac{X_{t+1}^0 \Big|_{\tau_\alpha=t+1}}{(1 - y_t) + (1 - c(X_{t,i})) y_t X_{t+1}^0 \Big|_{\tau_\alpha=t+1}}, \\ X_{t+1}^D &= X_{t+1}^0 \Big|_{\tau_\alpha=t+1} = \frac{(1 - y_t) X_{t+1}^0 \Big|_{\tau_\alpha>t+1}}{1 - (1 - c(X_{t,i})) y_t X_{t+1}^0 \Big|_{\tau_\alpha>t+1}}, \text{ and} \\ \delta_{t,i}^{\max} &= \min \left(\max \left(\frac{X_0^{**} - \xi_i}{\xi_i}, 0 \right), m_{t+1} \right), \end{aligned}$$

where x_0^{**} is the solution of $m_{t+1} - x_0^{**} [r_{t+1} (1 - z_t) (1 - y_t) + b(x_0^{**}, y_t) z_t (1 - y_t) + c(x_0^{**}) y_t] = 0$. As shown in Appendix D.3, the transition probabilities are

$$\begin{aligned} &\pi_{ij}^* (\delta_{t,i}) \tag{A12} \\ &= \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0 \Big|_{\tau_\alpha=t+1} < x_j, \tau_\alpha = t + 1, \tau_D > t + 1 \mid \tau_C > t, X_t \Big|_{\delta_t=0} = \xi_i \right] \\ &\cong \left(1 - \exp \left(-g \left(\frac{X_{t+1}^0 \Big|_{\tau_\alpha=t+1}}{(1 - y_t) + (1 - c(X_{t,i})) y_t X_{t+1}^0 \Big|_{\tau_\alpha=t+1}} \right) \right) \right) \exp(-h(\xi_j)) \\ &\times \left[\Phi(\varphi_{i,j-1}^* (\delta_{t,i})) - \Phi(\varphi_{i,j}^* (\delta_{t,i})) \right] \end{aligned}$$

with

$$\varphi_{i,j}^* (\delta_{t,i}) = \frac{\left[x_j^{-1} (1 - y_t) + \mu_{t+1,i} + (1 - c(X_{t,i})) y_t \right] X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}} \tag{A13}$$

and

$$\begin{aligned} &\pi_{ij}^{**} (\delta_{t,i}) \tag{A14} \\ &= \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0 \Big|_{\tau_\alpha>t+1} < x_j, \tau_\alpha > t + 1, \tau_D > t + 1 \mid \tau_C > t, X_t \Big|_{\delta_t=0} = \xi_i \right] \\ &\cong \exp \left(-g(\xi_j) - h \left(\frac{(1 - y_t) \xi_j}{1 - (1 - c(X_{t,i})) y_t \xi_j} \right) \right) \left[\Phi(\varphi_{i,j-1}^{**} (\delta_{t,i})) - \Phi(\varphi_{i,j}^{**} (\delta_{t,i})) \right] \end{aligned}$$

with

$$\varphi_{i,j}^{**} (\delta_{t,i}) = \frac{(x_j^{-1} + \mu_{t+1,i}) X_{t,i} - 1 - m_{t+1}}{\sigma_{t+1}}. \tag{A15}$$

The conditional expectations are approximated with

$$\begin{aligned} &\mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_C^*(t + 1, X_{t+1}^0)}{1 + w(X_t)} \mathbf{1}_{\tau_C=t+1} \mathbf{1}_{\tau_D>t+1} \mid X_t \Big|_{\delta_t=0} = \xi_i, \tau_C > t \right] \\ &\cong \sum_{j=1}^n \frac{(1 + \delta_{t,i}) \xi_i}{(1 + \delta_{t+1,j}^*) \xi_j} \frac{v_C^*(t + 1, \xi_j)}{1 + w((1 + \delta_{t,i}) \xi_i)} \pi_{ij}^* (\delta_{t,i}) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}_t^{\mathbb{P}} \left[\frac{X_t}{X_{t+1}} \frac{v_{NC}^*(t+1, X_{t+1}^0)}{1+w(X_t)} \mathbf{1}_{\tau_C > t+1} \mathbf{1}_{\tau_D > t+1} \mid X_t |_{\delta_i=0} = \zeta_i, \tau_C > t \right] \\ & \cong \sum_{j=1}^n \frac{(1+\delta_{t,i}) \zeta_i}{(1+\delta_{t+1,j}^{**}) \zeta_j} \frac{v_{NC}^*(t+1, \zeta_j)}{1+w((1+\delta_{t,i}) \zeta_i)} \pi_{ij}^{**}(\delta_{t,i}). \end{aligned}$$

Lastly,

$$\begin{aligned} \delta_{t,i}^{**} &= \arg \max_{\delta_{t,i} \in [0, \delta_{t,i}^{\max}]} A_t \left\{ \delta_{t,i} + (1-y_t) \sum_{j=1}^n \frac{(1+\delta_{t,i}) \zeta_i}{(1+\delta_{t+1,j}^{**}) \zeta_j} \frac{v_C^*(t+1, \zeta_j)}{1+w((1+\delta_{t,i}) \zeta_i)} \pi_{ij}^*(\delta_{t,i}) \right. \\ & \quad \left. + \sum_{j=1}^n \frac{(1+\delta_{t,i}) \zeta_i}{(1+\delta_{t+1,j}^{**}) \zeta_j} \frac{v_{NC}^*(t+1, \zeta_j) \mathbf{1}_{\tau_C > t+1}}{1+w((1+\delta_{t,i}) \zeta_i)} \pi_{ij}^{**}(\delta_{t,i}) \right\} \mathbf{1}_{\tau_D > t}. \\ v_{NC}^*(t, \zeta_i) &\cong \delta_{t,i}^{**} + (1-y_t) \sum_{j=1}^n \frac{(1+\delta_{t,i}^{**}) \zeta_i}{(1+\delta_{t+1,j}^{**}) \zeta_j} \frac{v_C^*(t+1, \zeta_j)}{1+w((1+\delta_{t,i}^{**}) \zeta_i)} \pi_{ij}^*(\delta_{t,i}^{**}) \\ & \quad + \sum_{j=1}^n \frac{(1+\delta_{t,i}^{**}) \zeta_i}{(1+\delta_{t+1,j}^{**}) \zeta_j} \frac{v_{NC}^*(t+1, \zeta_j)}{1+w((1+\delta_{t,i}^{**}) \zeta_i)} \pi_{ij}^{**}(\delta_{t,i}^{**}). \end{aligned} \tag{A16}$$

Appendix D.3. Proofs

Appendix D.3.1. Proof of Equation (A7)

$$\begin{aligned} & \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0 < x_j \mid \tau_C \leq t, X_t^0 = \zeta_i \right] \\ &= \mathbb{P}_t \left[x_{j-1} \leq \frac{X_{t,i}}{(1+R_{t+1}) - \mu_{t+1,i} X_{t,i}} < x_j \mid \tau_C \leq t, X_t^0 = \zeta_i \right] \\ &= \mathbb{P}_t \left[x_j^{-1} < \frac{(1+R_{t+1}) - \mu_{t+1,i} X_{t,i}}{X_{t,i}} \leq x_{j-1}^{-1} \mid \tau_C \leq t, X_t^0 = \zeta_i \right] \\ &= \mathbb{P}_t \left[\varphi_{ij}(\delta_{t,i}) < \frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \leq \varphi_{ij-1}(\delta_{t,i}) \mid \tau_C \leq t, X_t^0 = \zeta_i \right], \end{aligned}$$

where the $\varphi_{i,j}(\delta_{t,i})$ are provided in Equation (A8). The proof is completed by noting that

$$\frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \Big|_{\mathcal{F}_t}$$

is a standard normal random variable. Finally,

$$\begin{aligned} & \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0 < x_j, \tau_D > t+1 \mid \tau_C \leq t, X_t^0 = \zeta_i \right] \\ &= \mathbb{P}_t \left[\tau_D > t+1 \mid \tau_C \leq t, X_t^0 = \zeta_i, x_{j-1} \leq X_{t+1}^0 < x_j \right] \mathbb{P}_t \left[x_{j-1} \leq X_{t+1} |_{\delta_{t+1}=0} < x_j \mid \tau_C \leq t, X_t^0 = \zeta_i \right] \\ &\cong \exp(-h(\zeta_j)) \mathbb{P}_t \left[x_{j-1} \leq X_{t+1} |_{\delta_{t+1}=0} < x_j \mid \tau_C \leq t, X_t^0 = \zeta_i \right]. \end{aligned}$$

Appendix D.3.2. Proof of Equation (A12)

The probability that $X_{t+1}^0|_{\tau_a=t+1}$ lies between x_{j-1} and x_j is

$$\begin{aligned} & \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0|_{\tau_a=t+1} < x_j \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[x_{j-1} \leq \frac{(1 - y_t) X_{t,i}}{(1 + R_{t+1}) - (\mu_{t+1,i} - c(X_{t,i}) y_t + y_t) X_{t,i}} < x_j \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[x_j^{-1} \leq \frac{(1 + R_{t+1}) - (\mu_{t+1,i} - c(X_{t,i}) y_t + y_t) X_{t,i}}{(1 - y_t) X_{t,i}} < x_{j-1}^{-1} \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[\varphi_{i,j}^* (\delta_{t,i}) < \frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \leq \varphi_{i,j-1}^* (\delta_{t,i}) \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \Phi \left(\varphi_{i,j-1}^* (\delta_{t,i}) \right) - \Phi \left(\varphi_{i,j}^* (\delta_{t,i}) \right), \end{aligned}$$

where $\varphi_{i,j}^* (\delta_{t,i})$ is defined at Equation (A13).

The probability that $X_{t+1}^0|_{\tau_a=t+1}$ is contained between x_{j-1} and x_j while the conversion occurs without the firm default is

$$\begin{aligned} & \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0|_{\tau_a=t+1} < x_j, \tau_\alpha = t + 1, \tau_D > t + 1 \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[\tau_\alpha = t + 1 \mid x_{j-1} \leq X_{t+1}^0|_{\tau_a=t+1} < x_j, \tau_C > t, X_t^0 = \xi_i \right] \\ & \times \mathbb{P}_t \left[\tau_D > t + 1 \mid x_{j-1} \leq X_{t+1}^0|_{\tau_a=t+1} < x_j, \tau_C > t, X_t^0 = \xi_i \right] \\ & \times \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0|_{\tau_a=t+1} < x_j \mid \tau_C > t, X_t^0 = \xi_i \right] \\ \cong & \left(1 - \exp \left(g \left(\frac{X_{t+1}^0|_{\tau_a=t+1}}{(1 - y_t) + (1 - c(X_{t,i})) y_t X_{t+1}^0|_{\tau_a=t+1}} \right) \right) \right) \\ & \times \exp \left(-h(\xi_j) \right) \left[\Phi \left(\varphi_{i,j-1}^* (\delta_{t,i}) \right) - \Phi \left(\varphi_{i,j}^* (\delta_{t,i}) \right) \right]. \end{aligned}$$

Appendix D.3.3. Proof of Equation (A14)

$$\begin{aligned} & \mathbb{P}_t \left[x_{j-1} \leq X_{t+1}^0|_{\tau_a>t+1} < x_j \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[x_{j-1} \leq \frac{X_{t,i}}{(1 + R_{t+1}) - \mu_{t+1,i} X_{t,i}} < x_j \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[x_j^{-1} < \frac{(1 + R_{t+1}) - \mu_{t+1,i} X_{t,i}}{X_{t,i}} \leq x_{j-1}^{-1} \mid \tau_C > t, X_t^0 = \xi_i \right] \\ = & \mathbb{P}_t \left[\varphi_{i,j}^{**} (\delta_{t,i}) < \frac{R_{t+1} - m_{t+1}}{\sigma_{t+1}} \leq \varphi_{i,j-1}^{**} (\delta_{t,i}) \mid \tau_C > t, X_t^0 = \xi_i \right], \end{aligned}$$

where the $\varphi_{i,j}^{**} (\delta_{t,i})$ are provided in Equation (A15).

References

Albul, Boris, Dwight M. Jaffee, and Alexei Tchisty. 2013. Contingent convertible bonds and capital structure decisions. In *2013 Meeting Papers in Seoul (South Korea)*. Minneapolis: Society for Economic Dynamics, p. 682.

Altman, Edward I., and Vellore M. Kishore. 1996. Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal* 52: 57–64. [CrossRef]

Bolton, Patrick, and Frédéric Samama. 2012. Capital access bonds: Contingent capital with an option to convert. *Economic Policy* 27: 275–317. [CrossRef]

- Brigo, Damiano, João Garcia, and Nicola Pede. 2015. Coco bonds pricing with credit and equity calibrated first-passage firm value models. *International Journal of Theoretical and Applied Finance* 18: 1550015. [CrossRef]
- Chen, Nan, Paul Glasserman, Behzad Nouri, and Markus Pelger. 2013. *Cocos, Bail-in, and Tail Risk*. Working Paper. Washington, DC: Office of Financial Research, U.S. Treasury Department.
- Chen, Nan, Paul Glasserman, Behzad Nouri, and Markus Pelger. 2017. Contingent capital, tail risk, and debt-induced collapse. *The Review of Financial Studies* 30: 3921–69. [CrossRef]
- Cheridito, Patrick, and Zhikai Xu. 2015. A reduced-form contingent convertible bond model with deterministic conversion intensity. *Journal of Risk* 17: 18. [CrossRef]
- Chung, Tsz-Kin, and Yue-Kuen Kwok. 2016. Enhanced equity-credit modelling for contingent convertibles. *Quantitative Finance* 16: 1511–27. [CrossRef]
- De Spiegeleer, Jan, Stephan Höcht, Marquet Ine, and Wim Schoutens. 2017. Coco bonds and implied cet1 volatility. *Quantitative Finance* 17: 813–24. [CrossRef]
- Duffie, Darrell. 2010. *A Contractual Approach to Restructuring Financial Institutions*. Stanford: Hoover Institution, Stanford University.
- Duffie, Darrell, and David Lando. 2001. Term structures of credit spreads with incomplete accounting information. *Econometrica* 69: 633–64. [CrossRef]
- Flannery, Mark J. 2005. *No Pain, No Gain? Effecting Market Discipline via Reverse Convertible Debentures*. Oxford: Oxford University Press.
- Flannery, Mark J. 2014. Maintaining adequate bank capital. *Journal of Money, Credit and Banking* 46: 157–80. [CrossRef]
- Flannery, Mark J. 2016. Stabilizing large financial institutions with contingent capital certificates. *Quarterly Journal of Finance* 6: 1650006. [CrossRef]
- Glasserman, Paul, and Behzad Nouri. 2012. Contingent capital with a capital-ratio trigger. *Management Science* 58: 1816–33. [CrossRef]
- Hilscher, Jens, and Alon Raviv. 2014. Bank stability and market discipline: The effect of contingent capital on risk taking and default probability. *Journal of Corporate Finance* 29: 542–60. [CrossRef]
- McDonald, Robert L. 2013. Contingent capital with a dual price trigger. *Journal of Financial Stability* 9: 230–41. [CrossRef]
- Pennacchi, George. 2010. A Structural Model of Contingent Bank Capital. FRB of Cleveland Working Paper No. 10-04. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1595080 (accessed on 12 April 2019).
- Sundaresan, Suresh, and Zhenyu Wang. 2015. On the design of contingent capital with a market trigger. *The Journal of Finance* 70: 881–920. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Risks Editorial Office
E-mail: risks@mdpi.com
www.mdpi.com/journal/risks



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com



ISBN 978-3-03928-761-1