UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

ARDIT GJEÇI

THE DETERMINANTS OF NON-PERFORMING LOANS IN A BANKING SYSTEM

DOCTORAL DISSERTATION

Ljubljana, 2022

AUTHORSHIP STATEMENT

The undersigned Ardit Gjeçi, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB), author of this written final work of studies with the title: The Determinants of Non-performing Loans in a Banking System, prepared under supervision of prof. dr. Matej Marinč.

D E C L A R E

1. this doctoral dissertation to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this doctoral dissertation to be language-edited and technically in adherence with the SEB's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this doctoral dissertation in accordance with the SEB's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this doctoral dissertation could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this doctoral dissertation and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this doctoral dissertation and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this doctoral dissertation for the detection of content similarity with other written works, using similarity detection software that is connected with the SEB LU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this doctoral dissertation in the electronic form, the right of its reproduction, as well as the right of making this doctoral dissertation available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. to have acquired from publishers, to whom I have previously exclusively transferred material copyrights for articles, all the necessary permissions for the inclusion of articles in the printed and electronic forms of this dissertation. These permissions enable the University of Ljubljana to save this doctoral dissertation in the electronic form, to reproduce it, as well as to make it publicly available on the World Wide Web via the Repository of the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited;
11. my consent to publication of my personal data that are included in this doctoral dissertation and in this declaration, when this doctoral dissertation is published.

Ljubljana, February 7th, 2022 Author's signature: ________________________

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my mentor, prof. dr. Matej Marinč, for his encouragement, contributions of time, untiring effort, assistance, support, and valuable insights throughout the doctoral study.

I express my profound gratitude and heartfelt thanks for the financial support from the Erasmus Mundus project Green-Tech-WB: Smart and Green technologies for innovative and sustainable societies in Western Balkans (551984-EM-1-2014-1-ES-ERA MUNDUS-EMA2).

I would also like to extend my deepest gratitude to the committee members, prof. dr. Marko Košak, prof. dr. Igor Masten, and assoc. prof. dr. Orfea Dhuci.

I am also grateful to my close friends for all their support and motivation.

Finally, and most importantly, my heartfelt gratitude and thanks go to my father Vangjel, my mother Ylvije, and my brother Klaudio for their unconditional love, invaluable encouragement, and support throughout my studies.

THE DETERMINANTS OF NON-PERFORMING LOANS IN A BANKING SYSTEM

SUMMARY

The research purpose of this dissertation is to determine and assess different determinants of non-performing loans (NPLs) and their interaction with the performance of credit institutions. The global financial crisis of 2008 was characterized by a high level of NPLs. Thus, establishing special attention to the determinants of NPLs is crucial for the design of the legal and regulatory framework for lending and borrowing and the credit risk management framework (policies, procedures, methods and tools, models etc.). This dissertation has three objectives. The first objective is to examine the impact of corruption on the level of non-performing loans in the banking sector and analyse whether higher corruption is related to the elevated proportion of non-performing loans in the banking system. This is particularly pertinent because corruption has seriously impeded economic, political, and socially sustainable development. The second objective is to investigate additional factors (bank size, bank capitalization and profitability, and the presence of asset management companies) through which, NPLs may affect bank lending growth across EU and non-EU member countries. This allows us to consider and distinguish the effects of NPLs in different periods, before, during, and after the crisis, for the lending behaviour, in particular for loan growth. The third objective is to construct and evaluate a credit scoring model using the reject inference method to account for the rejected loan applicants in estimation of default probabilities. Moreover, a scoring model is proposed to be assessed by differentiating between parametric and non-parametric classification methods.

In Chapter 1, we aim to investigate the relationship between corruption and non-performing loans in a global context, which is relevant in view of the substantial economic and social costs of the last financial crisis. We construct a comprehensive dataset, covering 140 countries and 7,773 banks over the 2000-2016 period. We take into account several alternative indices as proxies for corruption levels as well as two alternative proxies for loan quality and combine bank-specific variables with institutional variables and macroeconomic data. Using this panel dataset, we test several hypotheses with respect to the relationship between corruption and non-performing loans. We analyse this relationship in-depth using bank-level data, as opposed to earlier country-level studies, which enables us to explore various possible channels through which corruption may interact with non-performing loans. In addition to the main relationship, we also test whether certain bank-specific (i.e., bank size, capitalization, and profitability), institutional (i.e., rule of law and regulatory framework,) or cultural (i.e., the level of cultural collectivism) characteristics modify the relationship between corruption and non-performing loans. Finally, we analyse the relationship between corruption and non-performing loans in different subperiods and use

alternative proxies, both for the level of corruption and for the quality of loans as a further robustness check.

In Chapter 2, we empirically examine the potential channels through which the level of non-performing loans affects lending growth. This chapter uses the unbalanced annual data from 42 EU and non-EU member countries over the period ranging between 2000 and 2017. In the empirical model, we check for bank-specific factors (i.e., bank size, bank capitalization, and profitability) and selected macroeconomic indicators (i.e., GDP growth and the real interest rate) and institutional factors (i.e., asset management companies). Different geographic regions are included in the study. Four subsample regressions were computed separately for banks in EU countries, non-EU countries, advanced economies, and emerging economies. We use loan loss provision (LLP) as an alternative credit risk variable alternative to NPLs to perform the robustness check of our results. Specifically, the results generated in chapter 2 reveal a statistically significant association between NPLs and lending growth. Furthermore, the results show that the effects of a higher level of NPLs on loan growth are more pronounced for larger and better-capitalized banks. Finally, we explain the impact of asset management companies on lending growth. We find that the interaction term between NPL and asset management companies (AMC) is statistically insignificant, suggesting no evidence related to loan growth.

In Chapter 3, we aim to contribute to the existing knowledge of scoring models that credit institutions use to assess a borrower's creditworthiness employing unique loan-level data from a candidate EU country. Compared to traditional credit analysis based predominantly on a bank officer's judgment, recent advanced/sophisticated credit scoring models consist of a broader range of characteristics to better evaluate the customer's creditworthiness. In some cases, the models were estimated using only approved borrowers from datasets, whereas rejected ones were not considered relevant. The question that arises here is whether using the same tools and techniques covering approved and rejected applications will improve the estimation of the credit scoring model while removing the biases within the model. This chapter employs rejected inference technique to examine the influence of rejected borrowers on the credit scoring models. We rely on three classification methods, logistic regression (LR), classification tree method (CART), and radial basis function (RBF), to assess and compare the diverse model performance with and without selection bias correction. The results confirm that the reject inference, augmentation through parcelling, improves the performance of scoring models indicating that the use of rejected borrowers' information is essential. Our findings also suggest that logistic regression achieves the highest average correct classification rate considering approved and rejected loan applications, 72.6% compared to 64.3% when only approved applications are taken into consideration, indicating an accuracy improvement of 8.3%. Furthermore, highest receiver operating curve increases from 68.2% to 79.8% resulting in an improvement of 11.6%.

Keywords: Artificial Neural Network, Bank Lending, Classification and Regression Tree, Corruption Perception Index, Credit Scoring, Economic Development, Global Financial Crisis, Logistic Regression, Macroeconomic Determinants, Non-performing Loans, Reject Inference.

DETERMINANTE NEDONOSNIH TERJATEV V BANČNEM SISTEMU

POVZETEK

Raziskovalni namen pričujoče disertacije je določiti in ovrednotiti različne determinante nedonosnih terjatev in njihov stik z delovanjem kreditnih institucij. Globalna finančna kriza leta 2008 je bila posledica visoke stopnje nedonosnih terjatev. Zato je poseben poudarek na determinantah nedonosnih terjatev ključnega pomena za ustanovitev pravnega in regulativnega okvira za posojilno in kreditno dejavnost ter okvira za upravljanje s tveganji (strategije, postopki, metode, pripomočki, modeli in tako dalje). Pričujoča disertacija ima tri cilje. Prvi cilj je preučiti vpliv korupcije na stopnjo nedonosnih terjatev v bančnem sektorju in analizirati, ali je višji odstotek korupcije povezan s povišano stopnjo nedonosnih terjatev v bančnem sistemu. Ta cilj je še posebej pomemben, saj je korupcija močno zavrla gospodarski, političen in družbeni trajnostni razvoj. Drugi cilj disertacije je preučiti dodatne dejavnike (velikost banke, bančno kapitalizacijo in dobičkonosnost ter prisotnost podjetij, ki upravljajo s premoženjem), skozi katere lahko nedonosne terjatve vplivajo na rast bančnih posojil, tako v državah članicah Evropske unije (EU) kot tudi v drugih državah. To nam omogoča, da lahko preučimo učinke nedonosnih terjatev v različnih časovnih obdobjih – pred, med in po krizi – na posojilno vedenje, predvsem pa na rast posojil. Tretji cilj disertacije je ustvariti in ovrednotiti model za zbiranje kreditnih točk, z uporabo metode zavrnilnega sklepa kot odgovorom na zavrnitev posojilne prošnje pri ocenjevanju verjetnosti neplačil. Poleg tega je predlagano, da je model pridobivanja kreditnih točk ovrednoten tudi na podlagi razločevanja parametrične in neparametrične metode klasifikacije.

V Poglavju 1 je naš cilj raziskati odnos med korupcijo in nedonosnimi terjatvami v globalnem kontekstu, ki je ključen v smislu znatnih ekonomskih in družbenih stroškov zadnje finančne krize. Za ta namen ustvarimo tudi razumljivo podatkovno bazo, ki pokriva 140 držav in 7.773 bank v obdobju 2000–2016. V zakup so vzeti tudi številni drugi kazalci, kot denimo pooblaščenci za prepoznavanje stopenj korupcije in tudi dva alternativna pooblaščenca za kakovost posojil in združene bančne ter institucionalne spremenljivke in sami makroekonomski podatki. S podano podatkovno bazo lahko preizkusimo številne predpostavke, z upoštevanjem povezave med korupcijo in nedonosnimi terjatvami. Hkrati v poglavju to povezavo analiziramo podrobneje, z uporabo podatkov na ravni bank, v nasprotju s predhodnimi raziskavami, ki so uporabljale podatke na ravni držav. To nam omogoča raziskavo številnih morebitnih kanalov, prek katerih je možna vzpostavitev povezave med korupcijo in nedonosnimi terjatvami. Poleg glavnega razmerja pa preizkušamo tudi, ali določene posebne bančne (velikost banke, kapitalizacija, dobičkonosnost), institucionalne (zakon pravnega in regulatornega okvira) ali kulturne (stopnja kulturnega kolektivizma) lastnosti vplivajo na odnos med korupcijo in nedonosnimi terjatvami. Nazadnje pa analiziramo tudi povezavo med korupcijo in nedonosnimi terjatvami

v različnih podobdobjih, z uporabo alternativnih podatkov, tako za prepoznavanje stopnje korupcije kot tudi za samo kakovost posojil, kot nadaljnji preizkus stabilnosti.

V Poglavju 2 z empiričnim pristopom preučujemo morebitne kanale, prek katerih stopnja nedonosnih terjatev vpliva na rast posojilne dejavnosti. V omenjenem poglavju so uporabljeni neuravnoteženi letni podatki 42 držav EU in držav nečlanic iz obdobja 2000–2017. V tem empiričnem modelu nadzorujemo določene bančne (velikost banke, bančno kapitalizacijo, dobičkonosnost) ter izbrane makroekonomske (denimo rast BDP-ja in realno obrestno mero) in institucionalne lastnosti (denimo podjetja, ki upravljajo s premoženjem). Vključene so različne zemljepisne regije. Izračunani so bili štirje posamezni podvzorci za banke v državah članicah EU, nečlanicah, v razvitih gospodarstvih in v gospodarstvih v vzponu. Uporabljena je rezervacija za izgubo pri posojilih kot alternativna spremenljivka kreditnega tveganja, namesto nedonosne terjatve, ki služi preverbi stabilnosti rezultatov. Rezultati v pričujočem poglavju razkrivajo statistično pomembno povezavo med nedonosnimi terjatvami in rastjo posojilne dejavnosti. Poleg tega razultati kažejo, da je večji učinek nedonosnih terjatev na posojilno dejavnost opaznejši pri večjih in bolje kapitaliziranih bankah. Nenazadnje pa preučimo tudi učinek podjetij, ki upravljajo s premoženjem, na rast posojilne dejavnosti. Izsledki raziskave kažejo, da je povezava med nedonosnimi terjatvami in podjetji za upravljanje s premoženjem statistično zanemarljiva in da ni dokazov za povezavo s posojilno rastjo.

V Poglavju 3 želimo nadgraditi že obstoječe poznavanje modelov za pridobivanje kreditnih točk, ki jih institucije uporabljajo za ovrednotenje kreditne sposobnosti kreditojemalca. To dosežemo z uporabo edinstvenih podatkov z ravni posojil države kandidatke za članstvo v EU. V primerjavi s tradicionalno kreditno analizo, osnovano predvsem na presoji bančnega uslužbenca, naprednejši/sodobnejši model za pridobivanje kreditnih točk sestoji iz širšega obsega lastnosti, z namenom boljšega ovrednotenja kreditojemalčeve kreditne sposobnosti. V nekaterih primerih ti modeli uporabljajo zgolj že odobrene kreditojemalce, medtem ko so zavrnjeni prosilci za kredit izločeni. Tu se poraja vprašanje, ali bo uporaba enakih orodij in tehnik za odobrene in zavrnjene prošnje za kredit dejansko izboljšala ovrednotenje modelov za pridobivanje kreditnih točk, medtem ko bodo vse pristranosti znotraj modela odstranjeni. Omenjeno poglavje vključuje tehniko zavrnilnega sklepa z namenom preučitve vpliva zavrnjenih kreditoprosilcev na model za pridobivanje kreditnih točk. Tu se opiramo na tri metode klasifikacije: logistično regresijo (LR), metodo drevesne klasifikacije (MDK) in radialno bazno funkcijo (RBF), z namenom ovrednotenja in primerjave razpršenih izvedb modelov in neizbirnih popravkov teženj. Rezultati kažejo, da zavrnilni sklep, povečevanje z razdeljevanjem, izboljšuje izvedbo modela za pridobivanje kreditnih točk, kar nakazuje, da je podatek o zavrnjenih kreditoprosilcih ključen. Izsledki kažejo, da logistična regresija doseže najvišji delež ustrezne klasifikacijske stopnje odobrenih in zavrnjenih prošenj za kredit – 72,6 % v primerjavi s 64,3 %, ko je govora zgolj o odobrenih prošnjah. To nakazuje na napredek v natančnosti v višini 8,3 %. Poleg tega najvišja operacijska krivulja naraste z 68,2 % na 79,8 %, s čimer je dosežen napredek v višini 11,6 %.

Ključne besede: bančna posojilna dejavnost, globalna finančna kriza, gospodarski razvoj, kazalec zaznavanja korupcije, klasifikacijsko in regresijsko drevo, logistična regresija, makroekonomske determinante, nedonosne terjatve, pridobivanje kreditnih točk, umetno nevronsko omrežje, zavrnilni sklep.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

Background of the Dissertation

Financial institutions play a crucial role in economic activity. The quality of bank loan portfolios is one of the biggest challenges the countries' banking systems face. After the global financial crisis of 2008, the quality of bank assets deteriorated mainly due to the increase of the level of non-performing loans (NPLs) and this led to a huge deterioration of the banks' financial positions bringing to light the fragility of the international banking system (Dimitrios, Helen, and Mike, 2016b; Ghosh, 2015; Louzis, Vouldis, and Metaxas, 2012). For instance, at the end of 2016, the NPL ratio rose to 47.0% for Cyprus, 37.0% for Greece, 17.5% for Italy, and around 12.7% for Portugal (ECB, Annual report, 2016), indicating that the problem of NPLs is important in several banking systems in the EU. Nkusu (2011), using annual data from 26 developed countries over the period 1998-2009, provides evidence that a 2.4 percentage point increase in NPLs is combined with a decline of credit in the private sector.

One of the sources of the NPL's rise, as it has been explained by many researchers, is the country's level of corruption and, more specifically, in the financial industry. For example, La Porta et al. (2003) using data from a sample of 17 Mexican banks, show that borrowers who have a connection with bank officials are 33% more likely to default and therefore have a 30% lower recovery rate comparing with the non-connected borrowers. Lízal and Kocenda (2001) maintain that corruption in the banking sector in the Czech Republic lead to an increase in the level of NPLs, which contributed towards the collapse of the banking system.

The level of NPLs in the Euro area increased during the crisis, reaching a low of 2.5% at the end of 2007 and rising to 7.7% at the end of 2013 (ECB, Annual report, 2016). This only declined somewhat to 6.7% by mid-2016 as a result of concerted policy actions taken in several of EU member countries. The pre-crisis double-digit credit growth plummeted and never fully recovered. Lowering lending standards in the face of rapid credit expansion may result in higher levels of NPLs in the future (Erdic and Abazi, 2014; Shahzad et al., 2019; Chavan and Gambacorta, 2019). Higher levels of NPLs bind bank capital, reduce bank net income, and restrict access to funding due to market participants' increased risk perception (Aiyar et al, 2015). The second relationship, which is particularly relevant in the aftermath of the global financial crisis of 2008-2010, is the focus of our study.

Credit institutions use credit scoring models to evaluate a customer's creditworthiness when they apply for a loan. These models play a significant role in identifying borrowers with a higher probability of default and, consequently, minimizing the default rate and the levels of NPLs. Updating scoring models using advanced methods and techniques is crucial for identifying potential default borrowers and protecting banks from loan defaults' undesirable effects.

Given the importance of the quality of banks' loans on financial stability and economic performance, this dissertation attempts to bridge the gaps in literature primarily from three perspectives:

1) by studying the relationship between corruption and non-performing loans and exploring various possible channels through which corruption may interact with non-performing loans;
2) by analyzing the effect size of non-performing loans on bank lending behaviour considering the banking systems of the EU and non-EU member countries;
3) by studying the influence of rejected borrowers on scoring models considering unique loan-level data on a commercial bank population from a candidate EU country.

Research Questions Addressed in the Dissertation

Chapter 1 of the dissertation focuses on the relationship between corruption and the level of non-performing loans and determines the channels through which corruption affects the level of NPLs. The following research questions are addressed in this chapter:

- *Research question 1.1:* How does corruption affect (positively or negatively) the level of non-performing loans (NPLs)?
- *Research question 1.2:* Is the link between corruption and bank size less pronounced for larger banks?
- *Research question 1.3:* Is the link between corruption and NPLs influenced by bank capital?
- *Research question 1.4:* Is the link between corruption and NPLs weaker in countries with a strong legal environment?
- *Research question 1.5:* Does corruption reduce the effectiveness of bank regulation in lowering the level of NPLs?
- *Research question 1.6:* Is the association between corruption and NPLs stronger in countries characterized by a high level of collectivism?

Chapter 2 of this dissertation investigates to what extent and under which conditions NPLs affect bank lending behaviour. Chapter 2 addresses the following research questions:

- *Research question 2.1:* How does the level of NPLs affect (positively or negatively) the bank lending behaviour?
- *Research question 2.2:* Is the relationship between NPL and loan growth influenced by bank size?
- *Research question 2.3:* Is the link between NPL and loan growth influenced by bank capital?

- *Research question 2.4:* Is the link between NPLs and loan growth less pronounced for banks in the presence of asset management companies?

Chapter 3 moves from the country-bank level to the firm-bank level for a selected commercial bank portfolio of a bank from a candidate EU country. Chapter 3 addresses the following research questions:

- *Research question 3.1:* Can the predictability of defaulted borrowers be improved by applying the reject inference technique in the credit scoring model?
- *Research question 3.2:* How do the classification techniques like logistic regression (LR), classification and regression tree (CART), and radial basis function (RBF) differ in terms of classification of defaulted borrowers?

Structure and Contents of the Dissertation

The doctoral dissertation is made up of three main chapters. The first chapter starts with an empirical analysis of the association between corruption and the level of non-performing loans and whether higher corruption is related to the elevated proportion of NPLs in the banking system. Section 1.3 is devoted to the theoretical framework and hypothesis development. Section 1.4 provides a description of the bank-level data in a global context, a description of the variables' respective definitions and sources, as well as the empirical model. Section 1.5 presents and discusses the empirical findings. Section 1.6 presents several robustness checks, and Section 1.7 gives the conclusions.

The second chapter starts by introducing the role of non-performing loans in bank lending behavior. In addition, it presents the main objective of this chapter and reports the results and conclusions of the study. Section 2.3 reviews the theoretical literature and presents hypothesis development. Section 2.4 describes the data, data sources, as well as the empirical model and variables. Section 2.5 includes empirical results. Section 2.6 presents robustness checks, and lastly, Section 2.7 concludes the chapter.

The third chapter evaluates the importance of using information from rejected borrowers when a comparative analysis of classification methods and credit scoring models is performed. It explains the role of the rejected inference approach in inferring the 'good' and 'bad' properties of those applicants who have not been accepted for a loan and reports the results and conclusions of the study. Section 3.2 discusses the theoretical background. Section 3.3 describes the data and variables. Section 3.4 outlines the research strategy. Section 3.5 discusses the results. Lastly, section 3.6 concludes the chapter.

The general discussion section presents an overall overview of this dissertation thesis. It starts by summarizing the main findings of the research studies, accompanied by the theoretical contributions of the dissertation to the non-performing loans research area.

1 CORRUPTION AND NON-PERFORMING LOANS[1]

1.1 Overview

This chapter empirically evaluates the impact of corruption on the level of non-performing loans (NPLs) using international bank-level data, spanning over the period 2000–2016 and across 140 countries. We find a positive and statistically significant relationship between corruption and NPLs. We also analyze the channels through which corruption affects NPLs. We find that the relationship between corruption and NPLs becomes more pronounced during and after the global financial crisis and is more pronounced for smaller banks. The association between corruption and NPLs is stronger in countries characterized by a high level of collectivism. The link between corruption and NPLs is higher where the legal environment is weak, and where economies are market-based compared to bank-based ones.

1.2 Introduction

The 2008–2010 global financial crisis pressured regulators, policymakers, and academics into a meticulous search for its drivers. In an extensive review of the evidence, Thakor (2018) concludes that insolvency risk was the main cause of the crisis. Bank balance-sheets deteriorated due to the financial market's related losses, but also due to increasing levels of non-performing loans (NPLs). In the U.S., the average proportion of NPLs per total gross loans increased from 1.4 percent in 2007 to 4.96 percent in 2009. In the EU, the level of NPLs increased from 2.4 percent in 2007 to 6.6 percent of total assets in 2016.[2] At the same time, several financial scandals caught public attention and called for further scrutiny of unsound practices in banking. In a 2015 survey, 47 percent of 1,200 financial services professionals in the U.K. and the U.S. claimed that it is necessary to engage in an illegal or unethical activity at least once to succeed and to gain an edge in the market.[3]

Excessive risk taking and failed management and board control functions were cited as causes of the failure of the U.K. bank HBOS that led to its acquisition by Lloyds and a government bailout (FCA and PRA, 2015). The regulator investigated the wider allegation of corruption following a criminal investigation in which two former HBOS bankers and four business associates were found guilty of corruption, money laundering, and fraud. In particular, the lead director of HBOS's impaired assets division was taking advantage of

---

[1] This chapter is co-authored with Matej Marinč. The authors would like to thank Iftekhar Hasan, Marko Košak, Igor Masten, Orfea Dhuci, Adalbert Winkler as our discussant, and the participants at INFINITI 2018 in Poznan, Poland and the 7th EBR Conference in Ljubljana for their valuable comments and suggestions. This chapter is published in *Economic and Business Review*.

[2] These data are available at: World Bank, IMF Financial Soundness Indicators, and ECB.

[3] University of Notre Dame's Mendoza College of Business and Labaton Sucharow LLP.http://www.labaton.com/en/about/press/Historic-Survey-of-Financial-Services-Professionals-Reveals-Widespread-Disregard-for-Ethics-Alarming-Use-of-Secrecy-Policies-to-Silence-Employees.cfm

small businesses by threatening to terminate loans unless a bribe was paid to the restructuring consultancy company led by his accomplices. This compelled HBOS to write off £266m in loans.[4] Other examples point to the anecdotal evidence between corruption and bad loans, including the arrest by the Indian Central Bureau of Investigation (CBI) in 2019 of eight senior executives from different financial institutions for taking bribes for a loan.[5] An older example is the bankruptcy of Hanbo Steel Industry Co. in South Korea where its founder was arrested for corruption offences, including bribing bankers and high-ranking politicians to keep granting loans to the second largest steelmaker in South Korea.[6]

The above examples indicate that the level of corruption can be intrinsically related to the proliferation of NPLs in banking. If corruption is widespread, corrupted bank officers might grant loans even to companies that do not fulfill loan requirements, with a subsequent decrease in loan portfolio quality. In a cross-country setting, the level of corruption in a particular country could be seen as a potential factor contributing to the surge of NPLs (Park, 2012).

To address this issue, we combine the bank financial data across 140 countries and 7,773 banks with a corruption index from Transparency International (2017) and several macroeconomic and institutional variables. We focus on the period 2000–2016, which comprises the pre-financial crisis, the financial crisis period, and the post-financial crisis period.

As a preliminary investigation, we plot the countries' mean values of the corruption index and the mean values of NPLs for the period 2000 to 2016 shown in Figure 1.1. Figure 1.1 indicates that there exists a positive relationship between the corruption index and the level of non-performing loans.

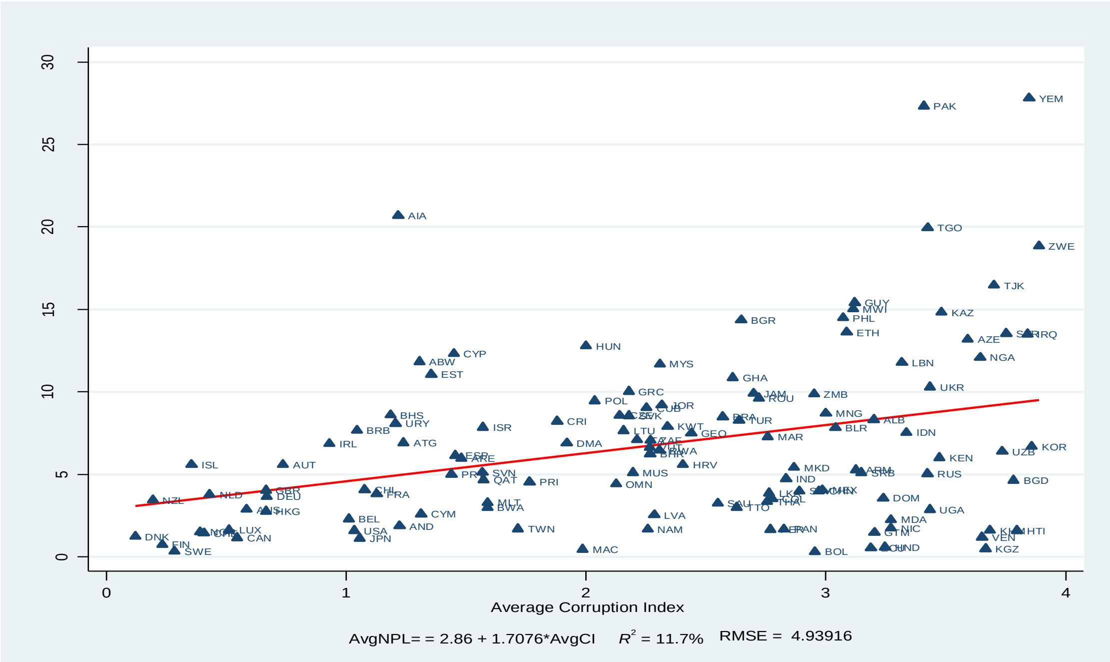
---

[4] Jane Croft, HBOS bankers bribed with sex parties guilty of corruption, *Financial Times*, 30 January, 2017. The article is available at: https://www.ft.com/content/4149dcd6-e70b-11e6-967b-c88452263daf

[5] Sangita Mehta, Bribe-for-loan scam: Time for more transparency, *The Economic Times*, updated 09 July 2019. The article is available at: https://economictimes.indiatimes.com/industry/banking/finance/banking/bribe-for-loan-scam-time-for-more-transparency/articleshow/7062623.cms

[6] Reuters, Hanbo Steel Founder Given 15 Years in Korean Scandal, *The New York Times,* 2 June, 1997. The article is available at: https://www.nytimes.com/1997/06/02/business/hanbo-steel-founder-given-15-years-in-korean scandal.html

Figure 1.1: Plot of the average corruption index and NPLs



Note. We use standard regression equation as an estimation method. The horizontal axis on the graph represents the average corruption index (AvgCI) and the vertical axis (AvgNPL) represents the average NPLs where the average is computed for different countries. *Source*: *Own work based on World Bank, and Fitch Connect data.*

We investigate whether higher corruption leads to lower loan quality measured by the level of NPLs. We find robust support that higher corruption is associated with higher levels of NPLs. This finding is statistically significant across several econometric specifications and robustness tests. The result is also economically significant. In particular, an increase in corruption of one standard deviation would lead to an expected increase in NPLs of 0.111 standard deviations.

Our study is closely related to the contribution by Goel and Hasan (2011), which shows that corruption is positively related to the level of NPLs. Whereas Goel and Hasan (2011) use cross-country data, we employ bank-level data in order to determine the channels through which corruption affects the level of NPLs.

When analyzing the channels through which corruption affects the level of NPLs, we find some evidence that the effect of corruption on NPLs is bigger for smaller banks than for larger banks. This is consistent with the explanation that smaller banks are less hierarchical and transfer bigger decision making powers to bank officers. An additional discretion then leads to a stronger link between the corruption index and NPLs.

We also find some evidence that the impact of corruption on NPLs is less pronounced for banks that are highly profitable. A potential reason for this result is that more profitable banks have more prudent lending policies and more effectively allocate funds and, consequently, can overcome high levels of corruption.

Our evidence shows that the impact of corruption on the level of NPLs is stronger for highly-collectivist countries. This indicates that the impact of corruption on the level of NPLs is stronger for countries where people tend to have less interdependent self-construal.

Our findings confirm the importance of the quality of the legal and institutional environment. We find that the well-functioning legal environment, evident in the strong rule of law in a given country, weakens the link between corruption and NPLs.

We investigate the effect of corruption on NPLs, before the crisis, during the crisis, and after the crisis. We show that the effect of corruption on the NPLs becomes more pronounced during the crisis period and in the post-crisis period. This is aligned with the view that during the financial crisis when the survival is at stake, the impact of corruption on bank lending is more pronounced, causing a deterioration of bank asset quality.

Hanousek, Shamshur, and Tresl (2019) show that corruption negatively affects firm efficiency but the negative effect is less pronounced for foreign controlled firms and for firms run by female CEOs. Wellalage, Locke, and Samujh (2018) find that corruption leads to increased credit constraints for small and medium enterprises in South Asia. They find that the effect of corruption varies between the male and female owners highlighting gender differences in the relationship between corruption and lending. Van Vu et al. (2018) find that various forms of corruption differently affect firm performance. Our findings complement their research by pointing to the mitigating role of firm size and the institutional environment contributing to the literature that discusses how ethical behavior is implemented in organizations (Batten, Loncarski, and Szilagyi, 2018; Batten, Lončarski, and Szilagyi, 2017).

Overall, our findings provide some relevant policy implications for the role of corruption as a potential factor that affects loan quality in the banking system. The results of our study indicate that measures taken against corruption have implications for financial stability.

This chapter is organized as follows. Section 1.3 provides the theoretical framework and hypothesis development. The data and research method are described in Section 1.4. Section 1.5 presents and discusses the results of the empirical analysis. Section 1.6 provides robustness checks. Section 1.7 concludes the article.

1.3 Theoretical framework and hypothesis development

The relationship between the legal environment and finance has gained prominence following the seminal article by La Porta et al. (1997) which analyzes how the quality of the legal and institutional framework affects the financial systems. Several articles find that corruption is negatively related to economic growth (see Mo, 2001; Lízal and Kocenda, 2001; Mauro, 1998) and that economic growth and firm failures are negatively related (Ghosh, 2015; Ali and Daly, 2010). Corruption can contribute to a reduction in growth, investments, and firm productivity in developed and developing countries (Méon and Weill, 2010; Mauro, 1995). Qi and Ongena (2019) show that firms engaged in bribery practices lose access to bank credit which impedes firm growth (see also Mohaddes, Raissi and Weber, 2017; Vithessonthi, 2016; Ghosh, 2015; Klein, 2013; Beck, Jakubik and Piloiu, 2013). Fan, Lin, and Treisman (2009) and Treisman (2000) show that highly developed countries have lower levels of corruption compared to developing countries. Corruption leads to weaker and less efficient banking systems (Chen, Jeon, Wang, and Wu, 2015; Park, 2012; Goel and Hasan, 2011).

In most countries that are greatly affected by corruption, even uncreditworthy firms could get a bank loan (for example, by bribing credit loans officers, Fungáčová, Kochanova and Weill, 2015),[7] which can subsequently lead to excessive corporate leverage (Jõeveer, 2013; Weill, 2011a) and loan delinquencies. Using cross-country data, Goel and Hasan (2011) find that countries with higher corruption are associated with a higher level of NPLs. Park (2012) finds that corruption aggravates the problems of bad loans in the banking sector. Similarly, findings in Bougatef (2015) confirm the positive link between corruption and NPLs in the Islamic banking system. Chen et al. (2015) analyze the impact of corruption on bank risk-taking behavior in 35 emerging economies during the period 2000-2012. They show that an increase in corruption is associated with more pronounced risk-taking behavior of banks.

Whereas the studies above provide an argument for the positive relationship between corruption and NPLs, it might be possible to construe an alternative explanation. If good borrowers do not receive the credit needed from banks or financial institutions due to market frictions, they might resort to corruptive practices. To avoid adverse selection issues, good firms may bribe to signal their quality to lenders, enabling good borrowers[8] to receive the credit needed and improving the efficiency of credit allocation. In this alternative explanation, corruption would reduce the level of NPLs. This leads to the following hypothesis:

Hypothesis 1.1: Corruption affects the level of NPLs.

Various determinants affect the quality of bank lending, as reflected in the level of NPLs. The internal determinants consist of bank size, management quality and efficiency, and bank capitalization (Dimitrios, Helen and Mike, 2016a,b; Louzis, Vouldis and Metaxas, 2012; Espinoza and Prasad, 2010; Hasan and Wall, 2004; Salas and Saurina, 2002; Berger and

[7] Chen, Liu, and Su (2013) find that bribery determines the access of private firms to bank credit in China (see also Weill (2011b)).

[8] Even though the firm fulfill the required conditions to obtain a loan, it might occur that a loan officer might require a bribe as an incentive to process the client's loan file, in the condition that his base salary is reduced due to volatile banks' operating income. Consequently, this bribery would increase the cost of the loan and the burden is borne solely by the borrower (see Beck, Demirguc-Kunt, Laeven and Maksimovic, 2006).

DeYoung, 1997). The external determinants consist of bank regulations and supervision (Barth, Caprio and Levine, 2004; Godlewski, 2004), law of enforcement (Boudriga, Boulila and Jellouli, 2009; Barth et al., 2004), and national culture (El Ghoul, Guedhami, Kwok and Zheng, 2016; Dheera-aumpon, 2019; Zheng, Ghoul, Guedhami and Kwok, 2013). Our aim is to analyze how internal and external determinants affect the link between corruption and NPLs.

The literature related to the organizational structure of banks emphasizes that large banks with more centralized and hierarchical structures can perform better when information can be hardened and easily transmitted across the hierarchical lines.[9] Whereas small banks have an advantage in soft information processing where discretion of a loan officer is of paramount importance (see Stein, 2002),[10] the problem is that discretion might be abused in environments with high levels of corruption. Skrastins and Vig (2015) show that more hierarchical organizations operated better in environments with high levels of corruption showing that hierarchy can restrain rent-seeking activities. This leads to the following hypothesis:

Hypothesis 1.2: The link between corruption and bank size is less pronounced for larger banks.

The related issue refers to the role of bank capital in the relationship between corruption and NPLs. One of the main roles of bank capital is to provide proper incentives to banks to internalize their risk taking strategy (see Admati and Hellwig, 2013). Thinking along these lines would posit that weakly capitalized banks have little incentive to engage in safe lending practices and would be prone to corruption. However, an alternative view is also possible. Weakly capitalized banks are exposed to intensive scrutiny from the bank supervisor. Anticipating heavy supervision, weakly capitalized banks could restrain corruption.

Hypothesis 1.3: The link between corruption and NPLs is influenced by bank capital.

The effect of corruption on NPLs is likely to be less pronounced in countries with a healthy legal environment and efficient regulatory systems (Danisman and Demirel, 2018). These countries are less likely to be adversely affected when the rules are better and more efficiently enforced. La Porta, Silanes, Shleifer, and Vishny (1998) and Levine (1998) show that countries with weaker legal structures to protect borrowers have a smaller number of performing banks in their economy. Moreover, Barth, Lin, Lin, and Song (2009) find that ownership of banks and firms, legal environment, and firm competition significantly reduce

---

[9] Larger banks have better-defined procedures and possess sophisticated capabilities for loan assessment (Hu, Li, and Chiu, 2004; Lis, Pagés, and Saurina, 2000).

[10] Canales and Nanda (2012) show that bank managers in a decentralized organizational structure are better at bank lending to small firms and firms with a plethora of soft information.

lending corruption. We analyze how the link between corruption and NPLs interacts with the quality of the legal and regulatory framework. This leads to the following hypothesis:

Hypothesis 1.4: The link between corruption and NPLs is weaker in countries with a strong legal environment.

Corruption alters the effectiveness of the banking system. Beck, Demirguc-Kunt, and Levine (2006) show that the most efficient strategy for reducing corruption in bank lending is to expose banks to private monitoring and discipline through the disclosure of accurate information. According to this view, banks would have to behave more cautiously and avoid bad practices in lending. In addition, countries with high-powered supervisory agencies tend to have lower levels of corruption (Beck, Demirguc-Kunt, and Levine, 2006; Demirguc-Kunt, Laeven, and Levine, 2004). Akins, Dou, and Ng (2017) find that timely loan loss recognition by banks reduces the level of corruption in lending. Awartani et al. (2016) show that better institution quality leads to longer maturity of corporate debt. We hypothesize that corruption also makes the regulatory framework less effective.

Hypothesis 1.5: Corruption reduces the effectiveness of bank regulation in lowering the level of NPLs.

The relationship between corruption and NPLs might be driven by the national cultural dimensions (i.e., individualism/collectivism, uncertainty avoidance, masculinity/femininity, and power distance) as construed by Hofstede (2001) and Hofstede (1983). In the extant literature, Zheng, Ghoul, Guedhami, and Kwok (2013) analyze the association between national culture and corruption in bank lending. They argue that compared to other cultural dimensions the role of national culture and collectivism in particular is an important factor affecting the corruption of bank officials. They confirm a positive and significant association between the level of collectivism and corruption of bank officials. El Ghoul, Guedhami, Kwok, and Zheng (2016) document that corruption is associated with collectivistic culture affecting bank lending, and Haider, Liu, Wang, and Zhang (2018) report that corruption impacts the relationship between financial constraints and firm performance. Triandis (2001) highlights individualism/collectivism as the most significant driver of cultural differences across countries.

Hypothesis 1.6: The association between corruption and NPLs is stronger in countries characterized by a high level of collectivism.

1.4 Data and methodology

1.4.1 Data description

The annual bank-level financial data are retrieved from the Fitch Connect database. We consider consolidated data.[11] Our final unbalanced panel data sample includes 7,773 banks from 140 countries over the period 2000–2016. We include in the database commercial, savings, and co-operative banks. Our bank and macroeconomic variables are 'winsorized' at 1% interval to mitigate the impact of extreme values and to exclude potential outliers.[12] We combine bank-level data, macroeconomic indicators, and institutional data from various sources. Macroeconomic data are obtained from the World Bank. Two indices of corruption are obtained from the Transparency International Corruption Perception Index and the World Bank. Table A.1 in Appendix A reports the definitions and the data sources of the variables used.

1.4.2 Empirical model

To analyze the effect of corruption on NPLs we estimate the following regression model.

$$\begin{aligned} NPL_{i,c,t} = \alpha &+ \beta NPL_{i,c,t-1} + \gamma CI_{c,t-1} + \delta Bank_{i,c,t-1} + \vartheta Macro_{c,t-1} \\ &+ \kappa Institutional_{c,t-1} + \mu Year_t \\ &+ \epsilon_{i,c,t} \end{aligned} \tag{1.1}$$

where the dependent variable is non-performing loans ratio $NPL_{i,c,t}$ measured by the loans and advances that are more than 90 days overdue divided by total loans for bank *i*, located in country *c*, at year *t*. We use $NPL_{i,c,t}$ as a proxy for banks' loan quality (Tarchouna, Jarraya and Bouri, 2017; Vithessonthi, 2016; Ghosh, 2015; Chaibi and Ftiti, 2015).

Our main explanatory variable is corruption index $CI_{c,t-1}$, which indicates the level of corruption in country *c*, in year *t-1*. We employ two alternative corruption indexes, the corruption perception index ($CPI_{c,t}$), obtained from the Transparency International Corruption Perception Index and the control of corruption of the World Bank. First, the score of the *CPI* ranges on a scale from 0 to 10. The higher the corruption in the country is, the lower the *CPI* score the country gets. For interpretation reasons, we define a new index Control of corruption, $CI_{c,t}$, where $CI_{c,t}=10-CPI_{c,t}$. Higher values of *CI* indicate higher levels of corruption. For example, the highest value of *CI* is 8.43 points for Bangladesh and the lowest is 0.61 points for Denmark. Second, we employ control of corruption (hereinafter CC) from the World Bank - Worldwide Governance Indicator. CC is widely used in related literature (Kaufmann, Kraay and Mastruzzi, 2010). The index ranges from the lowest value of -2.5 to the highest value of +2.5. The lower value represents the highest level of corruption

---

[11] For a standardized research on corruption, the inclusion of subsidiaries provides a better measure of the firm's propensity to corruption according to Pantzalis, Park, and Sutton (2008) and Zeume (2017). We have also performed the analysis on the unconsolidated data, obtaining qualitatively similar results.

[12] We removed negative values for non-performing loans, total assets, and loans to customer deposits.

and vice versa.[13] We follow Park (2012) and Goel and Hasan (2011) and compute WBCI as $WBCI_{c,t}$=(5-($CC_{c,t}$+2.5)) where $CC_{c,t}$ is control of corruption at time *t* four country *c*. This adjustment is employed for interpretation reasons such that higher values of WBCI are associated with lower values of control of corruption and higher values of corruption.[14] The highest value of WBCI was 3.98 points for Libya and the lowest for Denmark with 0.12 points. The average values of WBCI per each country are shown in Table A.2 in Appendix A. We include the following bank-specific variables as control variables. Bank $size_{i,c,t-1}$ is computed as the natural logarithm of total assets for bank *i*, located in country *c*, in year *t-1*.[15] $LTCD_{i,c,t-1}$ represents the ratio of total loans divided by total customer deposits for each bank *i*, located in country *c*, in year *t-1*.[16] $ROAA_{i,c,t-1}$ represents the ratio of the return on average assets for each bank *i*, located in country *c*, in year *t-1*.[17] The capitalization ratio, $Capitalization_{i,c,t-1}$, is the ratio of total equity divided by total assets for bank *i*, located in country *c*, in year *t-1*, as in Makri, Tsagkanos, and Bellas (2014), Klein (2013), and Louzis et al., (2012),[18] $LLP_{i,c,t}$ represents the ratio of the loan loss provisions divided by gross loans for each bank *i*, located in country *c*, in year *t*.[19]

We also include several macroeconomic indicators as controls. GDP $growth_{c,t-1}$ denotes the annual percentage growth rate of GDP in country *c*, at time *t-1*.[20] $UNEMP_{c,t-1}$ indicates the unemployment rate for a country *c*, at time *t-1*[21] and $GCF_{c,t-1}$ is computed as the growth

---

[13] For further detailed methodology on how the indicator is constructed, see Kaufmann et al. (2010).

[14] Whereas the corruption index described above presents the overall corruption in a country and is not specifically tight to the banking sector corruption, it is widely accepted that the corruption in the banking sector is highly correlated with the overall corruption in the economy; see Park (2012), Chen et al. (2015) and Goel and Hasan (2011).

[15] In the extant literature, the effect of bank size on the level of NPLs is ambiguous. On the one hand, Ranjan and Dhal (2003) and Salas and Saurina (2002) show that larger banks have more diversification possibilities which leads to lower NPLs. On the other hand, Chaibi and Ftiti (2015), Louzis et al. (2012), and Brei and Gadanecz (2012) argue that large banks may take excessive risks which can result in higher NPLs.

[16] The higher ratio of loans with respect to deposits indicates more aggressive lending practices of a bank resulting in easier loan granting and, therefore, a higher likelihood of establishing NPLs (Dimitrios, Helen, and Mike, 2016b). It indicates an increased risk appetite of banks with a potential for higher levels of non-performing loans.

[17] Return of average assets is used as a proxy for management's efficiency. We anticipate that the relationship between the return on average assets and non-performing loans is negative (see Vithessonthi, 2016; Dimitrios et al., 2016a).

[18] We anticipate that a higher level of capitalization pressures banks to avoid risky lending practices and subsequently negatively effects the level of NPLs. Makri, Tsagkanos and Bellas (2014), Salas and Saurina (2002), and Keeton and Morris (1987) show that the level of capital negatively affects the level of NPLs.

[19] We anticipate that the relationship between loan loss provisions and NPLs is positive since banks use higher levels of provisioning when they predict loan defaults (Chaibi and Ftiti, 2015; Boudriga, Boulila, and Jellouli, 2009).

[20] Love and Ariss (2014), Ghosh (2015), Ali and Daly (2010), and Treisman (2000) find that the GPD growth is negatively related to NPLs.

[21] A higher unemployment rate affects the borrowers' incomes and, hence, affects the delinquency rates (Al-Marhubi, 2000; Rinaldi and Sanchis-Arellano, 2006; Klein, 2013; Saha and Ben Ali, 2017; Dimitrios et al. 2016a).

capital formation as percentage of GDP in country *c,* at time *t-1*. $RIR_{c,t-1}$ presents the real interest rate in a country *c,* at time *t-1*.[22]

We also include several institutional and regulatory variables as controls. $RoL_{c,t-1}$ represents the rule of law in country *c* at time *t-1* and serves as a proxy for the quality of contract enforcement, property rights, the police, the courts, as well as the likelihood of crime and violence (Kaufmann et al., 2010). $Cap\_Str_{c,t-1}$ denotes capital stringency from Barth, Caprio, and Levine (2013) in country *c* at time *t-1*. $Year_t$ is the yearly dummy variable and is included in the model to control the time-specific effects (Petersen, 2009). $\varepsilon_{i,c,t}$ denotes the white noise error term. Hereinafter, we suppress the subscript in the variables.

To test our second, third, and sixth hypotheses, about the channels through which corruption might affect the level of non-performing loans, we estimate the following equation.

$$\begin{aligned} NPL_{i,c,t} = {} & \alpha + \beta NPL_{i,c,t-1} + \gamma CI_{c,t-1} + \delta Bank_{i,c,t-1} + \vartheta Macro_{c,t-1} \\ & + \propto CI * Bank\ Size_{i,c,t-1} + \tau CI * Capitalization_{i,c,t-1} \\ & + \omega CI * ROAA_{i,c,t-1} + \rho CI * CLT_{i,c} + \mu Year_t + \epsilon_{i,c,t} \end{aligned} \tag{1.2}$$

where the control variables are the same as in equation (1). CI*Bank Size denotes the interaction term between corruption and the natural logarithm of the total assets for bank *i*, located in country *c*, in year *t-1*. CI*Capitalization denotes the interaction term between corruption and the ratio of total equity divided by total assets for bank *i*, located in country *c*, in year *t-1*. CI*ROAA denotes the interaction term between corruption and return on average assets for each bank *i*, located in country *c*, in year *t-1*. CI*CLT denotes the interaction term between corruption and the level of collectivism (CLT) in a country *c*. To test our fourth and fifth hypotheses we estimate the following equation.

$$\begin{aligned} NPL_{i,c,t} = {} & \alpha + \gamma CI_{c,t-1} + \delta Bank_{i,c,t-1} + \vartheta Macro_{c,t-1} + \varphi RoL_{c,t-1} + \omega CI * RoL_{c,t-1} \\ & + \tau Cap_{Str_{c,t-1}} + \rho CI * Cap_{Str_{c,t-1}} + \mu Year_t \\ & + \epsilon_{i,c,t} \end{aligned} \tag{1.3}$$

where CI*RoL represents the interaction term between the corruption index and the rule of law, located in country *c*, in year *t-1*. Cap_Str, and the interaction term between the corruption index and a measure of capital stringency, located in country *c*, in year *t-1* and all other control variables are the same as in equation (1.1).

We apply different panel data estimation approaches. First, to account for unobserved heterogeneity across banks we setup the fixed effects and random effects (hereinafter, FE and RE) panel regression models with robust standard errors (Micco and Panizza, 2006; Carlson, Shanand Warusawitharana, 2013; Košak, Li, Lončarski and Marinc, 2015).[23]

[22] Chaibi and Ftiti (2015) and Castro (2013) find a positive influence of real interest rate on the NPLs.

[23] When using fixed and random effects panel regression methods, we drop the contemporaneous NPL variable from the regression equation in (1). We use the bank fixed effects model with robust standard errors, clustered at the country level.

Second, to take into consideration the time presence of NPLs, we include $NPL_{i,c,t-1}$ as the independent variable of NPL of bank *i*, located in country *c*, in year *t-1* and use the system GMM estimation to achieve consistent and efficient estimates (Blundell and Bond, 1998). In Table 1.1, we present the descriptive statistics, including the mean, standard deviation, and minimum and maximum values.

In the first part of the table, we present bank-specific variables. We observe a large disparity in the NPL rate between banks with a minimum of 0 percent and a maximum of 29.3 percent and the standard deviation of the NPL rate is 3.196 percent. We observe a high variation in the mean ratio of loans to customer deposits of 80.61 percent. Regarding management efficiency, ROAA ranges from a minimum of -6.0 percent to a maximum of 8.4 with a mean value of 1.11 percent during the period 2000-2016, whereas the standard deviation equals 1.35 percent, indicating that, on average, banks in our data are profitable, but some banks have financial difficulties. The mean value of capitalization and NPL equal to 11.66 percent and 1.81 percent, respectively. Loan loss provisions divided by total loans, LLP, has a mean of 0.49 percent.

In the second part of Table 1.1, we present the results for the macroeconomic indicators: growth rate of GDP, unemployment ratio, gross capital formation, and real interest rates. The mean annual percentage growth rate of GDP is 2.04 percent. We note that some countries have a negative GDP growth rate with a minimum value of -5.62 percent. The mean unemployment rate is 6.35 percent, the mean gross capital formation is 22.1 percent, and the mean real interest rate is 2.91 percent. In the third panel, we present the corruption index measured by the World Bank, Transparency International, and the International Country Risk Guide. In our sample, the mean of WBCI is 1.11, the mean of CI is 2.74, and the mean of ICRG is 1.95. The last column of Table 1.1 presents the maximum of a corruption index per each corruption measure. Higher values of these indexes indicate that various countries are suffering from severe corruption issues. The standard deviation of corruption index measured by the World Bank is near to the one documented in Chen et al. (2015) with a mean value of 0.704. In the fourth part of Table 1.1, we include institutional variables. The mean values of overall capital stringency, rule of law, regulatory quality, and the level of collectivism index are 4.98, 0.99, 1.09, and 11.2, respectively.

In Table 1.2, we present the correlations between our variables. In line with our empirical model we find that NPL is positively and significantly correlated with corruption, but negatively and significantly correlated with bank size and return on average assets (at the 1 percent significance level). The correlation between NPL and capitalization is positive and significant. Furthermore, the variable NPL is negatively and significantly correlated with GDP growth and gross capital formation as a percentage of GDP, but positively and statistically significantly correlated with the unemployment rate.

Table 1.1: Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Dependent variable					
NPL (%)	105,708	1.809	3.195	0.000	29.300
Bank specific variables					
Bank size (log)	109,178	5.422	1.905	2.229	14.319
Capitalization (%)	109,178	11.662	6.889	1.480	74.240
LTCD	108,156	80.606	46.020	5.900	864.97
ROAA (%)	108,429	1.107	1.349	-6.000	8.370
LLPGL (%)	106,431	0.490	0.955	-1.500	8.570
Crisis	109,178	0.182	0.386	0.000	1.000
Macroeconomic indicators					
GDP growth (%)	108,854	2.037	1.829	-5.619	10.636
Unemployment (%)	107,958	6.346	1.953	2.500	19.900
GCF (%)	108,697	21.074	2.513	14.428	42.894
RIR (%)	106,804	2.907	2.324	-9.633	29.120
Corruption perception index					
CI	108,653	2.736	0.862	0.000	9.600
WBCI	103,079	1.108	0.447	0.030	4.222
ICRG	108,522	1.953	0.569	0.000	5.500
Institutional controls					
CLT	106,462	11.202	10.819	9.000	94.000
Overall capital stringency	19,147	4.980	1.649	1.000	7.000
RoL	103,084	0.994	0.419	0.400	4.678
Economies classification	105,392	0.022	0.148	0.000	1.000

Note. The sample covers the period from 2000 to 2016. The bank variables are NPL - non-performing loans; Bank size - natural logarithm of total assets; Capitalization - total equity as a proportion of total assets; LTCD - total loans as a proportion of total customer deposits; ROAA - return on average assets; LLP - loan loss provision divided by gross loans. The macroeconomic indicators are: GFC - global financial crisis; GDP growth ratio; Unemployment ratio; GCF - gross capital formation as a percentage of GDP; RIR - real interest rate; CI - corruption perception index from Transparency International; WBCI - corruption perception index from World Bank; CI-ICRG - International Country Risk Guide; CLT - is an index of collectivism by using Hofstede data; Cap_Str - capital stringency; RoL - rule of law; Economies classification - is a dummy variable equal to 1 if the financial system is bank-based financial system or 0 if the financial system is market based financial system. *Source*: *Own work.*

Table 1.2: Correlation matrix

	NPL	CI	Bank Size	LTCD	ROAA	Capitaliz.	GDP growth	Unemp.	GCF	RIR
NPL	1									
CI	0.333^{***}	1								
Bank Size	0.221^{***}	0.415^{***}	1							
LTCD	0.0672^{***}	0.130^{***}	0.230^{***}	1						
ROAA	-0.240^{***}	0.104^{***}	0.142^{***}	0.0513^{***}	1					
Capitaliz.	0.00699^{*}	0.106^{***}	-0.0909^{***}	0.0608^{***}	0.0891^{***}	1				
GDP growth	-0.0417^{***}	0.263^{***}	0.112^{***}	0.0264^{***}	0.198^{***}	0.0390^{***}	1			
Unemp.	0.325^{***}	0.245^{***}	0.129^{***}	0.0112^{***}	-0.116^{***}	-0.00766^{*}	-0.294^{***}	1		
GCF	-0.150^{***}	0.227^{***}	0.146^{***}	0.0819^{***}	0.165^{***}	0.0399^{***}	0.543^{***}	-0.524^{***}	1	
RIR	-0.0171^{***}	0.112^{***}	0.00821^{**}	0.0669^{***}	0.0568^{***}	0.0497^{***}	0.0530^{***}	-0.283^{***}	0.319^{***}	1

Note. The sample covers the period from 2000 to 2016. The table reports the correlation matrix between the key variables, which are used in the model. NPL - non-performing loans; CI - corruption perception index from Transparency International; Bank size - natural logarithm of total assets; LTCD - total loans as a proportion of total customer deposits; ROAA - return on average assets; Capitalization - total equity as a proportion of total assets; GDP growth ratio; Unemp - unemployment ratio; GCF - gross capital formation as a percentage of GDP; RIR - real interest rate. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

1.5 Empirical results

1.5.1 Baseline results

We begin by estimating the regression in (1.1) using the fixed effects regression model (see Table 1.3).[24] The estimated regression coefficient between the level of corruption and NPL is positive and statistically significant. This supports the Hypothesis 1.1 and is consistent with the previous findings that corruption negatively affects loan quality measured by the NPL (Park, 2012; Goel and Hasan, 2011; Bougatef, 2015). The result is also economically significant. That is, an increase in corruption for one standard deviation would lead to an expected increase in NPL for 0.111 standard deviations (where 0.111 is computed as the estimated regression coefficient, associated with the level of corruption CI, 0.413, multiplied by the standard deviation of CI, 0.862, and divided by the standard deviation of NPL, 3.195).

The regression coefficients of control variables have the anticipated signs. The bank size is significantly and positively related to NPL. This implies that larger banks take larger risks potentially due to the associated government guarantees and their too-big-to-fail status, which leads to higher levels of NPLs. Our empirical evidence is consistent with the finding of Chaibi and Ftiti (2015) and Louzis et al. (2012). The return on average assets is negatively associated with NPL consistent with explanation and findings of Dimitrios et al. (2016a), Dimitrios, Helen, and Mike (2016b), and Baselga-Pascual et al. (2015). Additionally, capitalization is negatively and significantly related to NPL. The GDP growth is negatively but not significantly related to NPL whereas the unemployment rate is positively related to NPL. The GCF is negative but not significantly related to NPL. We also find that real interest rate is positively and statistically significantly related to NPL.

We use Hausman test (Hausman, 1978) to choose between the FE or RE estimator. From the chi-squared test statistics we conclude that the random effects model is rejected (with Prob>chi2(9)=0.0000<0.05). Hence, the FEs model should be employed rather than the RE model. In addition, we test for the time-fixed effects in the data and we confirm that year dummies should be included in the model (the Breuch – Pagan LM test shows the existence of the panel effect in the data for random effects). We estimate standard errors that are robust to heteroskedasticity.

The fixed effects model is prone to the endogeneity problem. The alternative explanation for our findings might be that a high level of NPLs creates a fertile ground for corruption to spread within the financial institutions. For example, corrupt loan officers can exclude the defaulted borrowers from the aberration of penalties (Boerner and Hainz, 2004). The presence of endogeneity in the model can lead to inconsistent and biased estimators. To address the endogeneity problem, we use the dynamic panel-data setup to account for potential endogeneity of our dependent variable $NPL_{i,c,t}$.

[24] To mitigate the reverse causality, we use one-year lag of each of the bank and macroeconomic variables.

Table 1.3: The relationship between corruption index and NPLs: the total sample with different estimation methods

Dependent variable: NPL	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	-0.620	-1.535		-0.392	-1.310	
	(-0.38)	(-1.63)		(-0.26)	(-1.41)	
Corruption level						
CI	0.413***	0.634***	0.992**			
	(3.21)	(6.76)	(2.52)			
WBCI				0.369***	0.771***	4.101***
				(3.78)	(2.88)	(2.71)
Bank-specific variables						
Lagged NPL			0.788***			0.746***
			(27.10)			(15.03)
Bank Size	0.186***	0.197***	0.105***	0.185***	0.213***	0.0507
	(6.28)	(13.17)	(2.59)	(5.64)	(11.13)	(0.77)
LTCD	0.00248	0.00270*	0.00339**	0.00259*	0.00293**	0.00240
	(1.62)	(1.89)	(2.19)	(1.76)	(2.12)	(1.23)
ROAA	-0.429***	-0.430***	-0.251***	-0.435***	-0.436***	-0.505**
	(-37.77)	(-36.24)	(-4.45)	(-35.63)	(-33.54)	(-2.08)
Capitalization	-0.0352***	-0.0317***	-0.0381**	-0.0351***	-0.0306***	0.0274
	(-21.60)	(-9.90)	(-2.03)	(-22.47)	(-9.76)	(0.43)
Macroeconomic indicators						
GDP growth	0.000390	0.000753	-0.616	0.00714	0.0167*	-1.121**
	(0.03)	(0.06)	(-1.32)	(0.50)	(1.74)	(-2.19)
Unemployment	0.219***	0.223***	1.101**	0.242***	0.253***	0.740**
	(5.26)	(6.71)	(2.40)	(5.09)	(7.11)	(2.05)
GCF	-0.0535	-0.0412	0.000751	-0.0354	-0.0193	0.282
	(-1.07)	(-1.17)	(0.00)	(-0.69)	(-0.50)	(0.87)
RIR	0.0979***	0.101***	-0.902***	0.0852***	0.0784***	-0.634***
	(3.57)	(4.28)	(-3.85)	(3.29)	(3.52)	(-3.22)
Coefficient Estimates	FE	RE	GMM	FE	RE	GMM
No. Obs.	102,905	102,905	102,420	97,314	97,314	96,850
Dummies Year	Yes	Yes	Yes	Yes	Yes	Yes
R-squared within	0.2004	0.1998		0.1983	0.1973	
Hansen J statistic (p-value)			21.08 (0.576)			22.13 (0.452)
AB test AR(2) (p-value)			-0.10 (0.923)			-0.67 (0.501)

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation methods are FE, RE and twostep Arellano-Bond system GMM. CI-corruption perception index; WBCI-corruption perception index from World Bank; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of country. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

We include the lagged variable $\mathrm{NPL}_{\mathrm{i,c,t-1}}$ as an independent variable, as in regression in (1.1) (see Baum, 2006; Chaibi and Ftiti, 2015; Tarchouna, Jarraya, and Bouri, 2017; Arellano and Bond, 1991). Furthermore, Blundell and Bond (1998) suggest relying on system generalized method of moments (GMM) estimator for dynamic panel data. This method provides unbiased and efficient estimates. As endogenous instrument, we consider the lagged dependent variable $NPL_{i,c,t-1}$. All other variables are treated as exogenous instruments. Moreover, we use the set of lags from 2 to 5 to mitigate the over-identification problem of endogenous instruments.

Furthermore, to account for appropriate instruments, the relevance condition states that the excluded instruments have to fulfill two conditions. They have to be correlated with the dependent variable (in our case NPL) and they have to be uncorrelated with the residuals (Baum, Schaffer, and Stillman, 2003). We also test for the overidentification restrictions by using Hansen's J statistic as well as the presence of the first order autocorrelation AR(1) and the second order autocorrelation AR(2) of the residuals in the dynamic model. We confirm the validity of the instruments chosen and no presence of the second order autocorrelation. This indicates that our GMM estimate is consistent. The regression coefficient of the first lag of NPLs is positive and significant (see column 3 of Table 1.3). The positive relationship indicates that NPLs are expected to increase when they have moved up the year before, as in Chaibi and Ftiti (2015).

The results reported in columns 4 and 6 of Table 1.3 indicate that the level of corruption is significantly and positively related to the NPLs even when we use an alternative corruption perception index from the World Bank. This confirms Hypothesis 1.1.

1.5.2 Channels through which corruption affects NPLs

Having identified the main determinants of NPLs, we analyze the interaction terms between the corruption index and bank-specific variables in order to investigate the channels through which corruption impacts the level of NPLs. We introduce four interaction terms: corruption and bank size, corruption and bank capitalization, corruption and ROAA, and corruption and collectivism. The results are presented in Table 1.4. As before, the level of corruption is positively and statistically significantly related to the level of NPLs.[25]

The interaction term between the corruption index and bank size[26] is negatively and significantly (for GMM estimation but not for FE estimation) related to the level of NPLs. We find that a one-unit increase in CI*Bank size leads to a reduction of NPLs by 0.0191 units. This indicates that the effect of corruption on the level of NPLs is less pronounced for

[25] The p-values of Hansen's J statistics in Table 1.4 imply that the instruments satisfy the orthogonality conditions for all GMM regressions.

[26] Skrastins and Vig (2015) approach to test the influence of bank size on the relation between corruption and NPLs. In their paper, the organizational hierarchy is measured by the managerial levels. We use the logarithm of total bank assets to test the impact.

larger banks than for smaller banks, providing some support for Hypothesis 1.2. The explanation may derive from the more hierarchical structures of larger banks that preclude corruptive bank practices, resulting in a weaker link between corruption and NPLs. The interaction term between corruption and capitalization has no significant effect on the level of NPLs. This indicates that the third hypothesis is not confirmed.

The interaction term between corruption and ROAA is negatively and significantly (for GMM estimation but not for FE estimation) related to the level of NPLs. We show that a one unit increase in CI*ROAA leads to a decrease in NPLs of 0.276 units. This implies that the effect of corruption on the level of NPLs is less pronounced for banks with high profitability. A potential explanation might be that as most profitable banks are less concerned with their revenue creation, they do not engage themselves in risky lending, causing a cutback in the level of NPLs.

We also analyze the association between corruption and NPLs by using a specific dimension of culture, namely collectivism, to control for heterogeneity in national cultures across countries.[27] The interaction term between corruption and the level of collectivism (CLT) is positively and significantly (for FE estimation but not for GMM estimation) related to the level of NPLs. We find that a one-unit increase in CI*CLT leads to an increase in NPLs of 0.0433 units. This indicates that the impact of corruption on the level of NPLs is stronger in countries where people tend to have less interdependent self-construal. This provides some support for our Hypothesis 1.6.

[27] Hofstede (2001) and Hofstede (1983) constructed four cultural dimensions: individualism/collectivism (IDV), uncertainty avoidance (UAI), masculinity/femininity (MAS), and power distance (PDI).

Table 1.4: Channels through which corruption affects NPL

Dependent variable NPL	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Intercept	-0.920		-0.711		-0.747		-2.714***	
	(-0.63)		(-0.41)		(-0.43)		(-4.81)	
Corruption level								
CI	0.517**	7.707***	0.444**	1.098***	0.441***	1.964***	0.295***	1.270
	(2.41)	(3.16)	(2.18)	(2.89)	(2.89)	(3.79)	(4.62)	(0.79)
Bank-specific variables								
Lagged NPL		0.752***		0.792***		0.774***		0.752***
		(25.32)		(30.27)		(26.69)		(35.18)
Bank Size	0.229**	2.008***	0.185***	0.0820**	0.183***	0.146**	0.178***	0.101**
	(2.30)	(2.95)	(6.37)	(2.13)	(6.34)	(2.15)	(7.40)	(2.38)
LTCD	0.00246	0.00353**	0.00250*	0.00422**	0.00247	0.00261	0.00169**	0.00235**
	(1.62)	(2.31)	(1.67)	(2.45)	(1.60)	(1.54)	(2.52)	(2.39)
ROAA	-0.429***	-0.295***	-0.428***	-0.247***	-0.362***	0.540*	-0.424***	-0.249***
	(-39.10)	(-4.96)	(-39.76)	(-4.83)	(-3.87)	(1.76)	(-22.34)	(-5.98)
Capitalization	-0.0352***	-0.0215	-0.0284	0.0570	-0.0348***	-0.0640**	-0.0343***	-0.0404**
	(-21.45)	(-1.11)	(-1.26)	(0.36)	(-25.40)	(-2.12)	(-9.35)	(-2.29)
Macroeconomic indicators								
GDP growth	-0.000325	-1.234**	0.000249	-0.525	-0.000294	-0.564	0.00174	-0.267
	(-0.02)	(-2.16)	(0.02)	(-1.09)	(-0.02)	(-0.74)	(0.24)	(-0.40)
Unemployment	0.219***	1.088**	0.220***	1.076**	0.220***	1.131**	0.236***	0.810
	(5.11)	(2.39)	(5.24)	(2.38)	(5.22)	(2.02)	(13.42)	(1.27)
GCF	-0.0522	0.457	-0.0532	0.171	-0.0510	-0.492	-0.00684	0.0615
	(-1.07)	(1.24)	(-1.06)	(0.53)	(-0.99)	(-0.74)	(-0.30)	(0.15)
RIR	0.0976***	-1.023***	0.0979***	-0.712**	0.0968***	-1.253***	0.109***	-0.610***
	(3.52)	(-3.61)	(3.56)	(-1.99)	(3.60)	(-2.76)	(14.10)	(-2.64)
Interaction terms								
CI*Bank Size	-0.0156	-0.698***						
	(-0.38)	(-2.77)						
CI*Capitalization			-0.00255	-0.0313				
			(-0.29)	(-0.61)				
CI*ROAA					-0.0231	-0.276**		

					(-0.63)	(-2.41)		
CI*CLT							0.0433***	-0.00924
							(6.06)	(-0.50)
Coefficient Estimates	FE	GMM	FE	GMM	FE	GMM	FE	GMM
No. Obs.	102,905	102,420	102,905	102,420	102,905	102,420	102,013	101,610
Dummies Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hansen J statistic (p-value)		16.38 (0.839)		30.17 (0.145)		21.10 (0.575)		48.51 (0.064)
AB test AR(2) (p-value)		0.09 (0.928)		0.29 (0.768)		-0.80 (0.425)		0.14 (0.887)

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation methods are pooled FE and twostep Arellano-Bond system GMM.CI-corruption perception index from Transparency International; Bank size - natural logarithm of bank assets; LTCD - total loans as a proportion of total customer deposits; ROAA - return on average assets. Capitalization - total equity as a proportion of total assets. CLT - the level of collectivism. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF - gross capital formation as a percentage of GDP; RIR - real interest rate. The regressions include Year dummies. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of country. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

1.5.3 Corruption and NPLs during and after the global financial crisis

We now evaluate how the global financial crisis affects the relationship between the corruption index and NPLs. We add the interaction terms of the corruption index with dummy variables that equal to 1 during the three sub-periods: before the crisis (2000-2007), during the crisis (2008-2010), and post-crisis (2011-2016), and 0 otherwise (see Allen et al., 2017).

Table 1.5: Corruption and the NPLs before, during, and after the global financial crisis

Dependent variable: NPL	**(1)**	**(2)**	**(3)**
Intercept	1.211	-0.332	0.776
	(0.89)	(-0.23)	(0.52)
Corruption level			
CI	0.292***	0.230**	0.299***
	(3.19)	(2.28)	(2.98)
Bank-specific variables			
Bank Size	0.123***	0.173***	0.145***
	(3.46)	(6.74)	(3.85)
LTCD	0.00248*	0.00232	0.00282**
	(1.73)	(1.48)	(2.00)
ROAA	-0.432***	-0.427***	-0.436***
	(-31.88)	(-36.24)	(-31.37)
Capitalization	-0.0385***	-0.0365***	-0.0360***
	(-21.22)	(-21.04)	(-19.35)
Macroeconomic indicators			
GDP growth	0.0116	0.0656	-0.130***
	(0.89)	(1.49)	(-3.57)
Unemployment	0.183***	0.260***	0.0903***
	(5.50)	(4.52)	(4.47)
GCF	-0.0606	-0.0609	-0.0453
	(-1.09)	(-1.28)	(-0.75)
RIR	0.0672***	0.0936***	0.0734***
	(5.77)	(4.17)	(5.00)
CI*BEFORE GFC	-0.344***		
	(-9.80)		
CI*DURING GFC		0.174**	
		(2.53)	
CI*AFTER GFC			0.380***
			(5.24)
Coefficient Estimates	FE	FE	FE
No. Obs.	102,905	102,905	102,905
Dummies Year	Yes	Yes	Yes

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation method is pooled FE. CI-corruption perception index; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of country. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

The estimation results in Table 1.5 show that the interaction term of corruption with before the crisis, CI*BEFORE GFC is negatively and significantly associated with NPLs. This suggests that before the global financial crisis, the effect of corruption on the NPLs is less pronounced. The interaction term of corruption index with the crisis dummy, CI*DURING GFC, is positively and significantly associated with NPLs. This implies that during the global financial crisis the effect of corruption on the NPLs is more pronounced. The interaction term of the corruption index with the post-crisis dummy, CI*AFTER GFC, is positively and significantly related to the NPLs, indicating that in the post-crisis period the positive effect of corruption on NPLs is more pronounced.

1.5.4 Corruption and NPLs in bank-based and market-based financial systems

We continue our investigation by exploring whether the impact of corruption on the NPLs is different in bank-based and market-based economies (see Demirguc-Kunt and Levine, 1999).

Table 1.6: Corruption and NPLs in bank-based and market-based economies

Dependent variable: NPL	**(1) Bank-based economy**	**(2) Market-based economy**
Intercept	12.16**	-2.339*
	(2.09)	(-1.83)
Corruption level		
CI	-0.652	0.434***
	(-1.29)	(5.00)
Bank-specific variables		
Bank Size	0.120	0.203***
	(0.60)	(11.08)
LTCD	0.00406	0.000917
	(1.21)	(1.64)
ROAA	-0.717**	-0.415***
	(-2.51)	(-26.85)
Capitalization	0.0708	-0.0335***
	(1.03)	(-14.55)
Macroeconomic indicators		
GDP growth	-0.0979	0.00827
	(-1.52)	(0.71)
Unemployment	-0.200	0.273***
	(-0.95)	(12.00)
GCF	-0.140	0.000791
	(-1.35)	(0.02)
RIR	0.0289	0.126***
	(0.49)	(11.92)
Coefficient Estimates	FE	FE
No. Obs.	788	100,901
Dummies Year	Yes	Yes

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation method is pooled FE. CI-corruption perception index; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of countries. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

Table 1.6 shows that corruption is positively related to NPL in the subsample of market-based economies while in the subsample of bank-based economies the association is statistically insignificant. Interestingly, the regression coefficient of the corruption index is larger in the market-based economies (0.434) than in the bank-based economies indicating that the effect of corruption on the level of NPLs is pronounced in the market-based economies.

1.5.5 The legal environment and the effectiveness of the regulatory framework

The relationship between the level of corruption and NPLs may be affected by cross-country differences, especially in the legal environment, and might impact the efficiency of the regulatory framework. In Table 1.7, we add the interaction term between the corruption index and the rule of law, CI*RoL.

The interaction term between the corruption index and the rule of law is negatively and significantly related to the NPLs (see Table 1.7). This confirms the view that in countries where the laws are better enforced, the impact of corruption on the NPLs becomes less pronounced. This finding is consistent with Hypothesis 1.4.

We also evaluate the impact of corruption on the effectiveness of capital regulation by including the level of capital stringency, Cap_Str, and the interaction term between the corruption index and a measure of capital stringency, CI*Cap_Str in the regression model. Capital stringency has no significant effect on the level of NPLs. This indicates that Hypothesis 1.5, is not confirmed.

Table 1.7: Corruption and NPLs and the quality of the legal and regulatory environment

Dependent variable: NPL	**(1)**	**(2)**
Intercept	0.111	6.124***
	(0.16)	(2.91)
Corruption level		
CI	1.261***	-0.572
	(13.96)	(-0.74)
Bank-specific variables		
Bank Size	0.153***	-0.241***
	(5.85)	(-4.55)
LTCD	0.00245***	-0.00113
	(2.98)	(-1.47)
ROAA	-0.441***	-0.500***
	(-21.81)	(-34.93)
Capitalization	-0.0350***	-0.0647***
	(-8.96)	(-23.73)
Macroeconomic indicators		
GDP growth	-0.0292***	-0.257
	(-2.64)	(-1.29)
Unemployment	0.202***	0.144

	(10.93)	(0.94)
GCF	-0.0430*	-0.153*
	(-1.67)	(-1.72)
RIR	0.00713	0.134
	(0.42)	(0.73)
Legal system quality		
RoL	-1.829***	
	(-7.40)	
CI*RoL	-0.288***	
	(-6.81)	
Regulatory effectiveness		
Cap_Str		0.370
		(1.20)
CI*Cap_Str		0.0147
		(0.17)
Coefficient Estimates	FE	FE
No. Obs.	91,657	18,187
Dummies Year	Yes	Yes

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation method is pooled FE. CI-corruption perception index; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of country. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

1.5.6 High-corrupt vs. low-corrupt group of countries

We now want to confirm that the relationship between the corruption index and NPL is positive for countries with high and low levels of corruption and countries with high and low levels of NPL. We use the average value as a reference unit to divide countries in the high-corrupt group (with the corruption index above the mean value of 2.736 and NPL above the mean value of 1.809) and the low-corrupt group (with the corruption index below the mean value of 2.736 and NPL below the mean value of 1.809). Table 1.8 shows that corruption is positively related to NPLs for countries in the high-corrupt group and for countries in the low-corrupt group. Furthermore, we find a significant and positive effect of corruption on the level of NPLs in both, high and low NPL countries. This implies that country-level corruption affects positively the level of NPL in the bank.

Table 1.8: Corruption and NPLs in low and high corrupt group of countries

Dependent variable: NPL	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	4.995***	0.296	3.576*	-0.368
	(3.03)	(0.93)	(1.81)	(-1.11)
Corruption index				
CI	0.403**	0.598***	0.554**	0.0928***
	(2.23)	(5.50)	(2.23)	(5.39)
Bank-specific variables				
Bank Size	0.0522	0.214***	0.0790	0.0571***
	(0.44)	(18.34)	(0.90)	(10.54)
LTCD	0.00679***	0.000632***	0.00604***	0.000243***
	(3.62)	(3.33)	(3.37)	(3.39)
ROAA	-0.614***	-0.342***	-0.527***	0.00862***
	(-14.11)	(-63.99)	(-13.74)	(11.03)
Capitalization	-0.0763***	-0.0235***	-0.0355***	-0.00139***
	(-6.33)	(-20.99)	(-3.37)	(-6.01)
Macroeconomic indicators				
GDP growth	-0.126**	-0.103***	-0.0425	-0.00182
	(-2.36)	(-12.17)	(-1.40)	(-0.55)
Unemployment	-0.000532	0.274***	0.168***	0.0394***
	(-0.01)	(95.08)	(3.37)	(5.55)
GCF	-0.101**	-0.129***	-0.113*	-0.00165
	(-2.31)	(-33.78)	(-1.89)	(-0.17)
RIR	0.0324	0.136***	0.0699**	0.0255***
	(1.36)	(11.36)	(2.42)	(5.50)
Coefficient Estimates	FE	FE	FE	FE
No. Obs.	20,452	82,453	27,908	74,997
Group	Low corrupt group	High corrupt group	Low NPL group	High NPL group

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation method is pooled FE. CI-corruption perception index; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of countries. ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively. *Source*: *Own work.*

1.6 Robustness checks

1.6.1 Alternative corruption index and loan quality indicator

As a robustness check, we also use the value of the corruption index based on the International Country Risk Guide $ICRG_{c,t}$. The score for the $ICRG_{c,t}$ ranges on a scale from 0 (least corrupt) to 6 (most corrupt). Table 1.9 provides the regression results using the alternative corruption indexes, the corruption index from *TI*, *CI*, and *ICRG*. The estimated results remain widely unchanged and corroborate our main finding that corruption is positively related to the NPLs.

Table 1.9 also provides the results of the alternative model in which we use loan loss provisions, LLP, as a loan quality indicator instead of NPL (following De Haan and Van Oordt, 2018; Tarchouna et al. 2017; Chaibi and Ftiti, 2015; Boudriga et al. 2009).

The results in Table 1.9 confirm a positive and significant relationship between the level of corruption index and LLP. This implies that corruption aggravates loan quality, which results in increased levels of LLPs. The effects of other variables on loan loss provisions remain similar to the basic model with NPLs as a dependent variable.

Table 1.9: Alternative measures of corruption and loan quality

Dependent variable:	**(1)** **NPL**	**(2)** **NPL**	**(3)** **LLP**	**(4)** **LLP**
Intercept	-0.186		-1.126***	
	(-0.13)		(-2.83)	
Corruption level				
CI			0.104***	0.637**
			(5.34)	(2.27)
ICRG	0.206***	2.415***		
	(5.97)	(3.46)		
Bank-specific variables				
Lagged NPL		0.798***		
		(24.94)		
Lagged LLP				0.182*
				(1.84)
Bank Size	0.212***	0.174***	0.0600***	0.0659***
	(6.84)	(4.08)	(4.05)	(3.83)
LTCD	0.00258*	0.00209*	0.00132***	0.00140***
	(1.90)	(1.81)	(3.62)	(3.16)
ROAA	-0.439***	0.0171	-0.113***	-0.0339
	(-37.13)	(0.37)	(-9.77)	(-0.99)
Capitalization	-0.0341***	0.0238***	0.0103***	0.00236
	(-22.49)	(3.44)	(17.53)	(0.45)
Macroeconomic indicators				
GDP growth	0.0208	-0.955**	-0.0941***	-0.640**
	(1.27)	(-2.45)	(-17.34)	(-2.01)
Unemployment	0.229***	0.300	0.0324***	0.00156
	(5.44)	(1.21)	(2.87)	(0.01)
GCF	-0.0508	0.0122	0.0349***	-0.0280
	(-1.02)	(0.06)	(3.67)	(-0.17)
RIR	0.0813***	-0.798***	0.0498***	-0.145
	(3.69)	(-4.11)	(4.13)	(-1.59)
Coefficient Estimates	FE	GMM	FE	GMM
No. Obs.	102,810	102,328	103,012	96,234
Hansen J statistic (p-value)		27.94 (0.218)		39.37 (0.144)
AB test AR(2) (p-value)		0.20 (0.840)		0.35 (0.724)

Note. The sample covers the period from 2000 to 2016. The dependent variables are is non-performing loans (NPL) and loan loss provision (LLP). The estimation methods are FE and twostep Arellano-Bond system GMM. CI-corruption perception index; ICRG-corruption index from the International Country Risk Guide; Bank size-natural logarithm of bank assets; LTCD-total loans as a proportion of total customer deposits; ROAA-return on average assets; Capitalization-total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF-gross capital formation as a percentage of GDP; RIR-real interest rate. The regressions include Year dummies. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of countries. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

1.6.2 Subsamples of commercial banks

We tested the robustness of our results based on the subsample of commercial banks (Table 1.10). The results remain similar to the ones in the basic model. The level of corruption is positively and significantly related to the level of non-performing loans in both estimation methods (FE and GMM) for bank lending. This confirms the view that corruption is widespread across different bank specializations.

Table 1.10: Corruption and NPLs for the subsample of commercial banks

Dependent variable: NPL	**(1)**	**(2)**
Intercept	-0.683	-0.408
	(-0.42)	(-0.27)
Corruption level		
CI	0.468***	
	(2.99)	
WBCI		0.390***
		(3.58)
Bank-specific variables		
Bank Size	0.199***	0.198***
	(6.16)	(5.50)
LTCD	0.00375**	0.00384**
	(2.16)	(2.30)
ROAA	-0.432***	-0.439***
	(-34.02)	(-31.92)
Capitalization	-0.0383***	-0.0384***
	(-19.53)	(-20.83)
Macroeconomic indicators		
GDP growth	-0.00552	0.00244
	(-0.34)	(0.15)
Unemployment	0.207***	0.233***
	(5.05)	(4.82)
GCF	-0.0582	-0.0371
	(-1.21)	(-0.74)
RIR	0.0993***	0.0858***
	(3.33)	(3.05)
Coefficient Estimates	FE	FE
No. Obs.	88,386	83,620
Dummies Year	Yes	Yes
R-squared within	0.1938	0.1919

Note. The sample covers the period from 2000 to 2016. The dependent variable is non-performing loans (NPL). The estimation method is FE. CI-corruption perception index from Transparency International; WBCI-corruption perception index from World Bank-World Development Indicators; Bank size - natural logarithm of bank assets; LTCD - total loans as a proportion of total customer deposits; ROAA - return on average assets. Capitalization - total equity as a proportion of total assets. The macroeconomic indicators are: GDP growth ratio; Unemployment ratio; GCF - gross capital formation as a percentage of GDP; RIR - real interest rate. The regressions include Year dummies. Robust-standard errors in parenthesis are clustered at the level of country. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

1.7 Conclusion

To unveil the relationship between corruption and loan quality, we combine an unbalanced panel data of 109,178 bank level observations from 140 countries over the period from 2000 to 2016 with macroeconomic and regulatory indicators and the indexes of corruption. Our main finding is that corruption is positively and significantly related to the level of NPLs indicating that corruption leads to reduced loan quality in the banking system.

We find that the relationship between corruption and NPLs is weaker for larger banks. Potential explanations reside in hierarchical practices of large banks and in regulatory scrutiny. Larger banks may be less exposed to corruption due to their highly hierarchical structures. Hierarchical structures give little discretion to loan officers, successfully preventing corruptive behavior.

Furthermore, we find evidence that countries with a high level of collectivism are associated with a higher level of NPLs. Given that collectivist cultures are characterized by group and social cooperation, our findings support the hypothesis that countries with a high level of collectivism are associated with higher levels of NPLs. We also analyze the role of the legal environment on the impact of corruption on loan quality. We find that a stronger rule of law makes the relationship between corruption and NPLs less pronounced.

2 NON-PERFORMING LOANS AND BANK LENDING BEHAVIOR[28]

2.1 Overview

This chapter empirically investigates whether the level of non-performing loans (NPLs) affects the bank lending behavior using the bank-level data across 42 countries, over a period spanning from 2000 to 2017. We find a negative and statistically significant relationship between NPL and bank loan growth. This impact was pronounced in EU, non-EU, advanced and emerging countries subsamples. We also examine the channels through which NPLs affect loan growth. Our results show that the association between NPL and loan growth is more pronounced for larger and well-capitalized banks. We find no evidence that justifies the negative association between NPLs and loan growth for banks operating in countries where asset management companies are present. In addition, our results are robust with respect to alternative measure of credit risk.

2.2 Introduction

The global financial crisis of 2008 exposed deteriorating asset quality of European banks as a possible constraint on financial intermediation and further expansion of credit. In an extensive review of the evidence, Aiyar et al. (2015) point out that many countries in the Southern parts of the Euro-area, as well as Eastern and South-eastern Europe experienced high levels of non-performing loans (NPLs), which impaired bank lending through profitability, capital, and funding channels.

The level of NPLs in the Euro area increased substantially during the crisis, from a low of 2.5% at end-2007 to as high as 7.7% at end-2013. This only declined somewhat to 6.7% by mid-2016 as a result of concerted policy actions taken in a number of EU member countries, notably on portfolio segments targeting and on- and off- balance sheet exposures (ECB, Annual report, 2016).[29] At the same time, double-digit pre-crisis credit growth plummeted and never fully recovered, with loan growth rates only surpassing 2.3 percent in December 2015 and rising to 4.7 percent in December 2016. These co-movements beg the question: to what extent and under which conditions do NPLs affect bank lending behavior?

Conceptually, the relationship between NPLs and bank lending runs in both directions. On the one hand, lowering lending standards associated with rapid credit expansion may lead to greater levels of NPLs in the future (Erdic and Abazi, 2014; Shahzad et al., 2019; Chavan

---

[28] This chapter is co-authored with Matej Marinč and Vasja Rant. The authors would like to thank Marko Košak, Igor Masten and Orfea Dhuci, and the participants at 9th EBR Conference in Ljubljana for their valuable comments and suggestions.

[29] https://www.ecb.europa.eu/pub/pdf/annrep/ar2016en.pdf. The main contributing countries on the level of NPL are: Cyprus a NPL level 47.0 percent, Greece 37.0 percent, Italy 17.5 percent and Portugal 12.7 percent.

and Gambacorta, 2019). Koudstaal and Wijnbergen (2012) found that banks with more troubled loan portfolios have the incitement to excessive risk-taking behavior in a study of U.S. banks from 1993 to 2010. Furthermore, Dell'Ariccia and Marquez (2006) further mention the interaction between the loan market's informational structure and bank lending volume, backing a positive association between credit growth and risky loans at the aggregated level. Klein (2013), along with Keeton (1999), shows that rapid loan growth leads to increase in loan losses. On the other hand, higher levels of NPLs tie down bank capital, reduce bank net income and narrow access to funding due to higher perception of risk by market participants (Aiyar et al, 2015). Ibrahim and Rizvi (2018) found that higher credit risk has lower credit growth, regardless of whether the banks are Islamic or conventional. These factors may in turn constrain future lending. The focus of our study is on this second relationship, which is particularly relevant in the aftermath of the last financial crisis.

Specifically, we empirically investigate whether a higher level of NPLs leads to a decrease in lending activity, using annual bank financial data across EU and non-EU member countries for 6,434 banks in the 2000 to 2017 period. Our results show a robust support that a higher level of NPLs is associated with a reduction in the level of lending. This finding is statistically significant across several econometric specifications and robustness test. The result is also economically significant. Specifically, an increase in the level of NPLs for one standard deviation is associated with an expected decrease in lending for 0.096 standard deviations.

Our study is closely related to the study by Cheisa and Mansilla-Fernández (2019), which investigates the impact of NPLs on the cost of capital, lending, and liquidity supply. Their results revealed that the cost of capital acts as a transmission channel for the negative effect of NPLs on lending supply and liquidity creation (see also Cucinelli, 2016; Vo, 2018; Chavan and Gambacorta, 2019). Our contribution is to look at additional factors through which NPLs may affect bank lending behavior, similar to the approach by Thornton and Di Tommaso (2020). Their study found that bank capital and profitability mitigate the negative effects of NPLs on credit expansion. On top of bank capitalization, we identify bank size and the presence of asset management companies, mandated with restructuring NPLs, as additional factors, conditioning the effects of NPLs on bank lending. Furthermore, we also expand the geographic and time dimension of research. Whereas earlier studies only used data for Euro zone banks, we include annual bank balance sheet data from European and non-European member countries over the period 2000-2017 and distinguish the effects of NPLs in different time periods (before, during and after the crisis).

We find that the negative association between NPLs and bank loan growth is stronger for larger and well-capitalized banks. A potential explanation could be that the effects of high levels of NPLs might increase the cost of capital, which might cause lowering bank lending activity.

We report no statistical evidence that asset companies' presence influences the link between non-performing loans and loan growth in a country. In addition, we show that the effect of NPL on lending activity is less pronounced during the pre-crisis period. From a policy perspective, these results point to facilitating roles of bank resolution tools as well as strengthened capital regulation in mitigating the effects of NPL overhang on the provision of credit.

This chapter is structured as follows. Section 2.3 presents the theoretical framework and hypothesis development. Section 2.4 discusses the data and variables used in our empirical analysis. Section 2.5 describes the model and method of estimation. Section 2.6 presents the empirical results, whereas Section 2.7 summarizes our findings and sets out our conclusions.

2.3 Theoretical Framework and Hypothesis Development

Our paper builds on and contributes to the banking literature studying the relationship between bank risk and lending behavior, in particular loan growth. Several papers highlighted this relationship, with the level of NPLs often used as a proxy for credit portfolio quality.[30] In a cross-country study, Louhichi and Boujelbene (2017) showed that the level of NPLs could be seen as a significant factor negatively affecting bank lending activity in addition to the relationship between bank capital and lending (see also Jeong and Jung, 2013). In a study of the Japanese banking system, Vithessonthi (2016) emphasized that the relationship between credit growth and NPLs may be time varying: positive before the financial crisis and negative afterwards. These results could imply that a lower quality of bank assets discourages loan growth whereas a higher quality encourages it. Vo (2018) found that bank lending behavior significantly depends on bank specific characteristics and macroeconomic factors in a study of Vietnamese commercial banks from 2006 to 2015. The effects of bank risk, measured by credit risk provisions, were insignificant in this study, however. Cucinell (2015) documented that credit risk, measured by NPLs and loan loss provisions, negatively impacts bank lending behavior using data from Italian listed and non-listed banks during the 2007 to 2013 period. Focusing on the lending behavior of Italian banks before and during the recent financial crisis, Cucinelli (2016) also showed that a majority of banks that grew faster before the financial crisis, subsequently faced higher levels of NPLs and implemented deeper cuts in lending activity during the global financial crisis. This again points to a negative relationship between NPLs and lending after the onset of the financial crisis. Using a panel dataset of banks in 18 Westerns European countries, Meriläinen (2016) found that decrease in lending growth due to the global financial crisis is also contingent upon the type of bank ownership. In particular, stakeholder oriented banks

[30] The level of non-performing loans is used as a proxy for credit portfolio quality (see Kuzucu and Kuzucu, 2019; Gulati, Goswami, and Kumar, 2019; Kabir, Worthington, and Gupta, 2015; Dimitrios, Helen, and Mike, 2016a; Klein, 2013; Ghosh, 2015; Louzis, Vouldis, and Metaxas, 2012; Salas and Saurina, 2002; among others).

(i.e. cooperatives and publically owned banks) experienced less pronounced swings in lending growth.

Whereas the studies above provide an argument for the negative relationship between NPLs and loan growth, it might be possible to construe an alternative explanation. For example, Louhichi and Boujelbene (2017) found a positive relationship between NPLs and loan growth during the financial crisis period in a subsample of conventional (as opposed to Islamic) banks. A potential reason could be that the tendency of banks to undertake risk and expand the loan portfolio increases in the case of higher levels of NPLs. In a similar fashion, Eisdorfer (2008) argued that due to a risk-shifting strategy, financial institutions facing financial contraction are predisposed to increase investment in risky projects or lower quality borrowers. Finally, Chavan and Gambacorta (2019) studied the procyclical behavior of the NPL ratio in the emerging economy of India. Their results revealed that a one-percentage point increase in loan growth is associated with an increase in the NPL ratio of 4,1%, with response being higher during economic expansions. The reviewed articles led us to the following hypothesis:

Hypothesis 2.1: Non-performing loans affect the bank lending behavior.

Bank size is considered another important determinant of bank lending decisions (Berger and Udell, 2006; among others). Several empirical articles provide different results regarding whether small and large banks react differently in low and high bank riskiness environment. On the one hand, there is evidence that bank size is negatively associated with credit risk and credit growth. Salas and Saurina (2002) found a negative relationship between bank size and bank credit risk for commercial banks in Spain using data on commercial and savings bank over the period 1985 to 1997. They attributed their results to the fact that larger commercial banks in Spain are substantially more geographically diversified relative to the savings banks. Chouchène, Ftiti, and Khiari (2017) found that bank size has a negative impact on bank lending using 85 French banks from 2005 to 2010 period. This is in line with Schnabl (2012) findings that an increase in the total asset by 10% causes a 0.1% decrease in bank lending. In a study of the effects of government ownership on NPLs, Hu, Li, and Chiu (2004) reported similar results, showing that bank size is negatively related to credit risk among forty Taiwanese commercial banks during the 1996 to 1999 period. Laeven, Ratnovski, and Tong (2016) showed that systemic risk is lower and standalone bank returns are higher in better capitalized banks, with this effect being especially pronounced for large banks (see also Rajan and Dhal, 2003). Also, Košak et al., (2015) documented that larger banks experienced lower credit growth rates than smaller banks during the years before the financial crisis (see also Peek and Rosengren, 1995). In a similar vein, Vo (2018) found that smaller banks in Vietnam were riskier during the financial crisis and generally experienced higher loan growth rates. This led the author to conclude that smaller banks may even increase lending in a time of crisis to compensate for the decrease in profitability. On the other hand, Stein (2002) argued that small banks have a comparative advantage in lending using soft information, whereas large banks dominate in lending using hard information. However, large and complex banks may rely on soft information in their lending decisions about small and medium sized enterprises when such information can be processed through technical expertise. Similarly, Berger and Udell (2006) pointed out that conclusions, which emphasize the disadvantage of large banks in lending to small businesses, may be misleading because they fail to take into account the complexity of the actual lending process, including diverse lending technologies and the overall financial structure. This literature narrative points to a possible positive relationship between bank size and lending. The effects of bank size on lending behavior are ambiguous (either positive or negative). This could mean that lending behavior is affected by the quality of the loan portfolio. This leads us to the following hypothesis:

Hypothesis 2.2: The relationship between NPL and loan growth is influenced by bank size.

The third strand of research pertains to the role of bank capital in determining bank lending behavior. The main role of bank capital is to help banks cover any losses and insulate them from the propagation of financial shocks and potential insolvency (Košak et al., 2015). From this point of view, more capitalized banks could expand their lending and experience rapid loan growth compared to less capitalized loans. Using data from the Reports of Income and Condition in the period 2001 to 2011, Carlson, Shan and Warusawitharana (2013) found that the positive association between capital ratios and bank lending growth is stronger for banks that experienced loan contraction as opposed to the banks that experienced loan growth. Focusing on 125 countries over the period 1998-2010, Deli and Hasan (2017) showed that in general, capital stringency has a negative effect on loan growth. However, this effect is less pronounced for well-capitalized banks and is completely offset for banks with an equity capital ratio equal to 11%. Similarly, Abdul-Karim et al. (2014) analyzed Islamic and conventional banks in 14 Organization of Islamic Conference (OIC) countries over the 1999-2009 period, and found that when banks are better capitalized, they are not constrained by regulatory capital adequacy ratio (CAR) requirements, and their lending and borrowing capacity are less impacted by changes in the level of capital. Gambacorta and Shin (2018) found that a 1-percentage point increase in the equity-to-total assets ratio is associated with a 0,6 percentage point increase in lending growth per year. These findings indicate that a larger capital base reduces the banks' financial constraint, allowing them to grant more loans to the economy (see also Cantú, Claessens and Gambacorta, 2019) and provide the grounds for our third hypothesis:

Hypothesis 2.3: The link between NPL and loan growth is influenced by bank capital.

Asset management companies (AMCs) are considered one of the main recovery and resolution tools to deal with problematic loan portfolios of distressed banks (Gandrud and Hallerberg, 2014). The objective of AMCs is to separate non-performing and healthy assets. Mandated with acquiring and managing banks' non-performing assets, AMCs have played a valuable role in bank rehabilitation in the euro area after the crisis (Lehmann, 2017). Lehmann (2017) argued that a large proportion of NPLs ties up banks' capital and discourages credit supply. Marinč et al. (2014) showed that the level of stringency of capital regulation varies substantially across the EU countries. Gandrud and Hallerberg (2014) showed that banks burdened by distressed and delinquent assets are unable to further supply loans to the economy. In this context, the purpose of AMCs is to stabilize the banking system, reduce the pressure on bank balance sheets, and enable the banks to restart financial intermediation. Based on the abovementioned arguments, we can expect AMCs to be a moderating factor in the relationship between credit risk and bank lending behavior, which leads us to our last hypothesis:

Hypothesis 2.4: The link between non-performing loans and loan growth is less pronounced for the banks in the presence of asset management companies.

2.4 Data and Methodology

2.4.1 Data Description

We conduct our analysis using annual bank-level data retrieved from the Fitch Connect database. The study period ranges from 2000 to 2017. Our sample includes commercial, savings and co-operative banks from the EU and non-EU member states.[31] We only use data on non-consolidated bank financial statements to consider the behavior of individual bank subsidiaries. We include active and inactive banks for which data are available for at least five consecutive years. We remove negative values for non-performing loans, total assets, and loans to customer deposits and 'winsorize' the remaining dataset at 1% interval to eliminate extreme values. Based on this initial screening, our final sample is an unbalanced panel[32] composed of 6,434 banks with 79,098 observations in 42 countries.[33] We augment bank-level data with macroeconomic indicators from various sources. Macroeconomic data are obtained from the World Bank (World Development Indicators and International Financial Statistics).[34]

Table 2.1 provides the descriptive statistics of the variables included in our analysis. First, we present results for the bank-specific variables. The average annual loan growth rate (ΔlogGL) in our sample is 6,4%, ranging from -74,4% to 105%. The mean share of impaired loans (NPL) is 6,378% with a standard deviation 8,531%, ranging from 0% to 50,25%

[31] Non-European Union member countries included in our sample are: Albania, Belarus, Bosnia and Herzegovina, Iceland, Israel, Kosovo, Macedonia, Moldova, Montenegro, Norway, Serbia, Switzerland, Turkey, and Ukraine. For the purpose of our analysis we treat these countries as one region.

[32] The panel is unbalanced because we do not have the data for all banks and for all years. All explanatory variables are lagged by one period to avoid endogeneity bias (see Cantero-Saiz, Sanfilippo-Azofra, Torre-Olmo and López-Gutiérrez, 2014).

[33] See Table B.3, in the Appendix B for a detailed information about the sample used in this study.

[34] These data are available at: http://databank.worldbank.org/data/source/world-development-indicators

percent. The average logarithm of total assets (Bank Size) is 7.139, ranging from 0.095 to 19.889. Its substantial variation underscores the importance of bank size differences in the banking system. Bank capitalization ratio has a mean of 11,698% with a standard deviation of 13,279%, the deposits to assets ratio has a mean 63,590%, whereas the average loan loss provision ratio is 0.801%. The average net interest margin (NIM) is 2,416% and the mean return on average equity (ROAE) is 9,084%, ranging from -0,06% to 9,29% and from -49,53% to 63,37%, respectively.

Second, we include two country-level indicators, namely GDP growth and the real interest rate (RIR), to control for the changes in macroeconomic conditions during the business cycle. The averages for GDP growth and RIR in our sample are 1.547% and 3.049%, respectively. Furthermore, we note that some countries have negative GDP growth with a minimum value of -5.69%, which is reflective of the fact that our sample period includes the global recession after the financial crisis. Finally, the averages of the bank type dummy variables denote the shares of each bank type (commercial, cooperative, and savings bank) in the total sample.

In Table 2.2, we present the correlations between our key variables. These correlations are generally in line with the underlying economic theory. Furthermore, correlations among our independent variables (all variables, except ΔlogGL) are below 0.8, implying no presence of multicollinearity among the variables in our base model (O'Brien, 2007). We recognize that NPL is significantly negatively correlated with loan growth, RIR, deposits to total assets ratio, and the return on average equity. The correlations between loan growth, bank size, and GDP growth are positive and highly significant. Capitalization is significantly positively correlated with the loan growth rate.

Table 2.1: Descriptive statistics

Variables	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
Dependent variable					
ΔlogGL (%)	68,927	6.362	20.807	-74.4	105
Bank specific variables					
Bank Size	79,060	7.139	2.435	0.095	19.889
NPL (%)	32,554	6.378	8.531	0	50.250
Capitalization (%)	79,069	11.698	13.279	0.920	87.110
Deposits/Assets (%)	75,274	63.590	21.774	0.220	92.250
ROAE (%)	74,153	9.084	13.533	-49.530	63.370
LLP (%)	68,441	0.801	1.791	-3.130	12.540
NIM (%)	73,732	2.416	1.372	-0.060	9.290
Crisis	79,098	0.166	0.372	0	1
Macroeconomic indicators					
GDP growth (%)	79,049	1.547	2.354	-5.619	8.240
RIR (%)	21,780	3.049	3.033	-12.283	11.800
Bank type variables					
Commercial dummy	78,944	0.363	0.481	0	1
Co-operative dummy	78,944	0.382	0.486	0	1
Savings dummy	78,944	0.255	0.436	0	1

Note. The sample covers the period from 2000 to 2017. The bank variables are the loan growth rate (ΔlogGL); Bank size - natural logarithm of total assets; NPL - non-performing loans; Capitalization - total equity as a proportion of total assets; Deposits/Assets- total deposits as a proportion of total assets; ROAE - return on average equity; LLP - loan loss provision divided by gross loans; NIM - net interest margin; Crisis - dummy variable for the global financial crisis. The macroeconomic indicators are: GDP growth ratio; RIR - real interest rate. Source: Fitch Connect and World Bank-World Development Indicators. The definitions of the variables are provided in Appendix B. *Source*: *Own work.*

Table 2.2: Correlation matrix among key variables

	ΔlogGL	**NPL**	**Bank Size**	**Deposits/ Assets**	**Capitaliz.**	**ROAE**	**NIM**	**GDP growth**	**RIR**
ΔlogGL	1.00								
NPL	-0.14***	1.00							
Bank Size	0.06***	0.06***	1.00						
Deposits/Assets	-0.01*	-0.14***	-0.12***	1.00					
Capitaliz.	0.08***	0.23***	-0.07***	-0.36***	1.00				
ROAE	0.06***	-0.37***	0.02***	0.04***	-0.07***	1.00			
NIM	0.14***	0.16***	-0.06***	0.06***	0.17***	0.03***	1.00		
GDP growth	0.11***	-0.11***	0.05***	0.02***	0.03***	0.14***	0.07***	1.00	
RIR	-0.05***	0.03***	-0.06***	0.06***	-0.09***	-0.03***	-0.01	-0.03***	1.00

Note. The sample covers the period from 2000 to 2017. The table reports the correlation matrix between the key variables which are used in the model. ΔlogGL - growth of gross loans; NPL - non-performing loans; Bank size - natural logarithm of total assets; Depos/Assets- total deposits as a proportion of total assets; Capitalization - total equity as a proportion of total assets; ROAE - return on average equity; NIM - net interest margin. The macroeconomic indicators are: GDP growth ratio; RIR - real interest rate. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

2.4.2 Empirical Model

Our baseline model investigates the impact of non-performing loans on loan growth, controlling for the effects of bank-specific variables and country-level indicators. In line with the recent literature on bank lending determinants in panel data studies (e.g. see Salas and Saurina, 2002; Košak et al., 2015), our model is specified as follows:

$$\Delta logLG_{i,j,t} = \alpha_i + \rho \Delta logLG_{i,j,t-1} + \beta NPL_{i,j,t-1} + \gamma Bank_{i,j,t-1} + \delta Macro_{j,t-1} + \partial BankType_{i,j,t} + \theta Year_t + \varepsilon_{i,j,t} \quad (2.1)$$

where: $\alpha_i$is the intercept, whereas ρ, β, γ, δ, ∂, and θ are the coefficient vectors. Our dependent variable, $\Delta logLG_{i,j,t}$, is the loan growth rate for bank *i*, located in country *j*, in year *t* (see Sanfilippo-Azofra et al., 2018; Košak et al., 2015; Laeven and Majnoni 2003; Gambacorta and Mistrulli, 2004). Our main explanatory variable, $NPL_{i,j,t-1}$, represents the non-performing loans ratio, measured by total impaired loans over total gross loans for bank *i*, located in country *j*, in year *t*. In line with Louzis et al., (2012), Tarchouna, Jarraya, and Bouri (2017), Vithessonthi (2016), and Ghosh (2015), we use the NPL ratio as a measure of credit portfolio quality. An increase in the value of the ratio represents a worsening quality of the loan portfolio. We expect the NPL ratio to have a negative effect on loan growth. We also include the one year lagged dependent variable ($\Delta logLG_{i,j,t-1}$) in our model to capture the persistence of loan growth rates.

We account for several bank-specific control variables in the $Bank_{i,j,t-1}$matrix. We use the natural logarithm of total assets for bank *i*, located in country *j*, in year *t* at the end of each year to capture $Bank\ size_{i,j,t}$. The expected relationship between bank size and lending is ambiguous and could be either positive (see Thornton and Di Tommaso, 2020; Abdul-Karim et al., 2014; Hau, Langfield, and Marques-Ibanez, 2013) or negative (see Fang et al., 2020; Vo, 2018; Košak et al., 2015; Puri, Rocholl, and Steffen, 2011) . $Capitalization_{i,j,t}$ is the ratio of equity over total assets for bank *i*, located in country *j*, in year *t*. According to Foos et al., (2010) and Kishan and Opiela (2006) banks that are more solvent play an important role in lending. Thus, a positive relationship is expected. $Deposits/Assets_{i,j,t}$is the ratio of total customer deposits over total assets for bank *i*, located in country *j*, in year. $ROAE_{i,j,t}$ is the return on average equity for each bank *i*, located in country *j*, in year *t*. We use ROAE as a proxy for management efficiency (Bonin, Hasan, and Wachtel, 2005) and expect its relationship with loan growth to be positive (Iwanicz-Drozdowska and Witkowski, 2016).[35] $NIM_{i,j,t}$ stands for the net interest margin for bank *i*, located in country *j*, in year *t*.

[35] We use ROAE due to the observed increase in the equity capital of banks on the global market (see Iwanicz-Drozdowska and Witkowski, 2016; Athanasoglou, Brissimis, and Delis, 2008; among others).

We use NIM as a measure of core bank profitability (Iwanicz-Drozdowska and Witkowski, 2016).

Additionally, we check for the impact of macroeconomic factors in the $Macro_{i,j,t-1}$matrix, which includes two country-level indicators. $GDP\ growth_{j,t}$ denotes the annual percentage growth rate of GDP in country *j*, in year *t* and serves as a proxy for credit demand as mentioned in Klein (2013) and Gambacorta and Mistrulli (2004). We expect a positive relationship between loan growth and GDP growth. $RIR_{j,t}$ is the real interest rate in a country *j*, at time *t*.

Finally, we account for different bank specializations in the $BankType_{i,j,t}$ matrix. The dummy variables included in this matrix distinguish commercial, co-operative and savings banks basted on ownership and organizational characteristics of banks (commercial dummy$_i$, co-operative dummy$_i$, and savings dummy$_i$); $Year_t$ denotes yearly dummy variables to control for unobserved time specific effects; $\varepsilon_{i,j,t}$ represents the error term.[36]

We setup two data estimation approaches. First, we use the fixed and random effects estimators with robust standard errors to account for unobserved heterogeneity across banks (Micco and Panizza, 2006; Carlson, Shan, and Warusawitharana, 2013). Second, we use system generalized method of moments (GMM) estimator for dynamic panel data (see Arellano and Bover, 1995; Blundell and Bond, 1998) to resolve endogeneity issues and to obtain consistent and efficient estimates. We use lags of 1 up to 3 as instruments for our explanatory variables in order to address the problem of endogeneity. We use the Hansen *J* specification test to test the overall validity of the instruments which confirms the consistency of our estimation results. We test the assumption of serially uncorrelated errors for the first order autocorrelation AR(1) and the second order autocorrelation AR(2) (see Arellano and Bond, 1991; Roodman, 2006; Baum, Schaffer, and Stillman, 2003; Baum, 2006).[37] We confirm that our instruments are relevant and no detectable of the AR(2) in the models, indicating that our GMM estimate coefficients are consistent and unbiased.

2.5 Results of regression analysis

2.5.1 Baseline results

We begin our analysis by estimating the baseline regression in (2.1) using both RE, FE and GMM models (see Table 2.3). We use Hausman test (Hausman, 1978) to choose between FE or RE estimator. Hausman specification test led to a rejection of the RE in favor of the FE specification (with Prob>chi2(25)=0.0000<0.05). The estimated regression coefficient between the NPL ratio and loan growth is negative and statistically significant in all

[36] The standard errors are robust and clustered at the bank level.

[37] The null hypothesis of the second order serial correlation AR(2) is that the errors in the first-differenced equation exhibit no second-order serial correlation.

specifications. The NPL coefficient ranges from 0.00235 for GMM to 0.00270 for FE. This result confirms the Hypothesis 2.1 and is consistent with the results of previous studies (Cucinelli, 2015; Cucinelli, 2016; Louhichi and Boujelbene, 2016). The result is also economically significant. For example, in model (3), a one standard deviation increase of the NPL ratio is associated with an expected 0.096 standard deviations decrease of the loan growth rate (where 0.096 equals to the estimated NPL regression coefficient, 0.00235, multiplied by the standard deviation of NPL, 8.531, and divided by the standard deviation of loan growth, 0.208). Furthermore, the positive and significant coefficient of the lagged dependent variable is also in line with expectations and reflects persistence in the growth of loans (Table 2.3, column 3).

Table 2.3: The relationship between NPL and loan growth: the total sample with different estimation methods

Dependent variable: ΔlogGL	**(1)**	**(2)**	**(3)**
Intercept	-0.00456	-0.376**	
	(-0.24)	(-2.45)	
Bank specific variables			
Lagged ΔlogGL $_{i,t-1}$			0.130***
			(4.93)
NPL$_{i,t-1}$	-0.00323***	-0.00270***	-0.00235***
	(-7.31)	(-4.78)	(-5.96)
Bank Size$_{i,t-1}$	0.00337**	0.0448**	0.00122
	(2.24)	(2.44)	(1.29)
Deposits/Assets$_{i,t-1}$	0.000603***	0.000607*	0.000535***
	(2.92)	(1.71)	(3.43)
Capitalization $_{i,t-1}$	0.00209***	0.00528***	0.000592
	(2.73)	(4.17)	(0.93)
ROAE$_{i,t-1}$	0.000702***	0.000501**	0.000578**
	(3.09)	(2.14)	(2.42)
NIM $_{i,t-1}$	0.0173***	0.0244***	0.0111***
	(5.10)	(4.86)	(3.56)
Macroeconomic indicators			
GDP growth $_{i,t-1}$	0.00766***	0.00878***	0.00688***
	(5.00)	(5.00)	(4.63)
RIR $_{i,t-1}$	-0.00872***	-0.00931***	-0.00649***
	(-6.29)	(-5.71)	(-4.53)
Coefficient Estimates	RE	FE	GMM
No. Obs.	12,295	12,295	11,353
R-squared within	0.1407	0.1503	
Hansen J statistic(p-value)			69.66(0.074)
AB test AR(2) (p-value)			-0.04(0.966)
Dummies Year	Yes	Yes	Yes
Dummy account standards	Yes	Yes	Yes

Note. The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL).The estimation methods are RE, FE, and GMM. The regressions include year dummies and dummy account standards. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of banks. ***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

Our results in Table 2.3 also confirm that both bank-specific variables and macroeconomic indicators are important determinants of bank lending behavior. Bank size is positively and significantly (at 5% level) related to loan growth in the static specification (the coefficient is insignificant, but still positive in the dynamic specification). This indicates that larger banks experience higher credit growth rates than small banks. The deposit to assets ratio coefficient is positive and significant in all estimations. The same is true for the capitalization coefficient. Higher capital ratios are associated with increased bank lending activity, as banks with more equity capital are less constrained by regulatory capital requirements. This finding is consistent with those of previous studies (see Foos et al., 2010). Return on average equity and net interest margin coefficients are significant and positive in all estimations. The last result reveals that increased bank profitability, which may be associated with effective management practices, is conducive for bank lending activity.

Next, we turn to the effect of macroeconomic indicators. The results in Table 2.3 suggest that loan growth is positively and highly significantly correlated with GDP growth, and negatively and highly significantly correlated with the real interest rate. This implies that better economic conditions positively affect credit growth. This result corroborates the empirical findings by Louhichi et al. (2017), Vithessonthi (2016), and Dell'Ariccia and Marquez (2006).

2.5.2 Channels through which NPL affects bank lending behavior

Having identified the main determinants of loan growth, we analyze the interaction terms between the non-performing loans and bank-specific variables in order to investigate the channels through which the level of NPLs impacts the loan growth. We introduce three interaction terms: non-performing loans and bank size, non-performing loans and bank capitalization, and non-performing loans and asset management companies. The results of these additional regressions are presented in Table 2.4. In line with our previous findings, the level of NPLs is negatively and significantly related to the level of loan growth. All bank-specific variables and macroeconomic indicators are similar to those in the basic model specification.

The interaction term between the level of NPLs and bank size is negatively and significantly related to the level of loan growth. This indicates that the effect of non-performing loans on the level of loan growth would lead banks to reduce their credit activity. Furthermore, results indicate that the effects of NPLs on bank lending are conditional upon bank size. The results reveal that the effects of higher level of NPLs are more pronounced with larger banks. The interaction term between the level of NPLs and capitalization is also negatively and significantly related to the level of loan growth. At first take, this result is counter-intuitive, since it implies that the negative effect of NPLs on loan growth rate is more pronounced for better capitalized banks. However, this may be due to the possibility that banks with higher levels of NPLs require more capital to cover loan loss provisions. In such a situation, more capitalized banks have to preserve their capital in order to cover their NPL losses and may be reluctant in providing risky long-term loans, providing ground to support our Hypothesis 2.3.

To control for the impact of asset management companies on lending growth, we also include an AMC dummy variable, which takes the value of 1 for countries that have an asset management company present in a given year or the value of 0 for countries without an asset management company in a given year. We interact this variable with the level of NPLs. We find that the interaction term between NPL and AMC is statistically insignificant, suggesting no evidence related to loan growth level. This finding does not provide support to Hypothesis 2.4.

Table 2.4: Channels through which NPL affects loan growth

Dependent variable: ΔlogGL	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intercept	-0.422^{***}		-0.362^{**}		-0.388^{**}	
	(-2.67)		(-2.24)		(-2.30)	
Bank specific variables						
Lagged $\Delta \text{logGL}_{i,t-1}$		0.130^{***}		0.119^{***}		0.123^{***}
		(4.92)		(4.51)		(4.70)
$\text{NPL}_{i,t-1}$	0.000980	-0.00196^{*}	-0.00158^{**}	-0.00125^{***}	-0.00274^{***}	-0.00234^{***}
	(0.82)	(-1.86)	(-2.49)	(-2.60)	(-4.75)	(-6.15)
$\text{Bank Size}_{i,t-1}$	0.0496^{***}	0.00165	0.0411^{**}	0.000403	0.0454^{**}	0.000499
	(2.63)	(1.17)	(2.15)	(0.43)	(2.26)	(0.50)
$\text{Deposits/Assets}_{i,t-1}$	0.000700^{*}	0.000537^{***}	0.000580	0.000445^{***}	0.000579	0.000487^{***}
	(1.96)	(3.44)	(1.57)	(2.84)	(1.53)	(3.09)
$\text{Capitalization}_{i,t-1}$	0.00530^{***}	0.000605	0.00678^{***}	0.00141^{*}	0.00558^{***}	0.000243
	(4.13)	(0.95)	(5.51)	(1.94)	(4.26)	(0.38)
$\text{ROAE}_{i,t-1}$	0.000496^{**}	0.000581^{**}	0.000495^{**}	0.000717^{***}	0.000418^{*}	0.000567^{**}
	(2.17)	(2.44)	(2.05)	(2.95)	(1.76)	(2.41)
$\text{AMC}_{i,t}$					0.0841	0.0110
					(1.46)	(0.26)
$\text{NIM}_{i,t-1}$	0.0246^{***}	0.0111^{***}	0.0246^{***}	0.0110^{***}	0.0263^{***}	0.0114^{***}
	(4.94)	(3.56)	(4.89)	(3.47)	(5.10)	(3.60)
Macroeconomic indicators						
$\text{GDP growth}_{i,t-1}$	0.00829^{***}	0.00684^{***}	0.00682^{***}	0.00586^{***}	0.00770^{***}	0.00595^{***}
	(4.64)	(4.60)	(4.00)	(3.92)	(4.56)	(3.96)
$\text{RIR}_{i,t-1}$	-0.00991^{***}	-0.00653^{***}	-0.00823^{***}	-0.00571^{***}	-0.00804^{***}	-0.00559^{***}
	(-6.02)	(-4.59)	(-4.92)	(-3.75)	(-4.87)	(-3.79)
Interaction term						
NPL*Bank Size	-0.000420^{***}	-0.0000444				

	(-3.52)	(-0.38)				
NPL*Capitalization			-0.000109***	-0.0000707***		
			(-5.10)	(-3.11)		
NPL*AMC					-0.000559	-0.00146
					(-0.08)	(-0.33)
Coefficient Estimates	FE	GMM	FE	GMM	FE	GMM
No. Obs.	12,295	11,353	11,720	10,813	11,720	10,813
R-squared within	0.1540		0.1590		0.1505	
Hansen J statistic(p-value)		69.63(0.075)		69.56(0.075)		70.19(0.068)
AB test AR(2) (p-value)		-0.04(0.968)		-0.43(0.666)		-0.29(0.769)
Dummies Year	Yes	Yes	Yes	Yes	Yes	Yes
Dummy account standards	Yes	Yes	Yes	Yes	Yes	Yes

Note. The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL). The estimation methods are FE and GMM. The regressions include year dummies and dummy account standards. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of banks. ***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

2.5.3 Bank lending behavior and NPLs in different regions

To assess the impact of non-performing loans on loan growth in different geographic regions, four subsample regressions were computed, separately for banks in EU countries, non-EU countries, advanced economies and emerging economies.[38] The model estimates are presented in Table 2.5 for each region based on the system GMM estimator. In all subsamples, the impact of non-performing loans on the level of bank lending is negative and statistically significant. This shows that the effect of non-performing loans remains a crucial problem for bank lending across different regions and economies.

The coefficient is also larger (in absolute terms) for the advanced economies, indicating that the advanced economies may be more prone to credit cycles as a result of NPLs.

Table 2.5: NPL and loan growth in different regions

Dependent variable: ΔlogGL	**(1) EU**	**(2) Non-EU**	**(3) Advanced**	**(4) Emerging**
Lagged $\Delta logGL_{i,t-1}$	0.168^{***}	0.0884^{*}	0.101^{***}	0.227^{***}
	(5.79)	(1.91)	(3.67)	(3.80)
Bank specific variables				
$NPL_{i,t-1}$	-0.00237^{***}	-0.00227^{***}	-0.00314^{***}	-0.00180^{***}
	(-4.72)	(-3.02)	(-6.63)	(-2.98)
Bank $Size_{i,t-1}$	0.000610	0.00479^{**}	0.00257^{**}	-0.000175
	(0.48)	(2.14)	(2.05)	(-0.08)
$Deposits/Assets_{i,t-1}$	0.000444^{**}	0.000216	0.000399^{**}	0.000457
	(2.52)	(0.63)	(2.28)	(1.25)
Capitalization $_{i,t-1}$	0.000841	0.000709	0.000155	0.00133
	(0.96)	(0.98)	(0.20)	(1.56)
$ROAE_{i,t-1}$	0.000296	0.00110^{***}	0.0000195	0.000849^{**}
	(0.92)	(4.11)	(0.08)	(2.40)
NIM $_{i,t-1}$	0.00367	0.0177^{***}	0.0117^{**}	0.00448
	(0.89)	(3.49)	(2.26)	(1.11)
Macroeconomic indicators				
GDP growth $_{i,t-1}$	0.00184	0.00735^{***}	0.00290	0.00456^{**}
	(0.71)	(2.69)	(0.90)	(2.28)
RIR $_{i,t-1}$	0.000359	-0.0108^{***}	-0.00180	-0.00773^{***}
	(0.12)	(-5.59)	(-0.42)	(-4.92)
Coefficient Estimates	GMM	GMM	GMM	GMM
No. Obs.	7,906	3,447	9,578	1,775
Hansen J statistic (p-value)	61.35 (0.229)	67.50 (0.103)	57.63 (0.343)	28.34 (0.393)
AB test AR(2) (p-value)	0.70 (0.484)	-0.31 (0.759)	-0.37 (0.710)	0.87 (0.386)
Dummies Year	Yes	Yes	Yes	Yes
Dummy account standards	Yes	Yes	Yes	Yes

Note: The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL). The estimation method is GMM. The regressions include year dummies and dummy account standards. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of banks. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

[38] In appendix, Table A.2 presents the country list for the EU, non-EU, advanced, and emerging economies.

2.5.4 Bank lending behavior and NPLs during and after the global financial crisis

The effect of the global financial crisis radically changed the perception of credit risk. To account for this structural break, we divide the analysis into a pre-crisis (2000-2007), crisis (2008-2010), and post-crisis (2011-2017) period to identify how the global financial crisis might have affected the association between NPLs and credit growth (see Allen, Jackowicz, Kowalewski, and Kozłowski, 2017).

Table 2.6: NPL and loan growth before, during, and after global financial crisis

Dependent variable: ΔlogGL	**(1) Before GFC**	**(2) During GFC**	**(3) After GFC**
Lagged ΔlogGL $_{i,t-1}$	0.126***	0.122***	0.122***
	(4.80)	(4.67)	(4.63)
Bank specific variables			
NPL$_{i,t-1}$	-0.00201***	-0.00232***	-0.00184***
	(-5.61)	(-6.11)	(-2.62)
Bank Size$_{i,t-1}$	0.000459	0.000459	0.000541
	(0.48)	(0.48)	(0.56)
Deposits/Assets$_{i,t-1}$	0.000498***	0.000489***	0.000474***
	(3.16)	(3.07)	(2.99)
Capitalization $_{i,t-1}$	0.000407	0.000275	0.000233
	(0.65)	(0.43)	(0.37)
ROAE$_{i,t-1}$	0.000626***	0.000561**	0.000561**
	(2.66)	(2.35)	(2.37)
NIM $_{i,t-1}$	0.0107***	0.0116***	0.0114***
	(3.37)	(3.64)	(3.56)
Macroeconomic indicators			
GDP growth $_{i,t-1}$	0.00631***	0.00590***	0.00584***
	(4.17)	(3.90)	(3.78)
RIR $_{i,t-1}$	-0.00556***	-0.00553***	-0.00563***
	(-3.65)	(-3.59)	(-3.66)
NPL*BEFORE GFC	-0.00413***		
	(-3.02)		
NPL*DURING GFC		-0.000659	
		(-0.96)	
NPL*AFTER GFC			-0.000726
			(-1.12)
Coefficient Estimates	GMM	GMM	GMM
No. Obs.	10,813	10,813	10,813
Hansen J statistic (p-value)	69.68 (0.072)	69.93 (0.071)	69.60 (0.075)
AB test AR(2) (p-value)	-0.28 (0.782)	-0.26 (0.793)	-0.41 (0.679)
Dummies Year	Yes	Yes	Yes
Dummy account standards	Yes	Yes	Yes

Note. The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL).The estimation method is GMM. The regressions include year dummies and dummy account standards. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of banks. ***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

The results of the estimated models are presented in Table 2.6. The effect of NPL on loan growth is negative and significant in all three subperiods. The interaction term of NPL with the pre-crisis dummy variable is significantly negative, indicating that before the crisis period the negative effect of NPLs on credit growth is more pronounced. This provides evidence that it could be a rapid credit expansion in a low NPL environment in the pre-crisis period. The interaction terms of NPL with the crisis and post-crisis period dummy are not significant.

2.6 Robustness checks

2.6.1 Alternative credit risk measurement

In our first robustness check, we use loan loss provision (LLP) as an alternative credit risk variable instead of NPLs. We find a negative and significant relationship between loan loss provision and loan growth.

Table 2.7: Estimation results of alternative credit risk

Dependent variable: ΔlogGL	**(1)**	**(2)**
Intercept	-0.0368*	-0.190**
	(-1.92)	(-2.13)
Bank specific variables		
$\text{LLP}_{i,t-1}$	-0.00477***	-0.00683***
	(-2.75)	(-3.46)
$\text{Bank Size}_{i,t-1}$	0.00225	0.0182*
	(1.49)	(1.72)
$\text{Deposits/Assets}_{i,t-1}$	0.000521**	0.000705**
	(2.49)	(1.97)
$\text{Capitalization}_{i,t-1}$	0.00181***	0.00343***
	(3.15)	(3.61)
$\text{ROAE}_{i,t-1}$	0.000704***	0.000433*
	(3.18)	(1.77)
$\text{NIM}_{i,t-1}$	0.0158***	0.0196***
	(5.18)	(4.07)
Macroeconomic indicators		
$\text{GDP growth}_{i,t-1}$	0.0103***	0.00844***
	(7.69)	(5.52)
$\text{RIR}_{i,t-1}$	-0.0102***	-0.0113***
	(-8.34)	(-7.96)
Coefficient Estimates	RE	FE
No. Obs.	15,381	15,381
R-squared within	0.133	0.136
Dummies Year	Yes	Yes
Dummy account standards	Yes	Yes

Note. The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL). The estimation methods are RE, and FE. The regressions include year dummies and dummy account standards. Robust-standard errors in parenthesis are clustered at the level of banks. ***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

The magnitude of the impact is higher than the NPL impact, and it is equal to 0.00683 in the fixed effect model.[39] This indicates that an increase in the level of LLP has a negative impact on bank lending activity. The finding is consistent with the results of Košak et al., (2015) and Cucinelli (2016). Other control variables remain unchanged under this alternative credit risk measurement specification relative to our baseline model. The results are presented in Table 2.7.

2.6.2 Subsample of commercial banks

Finally, we perform a robustness check using the subsample of commercial banks only. The results remain largely unchanged (Table 2.8).

Table 2.8: Estimation results of commercial bank

Dependent variable: ΔlogGL	**(1)**	**(2)**
Intercept	-0.527^{*}	
	(-1.94)	
Bank specific variables		
Lagged ΔlogGL $_{i,t-1}$		0.123^{***}
		(2.93)
NPL$_{i,t-1}$	-0.00308^{***}	-0.00237^{***}
	(-3.83)	(-4.03)
Bank Size$_{i,t-1}$	0.0429	0.00432^{**}
	(1.52)	(1.98)
Deposits/Assets$_{i,t-1}$	0.00141^{**}	0.000866^{***}
	(2.19)	(3.38)
Capitalization $_{i,t-1}$	0.00568^{***}	0.00127
	(3.58)	(1.34)
ROAE$_{i,t-1}$	0.000475	0.000333
	(1.27)	(0.74)
NIM $_{i,t-1}$	0.0269^{***}	0.0143^{***}
	(4.38)	(3.37)
Macroeconomic indicators		
GDP growth $_{i,t-1}$	0.00787^{***}	0.00548^{***}
	(3.52)	(3.00)
RIR $_{i,t-1}$	-0.00839^{***}	-0.00660^{***}
	(-4.80)	(-4.03)
Coefficient Estimates	FE	GMM
No. Obs.	4,073	3,807
R-squared within	0.163	
Hansen J statistic(p-value)		66.37(0.120)
AB test AR(2) (p-value)		-0.76(0.445)
Dummies Year	Yes	Yes
Dummy account standards	Yes	Yes

Note. The sample covers the period from 2000 to 2017. The dependent variable is the loans growth rate (ΔlogGL). The estimation methods are FE, and GMM. The regressions include year dummies and dummy account standards. AR(2) reports the p-values for the null hypothesis that the errors in the first regression exhibit no second-order serial correlation. The independent variables are lagged one period. Robust-standard errors in parenthesis are clustered at the level of banks. ***,**, and * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

[39] The impact of NPL on growth lending in the fixed effect model is 0.0027, see Table 2.4.

The non-performing loans are negatively and significantly related to bank lending in both estimation methods (FE and GMM). This confirms the view that commercial banks reduce their lending at higher NPL levels.

2.7 Conclusion

In this chapter, we evaluate the relationship between non-performing loans and bank lending behavior. Our analysis is based on a sample of 6,434 banks distributed across 42 EU and non-EU member countries, and observed over the period from 2000 to 2017.

We find a significant and negative impact of non-performing loans on bank lending, indicating that higher levels of NPLs are associated with a reduction of credit growth. Furthermore, our empirical results provide evidence that other bank-specific variables and macroeconomic determinants affect bank lending behavior. We find some evidence that the association between NPLs and bank lending is stronger for well-capitalized and larger banks. The better capitalized and larger the bank is, the higher the relationship between NPL and growth in lending. This findings suggest that well-capitalized and larger banks tend to have more stringent lending standards than weakly-capitalized and smaller banks. We also analyze the role of asset management companies on the impact of NPLs on loan growth. We find no statistical evidence that the link between NPLs and loan growth is less pronounced for banks operating in countries where an AMC is present.

3 ESTIMATING A SCORING MODEL USING REJECT INFERENCE WITH PARAMETRIC AND NON-PARAMETRIC METHODS[40]

3.1 Overview

Credit scoring models are worthwhile in the credit granting phase. These models generally are developed on a sample of accepted applications, while the rejected ones are avoided. Building scoring models utilizing only accepted applications without including the entire population leads the model to be biased. The reject inference specifically addresses the bias by appointing an inferred 'good' or 'bad' status to rejected applications. This study assesses the influence of rejected borrowers on scoring models, anticipating that predictive accuracy is improved. We extend previous study in two ways. First, we consider unique consumer loan-level data on a population of a commercial bank from a candidate EU country to assess the effectiveness of rejected applications. Second, we rely on three classification methods, logistic regression (LR), classification tree method (CART), and radial basis function (RBF), to assess and compare the diverse model performance with and without selection bias correction. Our results reveal that reject inference improves the performance of scoring models. We find that the LR model performs better than the CART and RBF in classification accuracy, ACC, and ROC curve.

3.2 Introduction

Commercial banks and other lending institutions use credit scoring models worldwide to assess loan applicants' creditworthiness. The borrowing history and financial circumstances of the borrowers are essential indicators of loan eligibility and default risk. Through credit scoring models lenders can correctly assign a credit application to either a 'good' loan that can pay back financial obligations or a 'bad' loan that has a high probability of defaults (Wang, Hao, Ma, and Jiang, 2011; Hand and Henley, 1997).

Traditional scoring models are built on a sample of accepted borrowers whereas rejected borrowers are not considered due to non-availability of information and unknown performance. A limitation of these models is that prediction based only on accepted borrowers may lead to bias in credit scoring (Kočenda and Vojtek, 2009). To mitigate this bias, reject inference has been proposed by academics (Bücker, van Kampen, and Krämer, 2013, Li, Tian, Li, Zhou, and Yang, 2017, and Sohn and Shin, 2006). Reject inference

[40] This chapter is co-authored with Matej Marinč and Igor Masten. The authors would like to thank Marko Košak and Orfea Dhuci, and the participants at International Conference on Applied Statistics and Econometrics (ICASE) 2019 in Tirana, Albania, the 8th EBR Conference 2019 in Ljubljana, Slovenia, as well as the Annual Event of Finance Research Letters, 2021 Virtual Conference for their valuable comments and suggestions.

specifies the problem of inferring the 'good' and 'bad' qualities of those customers who have not been given a loan (Sohn and Shin, 2006). The rationale for tagging the loans as 'good' or 'bad' is their performance and loan repayment capacity. To have a holistic scorecard for the entire application population, it is essential to infer the behavior of the borrowers with the rejected loans.

Many academic researchers such as Bücker, van Kampen, and Krämer (2013), Li, Tian, Li, Zhou, and Yang (2017), and Sohn and Shin (2006) have improved the prediction of their models by using reject inference in the absence of actual status of rejected applications in order to evaluate the default risk of the rejected borrowers. Various rejected inference techniques such as proportional assignment, simple augmentation, parcelling, extrapolation, bound and collapse, and fuzzy augmentation have also been employed by Anderson and Hardin (2013), Crook and Banasik (2004), Verstraeten and Poel (2004), Banasik and Crook (2007), Chen and Àstebro (2012), Bücker et al., (2013), among others.

Our work is closely related to Bekhet and Eletter (2014), who examine the efficacy of two credit scoring models[41] in the Jordanian commercial banks (see also Anderson and Hardin, 2013; Ince and Aktan, 2009). Whereas Bekhet and Eletter (2014) use Jordanian credit data, we employ a dataset from a commercial bank from a candidate EU country, and we add information of the rejected applications to identify which credit scoring model estimated by the logistic regression, CART and RBF classification methods improves the predictability of potential defaulters significantly. Additionally, we deal with the behavioral scoring model.[42]

We aim to contribute to this strand of literature in two ways. First, we employ unique consumer loan-level data from a commercial bank enabling us to statistically show that using information from the rejected borrowers can improve the scoring model's predictability. We provide valuable insights into the use of advanced statistical techniques and models in the credit scoring in the candidate EU country banking sector. Second, we explore and compare the performance of parametric (logistic regression) and non-parametric (CART and RBF) methods in the classification of the bank's potential defaulted borrowers.

We show that adding information of the rejected borrowers improves scoring models' accuracy, which unveils that reject inference augmentation through parcelling is a useful method to assess the rejected borrowers' default risk. Our results reveal that logistic regression is the reliable scoring model achieving the highest average correct classification and the highest receiver operating curve. Following reject inference augmentation, the average correct classification increases from 64.3% to 72.6% with an accuracy improvement

[41] The authors use radial basis function (RBF) and logistic regression (LR) model to evaluate credit applications.

[42] Bekhet and Eletter (2014) estimate an application scoring model, where the dependent variable is the credit decision, approved or rejected application. Instead, our dependent variable is the borrower behavior if the loan defaults or non-default.

of 8.3%. The receiver operating curve increases from 68.2% to 79.8% resulting in an improvement of 11.6%. Within the statistical model's framework, the traditional classification method in this study performs slightly better than the advanced classification methods. Overall, our findings provide some relevant practical implications that are beneficial for financial institutions, banks, and other loan providers' institutions to successfully assess their credit risk. The results of our study highlight that the borrowers' history and financial health are crucial for preventing financial loss.

The remainder of the chapter is organized as follows. Section 3.2 provides a theoretical background. In section 3.3, we describe the data and variables. In section 3.4, we describe the research methodology. In section 3.5, we present the main empirical results, and we conclude in section 3.6.

3.3 Theoretical Background

The principle of credit scoring is to trigger credit event, i.e., the probability that the borrower will repay the loan in the future. Our paper is related to two strands of literature. First, the methodological approach employed in credit score models can be broadly grouped into parametric methods (linear discriminant analysis and binary choice models, logit or probit) and non-parametric methods (artificial intelligence, genetic programming to classification and regression tree, CART). In the seminal study, Altman (1968) presents a corporate bankruptcy prediction scoring model based on financial and economic indicators obtained through the linear discriminant analysis (see the pioneering study of Beaver, 1966, among other authors). The linear discriminant analysis is a classification method in which a weight is assigned to an individual factor, enabling the borrowers' classification.

Since the study by Altman (1968), credit scoring has attracted the attention of many academics and financial institutions in recent years to upgrade their risk evaluation and develop adequate and powerful credit risk models (see Akkoç, 2012; Dinh and Kleimeier, 2007; Hasan and Zazzara, 2006; Roszbach, 2004). Jacobson and Roszbach (2003) maintain that the credit scoring prominently reduces the credit risk. Azam, Danish and Akbar (2012) analyze the significance of socio-economic factors in the decision process in a bank using a logistic regression. In the same vein, Abdou, El-Masry, and Pointon (2007) analyze the credit risk in Egyptian banks by using three parametric methodologies, particularly discriminant analysis, logit, and probit regression model. More recently, Ala’raj, Abbod, and Radi (2018) compare credit scoring models that employ logistic regression, artificial neural network, and support vector machine (SVM) analysis in the Jordanian banks. They point to the superiority of logistic regression over ANN and SVM.

Non-parametric statistical techniques, namely classification and regression tree (CART), multivariate adaptive regression splines (MARS), and support vector machines have recently been widely employed in credit scoring models (see Louzada, Ara, and Fernandes, 2016;

Hand and Henley, 1997). Lee T.-S., Chiu, Chou, and Lu (2006) estimate the scoring model by using CART and MARS on a credit card dataset. They find that both models are efficient in terms of effectiveness and sensitivity. Following the instructions from the Basel II Capital Accord, banks were required to develop and estimate three types of risk models based on the internal ratings (IRB) approach, probability of default (PD), the exposure at default (EAD), and the loss given default (LGD) for the credit risk capital. Bastos (2010) estimates loss-given-default by using the classification and regression tree model and the frictional response regression from a commercial bank in Portugal. The dataset contains data for 374 loans granted to the small and medium enterprises that defaulted between the periods June 1995 up to December 2000. He points out that the CART captures the effects on recovery rates due to the explanatory power that the model has. In addition, the model resulted to be fairly robust as an alternative approach for forecasting loss given default.

More recently, advanced statistical techniques such as artificial neural networks (ANN) (see for example Stefanowski and Wilk, 2001; Lee, Chiu, Lu, and Chen, 2002; Yim and Mitchell, 2005; Seow and Thomas, 2006; Abdou, Pointon, and El-Masry, 2008; Šušteršič, Mramor, and Zupan, 2009; Vellidoa, Lisboaa, and Vaughanb, 1999) and genetic programming (Teller and Veloso, 2000; McKee and Lensberg, 2002; Chen and Huang, 2003; Etemadi, Anvary Rostamy, and Dehkordi, 2009, Abdou, 2009) are considered when developing scoring models. On this subject, many academic researchers have compared traditional and advanced techniques (Malhotra and Malhotra, 2003; Lee T.-S and Chen, 2005; Ong, Huang, and Tzeng, 2005; Yap, Ong, and Husain, 2011, Ince and Aktan, 2009, among others). Bensic, Sarlija, and Zekic-Susac (2005) compared LR, probabilistic neural network, and CART decision tree for small-business lending, and found comparable conclusions in terms of decision accuracy. On the one hand, West (2000) found that neural networks scoring model were able to enhance scoring accuracy from 0.5% to 3% compared with LDA, LR, kernel density estimation, CART, and nearest neighbor. On the other hand, Bekhet and Eletter (2014) developed two credit scoring models on the Jordanian commercial banks. They found that radial basis function performs better in identifying higher-risk customers. Different scoring models evaluation criteria like average correct classification criteria (ACC), Type I and II errors, and receiver operating (ROC) curve are generally used on evaluating the effectiveness of a scoring models (Abdou, El-Masry, and Pointon, 2007; Abdou et al., 2008; Chandra and Varghese, 2009; Abdou and Pointon, 2011; Bekhet and Eletter, 2014; Li et al., 2017).

Second, our study is related to the literature that focuses on the importance of correcting sample selection bias using reject inference, which might influence the credit scoring model (Li et al. 2017; Crook, Edelman, and Thomas, 2007; Banasik and Crook, 2007). The credit scoring model should be developed based on complete applicants' data (both approved and rejected applications) and implemented to all future applicants in the lending process (Li et al. 2017).[43] The rejected borrowers have a determinant impact on a credit scoring model.

[43] Boyes, Hoffman, and Low (1989) show that both rejected and approved borrowers should be considered.

However, the rejected borrowers are typically not considered by the credit scoring models due to missing repayment information of the rejected applicants. Thus, Crook and Banasik (2004), who used extrapolation as a reject inference technique, reported a limited improvement of the model's predictive accuracy compared to the model developed only with the approved borrowers. While, Banasik et al. (2003), using a binary-choice model, found modest evidence of augmentation in enhancing the model efficiency. Kim and Sohn (2007), using small and medium enterprises data over the period from 1997 to 2012, found that reject inference improves the model effectiveness.

3.4 Data and Variables Description

3.4.1 Data

The data are provided by a commercial bank from a candidate EU country.[44] The database contains information such as a loan category (mortgage loans, consumer loans, and credit cards), borrowers' personal characteristics, financial variables, and the performance of granted loans. We exclude the dataset mortgage loans and credit card applications for two reasons. First, we do not possess information regarding their corresponding performance. Second, we consider only consumer loan applications. The total number of loan applications consists of 4,120, out of which 2,048 are approved applications and 2,072 are rejected. From the sample of approved loans 1,606 (or 38.9% of the sample and 78.4% of the granted loans) were classified as non-defaulters, and 442 (10.7% of the sample and 21.6% of the granted loans) were classified as defaulters. We have observed the borrowers' payment performance throughout a 12-month period for each loan. We consider a defaulted loan if a borrower is overdue for more than 90 days. The dependent variable (Good/Bad) is a dichotomous variable where it takes value 1 (one) for a non-default loan and 0 (zero), otherwise.

We recognize that our dataset is imbalanced because the number of non-defaulter is larger (78.4%) than the default customers (21.6%). The imbalanced dataset can affect the performance of the estimated models and significantly influence the classification accuracy. The imbalanced ratio (IR) is calculated as the fraction of several samples in the defaulters class to the number of observations in the non-defaulters class (López, Fernández, García, Palade, and Herrera, 2013). In our dataset, the IR ratio is 27.5%, which means a highly imbalanced dataset. Given the circumstances, different re-sampling techniques described in literature are used to deal with the imbalanced problem. They are categorized into three groups: undersampling, oversampling, and hybrids methods.[45] We use the oversampling method as the most frequent technique to deal with the imbalance problem (Marqués, García and, Sánchez, 2013, among others).

[44] Due to privacy concerns, the name of the bank is not disclosed.

[45] The goal of our study is not to describe the techniques used for the imbalanced problem. For further detailed methodology on how these techniques works, see Brown and Mues (2012); Batista, Prati, and Monard (2004); Verbeke, Dejaeger, Martens, Hur, and Baesens (2012); García, Marqués, and Sánchez (2012).

3.4.2 Description of Variables

The consumer loans dataset contains 14 independent variables namely: age, gender, marital status, residential status, current years at job, employment type, education type, contract type, loan maturity, interest rate, has other loans in the bank, branch recommendation, number of co-applicants, and net income in consideration. We use Wald-forward stepwise method to select the appropriate variables to be considered in the scoring model.[46] This method begins by including all the variables in the model and then removes one by one the variables that do not have enough discriminatory power (Lin, 2009). In our analysis the cut-off point used for significance is 0.05. After the selection procedure, we have 10 variables included in the respective models. The variables used to build the models are the same. We do not have missing data on the independent variables. Table 3.1 shows the descriptive statistics of the input variables used in the study. The average age of the borrower is 38.1 years old with the mean current year of job of 2.15 years. The mean consumer loan maturity is 55.6 months and the mean interest rate is 13.5%.

In Table 3.2 we present the correlation matrix of the key variables. The correlation between borrowers with the higher age level, more years at job, higher level of education, and more income have the lowest risk to default. Furthermore, borrowers who have one or more than one exposure in the banking system are at higher risk of defaulting to pay back the debt in contractual time. Furthermore, the correlation between interest rate and our dependent variable is negatively and significantly (-0.229). This indicates that borrowers which have higher interest rate are potentially at higher risk of defaulting in payment.

[46] In Appendix C, Table C.4 presents the results of selected variables.

Table 3.1: Descriptive statistics of the input variables

Variable name	Mean	St. Dev.	Min	Max	Definition
Good/Bad	0.50	0.50	0	1	Dummy. Takes value 1 for good loans (non-default) and 0 for bad loans (default)
Age	38.10	11.57	18	76	Categorical variable. Takes values 1 for 18-29, 2 for 30-40, 3 for 41-50 and 4 for +51
Gender	0.61	0.49	0	1	Dummy. Takes value 1 if the applicant is male and 0 otherwise
Marital status	0.66	0.47	0	1	Dummy. Takes values 1 if the applicant is married and 0 otherwise
Residential status	1.87	0.88	1	3	Categorical variable. Takes values owner, tenant and living with parents
Current years at job	2.15	1.21	1	5	Categorical variable. Years of experience in current job. Takes values 0-1 year, 2-3 years, 4-8 years, 9-15 years and +16 years
Employment type	1.74	0.44	1	2	Categorical variable. Takes value 1 if the applicant works in a public sector and 2 if the applicant works somewhere else.
Education type	2.83	0.42	1	3	Categorical variable. Takes value 1-less than high school, 2-high school and 3-university
Contract type	0.98	0.14	0	1	Dummy. Takes value 1 if the applicant has full time contract and 0 otherwise
Number of co-applicants	1.98	0.15	1	2	Number of co-applicants in the loan application
Loan maturity	55.6	31.13	6	300	Duration of the loan in months
Interest rate	13.5	2.36	3.33	15.9	Describes the loan interest rate for each applicant
Branch recommendation	1.04	0.20	1	2	Categorical variables. Takes value 1 if the branch gives positive recommendation and 2 otherwise for the loan application
Net income in consideration	7.70	4.47	0	14.11	Total net income into consideration for the decision process. The values are presented in natural logarithm of net income
Has other loan in the bank	0.95	0.21	0	1	Dummy. Takes value 1 if the applicant has the loan in the bank and 0 otherwise

Note. The descriptive statistics of the main variables used in the credit scoring. *Source*: *Own work.*

Table 3.2: Correlation matrix

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**
Age	1										
Gender	0.0712***	1									
Marital	0.536***	0.00693	1								
Current years at job	0.350***	-0.0124	0.218***	1							
Education type	-0.0940***	0.00819	-0.0764***	0.0110	1						
No. of co-applicants	-0.0465***	-0.0364**	-0.0465***	-0.0130	0.00782	1					
Loan maturity	0.0976***	0.0473***	0.0650***	0.0967***	0.0449**	-0.221***	1				
Interest rate	-0.0709***	0.0145	-0.0530***	-0.118***	-0.0267	0.0671***	-0.234***	1			
Has other loan in the bank	0.0452**	0.0208	0.0201	-0.0858***	-0.0434**	0.000233	-0.0288*	0.229***	1		
Net income	-0.0679***	-0.0209	-0.0429**	-0.128***	0.0538***	0.0226	0.00146	0.199***	0.0662***	1	
Good/Bad	0.133***	-0.0960***	0.141***	0.109***	0.0826***	-0.00683	0.0159	-0.229***	-0.136***	0.338***	1

Note. The table shows the correlation matrix of the main variables. The codification of the variables: A1-Age, A2-Gender, A3-Marital status, A4-Current years at job, A5-Education type, A6-Number of co-applicants, A7-Request duration, A8-Interest rate, A9-Has other loan in the bank, A10-Net income in consideration, A11-Good/Bad loans. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively. *Source*: *Own work.*

3.5 Research Methodology

We follow Meester (2000) and Bensic, Sarlija, and Zekic-Susac (2005) approach. We use augmentation through the parcelling approach to deal with the reject inference. In principle, this method divides the rejected loan applications into the same portion of good or bad borrowers in the given loans for each score range. We use the bank's calculated score values to create a score range for the rejected borrowers. The assignment of loans for each individual rejected borrower to 'good' or 'bad' class is random within each score range.

Table 3.3 displays the application of the parcelling approach for the rejected borrowers. The first four columns present the distribution (in number and percentages) among score ranges of the good disbursed loans and bad disbursed loans (of the accepted borrowers).[47] The total rejected application (column 5) represents the number of rejected borrowers for each specified score range. In the last two columns (6 and 7) it is presented the random assignment of the rejected applicants into good or bad performance based on the percentage of good and bad disbursed loans from the preliminary scorecard within each score range (Siddiqi, 2006).

Table 3.3: Parcelling method

Score Range	No. Bad (1)	No. Good (2)	Disbursed % Bad (3)	Disbursed % Good (4)	Total Rejected Application (5)	Reject Inference Bad (6)	Reject Inference Good (7)
0 - 17	8	14	36.36%	63.64%	102	37	65
18 - 30	96	182	34.53%	65.47%	580	200	380
31 - 43	180	616	22.61%	77.39%	652	147	505
44 - 57	135	601	18.34%	81.66%	558	102	456
58 - 71	22	182	10.78%	89.22%	169	18	151
72 - 85	1	11	8.33%	91.67%	11	1	10
Total	442	1,606	21.6%	78.4%	2,072	506	1,566

Source: Adapted from Siddiqi (2006) and own work.

To illustrate, there are 34.53% bad and 65.47% good disbursed loans in the score range 18-30 for the accepted population. From the 580 rejected borrowers in this score range, 34.53% (200 applicants) will be allocated to 'bad' and 65.47% (380 applicants) will be assigned as 'good'.

We proceed as follows. First, we use LR, CART, and RBF classification methods, initially for the accepted borrowers and, subsequently, for both accepted and rejected borrowers. Second, our dataset is randomly partitioned into a training sample (approximately 70%) and testing sample (approximately 30%) to have the same validation data to accomplish the

[47] This information is provided for the approved borrowers, if the borrowers are classified into good or bad class.

comparison (Boros et al., 2000). Third, we enable the comparability between classification methods in terms of classification accuracy and predictive power.

3.5.1 Logistic regression model

Logistic regression (LR) is the most extensively used statistical technique by researchers. It is simple to implement and understand (Abdou, Tsafack, Ntim and Baker, 2016; Lee T. S. et al., 2002; Akkoç, 2012). LR allows predicting that the probability of a dependent variable (having a given value) depends on a group of explanatory variables. This method is used for the analysis of multivariate data including binary variables. It estimates the linear association between the explanatory variables and the logit transformation of binary dependent variable y. Binary dependent variable y takes the value 0 (zero) if the loan defaults and 1 (one) if the loan does not default. The vector of borrower characteristics is denoted by x_i where i = 1,...,N. The probability of binary dependent variable of non-default loan can is therefore given by:

$$P(y = 1|x_i, \beta) = \frac{\exp(x_i\beta^T)}{1+\exp(x_i\beta^T)} \tag{3.1}$$

Considering this, the parameter vector β^T can be estimated consistently and efficiently by the maximum likelihood estimator (MLE). The MLE estimates the parameters by maximizing the loglikelihood function of β.

3.5.2 Classification and regression tree - CART

The classification and regression tree (CART) is a well-known method in data mining. The pioneer of the CART approach is Breiman et al. (1984). The CART method uses historical data to build a decision trees. The CART development indicates the division of the properties from each characteristic included in 'good' and 'bad' group risk. The CART method identifies the important and significant variables through the built tree and excludes the useless variables from the analysis (Brezigar-Masten and Masten, 2012). An additional advantage is that the estimated parameters are easy to be interpreted and implemented.

3.5.3 Artificial neural networks - Radial basis function

Artificial neural networks (ANN) provide advantages of statistical techniques and data mining tools used in various business applications. Recently, they are playing an essential role in supporting credit decisions. ANNs have been used in credit scoring due to their efficiency in modeling non-linear input-output relationships. In principle, a neural network model comprises several processing elements (neurons) interconnecting across different layers and associated through various relations or weights (Akkoç, 2012). Multilayer perceptron (MLP) and radial basis function (RBF) are particular types of artificial neural networks. We employ the RBF network instead of the MPL network because the RBF network performs better in approximation capability, accurate detection, and correct classification capacity (Zhu, Xie, Sun, Wang, and Yan, 2016; Gutiérrez et al., 2010). Besides, Bekhet, and Eletter (2014) show that the radial basis function provides the most accurate results in identifying the borrowers who may have a high probability of default. Baesens et al. (2003), using data from Benelux and UK financial institutions, find that neural network and least-squares support vector machines (LS-SVMs) can be considered superior in terms of correctly classified cases and the AUC compared with other techniques, LDA and LR. The RBF method is made up of three layers: the first layer feeds the network, the second is a hidden layer, which remaps inputs to make them linearly separable, and the output layer, which is a linear combination of radial basis functions of the inputs (Zhu et al., 2016; Bekhet and Eletter, 2014; Khashman, 2010; Xie, Yu, and Wilamowski, 2014; Huang, Chen, Hsu, Chen, and Wu, 2004; among others).

3.6 Empirical Results

To perform an appropriate comparison between models, we use three classification techniques, logistic regression (LR),[48] classification and regression tree (CART), and the radial basis function (RBF). This is supported by assessing the scoring models' predictive abilities using ACC, ROC curve, and Type I and II errors.

3.6.1 Credit scoring results between constructed models

Confusion matrix and the receiver operation curve (ROC) are widely used as techniques to assess the model's performance. The confusion matrix is the most common quantitative measure and has information regarding the actual and predicted classification performed by different approaches. This matrix shows the numbers of correctly allocated cases and those that are misclassified in different categories. In addition, the confusion matrix is defined as follows:

$$Confusion\ Matrix = \begin{Bmatrix} TruePositive\ (TP) & FalsePositive\ (FP) \\ FalseNegative\ (FN) & TrueNegative\ (TN) \end{Bmatrix}$$

where:

TP: The number of good applicants classified as good.
FP: The number of bad applicants classified as good.
FN: The number of good applicants classified as bad.
TN: The number of bad applicants classified as bad.

We consider Type I and Type II errors for the three chosen models. Type I error, also known as False Negative Rate (FNR), is the fraction of 0 class cases misclassified as belonging to the 1 class. It is defined as FNR=FN/(TP+FN). Type II error, also known as False Positive

[48] In appendix C, Table C.5 and Table C.6, we present the elasticity results of logistic regression.

Rate (FPR), is the fraction of 1 class misclassified as belonging to the 0 class. It is defined as FPR=FP/(TN+FP) (see Abdou, 2009; Abdou et al., 2016; among others).

Table 3.4 reports the classification accuracy rate for the approved loan applications models. The average correct classification (ACC) rate for logistic regression is greater than the accuracy rate for the classification and regression tree and the radial basis function (64.3%, 62.1%, and 62.6%) with an accuracy improvement of 2.2% comparing for the CART and 1.7% for the RBF, respectively. For the LR training sample, 429 borrowers with good credit are classified as bad credit borrowers, and 364 borrowers with bad credit are classified as good credit borrowers, for testing sample, 201 borrowers with good credit are classified as bad credit borrowers, and 152 borrowers with bad credit are classified as good credit borrowers.

Accuracy ranking of the classification methods shows that LR is the best followed by RBF, and CART. Table 3.5 displays the results of approved and rejected loan applications. We can see that logistic regression provides better results in terms of average correct classification. The performance of built credit scoring models can be ordered: logistic regression, CART, and RBF (72.6%, 71.6%, and 69.1%). For the LR training sample, 404 borrowers with good credit are classified as bad credit borrowers, and 574 borrowers with bad credit are classified as good credit borrowers. 169 borrowers with good credit are classified as bad credit borrowers, and 236 borrowers with bad credit are classified as good credit borrowers for testing sample. Table 3.4 and Table 3.5 confirm that the rejected applications' inclusion improved the model's accuracy.

3.6.2 Type I, II errors and receiver operation curve (ROC)

Type I error happens when a good credit borrower is wrongly defined as a bad credit borrower and Type II error happens when a bad credit borrower is considered as a good credit borrower. Addressing the problem from the lender perspective, lenders have to adopt different tools and techniques to reduce Type I error. Thus, financial institutions can benefit by avoiding granting loans to potential defaulters and reducing the loan default cost as much as possible. Table 3.4 presents the results for the approved loan applications. CART shows less Type I error compared to LR and RBF for training sample (30.4%, 37.8%, and 39.7%) and testing one (33.7%, 42.6%, and 39.4%). The LR has a lower level (32.8%) in the training set and testing sample (30.7%) in Type II error.

The CART model has a lower Type I error than LR (16.6% vs. 23.1%) for the training sample and (18.1% vs. 22.1%) for the testing sample (see Table 3.5). Type II error is lower for the LR model than for the RBF for training sample (32.3% vs. 42.2%), and also for the testing sample the LR's Type II error is approximately 31.9% and the CART's Type II error is 38.9%, whereas for the RBF Type II error is 39.7%. Overall, we find that the logistic regression has the highest average correct classification rate (72.6%). The radial basis function is the worst model since it has the lowest average correct classification rate (69.1%) and the highest Type II error for the training sample (42.2%).

The receiver operating curve (ROC) is a graph of the true positive rate as a function of the false positive rate at all possible cut-off values. The further the curve distances from the main diagonal, the better the model performance is. We compare the area under the receiver operating (AUC) characteristic curve results of the models. Abdou et al. (2016) presents some interpretations of the AUC in order to compare models classification: 0≤ AUC <0.6=fail; 0.6≤ AUC <0.7=poor; 0.7≤ AUC <0.8=fair; 0.8≤ AUC <0.9=good; and 0.9≤ AUC =excellent. From Table 3.5, both models are considered to be "fair" at classifying the overall applications. Overall accuracy from the best to worst classification is LR, CART, and RBF, respectively (79.8%, 77.8%, and 72.3%). In addition, we have an improvement in the ROC using rejected applications.[49]

Table 3.4: Comparing classification results, average correct classification, error rates, and ROC values – approved applications

Model	Sample	Classification rate (%)	ACC %	ROC %	Type I error (%)	Type II error (%)
LR	Training	64.7	64.3	68.2	37.8	32.8
	Testing	63.5			42.6	30.7
CART	Training	63.2	62.1	64.8	30.4	43.3
	Testing	59.5			33.7	47.0
RBF	Training	62.3	62.6	63.0	39.7	35.6
	Testing	63.2			39.4	34.3

Note. The classification results of LR, CART, and RBF for approved loan applications. LR - Logistic regression, CART-Classification and Regression Tree and RBF - Radial Basis Function. Classification rate - is the number of good correctly classified as good and number of bad correctly classified as bad divided by total number of applications for training or testing set; ACC - average correct classification rate; ROC - represent the areas under the ROC curve; Type I error - good loan misclassified as bad; Type II error - bad loan misclassified as good. *Source*: *Own work.*

Table 3.5 shows that the logistic regression performs significantly better than CART and RBF classification methods. The LR model is preferred in terms of accuracy by the ROC criterion. These results align well with other studies (Lee T.S. et al., 2006; Emrouznejad and Anouze, 2010). In short, we have proved that using rejected borrowers' information is of value in practice when we estimate a scoring model.

[49] In appendix C, Figure C.1, C.2, and C.3 present the areas under the ROC curve for each model.

Table 3.5: Comparing classification results, average correct classification, error rates, and ROC values - approved and rejected applications

Model	Sample	Classification rate (%)	ACC %	ROC %	Type I error (%)	Type II error (%)
LR	Training	72.2	72.6	79.8	23.1	32.3
	Testing	73.1			22.1	31.9
CART	Training	71.6	71.6	77.8	16.6	40.0
	Testing	71.7			18.1	38.9
RBF	Training	68.8	69.1	72.3	20.1	42.2
	Testing	69.7			21.1	39.7

Note. The classification results of LR, CART, and RBF for approved and rejected loan applications, including rejected inference technique. LR - Logistic regression, CART- Classification and Regression Tree and RBF - Radial Basis Function. Classification rate - is the number of good correctly classified as good and number of bad correctly classified as bad divided by total number of applications for training or testing set; ACC - average correct classification rate; ROC - represent the areas under the ROC curve; Type I error - good loan misclassified as bad; Type II error - bad loan misclassified as good. *Source*: *Own work.*

3.7 Conclusion

The study provides important insights into the credit scoring implications for a bank operating in a candidate EU country banking sector. The dataset used was provided by a commercial bank and contains borrowers' personal characteristics, financial information, and the performance of granted loans for consumer loan.

We investigate the rejected borrowers influence by using the reject inference technique to build an appropriate and powerfully predictive credit scoring model. Our article makes three significant contributions. First, it extends the existing literature on the scoring models by using a unique dataset from a commercial bank from a candidate EU country. Second, we show that considering rejected borrowers is crucially important for building accurate and unbiased credit scoring model. Reject inference and augmentation through the parcelling improves the predictive accuracy of the model. LR, CART, and RBF methods use the same variables as determinants, which implies that the methods of classification are robust and can be used to deal with the sensitive task of setting up a credit scoring. Third, by employing parametric and non-parametric methods, namely LR, CART, and RBF we explore the credit scoring models and, compare them for their efficiency and robustness in identifying 'good' and 'bad' borrowers. Overall, our results demonstrate that the logistic regression performs better than CART and RBF based on the average correct classification, ROC curve, and Type II errors.

GENERAL DISCUSSION AND CONCLUSION

This chapter's objective is to present in a structured manner the main findings of the research as well as the study and elaboration of theoretical and empirical data, which are the core focal points of this dissertation's main framework. The primary objective of this doctoral dissertation is to observe, examine, and evaluate the relationship between non-performing loans and corruption, as well as its impact on banks' lending pattern behavior. The second part of the chapter summarizes the main findings by explaining the individual research questions that were highlighted beginning of the dissertation. The last part elaborates briefly on the conclusion of the dissertation thesis.

Summary of the Main Findings

Chapter 1 examines the effect of corruption on the level of non-performing loans and whether higher corruption is related to the elevated proportion of non-performing loans in the banking system. This chapter is built around the research question of how corruption affects (positively or negatively) the level of non-performing loans (NPLs).

Chapter 1 employs an unbalanced panel data sample containing 7,773 banks from 140 countries over the period 2000–2016. We use two indices of corruption, the Transparency International Corruption Perception Index and the Control of Corruption Index of the World Bank. The results of the first chapter show that corruption has a statistically significant and positive impact on the level of NPLs through all models and several econometric specifications, and robustness tests. According to the result, higher level of corruption aggravates the loan quality in the banking system.

We also analyze the following research questions: Is the link between corruption and bank size less pronounced for larger banks? Is the link between corruption and NPLs influenced by bank capital?

More specifically, Chapter 1 investigates the potential channels through which corruption might affect loan quality through the introduction of interaction variables between the corruption index and bank-specific variables (i.e., corruption and bank size, corruption and bank capitalization, and corruption and ROAA). The results show that the effect of corruption on the level of NPLs is less pronounced for larger banks than for smaller banks. The explanation may derive from the more hierarchical structures of larger banks that preclude corruptive bank practices, resulting in a weaker link between corruption and NPLs. In addition, the results suggest that the level of NPLs is less pronounced for banks with high profitability indicating that as most profitable banks are less concerned with their revenue creation, they do not engage themselves in risky lending causing a cutback in the level of NPLs. We do not find significant evidence that the interaction term between corruption and capitalization affects the level of NPLs.

We also analyze whether there is the link between corruption and NPLs weaker in countries with a strong legal environment, whether corruption reduces the effectiveness of bank regulation in lowering the level of NPLs, and whether there exists a stronger association between corruption and NPLs in countries characterized by a high level of collectivism.

In further analysis we evaluate whether the level of corruption and NPLs may be affected by cross-country differences, especially in the legal environment, might impact the efficiency of the regulatory framework, and control for heterogeneity in national cultures across countries. The results confirm that in countries where the laws are better enforced, the impact of corruption on the NPLs becomes less pronounced. The interaction term between the corruption index and a measure of capital stringency is found to be a significant factor that is positively related to NPLs, indicating that less stringent capital regulation increases the levels of NPLs. In addition, we find that the interaction term between corruption and the level of collectivism is positively and significantly related to the level of NPLs.

In Chapter 2, we attempt to empirically investigate the relationship between non-performing loans and bank lending behavior. This chapter is built around the following research questions: how does the level of NPLs affect (positively or negatively) the bank lending behaviour?

In Chapter 2 we employ data from 42 EU and non-EU member countries in the period from 2000 to 2017. The results show that the relationship between the NPL ratio and loan growth is negative and statistically significant in all specifications.

We analyse whether the relationship between NPL and loan growth is influenced by bank size, whether the link between NPL and loan growth is influenced by bank capital and whether the link between NPLs and loan growth is less pronounced for the banks in the presence of asset management companies.

The results in Chapter 2 provide evidence that the interaction term between the level of NPLs and bank size is negatively related to the level of loan growth. Thus, this indicates that the effects of NPLs on bank lending are conditional upon bank size. The findings of Chapter 2 provide evidence that the effects of a higher level of NPLs are more pronounced for larger banks. In addition, the interaction term between the level of NPLs and capitalization is also negatively and significantly related to the level of loan growth. This may be due to the possibility that banks with higher levels of NPLs require more capital to cover loan loss provisions. Moreover, from a resolution tool standpoint to deal with problematic loan portfolios, we find that the interaction term between NPL and AMC is statistically insignificant, suggesting no evidence related to loan growth level.

Chapter 3 moves from country-bank level to firm-bank level data. We use unique consumer loan-level data from a commercial bank to study whether adding information of the rejected borrowers improves scoring models' accuracy and to explore and compare the performance of parametric (logistic regression) and non-parametric (CART and RBF) methods in the classification of the bank's potential defaulted borrowers. This chapter is built around the following research questions: Can the predictability of defaulted borrowers be improved by applying reject inference technique in the credit scoring model? How do the classification techniques like logistic regression (LR), classification and regression tree (CART), and radial basis function (RBF) differ in terms of classification of defaulted borrowers?

The results in this chapter show that considering rejected borrowers is crucially important for building an accurate and unbiased credit scoring model. Reject inference and augmentation through parcelling improves the predictive accuracy of the model. The LR, CART, and RBF methods use the same variables as determinants, which implies that the methods of classification are robust and can be used to deal with the sensitive task of setting up a credit scoring. Our results demonstrate that logistic regression performs better than the CART and RBF based on the average correct classification, ROC curve, and Type II errors.

Contributions

Chapter 1 empirically evaluates the channels through which corruption impacts the level of non-performing loans (NPLs) using international bank-level data, spanning the 2000 to 2016 period and across 140 countries. The results confirm that corruption is positively and significantly related to the level of NPLs, indicating that corruption leads to reduced loan quality in the banking system. Furthermore, the relationship between corruption and NPLs is weaker for larger banks. Potential explanations reside in the hierarchical practices of large banks and in regulatory scrutiny. The results indicate that larger banks may be less exposed to corruption due to their highly hierarchical structures. Hierarchical structures give little discretion to loan officers, successfully preventing corruptive behavior. However, countries with a high level of collectivism are associated with a higher level of NPLs. Given that collectivist cultures are characterized by group and social cooperation, it indicates that countries with a high level of collectivism are associated with higher levels of NPLs. The results show that in countries where the legal environment is stronger, the relationship between corruption and NPLs is less pronounced.

Chapter 2 examines the relationship between non-performing loans and bank lending behaviour. The results confirm a significant and negative impact of non-performing loans on bank lending, indicating that higher levels of NPLs are associated with a reduction in credit growth. Furthermore, the results from our empirical analysis provide evidence that other bank-specific variables and macroeconomic determinants affect bank lending behaviour. Moreover, the association between NPLs and bank lending is stronger for well-capitalized and larger banks. The better capitalized and the larger the bank is, the higher the relationship between NPL and lending growth. This finding suggests that well-capitalized and larger banks tend to behave more procyclically than weakly-capitalized and smaller banks, expanding their loan portfolio faster in times of low NPLs and contracting it further in times of high NPLs. In addition, the role of asset management companies in the impact of

NPLs on loan growth is also analyzed. The results indicate no statistical evidence that the link between NPLs and loan growth is less pronounced for banks operating in countries where an AMC is present.

Chapter 3 investigates the rejected borrower's influence by using the rejected inference technique to build an appropriate and powerfully predictive credit scoring model. Our study makes three significant contributions. First, it extends the existing literature on the scoring models by using a unique dataset from a commercial bank from a candidate EU country. Second, we show that considering rejected borrowers is crucially important for building an accurate and unbiased credit scoring model. The reject inference method (augmentation through parcelling) improves the predictive accuracy of the model. The LR, CART, and RBF methods use the same variables as determinants, which implies that the methods of classification are robust and can be used to deal with the sensitive task of setting up credit scoring. Third, by employing parametric and non-parametric methods, namely LR, CART, and RBF, we explore the credit scoring models and, compare them for their efficiency and robustness in identifying 'good' and 'bad' borrowers. Overall, our results demonstrate that logistic regression performs better than CART and RBF based on the average correct classification, ROC curve, and Type II errors.

Conclusion

This dissertation assesses the different drivers of non-performing loans (NPLs) in the banking sector. It proposes a novel method for credit quality evaluation and identification of the accurate classification method of retail loan defaults. It starts by examining the impact of corruption on the level of non-performing loans in banking from 2000 to 2016. Then, the key aspect of the analysis is identifying potential channels through which corruption impacts the level and the dynamics of NPLs. Next, the dissertation extends the empirical investigation into additional factors through which NPLs affect bank lending behaviour expanding the geographic and time dimension using bank-level data from European and non-European member countries over 2000 to 2017. Lastly, it concludes by moving from country-bank level to firm-bank level data for the case of a retail portfolio of a selected commercial bank from a candidate EU country considering different parametric and non-parametric classification methods for the identification of potential default borrowers.

A better understanding of the credit quality drivers and determinants of non-performing loans is crucial for the design of regulations and credit risk management methods and tools. The methodologies used in this dissertation add to the existing literature by filling a significant gap in this area of study and providing important insights that will benefit policymakers, regulators, and academic researchers.

REFERENCE LIST

1. Abdou, H. A. (2009). Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert Systems with Applications, 36(9),* 11402–11417.
2. Abdou, H. A., & Pointon, J. (2011). Credit scoring and decision making in Egyptian public sector banks. *International Journal of Managerial Finance, 5(4),* 391–406.
3. Abdou, H. A., Tsafack, M. D. D., Ntim, C. G., & Baker, R. D. (2016). Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems, 103,* 89–103.
4. Abdou, H., El-Masry, A., & Pointon, J. (2007). On the applicability of credit scoring models in Egyptian banks. *Banks and Bank Systems, 2(1),* 4–20.
5. Abdou, H., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications, 35(3),* 1275–1292.
6. Abdul-Karim, M., Hassan, M. K., Hassan, T., & Mohamad, S. (2014). Capital adequacy and lending and deposit behaviors of conventional and Islamic banks. *Pacific Basin Finance Journal, 28,* 58–75.
7. Admati, A., & Hellwig, M. (2013). The bankers' new clothes: What's wrong with banking and what to do about it. *Princeton University Press,* Princeton, NJ. Retrieved from https://www.gsb.stanford.edu/faculty-research/books/bankers-new-clothes-whats-wrong-banking-what-do-about-it
8. Aiyar, S., Bergthaler, W., Garrido, J. M., Ilyina, A., Jobst, A., & Kang, K. (2015). *A strategy for resolving Europe's problem loans.* International Monetary Fund Staff Discussion, Note 15/19.
9. Akins, B., Dou, Y., & Ng, J. (2017). Corruption in bank lending: The role of timely loan loss recognition. *Journal of Accounting and Economics, 63(2–3),* 454–478.
10. Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222(1),* 168–178.
11. Ala'raj, M., Abbod, M., & Radi, M. (2018). The applicability of credit scoring models in emerging economies: An evidence from Jordan. *International Journal of Islamic and Middle Eastern Finance and Management, 11(4),* 608–630.
12. Ali, A., & Daly, K. (2010). Macroeconomic determinants of credit risk: Recent evidence from a cross country study. *International Review of Financial Analysis, 19(3),* 165–171.
13. Allen, F., Jackowicz, K., Kowalewski, O., & Kozłowski, Ł. (2017). Bank lending, crises, and changing ownership structure in Central and Eastern European countries. *Journal of Corporate Finance, 42,* 494–515.
14. Al-Marhubi, F. A. (2000). Corruption and inflation. *Economics Letters, 66(2),* 199–202.

15. Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23(4),* 589–609.
16. Anderson, B., & Hardin, M. (2013). Modified logistic regression using the EM algorithm for reject inference. *International Journal of Data Analysis Techniques and Strategies, 5*(359–373).
17. Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58(2),* 277.
18. Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68(1),* 29–51.
19. Athanasoglou, P. P., Brissimis, S. N., & Delis, M. D. (2008). Bank-specific, industry-specific and macroeconomic determinants of bank profitability. *Journal of International Financial Markets, Institutions and Money, 18(2)*, 121–136.
20. Awartani, B., Belkhir, M., Boubaker, S., & Maghyereh, A. (2016). Corporate debt maturity in the MENA region: Does institutional quality matter? *International Review of Financial Analysis, 46,* 309–325.
21. Azam, R., Danish, M., & Akbar, S. S. (2012). *The significance of socioeconomic factors on personal loan decision a study of consumer banking local private banks in Pakistan.* MPRA paper. No. 42322. Retrieved from https://mpra.ub.uni-muenchen.de/42322/
22. Baesens, B., Van Gestel, T., Viaene, S. Stepanova, M., Suykens, J., Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, 627–635.
23. Banasik, J., & Crook, J. (2007). Reject inference, augmentation, and sample selection. *European Journal of Operational Research, 183(3),* 1582–1594.
24. Banasik, J., Crook, J., & Thomas, L. C. (2003). Sample selection bias in credit scoring models. *Journal of the Operational Research Society, 54(8)*, 822–832.
25. Barth, J. R., Caprio, G., & Levine, R. (2004). Bank regulation and supervision: What works best? *Journal of Financial Intermediation, 13(2),* 205–248.
26. Barth, J. R., Caprio, G., & Levine, R. (2013). Bank regulation and supervision in 180 countries from 1999 to 2011. *Journal of Financial Economic Policy, 5(2),* 111–219.
27. Barth, J. R., Lin, C., Lin, P., & Song, F. M. (2009). Corruption in bank lending to firms: Cross-country micro evidence on the beneficial role of competition and information sharing. *Journal of Financial Economics, 91(3),* 361–388.
28. Baselga-Pascual, L., Trujillo-Ponce, A., & Cardone-Riportella, C. (2015). Factors influencing bank risk in Europe: Evidence from the financial crisis. *North American Journal of Economics and Finance, 34,* 138–166.
29. Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking and Finance, 34(10)*, 2510–2517.
30. Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter, 6(1),* 20.

31. Batten, J. A., Loncarski, I., & Szilagyi, P. (2018). When Kamay met Hill: Organisational ethics in practice. *Journal of Business Ethics, 147(4),* 779–792.
32. Batten, J. A., Lončarski, I., & Szilagyi, P. G. (2017). Financial market manipulation, whistle-blowing and the common good: Evidence from the LIBOR scandal. 1–31. Available at SSRN: https://doi.org/http://dx.doi.org/10.2139/ssrn.2964917
33. Baum, Christopher F. (2006). An Introduction to Modern Econometrics Using Stata. Stata Corp.
34. Baum, Christopher F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal, 3(1)*, 1–31.
35. Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4,* 71–111.
36. Beck, R., Jakubik, P., & Piloiu, A. (2013). *Non-performing loans: What matters in addition to the economic cycle?* Working Paper Series. No. 515, 34. European Central Bank. Available at: http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1515.pdf
37. Beck, T., Demirguc-Kunt, A., & Levine, R. (2006). Bank supervision and corruption in lending. *Journal of Monetary Economics, 53*, 2131–2163.
38. Beck, T., Demirguc-Kunt, A., Laeven, L., & Maksimovic, V. (2006). The determinants of financing obstacles. *Journal of International Money and Finance, 25(6),* 932–952.
39. Bekhet, H. A., & Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance, 4(1),* 20–28.
40. Bensic, M., Sarlija, N., & Zekic-Susac, M. (2005). Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. *Intelligent Systems in Accounting, Finance & Management, 13(3)*, 133–150.
41. Berger, A. N., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance, 21(6),* 849–870.
42. Berger, A. N., & Udell, G. F. (2006). A more complete conceptual framework for SME finance. *Journal of Banking and Finance, 30(11)*, 2945–2966.
43. Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87(1)*, 115–143.
44. Boerner, K., & Hainz, C. (2004). *The political economy of corruption and the role of financial institutions.* Working Paper. No. 1293.
45. Bonin, J. P., Hasan, I., & Wachtel, P. (2005). Bank performance, efficiency, and ownership in transition countries. *Journal of Banking and Finance, 29,* 31–53.
46. Boros, E., Ibaraki, T., Mayoraz, E., Hammer, P., Kogan, A., & Muchnik, I. (2000). An implementation of logical analysis of data. *IEEE Transactions on Knowledge and Data Engineering, 12(2),* 292–306.
47. Boudriga, A., Boulila, N., & Jellouli, S. (2009). *Does bank supervision impact nonperforming loans : Cross-country determinants using aggregate data.* MPRA paper. No. 18068. Available at: https://mpra.ub.uni-muenchen.de/18068/

48. Bougatef, K. (2015). The impact of corruption on the soundness of Islamic banks. *Borsa Istanbul Review, 15(4),* 283–295.
49. Boyes, W., Hoffman, D., and Low, S. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics, 40(1)*, 3-14
50. Brei, M., & Gadanecz, B. (2012). *Public recapitalizations and bank risk: Evidence from loan spreads and leverage.* Working paper. Bank for International Settlements.
51. Breiman, L., Friedmann, H. L., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth International Group.
52. Brezigar-Masten, A., & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications, 39(11),* 10153–10159.
53. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39(3)*, 3446–3453.
54. Bücker, M., van Kampen, M., & Krämer, W. (2013). Reject inference in consumer credit scoring with nonignorable missing data. *Journal of Banking and Finance, 37(3)*, 1040–1045.
55. Canales, R., & Nanda, R. (2012). A darker side to decentralized banks: Market power and credit rationing in SME lending. *Journal of Financial Economics, 105(2)*, 353–366.
56. Cantero-Saiz, M., Sanfilippo-Azofra, S., Torre-Olmo, B., & López-Gutiérrez, C. (2014). Sovereign risk and the bank lending channel in Europe. *Journal of International Money and Finance, 47*, 1–20.
57. Cantú, C., Claessens, S., & Gambacorta, L. (2019). *How do bank-specific characteristics affect lending? New evidence based on credit registry data from Latin America*. BIS Working Paper. No. 798.
58. Carlson, M., Shan, H., & Warusawitharana, M. (2013). Capital ratios and bank lending: A matched bank approach. *Journal of Financial Intermediation, 22(4),* 663–687.
59. Castro, V. (2013). Macroeconomic determinants of the credit risk in the banking system: The case of the GIPSI. *Economic Modelling, 31(1),* 672–683.
60. Chaibi, H., & Ftiti, Z. (2015). Credit risk determinants: Evidence from a cross-country study. *Research in International Business and Finance, 33*, 1–16.
61. Chandra, B., & Varghese, P. P. (2009). Fuzzifying gini index based decision trees. *Expert Systems with Applications, 36(4)*, 8549–8559.
62. Chavan, P., & Gambacorta, L. (2019). Bank lending and loan quality: An emerging economy perspective. *Empirical Economics, 57(1),* 1–29.
63. Cheisa G., & Mansilla-Fernández, M. (2019). Disentangling the transmission channel NPLs-cost of capital-lending supply. *Applied Economics Letters, 26(16),* 1333-1338.
64. Chen, G. G., & Àstebro, T. (2012). Bound and collapse Bayesian reject inference for credit scoring. *The Journal of the Operational Research Society, 63(10),* 1374–1387.

65. Chen, M. C., & Huang, S. H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications, 24(4),* 433–441.
66. Chen, M., Jeon, B. N., Wang, R., & Wu, J. (2015). Corruption and bank risk-taking: Evidence from emerging economies. *Emerging Markets Review, 24*, 122–148.
67. Chen, Y., Liu, M., & Su, J. (2013). Greasing the wheels of bank lending: Evidence from private firms in China. *Journal of Banking and Finance, 37(7),* 2533–2545.
68. Chouchène, M., Ftiti, Z., & Khiari, W. (2017). Bank-to-bank lending channel and the transmission of bank liquidity shocks: Evidence from France. *Research in International Business and Finance, 39*, 940–950.
69. Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183(3),* 1447–1465.
70. Crook, J., & Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking and Finance, 28(4),* 857–874.
71. Cucinelli, D. (2015). The impact of non-performing loans on bank lending behavior: Evidence from the Italian Banking Sector. *Eurasian Journal of Business and Economics, 8(16),* 59–71.
72. Cucinelli, D. (2016). Can speed kill? The cyclical effect of rapid credit growth: evidence from bank lending behavior in Italy. *The Journal of Risk Finance, 17(5),* 562–584.
73. Danisman, G. O., & Demirel, P. (2018). Bank risk-taking in developed countries: the influence of market power and bank regulations. *Journal of International Financial Markets, Institutions and Money*.
74. De Haan, L., & Van Oordt, M. R. C. (2018). Timing of banks' loan loss provisioning during the crisis. *Journal of Banking and Finance, 87,* 293–303.
75. Deli, Y. D., & Hasan, I. (2017). Real effects of bank capital regulations: Global evidence. *Journal of Banking and Finance, 82*, 217–228.
76. Dell'Ariccia, G., & Marquez, R. (2006). Lending booms and lending standards. *Journal of Finance, 61(5),* 2511–2546.
77. Demirguc-Kunt, A., & Levine, R. (1999). *Bank-based and market-based financial systems: Cross-country comparisons.* World Bank Policy Research Working Paper. No. 2143.
78. Demirguc-Kunt, A., Laeven, L., & Levine, R. (2004). Regulations, market structure, institutions, and the cost of financial intermediation. *Journal of Money, Credit and Banking, 36(3),* 593–622.
79. Dheera-aumpon, S. (2019). Collectivism and connected lending. *Research in International Business and Finance, 48(May 2018),* 258–270.
80. Dimitrios, A., Helen, L., & Mike, T. (2016a). Determinants of non-performing loans: Evidence from euro-area countries. *Finance Research Letters,* 1–4.

81. Dimitrios, A., Helen, L., & Mike, T. (2016b). *Non-performing loans in the euro area: Are core-periphery banking markets fragmented?* Working paper. No. 219. Bank of Greece.
82. Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis, 16(5),* 471–495.
83. Eisdorfer, A. (2008). Empirical evidence of risk shifting in financially distressed firms. *Journal of Finance, 63(2),* 609–637.
84. El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Zheng, X. (2016). Collectivism and corruption in commercial loan production: How to break the curse? *Journal of Business Ethics, 139(2),* 225–250.
85. Emrouznejad, A., & Anouze, A. L. (2010). Data envelopment analysis with classification and regression tree - A case of banking efficiency. *Expert Systems, 27(4),* 231–246.
86. Erdinc, D., & Abazi, E. (2014). The determinants of NPLs in emerging Europe, 2000-2011. *Journal of Economics and Political Economy, 1(2),* 112-125.
87. European Central Bank (2016). Annual report. Obtained from https://www.ecb.europa.eu/pub/pdf/annrep/ar2016en.pdf
88. Espinoza, R., & Prasad, A. (2010). *Nonperforming loans in the GCC banking system and their macroeconomic effects.* Working Paper 10/224. International Monetary Fund.
89. Etemadi, H., Anvary Rostamy, A. A., & Dehkordi, H. F. (2009). A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications, 36(2/2),* 3199–3207.
90. Fan, C. S., Lin, C., & Treisman, D. (2009). Political decentralization and corruption: Evidence from around the world. *Journal of Public Economics, 93(1–2),* 14–34.
91. Fang, X., Jutrsa, D., Peria, S. M., Presbitero, A. F., & Ratnovski, L. (2020). Bank capital requirements and lending in emerging markets: The role of bank characteristics and economic conditions. *Journal of Banking and Finance.*
92. FCA, & PRA. (2015). The failure of HBOS plc (HBOS). Retrieved from https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/publication/hbos-complete-report
93. Foos, D., Norden, L., & Weber, M. (2010). Loan growth and riskiness of banks. *Journal of Banking and Finance, 34(12),* 2929–2940.
94. Fungáčová, Z., Kochanova, A., & Weill, L. (2015). Does money buy credit? Firm-level evidence on bribery and bank debt. *World Development, 68,* 308–322.
95. Gambacorta, L., & Mistrulli, P. E. (2004). Does bank capital affect lending behavior? *Journal of Financial Intermediation, 13(4),* 436–457.
96. Gambacorta, L., & Shin, H. S. (2018). Why bank capital matters for monetary policy. *Journal of Financial Intermediation, 35,* 17–29.
97. Gandrud, C., & Hallerberg, M. (2014). *Bad banks in the EU: The impact of Eurostat rules*. Bruegel Working Paper 2014/15. Obtained from https://www.econstor.eu/handle/10419/126729

98. García, V., Marqués, A. I., & Sánchez, J. S. (2012). Improving risk predictions by preprocessing imbalanced credit data. *Lecture Notes in Computer Science, 7664,* 68–75.
99. Ghosh, A. (2015). Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. *Journal of Financial Stability, 20,* 93–104.
100. Godlewski, C. J. (2004). Bank capital and risk taking in emerging market economies. *Journal of Banking Regulation, 6(2),* 128–145.
101. Goel, R. K., & Hasan, I. (2011). Economy-wide corruption and bad loans in banking: International evidence. *Applied Financial Economics, 21(7),* 455–461.
102. Gulati, R., Goswami, A., & Kumar, S. (2019). What drives credit risk in the Indian banking industry? An empirical investigation. *Economic Systems, 43(1),* 42–62.
103. Gutiérrez, P. A., Segovia-Vargas, M. J., Salcedo-Sanz, S., Hervás-Martínez, C., Sanchis, A., Portilla-Figueras, J. A., & Fernández-Navarro, F. (2010). Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises. *Omega, 38(5),* 333–344.
104. Haider, Z. A., Liu, M., Wang, Y., & Zhang, Y. (2018). Government ownership, financial constraint, corruption, and corporate performance: International evidence. *Journal of International Financial Markets, Institutions and Money, 53,* 76–93.
105. Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3),* 523–541.
106. Hanousek, J., Shamshur, A., & Tresl, J. (2019). Firm efficiency, foreign ownership and CEO gender in corrupt environments. *Journal of Corporate Finance, 59(15),* 344–360.
107. Hasan, I., & Wall, L. D. (2004). Determinants of the loan loss allowance: Some cross-country comparisons. *The Financial Review, 39(1),* 129–152.
108. Hasan, I., & Zazzara, C. (2006). Pricing risky bank loans in the new Basel II environment. *Journal of Banking Regulation, 7 (3-4),* 243-267.
109. Hau, H., Langfield, S., & Marques-Ibanez, D. (2013). Bank ratings: what determines their quality? *Economic Policy, 13(April),* 289–333.
110. Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46(6),* 1251–1271.
111. Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *Journal of International Business Studies, 14(2),* 75–89.
112. Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
113. Hu, J.-L., Li, Y., & Chiu, Y.-H. (2004). Ownership and nonperforming loans: Evidence from Taiwan's banks. *The Developing Economies, 42(3),* 405–420.
114. Huang, Z., Chen, H., Hsu, C., Chen, W., Wu, S., (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems. 37(4),* 543–558.

115. Ibrahim, M. H., & Rizvi, S. A. R. (2018). Bank lending, deposits and risk-taking in times of crisis: A panel analysis of Islamic and conventional banks. *Emerging Markets Review, 35,* 31–47.
116. Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management, 10*(3), 233–240.
117. Iwanicz-Drozdowska, M., & Witkowski, B. (2016). Credit growth in Central, Eastern, and South-Eastern Europe: The case of foreign bank subsidiaries. *International Review of Financial Analysis, 43,* 146–158.
118. Jacobson, T., & Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of Banking and Finance, 27(4),* 615–633.
119. Jeong, S. and Jung, H. (2013). Bank wholesale funding and credit procyclicality: Evidence from Korea. *Panoeconomicus, 5* 615–31.
120. Jõeveer, K. (2013). Firm, country and macroeconomic determinants of capital structure: Evidence from transition economies. *Journal of Comparative Economics, 41(1),* 294–308.
121. Kabir, N., Worthington, A., & Gupta, R. (2015). Comparative credit risk in Islamic and conventional bank. *Pacific-Basin Finance Journal, 34,* 327–353.
122. Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues.* Policy Research Working Paper 5430. World Bank.
123. Keeton, W. R. (1999). Does faster loan growth lead to higher loan losses? Federal Reserve Bank of Kansas city. *Economic Review, 84(2),* 57–75.
124. Keeton, W., & Morris, C. (1987). Why do banks loan losses differ? Federal Reserve Bank of Kansas city. *Economic Review,* 3–21.
125. Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications, 37(9),* 6233–6239.
126. Kim, Y., & Sohn, S. Y. (2007). Technology scoring model considering rejected applicants and effect of reject inference. *Journal of the Operational Research Society, 58(10),* 1341–1347.
127. Kishan, R. P., & Opiela, T. P. (2006). Bank capital and loan asymmetry in the transmission of monetary policy. *Journal of Banking and Finance, 30(1),* 259–285.
128. Klein, N. (2013). *Non-performing loans in CESEE: Determinants and impact on macroeconomic performance.* Working Paper, 13(72), 1. International Monetary Fund.
129. Kočenda, E., & Vojtek, M. (2009). *Default predictors and credit scoring models for retail banking.* CESIFO Working Paper. No. 2862, Category 12. Retrieved from www.CESifo-group.org/wp
130. Košak, M., Li, S., Lončarski, I., & Marinc, M. (2015). Quality of bank capital and bank lending behavior during the global financial crisis. *International Review of Financial Analysis, 37,* 168–183.

131. Koudstaal, M., & Wijnbergen, S.V., (2012). On risk, leverage, and banks: Do highly leveraged banks take on excessive risk? *SSRN Working Paper 2170008*.
132. Kuzucu, N., & Kuzucu, S. (2019). What drives non-performing loans? Evidence from emerging and advanced economies during pre- and post-global financial crisis. *Emerging Markets Finance and Trade, 55(8)*, 1694–1708.
133. La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *The Journal of Finance, 52(3),* 1131–1150.
134. La Porta, R., Silanes, F. L. De, Shleifer, A., & Vishny, R. W. (1998). Law and Finance. *Journal of Political Economy, 106(6)*, 1113–1155.
135. La Porta, R., Lopez-de-Silanes, F., & Zamarripa, G. (2003). Related lending. *The Quarterly Journal of Economics, 118(February),* 231–268.
136. Laeven, L., & Majnoni, G. (2003). Loan loss provisioning and economic slowdowns: Too much, too late? *Journal of Financial Intermediation, 12(2),* 178–197.
137. Laeven, L., Ratnovski, L., & Tong, H. (2016). Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking and Finance, 69,* S25–S34.
138. Lee, T. S., & Chen, I. F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications, 28(4),* 743–752.
139. Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications, 23(3),* 245–254.
140. Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis, 50(4),* 1113–1130.
141. Lehmann, A. (2017). Carving out legacy assets: a successful tool for bank restructuring? Policy Contribution. No. 9. Retrieved from http://www.europarl.europa.eu/RegData/etudes/IDAN/2017/587397/IPOL_IDA(2017)587397_EN.pdf
142. Levine, R. (1998). The legal environment, banks, and long-run economic growth. *Journal of Money, Credit and Banking, 30(3),* 596–613.
143. Li, Z., Tian, Y., Li, K., Zhou, F., & Yang, W. (2017). Reject inference in credit scoring using Semi-supervised Support Vector Machines. *Expert Systems with Applications, 74*, 105–114.
144. Lin, S. L. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications, 36(4),* 8333–8341.
145. Lis, S. F. de, Pagés, J. M., & Saurina, J. (2000). *Credit growth, problem loans and credit risk provisioning in Spain.* Working papers. Bank for International Settlements. Autumn Central Bank Economists' Meeting.
146. Lízal, L., & Kocenda, E. (2001). State of corruption in transition: Case of the Czech Republic. *Emerging Markets Review, 2(2),* 138–160.

147. López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences, 250*, 113–141.
148. Louhichi, A., & Boujelbene, Y. (2016). Credit risk, managerial behaviour and macroeconomic equilibrium within dual banking systems: Interest-free vs. interest-based banking industries. *Research in International Business and Finance, 38*, 104–121.
149. Louhichi, A., & Boujelbene, Y. (2017). Bank capital, lending and financing behaviour of dual banking systems. *Journal of Multinational Financial Management, 41(September 2008)*, 61–79.
150. Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science, 21(2)*, 117–134.
151. Louzis, D. P., Vouldis, A. T., & Metaxas, V. L. (2012). Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking and Finance, 36(4)*, 1012–1027.
152. Love, I., & Turk Ariss, R. (2014). Macro-financial linkages in Egypt: A panel analysis of economic shocks and loan portfolio quality. *Journal of International Financial Markets, Institutions and Money, 28(1)*, 158–181.
153. Makri, V., Tsagkanos, A., & Bellas, A. (2014). Determinants of non-performing loans: The case of Eurozone. *Panoeconomicus, 61(2)*, 193–206.
154. Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega, 31(2)*, 83–96.
155. Marqués, A. I., García, V., & Sánchez, J. S. (2013). On the suitability of resampling techniques for the class imbalance problem in credit scoring. *Journal of the Operational Research Society, 64(7)*, 1060–1070.
156. Marinč, M., Mojmir M., and Vasja, R. (2014). Dimensions of bank capital regulation: A cross-country analysis. *Panoeconomicus* 4, 415–39.
157. Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics, 110(3)*, 681–712.
158. Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics, 69(1)*, 263–279.
159. McKee, T. E., & Lensberg, T. (2002). Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research, 138(2)*, 436–451.
160. Meester S. (2000). Reject inference for credit scoring model development using extrapolation. Mimeo, CIT Group: New Jersey.
161. Méon, P. G., & Weill, L. (2010). Is corruption an efficient grease? *World Development, 38(3)*, 244–259.

162. Meriläinen, J. M. (2016). Lending growth during the financial crisis and the sovereign debt crisis: The role of bank ownership type. *Journal of International Financial Markets, Institutions and Money, 41*, 168–182.
163. Micco, A., & Panizza, U. (2006). Bank ownership and lending behavior. *Economics Letters, 93*, 248–254.
164. Mo, P. H. (2001). Corruption and economic growth. *Journal of Comparative Economics, 29(1)*, 66–79.
165. Mohaddes, K., Raissi, M., & Weber, A. (2017). Can Italy grow out of its NPL overhang? A panel threshold analysis. *Economics Letters*, *159*, 185–189.
166. Nkusu, M. (2011). *Nonperforming loans and macrofinancial vulnerabilities in advanced economies.* Working Paper 11/161. International Monetary Fund.
167. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity, 41(5)*, 673–690.
168. Ong, C., Huang, J., & Tzeng, G. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications, 29(1)*, 41–47.
169. Pantzalis, C., Park, J. C., & Sutton, N. (2008). Corruption and valuation of multinational corporations. *Journal of Empirical Finance, 15(3)*, 387–417.
170. Park, J. (2012). Corruption, soundness of the banking sector, and economic growth: A cross-country study. *Journal of International Money and Finance, 31(5)*, 907–929.
171. Peek, J., & Rosengren, E. (1995). The capital crunch: neither a borrower nor a lender be. *Journal of Money, Credit and Banking, 27(3)*, 625–638.
172. Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies, 22(1)*, 435–480.
173. Puri, M., Rocholl, J., & Steffen, S. (2011). Global retail lending in the aftermath of the US financial crisis: Distinguishing between supply and demand effects. *Journal of Financial Economics*, *100(3)*, 556–578.
174. Qi, S., & Ongena, S. (2019). Will money talk? Firm bribery and credit access. *Financial Management*, *48(1)*, 117–157.
175. Rajan, R., Dhal, S. (2003). Non-performing loans and terms of credit of public sector banks in India: an empirical assessment. *Reserve Bank of India Occasional Paper 24*, 81–121.
176. Rinaldi, L., & Sanchis-Arellano, A. (2006). *Household debt sustainability. What explains household non-performing loans? An empirical analysis.* Working Paper Series. No. 570. European Central Bank.
177. Roodman, D. (2006). How to do xtabond2: An introduction to difference and system GMM in Stata (No. 103). *Stata Journal*.
178. Roszbach, K. (2004). Bank lending policy, credit scoring, and the survival of loans. *The Review of Economics and Statistics, 86(4)*, 946–958.
179. Saha, S., & Ben Ali, M. S. (2017). Corruption and economic development: New evidence from the Middle Eastern and North African Countries. *Economic Analysis and Policy, 54*, 83–95.

180. Salas, V., & Saurina, J. (2002). Credit risk in two institutional regimes: Spanish commercial and savings banks. *Journal of Financial Services Research, 22,* 203–224.
181. Sanfilippo-Azofra, S., Torre-Olmo, B., Cantero-Saiz, M., & López-Gutiérrez, C. (2018). Financial development and the bank lending channel in developing countries. *Journal of Macroeconomics, 55,* 215–234.
182. Schnabl, P. (2012). The international transmission of bank liquidity shocks: Evidence from an emerging market. *Journal of Finance, 67(3),* 897–932.
183. Seow, H. V., & Thomas, L. C. (2006). Using adaptive learning in credit scoring to estimate take-up probability distribution. *European Journal of Operational Research, 173(3),* 880–892.
184. Shahzad, F., Fareed, Z., Zulfiqar, B., Habiba, U., & Ikram, M. (2019). Does abnormal lending behavior increase bank riskiness? Evidence from Turkey. *Financial Innovation, 5(1),* 37.
185. Skrastins, J., & Vig, V. (2015). *How organizational hierarchy affects information production.* Working Paper.
186. Sohn, S. Y., & Shin, H. W. (2006). Reject inference in credit operations based on survival analysis. *Expert Systems with Applications, 31(1),* 26–29.
187. Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring.* New Jersey: John Wiley & Sons, Inc.
188. Stefanowski, J., & Wilk, S. (2001). Evaluating business credit risk by means of approach-integrating decision rules and case-based learning. *International Journal of Intelligent Systems in Accounting, Finance & Management, 10(2),* 97–114.
189. Stein, J. C. (2002). Information production and capital allocation. *Journal of Finance, 57(5),* 1891–1921.
190. Šušteršič, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications, 36(3/1),* 4736–4744.
191. Tarchouna, A., Jarraya, B., & Bouri, A. (2017). How to explain non-performing loans by many corporate governance variables simultaneously? A corporate governance index is built to US commercial banks. *Research in International Business and Finance, 42(July),* 645–657.
192. Teller, A., & Veloso, M. (2000). Internal reinforcement in a connectionist genetic programming approach. *Artificial Intelligence, 120,* 165–198.
193. Thakor, A. V. (2018). Post-crisis regulatory reform in banking: Address insolvency risk, not illiquidity! *Journal of Financial Stability, 37,* 107–111.
194. Thornton, J., & Di Tommaso, C. (2020). The effect of non-performing loans on credit expansion: Do capital and profitability matter? Evidence from European banks. *International Journal of Finance & Economics.*
195. Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics, 76,* 399–457.
196. Triandis, H. C. (2001). Individualism collectivism and personality. *Journal of Personality, 69(6),* 907–924.

197. Van Vu, H., Tran, T. Q., Van Nguyen, T., & Lim, S. (2018). Corruption, types of corruption and firm financial performance: New evidence from a transitional economy. *Journal of Business Ethics, 148(4),* 847–858.
198. Vellidoa, A., Lisboaa, P. J. G., & Vaughanb, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications, 17(1)*, 51–70.
199. Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research, 218(1)*, 211–229.
200. Verstraeten, A. G., & Poel, D. Van Den. (2004). The impact of sample bias on consumer credit scoring performance and profitability. *The Journal of the Operational Research Society, 56(8)*, 981–992.
201. Vithessonthi, C. (2016). Deflation, bank credit growth, and non-performing loans: Evidence from Japan. *International Review of Financial Analysis, 45*, 295–305.
202. Vo, X. V. (2018). Bank lending behavior in emerging markets. *Finance Research Letters, 27(February),* 129–134.
203. Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications, 38(1)*, 223–230.
204. Weill, L. (2011a). Does corruption hamper bank lending? Macro and micro evidence. *Empirical Economics, 41(1),* 25–42.
205. Weill, L. (2011b). How corruption affects bank lending in Russia. *Economic Systems, 35(2),* 230–243.
206. Wellalage, N. H., Locke, S., & Samujh, H. (2018). Corruption, gender and credit constraints: Evidence from South Asian SMEs. *Journal of Business Ethics*, 1–14.
207. West, D. (2000). Neural network credit scoring models. *Computers and Operations Research, 27(11–12),* 1131–1152.
208. Xie, T., Yu, H., & Wilamowski, B. (2014). Comparison between traditional neural networks and radial basis function networks. IEEE, (September), 1194–1199.
209. Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications, 38(10)*, 13274–13283.
210. Yim, J., & Mitchell, H. (2005). Comparison of country risk models: Hybrid neural networks, logit models, discriminant analysis and cluster techniques. *Expert Systems with Applications, 28(1)*, 137–148.
211. Zeume, S. (2017). Bribes and firm value. *Review of Financial Studies, 30(5),* 1457–1489.
212. Zheng, X., Ghoul, S. El, Guedhami, O., & Kwok, C. C. Y. (2013). Collectivism and corruption in bank lending. *Journal of International Business Studies, 44(4),* 363–390.

213. Zhu, Y., Xie, C., Sun, B., Wang, G. J., & Yan, X. G. (2016). Predicting China’s SME credit risk in supply chain financing by logistic regression, artificial neural network and hybrid models. *Sustainability (Switzerland), 8(5)*, 433.

APPENDICES

Appendix A: Chapter 1

Table A.1: Definitions and data sources of variables

Variable	Definition	Source
Bank-specific variables		
NPL	Non-performing loans ratio (%). Measured as impaired loans divided by total loans.	Fitch Connect (2017)
Bank Size	The natural logarithm of total assets of bank *i* at time *t*.	Fitch Connect (2017)
LTCD	The ratio of loans to customer deposits	Fitch Connect (2017)
ROAA	Return on average assets ratio	Fitch Connect (2017)
Capitalization	The capitalization ratio represents the ratio of total equity to total assets in (%).	Fitch Connect (2017)
LLP	Loan loss provisions divided by total loans in (%).	Fitch Connect (2017)
Macroeconomic indicators		
GDP	Annual growth rate of GDP	World Bank (2017)
Unemployment	The unemployment ratio (%)	World Bank (2017)
GCF	Gross capital formation as percentage of GDP (%)	World Bank (2017)
RIR	Real interest rates, measured as the difference between nominal interest rate and inflation rate.	World Bank (2017)
Economies classification	Is a dummy variable equal to 1 if the financial system is bank-based financial system or 0 if the financial system is market-based financial system.	Demirguc-Kunt, A., & Levine, R. (1999) and own calculations
Corruption indexes		
CI	Corruption perception index ranges from 0 to 10. Transformed (10-$CI_{c,t}$). Higher values indicate more corruption.	Transparency International (2017)
WBCI	World bank corruption perception index. Ranges from -2.5 to +2.5. Transformed (5-($CI_{c,t}$+2.5)). Higher value indicates more corruption.	Kaufmann, Kraayand Mastruzzi (2010); World Bank (2017)
ICRG	Corruption index, higher value indicates higher corruption.	International County Risk Guide; https://www.prsgroup.com/explore-our-products/international-country-risk-guide/
Institutional variables		
Capital stringency (Cap_Str)	Indicates "whether the capital requirement reflects certain risk elements and deducts certain market value losses	Barth, Caprio, and Levine, 2013) Survey conducted

	from capital before minimum capital adequacy is determined". Higher values demonstrate grater capital stringency.	on 2003, 2007 and 2011.
Rule of law (RoL)	The indicator captures perceptions of "of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence".	Kaufmann, Kraayand Mastruzzi (2010); World Bank (2017)
Global Financial Crisis (GFC)	Takes value 1 for the period 2008-2010 and 0 otherwise.	Own calculations
CLT	We account for Hofstede's individualism index. For interpretation reasons, we define a new index named collectivism (CLT) as an index equal to 100 minus Hofstede's individualism (IDV). Higher values of CLT index indicating higher collectivism in the country.	Hofstede (2001); Zheng et al. (2013), and El Ghoul et al. (2016).

Table A.2: Number of banks and average corruption index in sample countries from the period 2000-2016

Country	**Bank s**	**Avg. of WBCI**	**Avg. of CI**	**Country**	**Banks**	**Avg. of WBCI**	**Avg. of CI**
Albania	4	3.20	7.05	Kuwait	14	2.34	5.51
Andorra	6	1.22		Kyrgyzstan	2	3.67	7.50
Anguilla	1	1.22		Latvia	1	2.28	6.00
Antigua and Barbuda	1	1.24		Lebanon	24	3.32	7.17
Argentina	1	2.87	7.00	Libya	1	3.98	8.23
Armenia	6	3.13	6.91	Liechtenstein	1	1.21	
Aruba	2	1.31		Lithuania	2	2.16	5.55
Australia	18	0.58	1.63	Luxembourg	8	0.51	1.57
Austria	28	0.74	2.09	Macau	1	1.99	4.48
Azerbaijan	12	3.59	7.60	Macedonia	1	2.87	7.30
Bahamas	10	1.18	2.96	Madagascar	1	2.74	7.10
Bahrain	16	2.27	4.94	Malawi	3	3.11	6.86
Bangladesh	3	3.78	8.43	Malaysia	5	2.31	5.02
Barbados	3	1.04	2.75	Malta	1	1.59	3.90
Belarus	9	3.04	7.11	Mauritius	3	2.20	5.04
Belgium	19	1.01	2.75	Mexico	62	2.99	6.71
Benin	1	3.09	7.50	Moldova	2	3.27	7.33
Bolivia	1	2.95	7.13	Monaco	1		
Botswana	2	1.59	3.90	Mongolia	1	3.00	6.20
Brazil	48	2.57	6.21	Morocco	2	2.76	6.57
Bulgaria	10	2.65	6.13	Myanmar	1	3.89	
Burundi	1	3.47	7.65	Namibia	1	2.26	5.90
Cambodia	1	3.68	7.86	Netherlands	27	0.43	1.27
Canada	98	0.55	1.47	New Zealand	14	0.19	0.77
Cayman Islands	12	1.31		Nicaragua	2	3.27	7.22
Chile	26	1.08	2.95	Nigeria	5	3.64	7.81
China	35	2.97	6.42	Norway	10	0.39	1.30
Colombia	12	2.78	6.29	Oman	1	2.13	4.89
Costa Rica	11	1.88	5.06	Pakistan	1	3.41	7.55
Croatia	4	2.40	6.32	Panama	40	2.82	6.49
Cuba	2	2.25	5.80	Peru	2	2.77	6.52
Curacao	9			Philippines	9	3.07	7.09
Cyprus	10	1.45	4.03	Poland	13	2.04	5.21
Czech Republic	6	2.14	5.76	Portugal	17	1.44	3.68
Denmark	4	0.12	0.61	Puerto Rico	5	1.76	4.10
Dominica	2	1.92	4.63	Qatar	7	1.57	3.46
Dominican Republic	5	3.24	6.94	Romania	10	2.72	6.44
Ecuador	4	3.19	7.08	Russian Federation	101	3.42	7.49
El Salvador	8	2.89	6.22	Rwanda	2	2.31	5.33
Estonia	6	1.36	3.29	Saint Lucia	1		2.90
Ethiopia	3	3.09	7.28	San Marino	1		
Finland	10	0.23	0.82	Saudi Arabia	11	2.55	5.80
France	81	1.13	3.09	Serbia	3	3.15	7.10
Georgia	13	2.44	5.99	Singapore	1	0.29	0.85
Germany	37	0.67	2.15	Slovakia	7	2.18	5.77
Ghana	6	2.61	6.00	Slovenia	8	1.57	4.13
Greece	6	2.18	5.69	South Africa	13	2.27	5.42
Guatemala	4	3.20	7.09	Spain	49	1.45	3.65
Guernsey	6			Sri Lanka	2	2.76	6.68
Guyana	1	3.12	7.04	Sweden	9	0.28	0.82

Haiti	3	3.80	8.22	Switzerland	19	0.41	1.25
Honduras	1	3.25	7.10	Syrian Arab Republic	7	3.75	7.77
Hong Kong	13	0.67	1.87	Taiwan	3	1.72	3.94
Hungary	18	2.00	4.97	Tajikistan	3	3.70	7.76
Iceland	14	0.36	0.99	Thailand	1	2.75	6.57
India	2	2.83	6.58	Togo	1	3.42	7.04
Indonesia	15	3.34	7.71	Trinidad and Tobago	6	2.63	6.34
Iran	1	3.22	7.10	Turkey	34	2.64	6.34
Iraq	3	3.84	8.33	Uganda	3	3.43	7.47
Ireland	18	0.93	2.55	Ukraine	26	3.43	7.66
Isle of Man	1			United Arab Emirates	20	1.48	3.59
Israel	3	1.57	3.56	United Kingdom	43	0.67	1.76
Italy	33	2.21	5.29	United States	6270	1.03	2.58
Jamaica	3	2.70	6.32	Uruguay	2	1.21	3.04
Japan	10	1.06	2.59	Uzbekistan	18	3.74	8.22
Jersey	2			Vanuatu	1	2.27	6.73
Jordan	9	2.32	5.09	Venezuela	4	3.65	7.86
Kazakhstan	16	3.48	7.40	Yemen	3	3.85	8.02
Kenya	5	3.47	7.57	Zambia	2	2.95	7.28
Korea (South)	3	3.86	5.25	Zimbabwe	1	3.89	7.93

Source: CI-Transparency International Corruption Perception Index; WBCI-corruption perception index from World Bank-World Development Indicators and own calculation.

Appendix B: Chapter 2

Table B.1: Definitions and data sources of variables

Variable	Definition	Source
Bank-specific variables		
ΔlogGL	The growth of the natural logarithm of gross loans	Fitch Connect (2019)
NPL	Non-performing loans ratio (%). Measured as total impaired loans divided by total gross loans	Fitch Connect (2019)
Bank Size	The natural logarithm of total assets of bank *i* at time *t*	Fitch Connect (2019)
Deposits/Assets	The ratio of customer deposits divided by total assets	Fitch Connect (2019)
ROAE	Return on average equity ratio	Fitch Connect (2019)
Capitalization	The capitalization ratio represents the ratio of total equity to total assets in (%)	Fitch Connect (2019)
LLP	Loan loss provisions divided by total loans in (%)	Fitch Connect (2019)
NIM	The ratio of net interest margin (%)	Fitch Connect (2019)
Macroeconomic indicators		
GDP	Annual growth rate of GDP	World Bank (2019)
RIR	Real interest rates measured as the difference between nominal interest rate and inflation rate	World Bank (2019)
Global Financial Crisis (GFC)	Takes a value of 1 for the period 2008-2010 and 0 otherwise	Own calculations
Bank type variables		
Commercial dummy	Takes a value of 1 if the bank is a commercial bank and 0 otherwise	Fitch Connect (2019)
Co-operative dummy	Takes a value of 1 if the bank is a co-operative bank and 0 otherwise	Fitch Connect (2019)
Savings dummy	Takes a value of 1 if the bank is a savings bank and 0 otherwise	Fitch Connect (2019)
Interaction terms		
NPL*Bank Size	The interaction term between non-performing loans and the natural logarithm of total assets bank size	Own calculations
NPL*Capitalization	The interaction term between non-performing loans and capitalization	Own calculations
NPL*AMC	The interaction term between non-performing loans and AMC. Takes a value 1 if in a country there is an AMC in a given year or 0 if in the country there is no AMC in a given year	Own calculations; Gandrud and Hallerberg (2014)

Table B.2: List of countries used in the study

Total sample	**EU**	**Non-EU**	**Advanced**	**Emerging**
Albania	Austria	Albania	Austria	Albania
Austria	Belgium	Bosnia and Herzegovina	Belgium	Belarus
Belarus	Bulgaria	Cyprus	Cyprus	Bosnia and Herzegovina
Belgium	Croatia	Iceland	Czech Republic	Bulgaria
Bosnia and Herzegovina	Czech Republic	Israel	Estonia	Croatia
Bulgaria	Denmark	Kosovo	Finland	Hungary
Croatia	Estonia	North Macedonia	France	Kosovo
Cyprus	Finland	Moldova	Germany	Lithuania
Czech Republic	France	Montenegro	Iceland	North Macedonia
Denmark	Germany	Serbia	Ireland	Moldova
Estonia	Greece	Switzerland	Italy	Montenegro
Finland	Hungary	Turkey	Latvia	Romania
France	Ireland	Ukraine	Malta	Serbia
Germany	Italy		Netherlands	Turkey
Greece	Latvia		Norway	Ukraine
Hungary	Lithuania		Poland	
Iceland	Luxemburg		Portugal	
Ireland	Malta		Slovakia	
Israel	Netherlands		Slovenia	
Italy	Poland		Spain	
Kosovo	Portugal		Sweden	
Latvia	Romania		Switzerland	
Lithuania	Slovakia		United Kingdom	
Luxemburg	Slovenia			
North Macedonia	Spain			
Malta	Sweden			
Moldova	United Kingdom			
Montenegro				
Netherlands				
Norway				
Poland				
Portugal				
Romania				
Serbia				
Slovakia				
Slovenia				
Spain				
Sweden				
Switzerland				
Turkey				
Ukraine				
United Kingdom				

Table B.3: Number of banks and observations used in the study, 2000-2017

Country	**Banks**	**Obs.**	**Country**	**Banks**	**Obs.**
Albania	15	165	Latvia	24	320
Austria	657	6,279	Lithuania	13	165
Belarus	23	204	Luxemburg	120	1,444
Belgium	70	844	North Macedonia	9	107
Bosnia and Herzegovina	26	383	Malta	13	169
Bulgaria	25	329	Moldova	15	171
Croatia	41	591	Montenegro	13	141
Cyprus	18	184	Netherlands	49	562
Czech Republic	31	377	Norway	173	2,031
Denmark	120	1,531	Poland	117	969
Estonia	7	85	Portugal	133	992
Finland	82	544	Romania	30	380
France	349	4,362	Serbia	40	473
Germany	1,974	28,977	Slovakia	17	194
Greece	20	216	Slovenia	18	242
Hungary	129	863	Spain	249	2,653
Iceland	14	116	Sweden	111	1,431
Ireland	37	426	Switzerland	416	5,544
Israel	9	146	Turkey	70	600
Italy	742	10,010	Ukraine	187	1,127
Kosovo	4	49	United Kingdom	224	2,702

Appendix C: Chapter 3

Credit scoring results of logistic regression

Table C.1: Classification table of logistic regression (LR)

LR		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**746**	364	67.2	**344**	152	69.3
	Good	429	**705**	62.1	201	**271**	57.4

Note. The total number of applications is 3,212. We have divided the sample in two subsamples: 2,244 (approximately 70% are used for training) and 968 (approximately 30% are used for testing). The results presented in Table C.1 are for approved loan applications. *Source*: *Own work.*

Table C.1.1: Classification table of logistic regression (LR)

LR		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**1202**	574	67.7	**502**	236	68.0
	Good	404	**1344**	76.9	169	**597**	77.9

Note. The results presented in Table C.1.1 are for both, approved and rejected loan applications. We have divided the sample in two subsamples: 3,524 (approximately 70% are used for training) and 1,504 (approximately 30% are used for testing). *Source*: *Own work.*

Credit scoring results of classification and regression tree

Table C.2: Classification table of classification and regression tree (CART)

CART		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**629**	481	56.7	**263**	233	53.0
	Good	345	**789**	69.6	159	**313**	66.3

Note. The total number of applications is 3,212. We have divided the sample in two subsamples: 2,244 (approximately 70% are used for training) and 968 (approximately 30% are used for testing). The results presented in Table C.2 are for approved loan applications. *Source*: *Own work.*

Table C.2.1: Classification table of classification and regression tree (CART)

CART		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**1065**	711	60.0	**451**	287	61.1
	Good	291	**1457**	83.4	139	**627**	81.9

Note. The results presented in Table C.2.1 are for both, approved and rejected loan applications. We have divided the sample in two subsamples: 3,524 (approximately 70% are used for training) and 1,504 (approximately 30% are used for testing). *Source*: *Own work.*

Credit scoring results of artificial neural networks - radial basis function

Table C.3: Classification table of radial basis function (RBF)

RBF		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**715**	395	64.4	**326**	170	65.7
	Good	451	**683**	60.2	186	**286**	60.6

Note. The total number of applications is 3,212. We have divided the sample in two subsamples: 2,244 (approximately 70% are used for training) and 968 (approximately 30% are used for testing). The results presented in Table C.3 are for approved loan applications. *Source*: *Own work.*

Table C.3.1: Classification table of radial basis function (RBF)

RBF		Predicted					
		Training sample			Testing sample		
		Credit performance		% Correct	Credit performance		% Correct
		Bad	Good		Bad	Good	
Credit performance	Bad	**1027**	749	57.8	**445**	293	60.3
	Good	351	**1397**	79.9	162	**604**	78.9

Note. The results presented in Table C.3.1 are for both, approved and rejected loan applications. We have divided the sample in two subsamples: 3,524 (approximately 70% are used for training) and 1,504 (approximately 30% are used for testing). *Source*: *Own work.*

Table C.4: Estimation results of logistic regression

Dependent variable: Good/Bad	**(1)**
Age	0.0364***
	(0.00682)
Gender	-0.0876***
	(0.0124)
Marital status	0.0947***
	(0.0151)
Current years at job	0.0313***
	(0.00544)
Education type	0.0772***
	(0.0146)
Number of co-applicants	0.0752***
	(0.0136)
Loan maturity	-0.00181***
	(0.000229)
Interest rate	-0.0615***
	(0.00277)
Net income in consideration	0.0463***
	(0.00139)
Has other loan in the bank	-0.223***
	(0.0300)
Constant	0.884***
	(0.0647)
R-squared	0.268

Note. The table reports the estimated results of the foreword likelihood approach, where the dependent variable is equal to 1 (one) if the loan non-default and 0 (zero) otherwise. ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively. *Source*: *Own work.*

Table C.5: The coefficient estimates of logistic regression

Dependent variable: Good/Bad		**(1) Marginal effects**
Age	30-40	0.139***
(18-29)		(0.0250)
	41-50	0.202***
		(0.0292)
	+51	0.280***
		(0.0307)
Gender	Male	-0.0702***
(Female)		(0.0192)
Marital status		-0.0419*
(Married)		(0.0239)
Current years at job	2-3 years	0.0424*
(0-1 years)		(0.0251)
	4-8 years	0.0917***

		(0.0242)
	9-15 years	0.0561
		(0.0355)
	+16 years	0.0438
		(0.0510)
Education type	High school	-0.0790
(Less than high school)		(0.0991)
	University	-0.0505
		(0.0952)
Number of co-applicants		0.134***
		(0.0221)
Loan maturity		-0.00229***
		(0.000388)
Interest rate		-0.0534***
		(0.00532)
Net income in consideration		0.00172
		(0.00447)
Has other loan in the bank	Yes	-0.0720
(No)		(0.0454)
No. Obs.		3,212

Note. The table reports margins effects of logistic regression. The estimated equation includes the approved loans. The dependent variable is a dummy equal to 1 (one) if the loan non-default and 0 (zero) otherwise. For categorical variables the reference class is given in parenthesis. Standard errors are in parentheses and ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

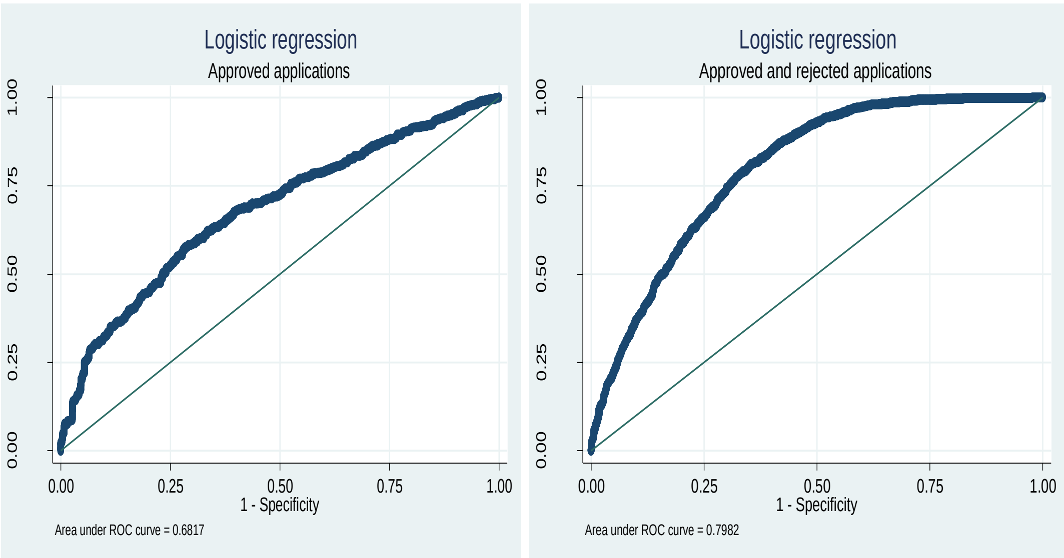
Table C.6: The coefficient estimates of logistic regression

Dependent variable: Good/Bad		**(1) Marginal effects**
Age	30-40	0.119***
(18-29)		(0.0226)
	41-50	0.128***
		(0.0269)
	+51	0.173***
		(0.0285)
Gender	Male	-0.117***
(Female)		(0.0172)
Marital status		0.117***
(Married)		(0.0211)
Current years at job	2-3 years	0.107***
(0-1 years)		(0.0225)
	4-8 years	0.118***
		(0.0218)
	9-15 years	0.144***
		(0.0316)
	+16 years	0.0812*
		(0.0446)
Education type	High school	0.0427

(Less than high school)		(0.0619)
	University	0.166***
		(0.0589)
Number of co-applicants		0.107***
		(0.0192)
Loan maturity		-0.00260***
		(0.000333)
Interest rate		-0.0893***
		(0.00433)
Net income in consideration		0.0691***
		(0.00263)
Has other loan in the bank	Yes	-0.336***
(No)		(0.0489)
No. Obs.		5,028

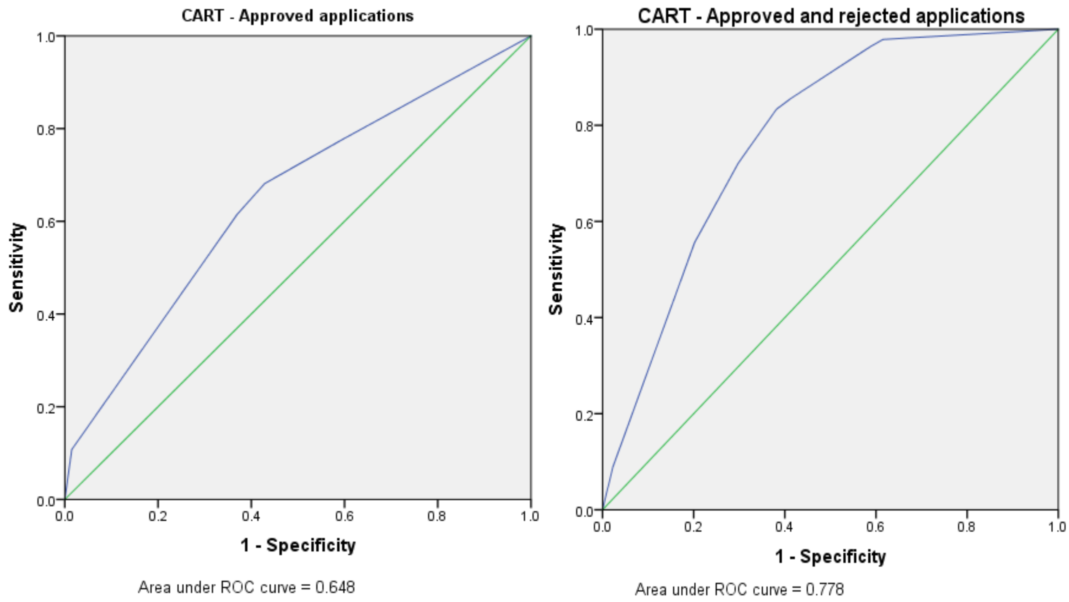
Note. The table reports the margin effects of logistic regression. The estimated equation includes the approved and rejected loan applications, where the status information of the rejected applications is estimated using the rejected inference technique. The dependent variable is a dummy equal to 1 (one) if the loan non-default and 0 (zero) otherwise. For categorical variables the reference class is given in parenthesis. Standard errors are in parentheses and ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively. *Source*: *Own work.*

Figure C.1: Receiver operating curve for logistic regression



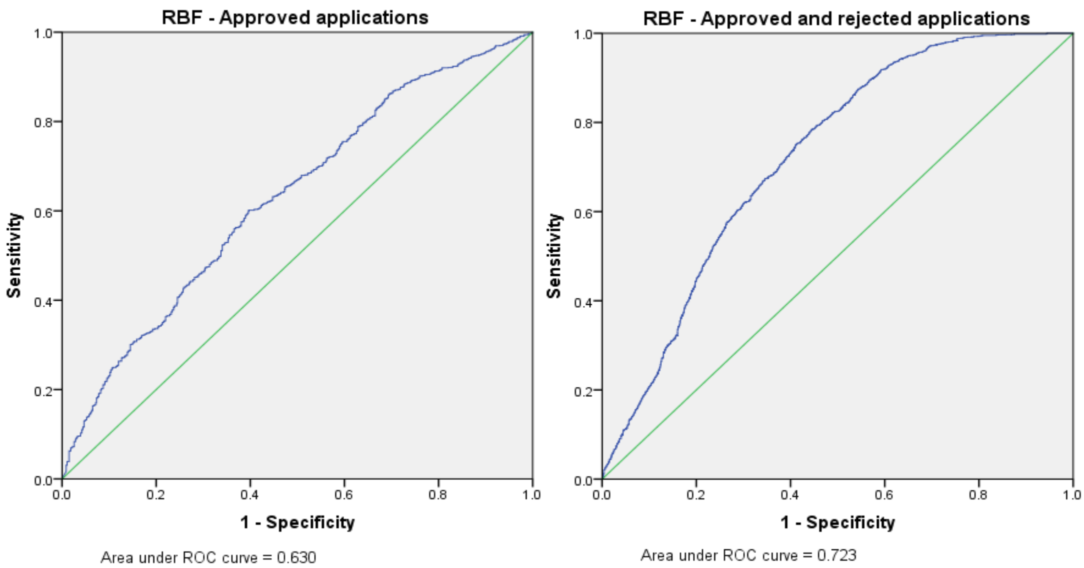
Source: *Own work.*

Figure C.2: Receiver operating curve for CART



Source: *Own work.*

Figure C.3: Receiver operating curve for RBF



Source: *Own work.*

Appendix D: Summary in Slovenian language / Daljši povzetek disertacije v slovenskem jeziku

Ozadje disertacije

Finančne institucije v gospodarski dejavnosti igrajo ključno vlogo. Kakovost portfeljev bančnih posojil predstavlja enega največjih izzivov v zgodovini držav. Po globalni finančni krizi leta 2008 je kakovost bančnih sredstev, predvsem zaradi porasta stopnje nedonosnih terjatev, močno upadla, kar je vodilo v globok upad bančnega finančnega položaja in posledično razkrilo močno ranljivost mednarodnega bančnega sistema (Dimitrios, Helen, and Mike, 2016b; Ghosh, 2015; Louzis, Vouldis, and Metaxas, 2012).

Ob koncu leta 2016, denimo, je odstotek nedonosnih terjatev Cipra narasel za 47 %, Grčije za 37 %, Italije za 17.5 % in Portugalske za približno 12.7 % (ECB, Annual report, 2016), kar nakazuje na to, da so nedonosne terjatve ključne v številnih bančnih sistemih v Evropski uniji. Nkusu (2011) z uporabo letnih poročil 26 razvitih državah za obdobje med 1998 in 2009 priskrbi dokaze, da je 2.4 % rast povezana z upadom kreditov v zasebnem sektorju.

Eden od virov rasti nedonosnih terjatev, kot razlagajo številni znanstveniki, je rast korupcije v državi, natančneje v finančni industriji. La Porta et al. (2003) uporablja podatke iz vzorca 17 mehiških bank, iz česar izpeljuje, da obstaja za posojilojemalce, ki so povezani z bančnimi uslužbenci, kar 33 % možnost neplačila, torej kar 30 % nižja stopnja izterjave plačil v primerjavi s tistimi posojilojemalci, ki nimajo osebne povezave z bančnimi uslužbenci.

Lízal in Kocenda (2001) dodajata, da je korupcija v bančnem sektorju na Češkem vodila v povišano stopnjo nedonosnih terjatev, kar je botrovalo propadu bančnega sistema.

Stopnja nedonosnih terjatev v Evroobmočju se je v času krize povečala in na koncu leta 2007 dosegla nizko stopnjo 2.5 %, ob koncu leta 2013 pa kar 7.7 % (ECB, Annual report, 2016). Stopnja se je malce znižala šele v sredini leta 2016, in sicer je padla na 6.7 %, kot posledica skupnih političnih dejanj, uresničenih v številnih državah članicah EU.

Dvojna kreditna rast, ki je bila prisotna pred obdobjem krize, je posledično upadla in si pravzaprav nikdar ni v celoti opomogla. Znižani izposojevalni standardi bodo v prihodnosti zaradi povečane kreditne širitve najbrž botrovali povišani stopnji nedonosnih terjatev (Erdic and Abazi, 2014; Shahzad et al., 2019; Chavan and Gambacorta, 2019).

Povišana stopnja nedonosnih terjatev zateguje bančni kapital, zmanjšuje neto prihodek in onemogoča dostop do financiranja zaradi povečanega vidika tveganja akterjev na trgu (Aiyar et al, 2015). V središču pričujoče raziskave je tudi drugo razmerje, ki je ključno v obdobju po globalni finančni krizi, med leti 2008 in 2010.

Kreditne institucije uporabljajo model za zbiranje kreditnih točk za ocenjevanje prosilčeve kreditne sposobnosti. Ti modeli igrajo ključno vlogo v prepoznavanju prosilcev z višjim odstotkom verjetnosti neplačila in posledično zmanjšujejo stopnjo neplačil in nedonosnih terjatev. Tako je ključno posodabljanje modelov za zbiranje kreditnih točk z uporabno naprednih metod in tehnik, z namenom prepoznavanja morebitnih prosilcev z višjo stopnjo verjetnosti neplačila in posledično zaščite banke pred neželenimi učinki neplačil.

Ker kakovost bančnih posojil znatno vpliva na finančno stabilnost in gospodarsko dejavnost, skuša pričujoča disertacija zapolniti vrzeli v že obstoječi literaturi, česar se loti na tri načine:

1) s preučevanjem odnosa med korupcijo in nedonosnimi terjatvami ter z raziskovanjem številnih morebitnih kanalov, skozi katere se korupcija morda povezuje z nedonosnimi terjatvami;
2) z analizo učinka nedonosnih terjatev na posojilno vedenje bank v bančnem sistemu EU in v bančnem sistemu držav nečlanic
3) s preučevanjem vpliva zavrnjenih prosilcev za kredit na model za pridobivanje kreditnih točk, z upoštevanjem edinstvenih podatkov na ravni posojil komercialnih bank iz države kandidatke za članstvo v EU.

Raziskovalna vprašanja v pričujoči disertaciji

Poglavje 1 se osredotoča na odnos med korupcijo in stopnjo nedonosnih terjatev in določa kanale, skozi katere korupcija vpliva na stopnjo nedonosnih terjatev. V Poglavju 1 so izpostavljena naslednja raziskovalna vprašanja:

- Raziskovalno vprašanje 1.1: Kako (pozitivno ali negativno) korupcija vpliva na stopnjo nedonosnih terjatev?
- Raziskovalno vprašanje 1.2: Ali je povezava med korupcijo in velikostjo banke manj očitna pri večjih bankah?
- Raziskovalno vprašanje 1.3: Ali bančni kapital vpliva na povezavo med korupcijo in nedonosnimi terjatvami?
- Raziskovalno vprašanje 1.4: Je povezava med korupcijo in nedonosnimi terjatvami šibkejša v državah z močnim pravnim okvirom?
- Raziskovalno vprašanje 1.5: Ali korupcija zmanjšuje učinkovitost bančnega upravljanja/zniževanja stopnje nedonosnih terjatev?
- Raziskovalno vprašanje 1.6: Ali je povezava med korupcijo in nedonosnimi terjatvami močnejša v državah z visoko stopnjo kolektivizma?

Poglavje 2 preučuje obseg in pogoje, v katerih nedonosne terjatve vplivajo na posojilno vedenje bank. V Poglavju 2 so izpostavljenja naslednja raziskovalna vprašanja:

- Raziskovalno vprašanje 2.1: Kako (pozitivno ali negativno) stopnja nedonosnih terjatev vpliva na posojilno vedenje bank?
- Raziskovalno vprašanje 2.2: Ali velikost banke vpliva na povezavo med nedonosnimi terjatvami in rastjo posojil?

- Raziskovalno vprašanje 2.3: Ali bančni kapital vpliva na povezavo med nedonosnimi terjatvami in rastjo posojil?
- Raziskovalno vprašanje 2.4: Ali je povezava med nedonosnimi terjatvami in rastjo posojil manj izrazita pri bankah, kjer so prisotna podjetja za upravljanje s sredstvi?

Poglavje 3 se s stopnje državnih bank preusmeri na stopnjo bank na ravni podjetij, z izbranim portfeljem komercialnih bank iz države kandidatke za članstvo v EU. Poglavje 3 tako naslavlja naslednji raziskovalni vprašanji:

- Raziskovalno vprašanje 3.1: Ali se lahko predvidljivost prosilcev z višjo stopnjo verjetnosti neplačila izboljša z uporabo tehnike zavrnilnega sklepa pri modelu za zbiranje kreditnih točk?
- Raziskovalno vprašanje 3.2: Kako se klasifikacijske tehnike, kot so logistična regresija (LR), klasifikacija in drevesna regresija (KDR) ter radialna bazna funkcija (RBF) razlikujejo v smislu klasifikacije prosilcev z višjo stopnjo verjetnosti neplačila?

Struktura in vsebina disertacije

Doktorska disertacija sestoji iz treh ključnih poglavij. Prvo poglavje se začenja z empirično analizo povezave med korupcijo in stopnjo nedonosnih terjatev ter z oceno, ali je višja stopnja korupcije povezana s povišano stopnjo nedonosnih terjatev v bančnem sistemu. Podpoglavje 1.3 je posvečeno teoretičnemu okvirju in razvoju hipoteze. Podpoglavje 1.4 priskrbi opis podatkov na ravni bank v globalnem okvirju, opredelitev spremenljivk in virov, predstavi pa tudi empirični model. Podpoglavje 1.5 predstavi in razišče empirične izsledke, medtem ko se Podpoglavje 1.6 posveti ponovljivosti podatkov. Podpoglavje 1.7 ponudi zaključek.

Drugo poglavje se začne s predstavitvijo vloge nedonosnih terjatev v posojilnem vedenju banke. Poleg tega predstavi ključne cilje poglavja in ponudi rezultate ter sklepe študije. Podpoglavje 2.3 predstavi teoretičen okvir in ponudi potrditev/zavrnitev hipoteze. Podpoglavje 2.4 opiše podatke, vire podatkov kot tudi empirični model in spremenljivke. Podpoglavje 2.5 ponudi empirične rezultate. Podpoglavje 2.6 se posveča ponovljivosti podatkov, medtem ko Podpoglavje 2.7 sklene drugo poglavje.

Poglavje 3 oceni pomen uporabe informacij zavrnjenih prosilcev pri uporabi primerjalne analize klasifikacijskih metod in modela za zbiranje kreditnih točk. Pojasni tudi vlogo tehnike zavrnilnega sklepa pri ocenjevanju pozitivnih in negativnih lastnosti zavrnjenih prosilcev in ponudi rezultate ter sklepe raziskave. Podpoglavje 3.2 razišče teoretično ozadje, Podpoglavje 3.3 opiše podatke in spremenljivke, medtem ko Podpoglavje 3.4 oriše raziskovalno strategijo. Podpoglavje 3.5 ponudi rezultate, Podpoglavje 3.6 pa sklene Poglavje 3.

Splošna razprava predstavi celotno bistvo disertacije, začenši s povzemanjem ključnih ugotovitev raziskave ter z razlago teoretičnih doprinosov disertacije k raziskovalnemu področju nedonosnih terjatev.

Namen pričujočega poglavja je strukturirano predstaviti ključne izsledke študije kot tudi samo študijo ter nadaljnja pojasnila teoretičnih in empiričnih podatkov, ki so pravzaprav ključna izhodišča glavnega okvira pričujoče disertacije.

Primarni cilj te doktorske disertacije je opazovati, preučiti in ovrednotiti razmerje med nedonosnimi terjatvami in korupcijo kot tudi vplive slednje na vzorec bančnoposojilnega delovanja. Drugi del poglavja povzema ključne izsledke in pojasnjuje posamezna raziskovalna vprašanja, izpostavljena na začetku pričujočega dela. V zadnjem delu so podane sklepne misli doktorske disertacije.

Povzetek ključnih izsledkov

Poglavje 1 preučuje učinke korupcije na nedonosne terjatve in ugotavlja, ali je višja stopnja korupcije povezana s povišanim deležem nedonosnih terjatev v bančnem sistemu. Pričujoče poglavje je zgrajeno na raziskovalnem vprašanju:

‘Kako korupcija vpliva na nedonosne terjatve (pozitivno ali negativno)?’

Poglavje 1 vključuje neuravnotežene podatke 7.773 bank iz 140 držav v obdobju 2000–2016. Uporabljena sta dva kazalca korupcije: transparentni mednarodni korupcijski kazalec in nadzorni korupcijski kazalec Svetovne banke. Rezultati prvega dokumenta kažejo, da ima korpucija statistično pomemben in pozitiven učinek na stopnjo nedonosnih terjatev, na osnovi vseh modelov in številnih ekonometričnih specifikacij in preizkusov stabilnosti. Glede na izsledke višja stopnja korupcije poslabša kakovost posojil v bančnem sistemu.

Poleg tega v disertaciji analiziramo tudi naslednje raziskovalno vprašanje:

‘Je povezava med korupcijo in velikostjo bank pri večjih bankah manj očitna? Ali bančni kapital vpliva na razmerje med korupcijo in nedonosnimi terjatvami?’

Poglavje 1 tako preučuje morebitne kanale, prek katerih lahko korupcija vpliva na kakovost posojil z vpeljavo interakcijskih spremenljivk, stika med korupcijskim kazalcem in specifičnimi spremenljivkami, vezanimi na banke (denimo korupcija in velikost banke, korupcija in kapitalizacija banke, korupcija in ROAA).

Rezultati kažejo, da je vpliv korupcije na stopnjo nedonosnih terjatev pri večjih bankah manj očiten kot pri majhnih. Razlog za to morda izvira iz hierarhičnih struktur večjih bank, ki onemogočajo koruptivne prakse, kar posledično vodi v šibkejšo povezavo med korupcijo in nedonosnimi terjatvami. Poleg tega rezultati kažejo, da je stopnja nedonosnih terjatev nižja pri bankah z višjo stopnjo dobičkonosnosti, kar nakazuje, da so najbolj dobičkonosne banke manj obremenjene z ustvarjanjem prihodka, da se ne vpletajo v tvegano posojilno delovanje, kar posledično vodi v nižjo stopnjo nedonosnih terjatev. Rezultati ne prinašajo dokazov, da razmerje med korupcijo in kapitalizacijo vpliva na stopnjo nedonosnih terjatev.

V delu prav tako analiziramo, ali je povezava med korupcijo in nedonosnimi terjatvami šibkejša v državah z močnim pravnim okvirom, ali korupcija zmanjšuje učinkovitost bančnih regulacij v zmanjševanju stopnje nedonosnih terjatev in ali je povezava med korupcijo in nedonosnimi terjatvami močnejša v državah, kjer je prisotna višja stopnja kolektivizma.

V nadaljevanju ocenimo, ali lahko na stopnjo korupcije in nedonosnih terjatev vplivajo tudi meddržavne razlike, denimo v pravnem okolju, kar morebiti vpliva na učinkovitost regulatornega okvira, ali pa nadzor heterogenosti v nacionalnih kulturah po državah. Rezultati potrjujejo, da je v državah, kjer je pravna podlaga trdnejša, učinek korupcije na nedonosne terjatve manj izrazit. Interakcija med korupcijskim kazalcem in merilom kapitalske strogosti se je izkazala kot ključni dejavnik, ki je pozitivno povezan z nedonosnimi terjatvami, kar nakazuje, da manj stroga regulacija kapitala povečuje stopnjo nedonosnih terjatev. Poleg tega smo dognali, da je interakcija med korupcijo in stopnjo kolektivizma pozitivno povezana s stopnjo nedonosnih terjatev.

V Poglavju 2 skušamo empirično raziskati razmerje med nedonosnimi terjatvami in posojilno dejavnostjo bank. Poglavje je osnovano na naslednjem raziskovalnem vprašanju:

‘Kako stopnja nedonosnih terjatev vpliva na posojilno dejavnost bank (pozitivno ali negativno)?’

V poglavju uporabimo podatke 42 držav članic in nečlanic EU za obdobje 2000–2017. Rezultati kažejo, da je razmerje med nedonosnimi terjatvami in rastjo posojilne dejavnosti negativno in statistično gledano nepomembno v vseh ozirih.

Analiziramo, ali na povezavo med nedonosnimi terjatvami in rastjo posojilne dejavnosti vpliva tudi velikost bank, ali na razmerje med nedonosnimi terjatvami in rastjo posojilne dejavnosti vpliva bančni kapital in ali je omenjena povezava med nedonosnimi terjatvami in rastjo posojilne dejavnosti manj izrazita v prisotnosti podjetij za upravljanje s premoženjem.

Rezultati Poglavja 2 ponujajo dokaze, da je stik med stopnjo nedonosnih terjatev in velikostjo bank negativno povezan z rastjo posojilne dejavnosti. Rezultati tako kažejo, da je učinek nedonosnih terjatev na bančno posojilno dejavnost odvisen od velikosti bank. Izsledki poglavja kažejo, da je vpliv višje stopnje nedonosnih terjatev bolj očiten pri večjih bankah. Poleg tega je stik med nedonosnimi terjatvami in kapitalizacijo negativno povezan s stopnjo posojilne dejavnosti. Razlog je morda to, da banke z višjo stopnjo nedonosnih terjatev potrebujejo več kapitala za kritje rezervacij za izgube pri posojilih. Z vidika razrešitve izzivov spornih posojilnih portfeljev ugotavljamo, da je stik med nedonosnimi terjatvami in podjetji za upravljanje s premoženjem zanemarljiv, kar nakazuje, da ni dokazov, da je povezan z rastjo posojilne dejavnosti.

Poglavje 3 se s podatkov na ravni držav premakne k podatkom na ravni podjetij/bank. Uporabljeni so edinstveni potrošniško-posojilni podatki komercialnih bank, z namenom preučitve, ali dodajanje podatkov zavrnjenih kreditoprosilcev dejansko izboljšuje natančnost modela za pridobivanje kreditnih točk, in z namenom raziskave in primerjave izvedbe parametričnih (logistične regresije) in neparametričnih (MDK in RBF) metod klasifikacije morebitnih posojilojemalcev neplačnikov. Poglavje je osnovano na naslednjem raziskovalnem vprašanju:

'Ali je lahko predvidljivost morebitnih posojilojemalcev neplačnikov izboljšana z uporabo tehnike zavrnilnega sklepa v modelu za pridobivanje kreditnih točk? Kako se klasifikacijske tehnike, kot denimo logistična regresija (LR), drevesna regresija (MDK) in radialna bazna funkcija (RBF), razlikujejo v smislu klasifikacije posojilojemalcev neplačnikov?'

Rezultati v tem poglavju kažejo, da je vključitev zavrnjenih kreditoprosilcev ključna za izgraditev natančnega in nepristranskega modela za pridobivanje kreditnih točk. Zavrnilni sklep in povečevanje z razdeljevanjem izboljšujeta predvidljivo natančnost modela. Metode LR, MDK in RBF uporabljajo enake spremenljivke kot determinante, kar nakazuje, da so metode klasifikacije stabilne in so lahko uporabljene tudi pri občutljivih nalogah, kot je izgradnja modela za pridobivanje kreditnih točk. Rezultati kažejo tudi, da je logistična regresija učinkovitejša kot MDK ali RBF, glede na povprečno ustrezno klasifikacijo, krivuljo ROC in napake tipa II.

Doprinos

Poglavje 1 empirično ovrednoti kanale, prek katerih korupcija vpliva na stopnjo nedonosnih terjatev z uporabo podatkov na ravni bank iz obdobja 2000–2016 iz 140 držav. Rezultati potrjujejo, da je odnos pozitiven, da obstaja znatna povezava s stopnjo nedonosnih terjatev, kar nakazuje, da korupcija vodi v zmanjšano kakovost posojil v bančnem sistemu.

Medtem pa je povezava med korupcijo in nedonosnimi terjatvami manjša pri večjih bankah. Morebitna razlaga zapisanega leži v hierarhičnih praksah večjih bank in v njihovem regulatornem pregledu. Rezultati kažejo, da so večje banke povečini manj izpostavljene korupciji, predvsem zaradi izrazito hierarhične strukture.

Slednje ponujajo uradnikom, ki skrbijo za posojila, več zasebnosti, s čimer uspešno preprečujejo korupcijo. Na drugi strani pa so države z visoko stopnjo kolektivizma povezane z višjo stopnjo nedonosnih terjatev. Ker so kolektivistične kulture opredeljene s skupinami in družbenim sodelovanjem, to kaže, da imajo države z visoko stopnjo kolektivizma tudi višji delež nedonosnih terjatev. Izsledki raziskave kažejo, da je v državah, kjer je pravni okvir močnejši, povezava med korupcijo in nedonosnimi terjatvami šibkejša.

Poglavje 2 preučuje odnos med nedonosnimi terjatvami in posojilnim delovanjem banke. Rezultati potrjujejo pomemben negativen vpliv nedonosnih terjatev na posojilno vedenje bank, kar nakazuje, da je višja stopnja nedonosnih terjatev neodtujljivo povezana z zmanjšanim obsegom kreditne rasti. Empirični izsledki ponujajo dokaze, da tudi druge specifične spremenljivke, vezane na banke, in makroekonomske determinante vplivajo na posojilno delovanje bank.

Poleg tega je povezava med nedonosnimi terjatvami in posojilno dejavnostjo močnejša pri bolje kapitaliziranih in večjih bankah. Večja kot je banka, bolje, kot je kapitalizirana, močnejša je povezava med nedonosnimi terjatvami in posojilno dejavnostjo. Ti izsledki torej kažejo, da bolje kapitalizirane in večje banke običajno delujejo bolj prociklično kot šibkeje kapitalizirane in manjše banke.

Večje banke tudi hitreje širijo svoj posojilni portfelj v času, ko je stopnja nedonosnih terjatev nizka, in ga hitreje krčijo, ko je stopnja nedonosnih terjatev višja. Poleg tega je analizirana tudi vloga podjetij za upravljanje s premoženjem in njihov vpliv na nedonosnih terjatev. Rezultati ne prinašajo nobenih statističnih dokazov, da je povezava med nedonosnimi terjatvami in rastjo posojilne dejavnosti šibkejša v bankah, ki poslujejo v državah, kjer so prisotna tudi podjetja za upravljanje s premoženjem.

Poglavje 3 raziskuje vpliv zavrnjenega prosilca za kredit z uporabo tehnike zavrnilnega sklepa, z namenom izgraditve ustreznega in močnega, predvidljivega modela za pridobivanje kreditnih točk. Pričujoči prispevek tako poskrbi za tri ključne doprinose. Najprej razširi že obstoječo literaturo o modelih za pridobivanje kreditnih točk z uporabo edinstvene podatkovne baze komercialnih bank države kandidatke za članstvo v EU.

Nadalje ugotavljamo, da je ključnega pomena, da se upošteva tudi zavrnjene prosilce za kredit, da se izgradi natančen in nepristranski model za pridobivanje kreditnih točk. Metoda zavrnilnega sklepa (povečevanje z razdeljevanjem) izboljšuje predvidljivo natančnost modela. Metode LR, MDK in RBF uporabljajo enake spremenljivke kot determinante, kar nakazuje, da so metode klasifikacije stabilne in so lahko uporabljene tudi za občutljive naloge, kot je postavitev modela za pridobivanje kreditnih točk.

Nenazadnje pa z uporabo parametričnih in neparametričnih metod, kot so LR, MDK in RBF, raziščemo modele za pridobivanje kreditnih točk in primerjamo tako njihovo učinkovitost kot tudi stabilnost, za namene prepoznavanja 'dobrih' in 'slabih' kreditojemalcev. Rezultati tako kažejo, da ima logistična regresija boljše učinke kot MDK in RBF, glede na povprečno ustrezno klasifikacijo, krivuljo ROC in napake tipa II.

Zaključek

Pričujoča disertacija ovrednoti različna gibala nedonosnih terjatev v bančnem sektorju. Predlaga nove metode za vrednotenje kreditne kakovosti in prepoznavanje natančne klasifikacijske metode neplačil maloprodajnih posojil. Izhodiščna točka je preučevanje vpliva korupcije na stopnjo nedonosnih terjatev v bančništvu v obdobju 2000–2016. Ključni vidik analize je prepoznati morebitne kanale, prek katerih korupcija lahko vpliva na stopnjo in dinamiko nedonosnih terjatev.

V nadaljevanju disertacija razširi empirično raziskavo z dodajanjem dejavnikov, prek katerih nedonosne terjatve vplivajo na posojilno delovanje bank, in razširi zemljepisno in časovno dimenzijo z uporabo podatkov z bančne ravni držav članic in nečlanic EU v obdobju 2000–2017.

Disertacija naposled zaključi s premikom s podatkov na državni ravni na podatke na ravni podjetij/bank za primere maloprodajnih portfeljev izbranih komercialnih bank iz države kandidatke za članstvo v EU. Tu so upoštevani različni parametrični in neparametrični klasifikacijski pristopi prepoznavanja morebitnih neplačnikov posojil.

Globlje razumevanje gibal kreditne kakovosti in determinant nedonosnih terjatev je ključno za ustvarjanje regulacijskega modela in metod ter pripomočkov za upravljanje s kreditnimi tveganji. V disertaciji je uporabljena metodologija, ki doprinese k že poznani literaturi z zapolnitvijo ključnih vrzeli na tem področju in s tem, da priskrbi vpogled, ki koristi tako snovalcem strategij kot tudi regulatorjem in akademskim raziskovalcem.