UNIVERSITY OF LJUBLJANA
SCHOOL OF ECONOMICS AND BUSINESS

**URSULA SLAPNIK**

# TEXTUAL SENTIMENT IN SOVEREIGN CREDIT RATING REPORTS

**DOCTORAL DISSERTATION**

Ljubljana, 2021

## AUTHORSHIP STATEMENT

The undersigned Ursula Slapnik, a student at the University of Ljubljana, School of Economics and Business, (hereafter: SEB LU), author of this doctoral dissertation with the title Textual Sentiment in Sovereign Credit Rating Reports (Sentiment besedil v poročilih o bonitetnih ocenah držav), prepared under the supervision of prof. dr. Igor Lončarski.

I further declare

- this doctoral dissertation to be based on the results of my own research;
- the printed form of this doctoral dissertation to be identical to its electronic form;
- the text of this doctoral dissertation to be language-edited and technically in adherence with the SEB LU'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 LU's Technical Guidelines for Written Works;
- 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;
- to be aware of the consequences a proven plagiarism charge based on this doctoral dissertation could have for my status at the SEB LU in accordance with the relevant SEB LU Rules;
- 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;
- 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;
- 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;
- 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;
- 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, June $16^{th}$, 2021 Author's signature: ____________________

**TEXTUAL SENTIMENT IN SOVEREIGN CREDIT RATING REPORTS**


**Abstract**

Sovereign credit ratings are essential for a country in international financial markets, since they imply the country's credit risk and thus affect the issuer's borrowing cost. They are predominantly assigned by the three major credit rating agencies, namely Standard & Poor's, Fitch and Moody's, by taking into account both quantitative and qualitative factors. The agencies emphasise that their sovereign credit ratings are merely an opinion (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017). In this dissertation, we apply a novel approach to identifying the qualitative judgement or opinion of the rating committee in sovereign credit ratings. We extend the traditional frameworks with new measures - sentiment and subjectivity scores - obtained by textual sentiment analysis methods. The main objective is to offer a comprehensive analysis of sovereign credit rating reports in terms of textual sentiment and subjectivity, and of their role in determining sovereign credit ratings and changes in the ratings, as well as their relationship with sovereign debt markets.

Overall, we find that textual sentiment significantly contributes to explaining sovereign credit ratings. Furthermore, evidence suggests that qualitative judgement of the rating committee is captured by the subjectivity score. There are meaningful differences in textual sentiment between advanced economies and emerging markets, which we attribute to the difference in the general perception of two groups of countries. We also observe a significant difference in subjectivity measures for emerging markets compared to advanced economies, which indicates that the credit rating agencies attach different weights to the rating committee's qualitative judgement for the two groups of countries. We analyse the behaviour of textual sentiment and subjectivity before and after the global financial crisis. We notice a change in sentiment after the crisis, which we ascribe to the general negative economic environment that lingered for several years. However, we do not detect any difference in subjectivity, meaning that the crisis did not disrupt the way the credit rating committee conveys its judgement.

We also explore how we can better model and predict rating transitions using additional textual information. The results show that, on average, sentiment measures perform better than subjectivity measures in predicting rating changes. The improvement is more notable for downgrades than upgrades. We register higher sensitivity scores for sentiment measures compared to subjectivity measures, on average. Dictionary-based methods appear to outperform machine learning algorithms for sentiment measures, while the reverse is true for subjectivity indicators. We examine model perfor-

mance with two philosophical frameworks for assigning sovereign credit ratings in mind, namely the point-in-time and through-the-cycle approach. We confirm that credit rating agencies follow the through-the-cycle rating philosophy by taking a longer horizon into account.

We examine the relationship between textual sentiment and subjectivity measures and CDS markets. By focusing on the narrow window of three days surrounding the rating announcement, we conclude that the value market participants ascribe to sovereign credit rating reports is fairly limited. However, we do observe a positive effect of subjectivity measures on CDS spreads during a positive credit rating announcement. The results also indicate, that markets marginally value Moody's qualitative judgement expressed in the reports above the one expressed by the remaining two agencies. We do not observe meaningful discrepancies in market reactions between advanced economies and emerging markets. Finally, by examining the special role of the transition between investment and speculative grade rating classes, we register significant returns at the crossover from speculative to investment grade.

We believe this dissertation offers important scientific contributions, both methodological and practical. To the best of our knowledge, we carry out the most comprehensive analysis of sovereign credit rating reports thus far by using novel and advanced approaches, i.e. textual sentiment analysis. These methods have been relatively widely utilised in corporate finance, but not in international finance, and especially not in connection to sovereign credit ratings, where such studies are almost non-existent. We thus fill an important gap in the literature by addressing this issue. Additionally, apart from textual sentiment measures already present in textual analysis literature, we also introduce textual subjectivity indicators. We believe this is a unique approach to identifying opinion in sovereign credit ratings.

Furthermore, we believe this dissertation makes a considerable contribution to the understanding of sovereign credit ratings. This has practical implications for market participants, from private and institutional investors to borrowers. Sovereign credit ratings and changes in the ratings can have a substantial effect on returns and cost of financing, which is especially true for the lower rating classes. By using textual sentiment and subjectivity indicators, we can improve the predictability of changes in sovereign credit ratings, which allows issuers to revise their debt financing schedules and reduce the cost of financing, as well as allows investors to optimise their portfolios and seek higher returns.

*Keywords: sovereign credit ratings, sovereign credit rating reports, textual sentiment analysis, soft information, subjectivity, rating transitions, market reactions*

# SENTIMENT BESEDIL V POROČILIH O BONITETNIH OCENAH DRŽAV


## Povzetek

Bonitetne ocene so bistvenega pomena za državo pri dostopu do mednarodnih finančnih trgov, saj odražajo kreditno tveganje države in tako vplivajo na izdajateljeve stroške financiranja. Pretežno jih dodeljujejo tri glavne bonitetne agencije, in sicer Standard & Poor's, Fitch in Moody's, ki pri tem upoštevajo tako kvantitativne kot tudi kvalitativne dejavnike. Agencije poudarjajo, da so njihove bonitetne ocene držav zgolj mnenje (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017). V tej disertaciji uporabljamo nov pristop k ugotavljanju prisotnosti kvalitativne presoje ali mnenja bonitetne komisije v bonitetnih ocenah držav. Tradicionalne pristope razširimo z novimi kazalniki - merami besedilnega sentimenta in subjektivnosti - pridobljenimi s besedilnimi metodami analize sentimenta. Glavni cilj je ponuditi izčrpno analizo poročil o bonitetnih ocenah držav glede na besedilni sentiment in subjektivnost, ter njihovo vlogo pri določanju bonitetnih ocen in sprememb bonitetnih ocen držav, pa tudi povezav z dolžniškimi trgi.

Na splošno ugotavljamo, da besedilni sentiment bistveno prispeva k razlagi bonitetnih ocen držav. Poleg tega dokazi kažejo, da mere subjektivnosti zajemajo kakovostno presojo bonitetne komisije. Med razvitimi gospodarstvi in razvijajočimi se trgi obstajajo pomembne razlike v besedilnem sentimentu, kar pripisujemo razliki v splošnem zaznavanju teh skupin držav. Opažamo tudi znatno razliko v merah subjektivnosti za razvijajoče se trge v primerjavi z razvitimi gospodarstvi, kar kaže na to, da bonitetne agencije kvalitativni presoji bonitetne komisije za obe skupini držav pripisujejo različne uteži. Analiziramo vedenje besedilnega sentimenta in subjektivnosti pred in po svetovni finančni krizi. Po krizi opažamo spremembo v sentimentu, ki jo pripisujemo splošnemu negativnemu gospodarskemu okolju, ki je trajalo več let. Vendar ne zaznavamo nobene razlike v subjektivnosti, kar pomeni, da kriza ni spremenila načina oblikovanja presoje bonitetne komisije.

Raziskujemo tudi, kako lahko z dodatnimi besedilnimi informacijami bolje modeliramo in napovedujemo spremembe bonitetnih ocen. Rezultati kažejo, da se pri napovedovanju sprememb bonitetnih ocen v povprečju mere sentimenta odrežejo bolje kot mere subjektivnosti. Izboljšanje je opaznejše pri znižanju kot pri zvišanju. V povprečju zabeležimo višjo pravilnost napovedi pri merah sentimenta v primerjavi z merami subjektivnosti. Zdi se, da slovarske metode prekašajo algoritme strojnega učenja za mere sentimenta, medtem ko velja obratno za mere subjektivnosti. Uspešnost modela preučujemo v dveh filozofskih okvirih za določanje bonitetnih ocen držav, in sicer

pristop točkovne ocene in skozi cikel. Potrjujemo, da bonitetne agencije upoštevajo filozofijo ocenjevanja skozi celoten cikel z upoštevanjem daljšega obdobja.

Preučujemo tudi povezavo med merami besedilnega sentimenta in subjektivnosti ter trgi s posli kreditnih zamenjav državnih vrednostnih papirjev (angl. credit default swaps, CDS). Če se osredotočimo na ozko okno treh dni okoli objave bonitetne ocene, sklepamo, da tržni udeleženci poročilom o bonitetni oceni držav pripisujejo precej omejeno vlogo. Vendar pa opažamo pozitiven učinek mer subjektivnosti na CDS razmike po pozitivni boniteti objavi. Rezultati tudi kažejo, da trgi nekoliko bolje ocenjujejo kakovostno presojo agencije Moody's, kot je izražena v poročilih, v primerjavi s poročili preostalih dveh agencij. Ne opažamo pomembnih razlik v tržnih reakcijah med razvitimi gospodarstvi in razvijajočimi se trgi. S preučitvijo specifičnega prehoda med naložbenimi in špekulativnimi bonitetnimi razredi zaznamo statistično značilne donose pri prehodu iz špekulativnega v naložbeni razred.

Menimo, da ta disertacija pomembno prispeva k znanosti, tako z metodološkega kot praktičnega vidika. Kolikor nam je znano, ponujamo najobsežnejšo analizo poročil o bonitetnih ocenah držav z uporabo novih in naprednih pristopov, tj. besedilne analize sentimenta. Te metode se sorazmerno pogosto uporabljajo v poslovnih financah, ne pa tudi v mednarodnih financah, še posebej pa ne v povezavi z bonitetnimi ocenami držav, kjer takšnih študij skorajda ni. Tako zapolnjujemo pomembno praznino v literaturi. Poleg mer besedilnega sentimenta, ki so že prisotne v literaturi besedilnih analiz, uvajamo tudi mere besedilne subjektivnosti. Menimo, da je to edinstven pristop k zaznavanju mnenja v bonitetnih ocenah držav.

Poleg tega menimo, da ta disertacija pomembno prispeva k razumevanju bonitetnih ocen držav. To ima praktične posledice za udeležence na trgu, od zasebnih in institucionalnih vlagateljev do posojilojemalcev. Bonitetne ocene držav in spremembe le-teh lahko bistveno vplivajo na donose in stroške financiranja, kar še posebej velja za nižje bonitetne razrede. Z uporabo mer besedilnega sentimenta in subjektivnosti lahko izboljšamo predvidljivost sprememb bonitetnih ocen držav, kar izdajateljem omogoča, da prilagodijo časovnico financiranja dolga in zmanjšajo stroške, hkrati pa vlagateljem omogoča, da optimizirajo svoje portfelje in iščejo višje donose.

*Ključne besede: bonitetne ocene držav, poročila o bonitetnih ocenah držav, analiza besedilnega sentimenta, mehke informacije, subjektivnost, prehodi med bonitetnimi ocenami, tržne reakcije*

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF APPENDICES

# Introduction

Sovereign credit ratings play an important role in international finance, as they determine the country's credit risk and consequently affect the issuer's borrowing cost on international financial markets. Indirectly, they also have an impact on the cost of financing of residents, both individual and corporate entities. Sovereign credit ratings are mostly assigned by the three major credit rating agencies, namely Standard & Poor's, Fitch and Moody's, which take into account both quantitative and qualitative factors. They specifically stress that their sovereign credit ratings are only an opinion (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017).

Existing studies examining sovereign credit ratings can predominantly be separated into three groups. The first tries to identify the determinants of sovereign credit ratings (e.g. Afonso, 2003; Afonso, Gomes, & Rother, 2009; Cantor & Packer, 1996; Özturk, 2014). They argue that ratings consist of hard and soft information. The hard part of the rating is mostly based on macroeconomic data, e.g. the level of GDP per capita, real GDP growth, external debt, public debt and the government budget balance. The soft part of the rating is more difficult to measure, but several proxies for political risk and institutional quality efficiently capture the prevailing environment in a particular country. A significant part of these studies argues that credit rating agencies unjustly assign lower ratings to emerging markets relative to their advanced counterparts, resulting in biased sovereign credit ratings (De Moor, Luitel, Sercu, & Vanpée, 2018; Fuchs & Gehring, 2017; Zheng, 2012). The second group studies rating transitions and tries to improve their predictability (e.g. Hill, Brooks, & Faff, 2010; Hu, Kiesel, & Perraudin, 2002). The third group focuses on the impact of sovereign credit ratings on movements in government bond yields, and CDS spreads (Afonso, Furceri, & Gomes, 2012; Drago & Gallo, 2016; Gande & Parsley, 2005; Ismailescu & Kazemi, 2010; Kiff, Novak, & Schumacher, 2012). Consequently, this dissertation consists of three main parts, one corresponding to each group, namely Chapters 4, 5 and 6.

First, while prior literature tried to identify the determinants of sovereign credit ratings using various classical estimation techniques, a large part of the ratings was still left unexplained, and the percentages of correctly predicted sovereign credit ratings were relatively low. We believe that the last (missing) part of the rating represents the qualitative judgement of the rating committee, which they express in sovereign credit rating reports. We thus propose an alternative approach. Due to the lack of prior evidence on the information value of sovereign credit rating reports, the objective is to exploit this under-utilised source of qualitative data to gain new insights into the formation of credit ratings. We will analyse reports issued by Standard & Poor's, Fitch and Moody's credit rating agency by applying textual analysis methods and explore to what extent different sentiment and subjectivity measures relate to the ratings. Taking into account the before mentioned studies on bias in sovereign credit ratings, we focus our analysis on the comparison of two groups of countries: emerging markets and advanced economies. We hypothesise that emerging markets will have higher sentiment and subjectivity scores than advanced economies, because data is limited and potentially unreliable, which leads to the qualitative judgement of the credit rating committee having a more significant role in the rating process. Furthermore, we hypothesise the sentiment and subjectivity scores will likely change after the global financial crisis of 2008, because of increased demand for transparency of the rating process and recent criticism of the credit rating agencies that they inflated particular sovereign credit ratings (Agarwal, Chen, Sim, & Zhang, 2019; Gaillard, 2012), leading to more realistic perceptions of country risk.

Second, changes in sovereign credit ratings have significant economic consequences, as they influence borrowing costs (Alsakka & ap Gwilym, 2013; Eijffinger, 2012). More problems arise due to spillovers across markets (Alsakka & ap Gwilym, 2013). Consequently, predicting future downgrades or upgrades is important for a country as well as its economic and financial partners. This is especially critical for emerging markets, which in general, have relatively low ratings and higher cost of financing. We thus focus specifically on downgrades and upgrades. As already established, traditional approaches to sovereign credit rating analysis are limited as they fail to identify the qualitative judgement of the rating committee. Since the rating change is explained in the sovereign credit rating report, we hypothesise that textual sentiment and subjectivity measures will be able to capture the predominant view, and will thus help in boosting the predictability of downgrades and upgrades.

Third, apart from sovereign credit rating changes, credit rating agencies also issue rating outlooks and watches, which also have the potential to influence bond yields and CDS returns. Many studies note that the markets react to all kinds of rating announcements, especially when ratings cross the border between investment and speculative grade rating classes (Drago & Gallo, 2016; Ismailescu & Kazemi, 2010; Kiff et al., 2012). Naturally, we seek to determine whether markets also react to sovereign credit rating reports, harbouring additional qualitative information, which is an area that has scarcely been explored.

The main objective of this dissertation is to offer a comprehensive analysis of sovereign credit rating reports in terms of textual sentiment and subjectivity, and of their role in determining sovereign credit ratings and changes in the ratings, as well as their relationship with sovereign debt markets.

### Research questions

We aim to address the following research problem: *To what extent do sentiment and subjectivity affect sovereign credit ratings and markets across countries and over time?*

In order to shed more light onto this, we will focus on three main research questions that are all primarily related to sentiment or opinion (subjectivity) extracted from sovereign credit rating reports:

- *To what extent do textual sentiment and subjectivity explain differences between implied and actual credit ratings of advanced economies and emerging markets overall, as well as before and after the global financial crisis?*
- *How can we better model and predict rating transitions using additional textual information?* and
- *Do markets value additional information extracted from sovereign credit rating reports?*

First, we find that both textual sentiment and subjectivity scores help in explaining sovereign credit ratings. Specifically, we conclude that textual sentiment reflects the general perception (economic and political climate) of a country, which can be either negative, positive or even neutral. More importantly, we find that qualitative judgement of the rating committee is actually expressed in the credit rating reports and manifested in the subjectivity score. We also find significant differences in subjectivity scores of advanced economies and emerging markets, suggesting distinct levels of the qualitative judgement of the rating committee applied in the case of both groups of countries.

Second, we observe a superior performance of sentiment measures compared to subjectivity scores in terms of classification accuracy and potential rating transition predictions. The improvement in correct classification of true positives is more evident for downgrades than for upgrades. The impact of subjectivity scores is fairly limited, suggesting that the qualitative judgement of the rating committee does not help in predicting rating transitions.

Third, we generally conclude that textual sentiment and subjectivity scores do not impact CDS spreads in a significant way. Nevertheless, we find limited evidence that subjectivity measures positively influence bond market movements during a positive rating announcement, which suggests that markets occasionally identify additional informational value in sovereign credit rating reports beyond credit ratings alone.

### Scientific contribution

We believe this dissertation provides meaningful scientific contributions, both methodological and practical. First, to our knowledge, we implement the most comprehensive analysis of sovereign credit rating reports thus far by using novel and under-utilised approaches, namely textual sentiment analysis. While these methods have been relatively extensively used in corporate finance, similar studies in international finance, specifically in connection to sovereign credit ratings, are practically non-existent. We thus fill an important gap in the literature by addressing this issue. Furthermore, in addition to textual sentiment measures already present in some of the literature, we introduce textual subjectivity indicators, a novel approach to identifying opinion in sovereign credit ratings. To our knowledge, we are the first to use this definition.

Second, we believe this dissertation significantly contributes to the understanding of sovereign credit ratings, which has practical implications for all market participants, from various investors to borrowers. Sovereign credit ratings and changes in the ratings can considerably affect returns and cost of financing, especially for the lower rating classes. Specifically, a profound understanding of sovereign credit ratings and their formation is important for (i) investors, allowing them to reduce information asymmetries, (ii) borrowing governments, allowing them to implement reforms and take action to ensure a more favourable rating, thus reducing the cost of borrowing, and (iii) financial and other institutions holding government bonds, who usually have strict balance sheet restrictions that depend on the sovereign credit rating. Additionally, the improved ability to predict changes in sovereign credit ratings using textual sentiment and subjectivity indicators allows issuers to adapt their debt financing schedules and reduce the cost of financing, as well as allows investors to optimise their portfolios and seek higher returns.

### Data

The main concepts examined in this dissertation are constructed using the textual analysis approach. Specifically, we apply textual sentiment extraction methods to sovereign credit rating reports and establish three textual sentiment and three subjectivity measures, which form the basis for addressing all three research questions. We apply both dictionary-based methods and machine learning approaches to Rating Action reports and Full Rating reports by Standard & Poor's and Fitch, and Rating Action reports by Moody's.

We obtain sovereign credit ratings from Thomson Reuters Eikon, focusing on long-term foreign currency sovereign ratings assigned by Standard & Poor's, Fitch and Moody's to a wide range of countries, both advanced economies and emerging markets. The countries are rated both as investment and speculative grades.

Additionally, we assemble an exhaustive dataset with macroeconomic and fiscal strength variables (IMF World Economic Outlook Database, World Bank), institutional strength and political risk variables (International Country Risk Guide), and economic and cultural proximity variables (OECD, CEPII, World Religion Data) at annual level for the first part of the dissertation (Chapters 4 and 5). We also acquire daily CDS spreads and other high-frequency financial variables (Thomson Reuters Datastream, Federal Reserve Bank of St. Louis) for the last part of the dissertation (Chapter 6).

**Structure of the dissertation**

The dissertation is structured as follows. Chapter 1 offers an extensive review of existing literature on credit rating agencies, sovereign credit ratings, textual sentiment analysis, and market reactions. In Chapter 2, we present the data and the natural language processing methodological framework. We apply textual sentiment analysis methods to sovereign credit rating reports in Chapter 3. We also look at the differences between credit rating agencies in terms of assigned ratings, as well as textual sentiment and subjectivity measures. We explore the determinants of sovereign credit ratings, the role of textual sentiment and subjectivity, and potential discrepancies between (i) advanced economies and emerging markets, and (ii) the period before and after the global financial crisis in Chapter 4. In Chapter 5, we focus on the predictability of rating transitions or how additional information from sovereign credit rating reports improves the accuracy of predictions. We base the analysis on two distinct rating philosophies, namely point-in-time and through-the-cycle. In Chapter 6, we shift the attention to sovereign CDS markets, specifically the reaction to sovereign credit rating announcements and textual sentiment and subjectivity extracted from the sovereign credit rating reports. We conclude in the last chapter. In the appendix, we introduce additional sets of results not included in the main part of the dissertation and extended abstract in the Slovene language.

# 1 Sentiment and sovereign credit ratings: a review of existing evidence

## 1.1 Credit rating agencies

Credit rating agencies (CRAs) play an important role in domestic and international financial markets as they provide credit risk assessment of governments, financial and non-financial firms and alleviate the information asymmetry problem. Kiff et al. (2010) argue that CRAs' contribution is threefold: (i) they independently assess the ability of issuers to repay their debt obligations, which reduces information cost, widens the span of potential investors and supports market liquidity, i.e. they provide information services; (ii) they monitor issuers, which results in corrective actions in order to avoid downgrades, i.e. they provide monitoring services[1]; and (iii) they give certification to issues when ratings are embedded in regulatory capital requirements and thresholds, i.e. they provide certification services. They note that the ratings only relate to credit risk, whereas other risks (e.g. market or liquidity risk) are not taken into account.

White (2010) states that there are around 150 CRAs, but the three largest, namely Standard & Poor's (S&P), Fitch and Moody's, represent about 95% of the market. The big three have the longest history of ratings, are truly global and are covering the global sovereign credit market, while others cover regions or specific products. Standard & Poor's and Fitch use AAA, AA, A and BBB for investment grade credit ratings, and BB, B, CCC, CC, C and D for speculative grade credit ratings. Moody's uses Aaa, Aa, A and Baa for the first group of ratings, and Ba, B, Caa, Ca and C for the second group. All three use modifiers to further categorise the ratings: Standard & Poor's and Fitch use pluses and minuses (e.g. A+ or A-), while Moody's uses numbers (A1 and A3) (Kiff et al., 2010). Hill et al. (2010) note that, in addition to credit ratings, CRAs also issue 'outlooks' and 'watchlist', signalling a motive for a potential future rating change. They add that outlooks reflect medium-term developments (up to two years), while watchlists focus more on a shorter horizon (up to three months), and that both have strong predictive power of future rating changes.

Given that CRAs assess credit risk and that both individual and institutional investors rely heavily on their assessment, CRAs significantly affect the interest rate on issuer's debt. Due to this important role de Haan and Amtenbrink (2011) argue that, although ratings should be as objective and reliable as possible, their current business model unveils several weaknesses. Specifically, in the early 1970s, credit rating agen-

[1] The monitoring effect theory was first proposed by Boot, Milbourn, and Schmeits (2006)

cies replaced the ‘investor pays’ business model with the ‘issuer pays’ model (White, 2010). The ‘issuer pays’ model requires that issuers pay CRAs for the so-called solicited credit ratings. Although CRAs also issue unsolicited or self-initiated ratings, these rely on publicly available data and are thus less reliable or accurate. The ‘investor pays’ model raises potential conflicts of interest, since CRAs may be tempted to issue more favourable ratings to satisfy their customers (de Haan & Amtenbrink, 2011; Pagano & Volpin, 2010; White, 2010). White (2010) and Pagano and Volpin (2010) explain how this conflict led to the subprime mortgage crisis in 2007-08. On the other hand, de Haan and Amtenbrink (2011) argue that CRAs have to look after their reputation and cannot afford to have their ratings perceived as incredible by investors[2].

Not only were CRAs under scrutiny due to their role in the global financial crisis, but they were also criticised because they were unable to predict the Asian crisis and even deepened the crisis with extensive downgrades in the middle of financial turmoil (de Haan & Amtenbrink, 2011). As a response, Standard & Poor’s focused more on transparency issues and quality of (fiscal) data. A similar story happened in the recent euro area sovereign debt crisis when they aggravated the fiscal problems of Greece, Italy, Spain and Portugal (Kiff et al., 2010).

## 1.2 Determinants of sovereign credit ratings

Governments generally seek credit ratings to gain access to international capital markets, where institutional investors are limited to investing in rated (most often even investment grade rated) securities (Cantor & Packer, 1996). Credit rating agencies estimate countries’ creditworthiness (the relative likelihood that a borrower will default on its obligations) by assigning sovereign credit ratings. The rating committee takes into account key economic factors that, together with some qualitative judgement, determine the creditworthiness in order to assign sovereign credit rating (Reusens & Croux, 2017). According to Kiff et al. (2010), the relative importance of these factors is not fixed but can vary over time, depending on new information and economic environment. Reusens and Croux (2017) point out that even though the CRAs publicly disclose the components of sovereign credit ratings, the rating committee typically makes additional arbitrary modifications to it.

While the official rating methodologies of CRAs are generally similar, some subtle differences demonstrate the role of qualitative assessment and subjectivity in sovereign credit ratings. Credit rating agencies all note that their ratings are merely an opinion.

[2]For a comprehensive overview of the economics of credit rating agencies see Sangiorgi and Spatt (2017).

According to Standard & Poor's (2017), they assess five main factors that form the foundation of their sovereign credit analysis: institutional and governance effectiveness and security risks (reflected in the institutional assessment), economic structure and growth prospects (economic assessment), external liquidity and international investment position (external assessment), fiscal performance and flexibility as well as debt burden (fiscal assessment), and monetary flexibility (monetary assessment). They assign an assessment to each of the five factors on a six-point numerical scale, where 1 is the strongest, and 6 the weakest. Each assessment is based on an array of quantitative factors and qualitative considerations. Despite increased transparency of the (objective) credit rating procedure, they state that qualitative assessment still plays a significant part in the process, as they consider various adjustments, trends and other factors that can cause a deviation from the indicative rating. Similarly, Moody's (2016) claim that the initial sovereign credit rating is based on four key factors: economic strength, institutional strength, fiscal strength, and susceptibility to event risk. The total number of sub-factor indicators is 14. These indicators are mapped to one of 15 ranking categories, ranging from Very High Plus (VH+) to Very Low Minus (VL-). Despite revealing the importance of these factors, they stress that the actual weights may differ due to qualitative reasoning. Fitch's (2017) approach to sovereign credit ratings is also a combination of quantitative and qualitative judgements. They rely on four analytical pillars: structural features, macroeconomic performance, policies and prospects, public finances and external finances, where structural features usually have the highest weights. They employ its 'Sovereign Rating Model' as the starting point for assigning sovereign ratings. It is a multiple regression-based rating model that makes use of historical, current and forward-looking data for 18 key variables. However, since no model perfectly captures all the relevant information, the rating committee also adjusts for factors not reflected by the model. Despite increased demand for transparency of the credit rating process, discrepancies and lack of clarity remain (Eijffinger, 2012; Kiff et al., 2010; Reusens & Croux, 2017).

This suggests that sovereign credit ratings are driven by a combination of hard and soft information, as has been argued in several studies on the determinants of sovereign credit ratings (e.g. Afonso, 2003; Afonso et al., 2009; Butler & Fauver, 2006; Cantor & Packer, 1996; Özturk, 2014).

Cantor and Packer (1996) were the first to investigate the determinants and impact of sovereign credit ratings assigned by Moody's and Standard & Poor's for 49 countries in September 1995. They state that identifying the relationship between CRA's criteria and actual ratings is difficult, in part because some of the criteria are not quantifiable. A total of eight factors are identified as possible determinants in assigning a country's credit rating: per capita income, GDP growth, inflation, fiscal balance, external balance, external debt, level of economic development and default history. They find that all, but fiscal and external balance, appear to play an important role in determining a country's credit rating[3]. They explain the relationship between each determinant and a country's ability and willingness to service its debt. First, per capita income represents the potential tax base of the borrowing country. The greater it is, the greater will be the government's ability to repay debt. It can also be used as a proxy for the level of political stability or other factors. Second, a high rate of GDP growth implies that country's current debt will become easier to service over time. Third, a higher inflation rate indicates there may be structural problems in the government's finances, i.e. the government (central bank) expands monetary base in order to service or pay off the debt in local currency. This can lead to political instability. Fourth, higher external debt correlates to a higher risk of default. Fifth, rating agencies appear to have a threshold for economic development, i.e. once countries reach a certain income or level of development, they may be less likely to default. Finally, default history is important. Therefore a country that has defaulted on debt in the recent past is widely perceived as a high credit risk. Even though fiscal and external balances are not statistically significant in their results, they may still influence the rating, where both, a larger fiscal and external balance, lead to a higher risk of default.

Kiff et al. (2010) state that in addition to assessing the government's ability to repay its debt, credit rating agencies also assess its willingness to repay. This means that they contemplate the risk of default if the government can repay the debt but is unwilling to do so due to overwhelming social and political costs. For this reason, credit rating agencies also gauge several qualitative factors, such as institutional strength, political stability or fiscal and monetary flexibility. They argue that Standard & Poor's and Fitch focus more on willingness to repay compared to Moody's. Butler and Fauver (2006) examine the cross-sectional determinants of sovereign credit ratings and how the efficiency of a country's legal and political institutions affects its sovereign credit rating. Using a broad cross-section of 86 developed and emerging markets in March 2004, their findings suggest that the quality of country's legal and political institutions has a strong positive effect on sovereign credit ratings, whereby controlling for factors such as GDP per capita, inflation, foreign debt per GDP, previous defaults, and general development.

Interestingly, the methodologies of credit rating agencies seem to have changed after the recent financial crisis. Amstad and Packer (2015) find that more attention is now

[3](Afonso, 2003) extends their analysis and confirms these results by analysing credit ratings of 81 developed and developing countries assigned to those countries by Standard & Poor's and Moody's in June 2001.

given to monetary policy regimes, currency internationalisation, financial cycles, event risk and general economic growth.

Vernazza and Nielsen (2015) state that credit ratings comprise of objective and subjective components. They break down the credit ratings to objective and subjective part. The former is a fitted value from an OLS regression of ratings on ten explanatory variables, while the latter is the corresponding residuals. They try to predict short- and long-term defaults and conclude that while the objective part can predict sovereign defaults, the subjective part is not. Specifically, when analysing the probability of default within three years, they find the 'subjective' component is biasing default predictions in the wrong direction.

The general finding of prior research shows that identification of the hard information part of a credit rating is relatively straightforward and based on relevant macroeconomic and fiscal variables. Additionally, to some extent, soft information can be proxied by political risk variables. However, in order to fully identify the soft information content and its effect, a more complex research approach (design) is needed. This part of the ratings to a large extent reflects the sentiment or perception (interpretation) of the rating committee. Thus, two countries with similar exposure to macroeconomic shocks and comparable political risk may have substantially different credit ratings. However, the relative importance of the soft part is ambiguous (Amstad & Packer, 2015; Bruner & Abdelal, 2005; Luitel, Vanpée, & De Moor, 2016). We review the existing literature examining the soft part in the next subsection.

## 1.3 The importance of qualitative judgement in credit ratings

There already exists a fair body of literature examining the qualitative judgement or soft information in sovereign credit ratings. A large part of that literature discusses the impact of qualitative judgement from the perspective of subjectivity that leads to biased sovereign credit ratings.

Prior literature dealing with the so-called 'bias' in sovereign credit ratings refers to the following three types of biases: (i) rating agencies assign a higher rating to their home country relative to foreign countries, i.e. a home bias, (ii) rating agencies favour countries that are close to them, i.e. a proximity bias and (iii) rating agencies underrate developing countries, i.e. a foreign bias (Luitel et al., 2016).

Luitel et al. (2016) compare historical ratings issued by US-based (Moody's, Fitch and Standard & Poor's) and Chinese-based (Dagong) rating agencies and note that the sovereign ratings differ depending on the rating agency home country. More specifically, Dagong rates China higher than Fitch, and Fitch rates the US higher than Dagong. This is supported by Fuchs and Gehring (2017) who compare 143 countries rated by nine different agencies from six countries. They find that agencies assign higher ratings to their home countries, those with similar cultural interests and those to which home countries have the highest risk exposure. Similarly, Zheng (2012) compares ratings by Standard & Poor's and a Chinese agency Dagong. They argue that Dagong tends to rate non-Western countries higher than Standard & Poor's. Different perceptions can explain this discrepancy. On the other hand, Güttler and Wahrenburg (2007) study credit ratings by Moody's and Standard & Poor's and find no evidence of home preference and even detect a more conservative approach to US issuers compared to non-US issuers.

Concerning the proximity bias, Luitel et al. (2016) find evidence that US rating agencies favour countries, which have stronger geopolitical and trade ties with the US, where the same holds for Dagong and East Asian countries, except Japan. Very closely related to home bias is foreign bias since countries connected to the US are mainly the developed countries. Luitel et al. (2016) notice that emerging markets receive relatively low ratings and very frequent rating downgrades, giving as an example the sovereign ratings of the East Asian countries in the period between 1995 and 1999. Studies by Fuchs and Gehring (2017), Gültekin-Karakaş, Hisarcıklılar, and Öztürk (2011), Özturk (2014) and De Moor et al. (2018) also find evidence of the so-called 'foreign bias'. Gültekin-Karakaş et al. (2011) show the discrepancies between developed and developing countries, indicating that rating agencies disfavour emerging markets relative to developed markets. De Moor et al. (2018) find that the qualitative judgement component of a credit rating is substantial, especially for the lower rating classes (thus mainly emerging markets), and that it helps to predict imminent sovereign defaults. Özturk (2014) finds that common language influences sovereign credit ratings upwards.

According to Fuchs and Gehring (2017) and Zheng (2012), the degree of foreign bias in sovereign ratings varies across agencies due to different weights being applied to subjective judgement. Cantor and Packer (1996) state that analysts may face several barriers when assessing country's political and economic status, which is, as Luitel et al. (2016) point out, especially true for emerging markets, where the data is usually limited and of questionable quality. This leads to analysts having to rely more on their subjective judgement for such countries compared to the developed markets.

On the other hand, Afonso and Jalles (2019) and Klusak, Thornton, and Uymaz (2019) explore a different type of bias, namely personal bias. Afonso and Jalles (2019) focus on finance ministers' characteristics in 26 EU countries between 1980 and 2012 and ratings by all three credit rating agencies. They find that certain characteristics, such as the finance minister being female or having a finance or 'hard sciences' degree, lead to a more favourable sovereign credit rating. Similarly, Klusak et al. (2019) ask whether personal connections improve sovereign credit ratings by analysing sovereign credit ratings assigned to European countries by Standard & Poor's between 2000 and 2017. They show that a personal relationship between credit rating agencies' senior executives and the prime minister in a particular country manifests itself in a more positive sovereign credit rating.

Studies in the previous subsection predominantly use macroeconomic explanatory variables, whereas some researchers (see e.g. Mellios & Paget-Blanc, 2006; Özturk, 2014) argue that political risk variables should be taken into account as well in order to avoid bias. Mellios and Paget-Blanc (2006) start by including a corruption perception index as a proxy for political risk, which has a strong influence on ratings. Özturk (2014) builds on this finding by using six different governance indicators for political risk. He explores the inaccuracy (i.e. negative bias) of sovereign credit ratings by examining cross country variations in the quality of institutions. He finds a positive relationship between institutional quality and sovereign credit ratings, which is mainly captured by an indicator of government effectiveness and regulatory framework quality.

A question arises, whether it is possible to disentangle bias and soft information? While some studies do not distinguish between those terms (e.g. Özturk, 2014), others make a clear distinction (e.g. De Moor et al., 2018). We believe the latter approach is more appropriate and thus hypothesise that the soft information represents objectively unobservable factors, but can affect the country's ability to repay its debt, e.g. governance and institutional quality. Several proxies for these factors are available (e.g. ICRG[4], WGI[5]), but are usually based on expert or public opinion and thus subjective. We also identify factors that may affect the rating committee's decision, but by definition, do not affect the country's creditworthiness and are thus potential sources of bias, e.g. economic and cultural proximity. Qualitative judgement of the rating committee is thus defined as a subjective interpretation of soft information and may contain potential bias. For example, if the rating committee takes into account the level of corruption in a particular country when assigning credit ratings, that falls into the soft information category and does not necessarily lead to biased sovereign credit ratings. On the other hand, if the rating committee (albeit unintentionally) weighs in cultural similarities or differences, that is considered as a bias.

[4]International Country Risk Guide by PRS Group.
[5]World Governance Indicators by World Bank.

## 1.4 Sentiment in the credit rating reports or other informational content

Having established that qualitative judgement represents an important component in sovereign credit ratings, we now turn to its measurement and determinants. A body of literature already exists on the impact of sentiment or tone (qualitative information) in corporate credit rating reports on corporate equity valuation (Kearney & Liu, 2014; Loughran & McDonald, 2016). However, the evidence on the impact of sovereign credit rating reports on the sovereign debt market is practically non-existent. Sovereign credit ratings can have economically more important consequences than firm-level credit ratings since they can affect the efficiency and stability of capital markets within and across countries. To our knowledge, only one study applies textual analysis methods to these reports. Agarwal et al. (2019) analyse sovereign credit rating reports issued by Moody's in sovereign credit default swap (CDS) markets across 62 countries from 2003 to 2013 using Naïve Bayesian algorithm. They classify each sentence of every report into different linguistic tone category (positive, negative, and neutral) and find that a negative tone in the reports gives additional information not detected in credit ratings alone. Their finding is substantial, as it shows a new determinant of sovereign credit risk that is not captured by the usual quantitative credit rating analysis. They also sort sentences into six different content categories (macroeconomic, public and external finance, debt dynamics, financial sector, political and institutional, and others) and show that content on negative debt dynamics is the most informative.

Liu (2014) also employs textual analysis but focuses on news announcements (concentration and volume). Specifically, they look at selected European countries[6] most affected by the debt crisis of 2009-2012. Additionally, the relationship between sentiment and sovereign bond yield spreads is analysed. They find evidence that increased media pessimism and higher volume of news give additional information not picked up by existing determinants of yield spreads and that they help predict the widening of yield spreads. No earlier studies exist on the impact of textual sentiment on sovereign yield spreads, but a few related studies examine the effect of credit rating announcements (e.g. Afonso et al., 2012). While exploring media pessimism and its connection to debt markets resulted in significant findings, these are very limited (Liu, 2014). A similar analysis can be done in times of crises when there is abundant news available; otherwise, relevant news stories are scarce. On the other hand, credit rating agencies regularly issue rating reports, thus giving the opportunity to analyse the sentiment in non-crisis times, as well as exploring both positive and negative sentiment.

[6]Namely, Portugal, Italy, Ireland, Greece, Spain or the so-called PIIGS countries.

Broadening the scope to credit rating reports in general, a bit more research exists. Agarwal, Chen, and Zhang (2016) examine Standard & Poor's corporate credit rating action reports using the bag-of-words (dictionary-based) approach and Naïve Bayes algorithm. Evidence suggests that net linguistic tone is negatively related to abnormal returns and can predict rating changes. Kiesel (2016) verifies and extends their analysis by looking at the informational content of 3365 Moody's credit rating reports (rating changes, watchlist announcements, and outlooks). He finds a relation between the tone of credit rating reports and equity or CDS markets, where a negative sentiment in the text results in the negative market reaction. Löffler, Norden, and Rieber (2018) investigate whether and how the linguistic tone of Moody's rating reports affects the stock market in the United States and find significant short-term market impact of net tone. They conclude that investors overreact to the net tone of rating reports.

## 1.5 Rating philosophy

Credit rating agencies have been criticised extensively in the past and accused of being procyclical. Ferri, Liu, and Stiglitz (1999) argue that rating changes were delayed during the East Asian crisis, i.e. were downgraded when it was already too late, causing a deepening of the crisis. On the other hand, the ratings did not increase sufficiently after the crisis, i.e. were upgraded too late. Additionally, Mora (2006) finds evidence of ratings lagging behind financial markets, which is more evidence that credit rating agencies are (or at least were) not as forward-looking as they claim to be. Similarly, Kaminsky and Schmukler (2002) find that downgrades occurred after the markets started crashing. The East Asian crisis failure is not an isolated event, as Kiff et al. (2012) argue that this is also the case of the last financial crisis, especially the downgrades of European sovereigns. Kiff et al. (2010) conclude that the credit rating agencies' intent to smooth the rating changes makes them exposed to procyclical cliff effects.

Credit rating agencies can follow different rating philosophies for the incorporation of macroeconomic effects in credit ratings: through-the-cycle (TTC) and point-in-time (PIT), where the former looks over the whole economic cycle (i.e., longer horizon) and the latter reflects currently available information (i.e., shorter horizon) (Basel Committee Basel Committee on Banking Supervision, 2005). White (2010) contemplates the credit rating agencies' slow adjustment and notes that they provide a long-term perspective, i.e. they rate through the cycle, which means they will always have a delayed response to a persistent downturn or progress.

Kiff et al. (2010) argue that rating stability is one of the objectives of credit rating agencies and note that the goal is to ensure that higher rating grades are more stable compared to lower rating grades. One way of accomplishing this goal is to rate through-the-cycle and thus to avoid procyclicality. Kiff, Kisser, and Schumacher (2013) compare the two approaches and find that ratings are less likely to decline under the TTC approach compared to the PIT approach. As Kiff et al. (2012) point out, ratings are based on the probability that the issuer will withstand potential turmoil and should not be changed unless fundamentals change (through-the-cycle). Micu, Remolona, and Wooldridge (2006) also note that rating decisions are usually not influenced by temporary events and are thus often driven by stale information. Taking this into consideration, a recession or tightening should not cause a downgrade. Kiff et al. (2012) note that susceptibility to cycles affects the rating, but not the current situation (point-in-time). One could argue that in the above examples, the credit rating agencies reacted more in line with the point-in-time philosophy rather than through-the-cycle. Kiff et al. (2013) conclude that it would be optimal if the credit rating agencies followed the TTC approach, but would instantly adjust the rating in the event of a breach in initial forecasts. Kiff et al. (2012) argue that in light of above mentioned excessive downgrades, credit rating agencies established new methodologies that extended the TTC criteria to what they call 'through-the-crisis' criteria, using different hypothetical stress scenarios corresponding to different rating categories. Credit rating agencies use this to determine how much stress governments can endure before defaulting.

Given the above mentioned ambiguity of credit rating agencies assigning ratings following the point-in-time (PIT) or through-the-cycle (TTC) philosophy, we propose two approaches: one representing the point-in-time concept, where we take into account current values only, and one representing the through-the-cycle concept, where we also consider past and future values. If credit rating agencies follow the through-the-cycle approach, the classification accuracy should outperform the point-in-time approach.

## 1.6 Rating transitions and market reactions

A part of existing research tries to estimate transition matrices, specifically, the default probabilities for each rating class and the probabilities of transition between them (e.g. Fuertes & Kalotychou, 2007; Hill et al., 2010; Hu et al., 2002). However, as Hill et al. (2010) argue, the availability of data due to short time series, especially for emerging markets, poses a limitation when conditioning transitions between ratings at the sovereign level. Some researchers address this problem by constructing rating histories to augment the dataset (Fuertes & Kalotychou, 2007; Hu et al., 2002). We believe this approach is potentially problematic because it assumes the underlying model predicting the missing ratings is the true model. Existing evidence shows that most models have limited classification accuracy of correct rating predictions (Özturk, 2014; Reusens & Croux, 2017). Fuertes and Kalotychou (2007) test three alternative estimators of sovereign transition matrices and identify biases. Others avoid this by only focusing on upgrades, downgrades and no change in ratings (Purda, 2007). We take the latter approach, partly because of the above mentioned concerns, but also because we want to focus the attention on textual sentiment measures and their comparison, and not on transition matrices themselves. A potential drawback of this approach is that, on average, the probability of a rating change is higher for lower rating classes (Hill et al., 2010). We control for this by including credit ratings in the model. Additionally, Hill et al. (2010) find that credit watch and outlook have, on average, the relatively strong predictive power of rating changes. We thus include the outlook variable in our analysis. Purda (2007) finds that upgrades are relatively more difficult to predict than downgrades. We thus expect to achieve higher classification accuracy for downgrades compared to upgrades.

Alsakka and ap Gwilym (2010) study the interactions between credit rating agencies. Overall, they find higher probabilities of upgrades and lower probabilities of downgrades for sovereigns with a recent upgrade by another agency and vice versa for recent downgrades. They find that Standard & Poor's exhibits the least dependence on other agencies. They conclude that Moody's appears to be the first mover in sovereign upgrades, whereas Standard & Poor's leads Moody's rating downgrades.

Another important issue is whether sovereign credit ratings or changes in these ratings influence the markets. Several studies exist, mostly focusing on CDS spreads. The CDS spreads measure the market price of creditworthiness, where lower spreads are predominantly associated with higher ratings. A CDS is a financial contract under which an agent buys or sells risk protection against the credit risk associated with a particular reference entity. For a fee, usually expressed as a spread, the protection seller agrees to make a contingent payment to the buyer in the event of a default or other specified credit event. Therefore, the spread can be viewed as a reflection of the market's perception of the reference entity's credit quality (Kiff et al., 2010). One may thus claim that CDS spreads are an efficient and suitable measure of creditworthiness and that credit ratings are redundant.

However, Kiff et al. (2012) argue that credit rating agencies provide added value apart from the already available public information and therefore have a significant role in international markets. On the one hand, CDS spreads are potentially unstable, as they react to a myriad of factors that may change daily. On the other hand, credit ratings are relatively stable since credit rating agencies assign ratings through-the-cycle.

Rating stability is important at a systemic level since rating downgrades (especially from investment to speculative grade) can be related to liquidation and price falls (Eijffinger, 2012). More problems arise due to spillovers across markets (Alsakka & ap Gwilym, 2013).

The empirical findings are somewhat puzzling. Kiff et al. (2010) find that sovereign credit ratings influence markets, although they find stronger evidence of the effects of credit rating warnings, such as outlooks, than rating changes themselves. Nevertheless, they note significant effects of rating changes on markets when ratings cross the investment grade threshold. They also detect a general increase in spreads during the global financial crisis and a widening of the dispersion of spreads at the lowest rating classes after 2007. They note that this may mean additional discrimination among different risk profiles by the market compared to the credit rating agencies, especially among the worst-rated countries. Sy (2002) reaches a similar conclusion. Additionally, as expected, they find a negative relationship between sovereign spreads and ratings, with higher ratings being associated with lower spreads.

Differently, Drago and Gallo (2016) identify significant market reactions to downgrades and upgrades in euro area CDS markets, whereas they find no evidence of market reactions to credit rating warning announcements. Ismailescu and Kazemi (2010) examine the relationship between emerging markets sovereign CDS spreads and rating changes. They observe a significant effect of positive events, but not negative events. Similarly, Afonso et al. (2012) analyse European bond and CDS markets and also find significant market reactions of both bond yields and CDS spreads to rating changes. However, negative events appear to affect markets more than positive events. Kiff et al. (2012) confirm their previous results when they find significant effects of upgrades and downgrades of in and out of investment grade category on CDS spreads. Additionally, they show that negative credit warnings have the most significant impact on CDS spreads.

Interestingly, Rodríguez, Dandapani, and Lawrence (2019) examine the reverse relationship and find that the variation in average sovereign credit ratings in a given year can be explained by average CDS spreads over the previous three years. Additionally, they find that while changes in CDS spreads can predict sovereign credit rating events, rating changes cannot.

Sovereign credit ratings do not affect only debt markets of a particular country, but also other market players, as well as other countries through spillover effects. Brooks, Faff, Hillier, and Hillier (2004) examine the reaction of stock returns to sovereign rating changes and find that while downgrades have a strong negative impact on stock returns, little evidence exists on abnormal returns following upgrades. Gande and

Parsley (2005) study the spillover effects of sovereign credit rating changes on sovereign credit spreads and find asymmetric effects. They observe that negative rating events are related to an increase in spreads, while the effects of positive rating events on sovereign spreads are negligible. Specifically, they find that a one-notch sovereign downgrade is related with a 12 basis point increase in spreads of sovereign bonds of other countries. Similarly, Arezki, Candelon, and Sy (2011) and Drago and Gallo (2016) focus on spillover effects of selected European countries' downgrades by examining sovereign CDS spreads and stock market indices. They find that a sovereign downgrade affects both domestic and other euro area financial markets. Arezki et al. (2011) also note that the consequences depend on the type of announcements, the downgraded country and the credit rating agency.

Furthermore, financial markets may react differently to rating changes made by different credit rating agencies. Brooks et al. (2004) present evidence of an unequal reaction to sovereign credit rating changes of different credit rating agencies. Specifically, they find that only a downgrade by Standard & Poor's or Fitch causes a significant market reaction, whereas only upgrades by Moody's are related to abnormal returns.

Finally, to the best of our knowledge, Agarwal et al. (2019) are the only ones analysing the relationship between sentiment or tone in Moody's reports and sovereign CDS spreads. They argue that CDS spreads are most affected by negative sentiment in relation to the 'debt dynamics' category, which also helps in predicting future downgrades. Turning to corporates, Agarwal et al. (2016) find evidence of a significant negative relationship between net sentiment or tone and abnormal returns. Similarly, Kiesel (2016) finds that the sentiment or tone of the rating report has a significant impact both on stock returns and CDS spreads, especially for negative sentiment, resulting in negative market returns and rising CDS spreads. These studies shed new light on the formation of sovereign credit ratings and highlight the importance of alternative sources of information for investors, issuers, and other users apart from sovereign credit ratings alone.

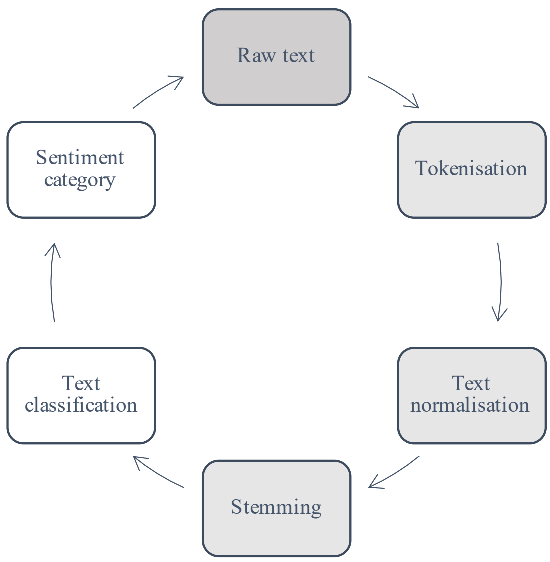
Clearly, rating transitions have significant market consequences. It is therefore vital to develop a more efficient prediction model, which is the sole purpose of Chapter 5. Furthermore, given that there is insufficient existing evidence of the relationship between sovereign CDS spreads, changes in credit ratings and sentiment, we plan to examine the relation further. In Chapter 6, we will first analyse the abnormal CDS spreads around the credit rating event, namely a change in the sovereign credit rating or rating warning, and investigate whether our sentiment analysis measures significantly affect these spreads.

# 2 Data and methodology

## 2.1 Natural language processing

Textual analysis, also known as content analysis, computational linguistics and natural language processing, was first defined by Stone, Dunphy, Smith, and Ogilvie (1966) as any technique that enables inference by objectively and systematically identifying specified characteristics within the text. It has a long history in various disciplines that escalated with the technological development on the one hand and digitalisation and availability of texts on the other. The field of finance has generally been limited to quantitative data and numerical financial analysis. With the growing availability of news articles, corporate filings (e.g. annual reports), and even social media posts (e.g. tweets), the finance field adopted the textual analysis approach as well, turning the attention to qualitative data as well. Many approaches exist, such as content analysis, opinion mining and sentiment analysis. The detailed analysis of various techniques is beyond the scope of this dissertation. The focus is on sentiment analysis. The complete process is shown in Figure 1. The first stage of any textual analysis approach requires several steps, i.e. the raw texts need to be preprocessed (light grey boxes). We go through these steps with examples in the following subsection. At the second stage, we classify cleaned text using text classification approaches described in subsection 2.1.2.

*Figure 1.* The sentiment analysis process




*Source:* Own work

### 2.1.1 Preprocessing

Preprocessing or ‘cleaning the data’ is an essential part of the textual analysis procedure. It entails various techniques in order to transform the initial (unprocessed) text, which is usually not structured or standardised, into standard, well-defined components, which we can use as inputs for further analysis. Thorough preprocessing leads to better results at the text classification stage (e.g. increases classification accuracy), while neglecting the preprocessing stage can result in inefficient and irrelevant output (Bird, Klein, & Loper, 2009; Sarkar, 2016).

Relying on Bird et al. (2009) and Sarkar (2016), we describe the steps taken and provide examples based on the following text:

```
text = "In 2018, Slovenia posted a general government surplus of 0.7% of GDP.
We think that fiscal prudency will remain a political anchor going forward.
Therefore, we project Slovenia will post at least balanced budgetary outcomes
over our forecast horizon."
```

**Tokenisation.** Tokenisation is a process of breaking down the text into tokens. We analyse text at the sentence and word level. Hence we work with the sentence and word tokens. At the first stage, we split the text into sentences based on the period delimiter (.) using the NLTK library in Python:

```
[‘In 2018, Slovenia posted a general government surplus of 0.7\\% of GDP.’,
 ‘We think that fiscal prudency will remain a political anchor going forward.’,
 ‘Therefore, we project Slovenia will post at least balanced budgetary outcomes

over our forecast horizon.’]
```

and, at the second stage, into words :

```
[‘In’, ‘2018’, ‘,’, ‘Slovenia’, ‘posted’, ‘a’, ‘general’, ‘government’,
 ‘surplus’, ‘of’, ‘0.7\\’, ‘%’, ‘of’, ‘GDP’, ‘.’, ‘We’, ‘think’, ‘that’,
 ‘fiscal’, ‘prudency’, ‘will’, ‘remain’, ‘a’, ‘political’, ‘anchor’, ‘going’,
 ‘forward’, ‘.’, ‘Therefore’, ‘,’, ‘we’, ‘project’, ‘Slovenia’, ‘will’, ‘post’,
 ‘at’, ‘least’, ‘balanced’, ‘budgetary’, ‘outcomes’, ‘over’, ‘our’, ‘forecast’,
 ‘horizon’, ‘.’]
```

Word tokenisation is especially important for the remainder of the process, as we clean the text at word level and join it back into sentences for the analysis at the sentence level.

**Text normalisation.** Text normalisation includes a series of steps that further strip the data of unnecessary tokens and characters. We strip each token (i.e. word) by keeping only letters but deleting numbers, special characters and white space, and convert them into lower case. Next, we remove stop words. These are words that have little or no significance, such as 'a', 'the', 'an', etc. In the process we reduce the number of tokens from 45 to 24:

```
['slovenia', 'posted', 'general', 'government', 'surplus', 'gdp', 'think',
 'fiscal', 'prudency', 'remain', 'political', 'anchor', 'going', 'forward',
 'therefore', 'project', 'slovenia', 'post', 'least', 'balanced', 'budgetary',
 'outcomes', 'forecast', 'horizon']
```

**Stemming.** Stemming is a process of reducing the words to stems, i.e. the base form of the word, by getting rid of affixes, such as prefixes, suffixes, etc. This helps us to standardize words, which improves the accuracy of classifiers. The result of the process is:

```
['slovenia', 'post', 'gener', 'govern', 'surplu', 'gdp', 'think', 'fiscal',
 'prudenc', 'remain', 'polit', 'anchor', 'go', 'forward', 'therefor', 'project',
 'slovenia', 'post', 'least', 'balanc', 'budgetari', 'outcom', 'forecast',
 'horizon']
```

The alternative process is lemmatisation, where the words are transformed to root words (lemmas). While word stems are not necessarily lexicographically correct, word roots will always be present in the dictionary. The process is much slower than stemming due to additional steps involved. This is the main reason we prefer stemming to lemmatisation.

### 2.1.2 Text classification

Text classification or categorisation is a vital step in sentiment analysis, as it enables texts to be classified (categorised) in a respective (predefined) sentiment class (category) based on characteristics or features of texts. We can classify words, sentences or complete documents. Several text classification approaches exist, both dictionary-based and machine learning techniques. In this subsection, we review these approaches in general (technical) sense, while an application to sentiment analysis is described in the following chapter.

**Dictionary-based approach.** With dictionary-based methods a computer processes the text and classifies words, phrases or sentences into groups based on a user-defined dictionary or list (Li, 2010). It is also known as the 'bag-of-words' approach because texts can be viewed as the bag of words, and the structure along with any linear ordering of words within the context is ignored (Manning & Schütze, 1999). Sarkar (2016) argues it is one of the simplest but at the same time most powerful methods. He defines it as the process of transforming text into vectors in a way that each document is transformed into a vector, representing the frequency of all specific words that exist in the document vector space for that particular document.

The process of using the dictionary-based approach to extract sentiment from reports is the following. The first step is to obtain the appropriate texts which form the corpus. Next, the dictionary (and categories) and the textual analysis program are selected. Subsequently, assuming that the preprocessing stage has been completed, classification based on the user-defined dictionary is done by running the program. Finally, specific measures are constructed, which, together with other variables, are used for further analysis.

**Machine learning.** Machine learning, pioneered by computer scientists and mathematicians, relies on statistical techniques to infer the content of texts and to classify them based on statistical inference (Li, 2010). Sarkar (2016) defines classification algorithms as supervised machine learning algorithms that classify, categorise or label text based on past observations.

The main steps of the machine learning approach to classify text are: First, a part of the complete corpus of text is specified as the training set. Each sentence in this set is manually categorised, where the categories or classifications are predefined. After preprocessing, a classification algorithm is trained on the training set. The algorithm learns the classification rules (or grammar) from the pre-classified data set. The test set is used to determine the accuracy of classifications. Given that the accuracy is satisfactory, the algorithm applies these rules out-of-sample to the whole corpus.

**Naïve Bayes algorithm.** Our algorithm of choice is Naïve Bayes, which is one of the oldest and most established algorithms for text analysis. In Bayes classification, prior probabilities of the classification are determined based on the training set. They are then used together with additional cases to determine the classification of these cases from the posterior probabilities of them falling into the specified categories (Das, 2014). It is based on the Bayes theorem with an additional (naïve) assumption that each feature is independent of the others. A feature is a unique, measurable characteristic or property for each observation in a dataset (Sarkar, 2016).

Let's denote a response class variable $y$ and a set of $n$ features in the form of a feature vector $\{x_1, x_2, ..., x_n\}$. Using Bayes theorem, the probability of the existence of $y$ given the features as:

$$P(y|x_1, x_2, ..., x_n) = \frac{P(y) \times P(x_1, x_2, ..., x_n|y)}{P(x_1, x_2, ..., x_n)} \tag{1}$$

assuming that $P(x_i|y, x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n) = P(x_i|y)$ and that for all $i$, we can illustrate this as:

$$P(y|x_1, x_2, ..., x_n) = \frac{P(y) \times \prod_{i=1}^{n} P(x_i|y)}{P(x_1, x_2, ..., x_n)} \tag{2}$$

where $i$ ranges from 1 to $n$. Since $P(x_1, x_2, ..., x_n)$ is constant, it follows that $P(y|x_1, x_2, ..., x_n)$ and $P(y) \times \prod_{i=1}^{n} P(x_i|y)$ are proportional. This means that, given the independence assumption among features, the conditional distribution over the class variable, which is to be predicted, $y$ can be represented as:

$$P(y|x_1, x_2, ..., x_n) = \frac{1}{Z} P(y) \times \prod_{i=1}^{n} P(x_i|y) \tag{3}$$

where $Z = p(x)$ is a constant scaling factor dependent on the feature variables. Finally, the classifier is denoted as a function that can assign a predicted class label $\hat{y} = C_k$, for some $k$ as:

$$\hat{y} = \underset{k \in \{1,2,...,K\}}{\arg\max}\ P(C_k) \times \prod_{i=1}^{n} P(x_i|C_k) \tag{4}$$

where $C_k$ is a pre-defined class, $k = 1, 2, ..., K$ and $K$ is the number of classes/categories (Sarkar, 2016).

Sarkar (2016) notes that although the classifier is very simple and based on several assumptions, it still performs surprisingly well in many classification-related problems, including multi-class classifications. Additionally, it trains very fast and offers satisfactory results even when training data is lacking. Finally, it outperforms other models when we have a lot of features (dimensionality problem) since it ensures that each distribution is independently estimated as a single dimension distribution.

## 2.2 Sovereign credit ratings

In Chapters 4 and 5, we focus on sovereign credit ratings: which are the determinants, how they change and, most importantly, how they relate to textual sentiment measures. Sovereign credit ratings, assessing credit risk, are mainly assigned by three most prominent credit rating agencies, namely Standard & Poor's, Moody's and Fitch. Historical credit ratings are available at Thomson Reuters Eikon, with the earliest sovereign credit rating being assigned just before the World War I by Moody's. However, most countries got their first time rating in the 1990s or 2000s. The list of countries included in the analysis is provided in Table 1.

*Table 1.* List of countries in the sample: the upper panel lists countries rated by all credit rating agencies, and the lower panel offers a list of countries rated by one or two credit rating agencies

| **Country** | **Country** | **Country** | **Country** | **Country** |
|---|---|---|---|---|
| Angola | Czech Republic | Israel | Norway | Switzerland |
| Argentina | Denmark | Italy | Pakistan | Thailand |
| Australia | Dominican Republic | Jamaica | Panama | Turkey |
| Austria | Ecuador | Japan | Papua New Guinea | Uganda |
| Azerbaijan | Egypt | Kazakhstan | Paraguay | Ukraine |
| Bangladesh | El Salvador | Kenya | Peru | United Kingdom |
| Belarus | Estonia | Korea | Philippines | United States |
| Belgium | Ethiopia | Latvia | Poland | Uruguay |
| Bolivia | Finland | Lebanon | Portugal | Venezuela |
| Brazil | France | Lithuania | Romania | Vietnam |
| Bulgaria | Germany | Luxembourg | Russia | Zambia |
| Cameroon | Ghana | Malaysia | Saudi Arabia | |
| Canada | Greece | Malta | Serbia | |
| Chile | Guatemala | Mexico | Singapore | |
| China | Hong Kong | Morocco | Slovakia | |
| Colombia | Hungary | Mozambique | Slovenia | |
| Congo, Republic of | Iceland | Netherlands | South Africa | |
| Costa Rica | India | New Zealand | Spain | |
| Croatia | Indonesia | Nicaragua | Sri Lanka | |
| Cyprus | Ireland | Nigeria | Sweden | |
| **S&P** | **Fitch** | | **Moody's** | |
| Albania | Armenia | Tunisia | Albania | Moldova |
| Bahamas, The | Gabon | | Armenia | Namibia |
| Botswana | Gambia, The | | Bahamas, The | Tunisia |
| Gabon | Malawi | | Botswana | |
| Honduras | Moldova | | Honduras | |
| Jordan | Namibia | | Jordan | |

*Source:* Standard & Poor's, Fitch, Moody's, own calculations

We focus on long-term foreign currency sovereign ratings assigned by the three credit rating agencies, namely to 97 countries from 2002 to 2018 by Standard & Poor's, to 98 countries from 1999 to 2018 by Fitch, and to 100 countries from 1995 to 2018 by

Moody's, due to the availability of sovereign credit rating reports for these periods. CRAs generally review the assigned ratings once a year, except under extreme circumstances (e. g. a country is in selective default), and either affirm or change the rating. For this reason we are working with yearly data. For the few occurrences when the ratings are changed more than once in a calendar year, we take the last assigned rating in that year. There are 35 advanced countries in all three samples, and 62 emerging countries in the Standard & Poor's sample, 63 emerging countries in the Fitch sample, and 65 emerging countries in the Moody's sample[7].

*Table 2.* Rating scales of Standard & Poor's, Fitch and Moody's with a corresponding ordinal numerical scale

| S&P | Fitch | Moody's | Num. scale |
|---|---|---|---|
| Credit ratings | | | |
| AAA | AAA | Aaa | 21 |
| AA+ | AA+ | Aa1 | 20 |
| AA | AA | Aa2 | 19 |
| AA- | AA- | Aa3 | 18 |
| A+ | A+ | A1 | 17 |
| A | A | A2 | 16 |
| A- | A- | A3 | 15 |
| BBB+ | BBB+ | Baa1 | 14 |
| BBB | BBB | Baa2 | 13 |
| BBB- | BBB- | Baa3 | 12 |
| BB+ | BB+ | Ba1 | 11 |
| BB | BB | Ba2 | 10 |
| BB- | BB- | Ba3 | 9 |
| B+ | B+ | B1 | 8 |
| B | B | B2 | 7 |
| B- | B- | B3 | 6 |
| CCC+ | CCC+ | Caa1 | 5 |
| CCC | CCC | Caa2 | 4 |
| CCC- | CCC- | Caa3 | 3 |
| CC | CC | Ca | 2 |
| C, SD, D | C, DDD, DD, RD, D | C | 1 |
| Outlook | | | |
| Positive | Positive | Positive | 0.5 |
| Stable | Stable | Stable | 0 |
| Negative | Negative | Negative | -0.5 |

*Source:* Standard & Poor's, Fitch, Moody's, own calculations

The countries are rated both as investment grade (ratings AAA[8]/Aaa[9] through BBB-/Baa3) and speculative grade (ratings BB+/Ba1 through D/C), with AAA/Aaa being the highest possible rating and D/C the lowest. For quantitative analysis, as is accustomed in previous work on credit ratings, the ratings are transformed to an ordinal numerical scale ranging from 1 to 21, with 21 corresponding to AAA/Aaa rating and 1 corresponding to D/C rating. Table 2 shows the rating scales and their numerical

[7]Based on the IMF classification.
[8]Standard & Poor's/Fitch credit rating scale.
[9]Moody's credit rating scale.

counterparts.

The dataset for the first part of the dissertation consists of variables, described in Table 3. In addition to the traditional macroeconomic explanatory variables, we include country risk indicators. To control for potential bias discussed in the existing literature (Fuchs & Gehring, 2017; Gültekin-Karakaş et al., 2011; Luitel et al., 2016; Özturk, 2014), additional variables are added, namely (economic and cultural) proximity variables in line with De Moor et al. (2018). We look at these in more detail in one of the following subsections. The summary statistics is provided in Table 4.

*Table 3.* Analysis of sovereign credit ratings: Definitions and sources of variables

| **Variable** | **Description** | **Source** |
|---|---|---|
| **Macroeconomic and fiscal strength** | | |
| Credit rating | Long-term issuer default rating (foreign) | Thomson Reuters Eikon |
| Downgrade | Dummy variable: 1 if downgrade in year t, zero otherwise | Thomson Reuters Eikon |
| Upgrade | Dummy variable: 1 if upgrade in year t, zero otherwise | Thomson Reuters Eikon |
| Outlook | Credit rating outlook (Positive, Negative, Stable) | Capital IQ, Fitch Connect, Moody's |
| GDP per capita | Nominal GDP in 000 USD divided by midyear population | IMF World Economic Outlook Database |
| Real GDP growth | Yearly real GDP growth rate | IMF World Economic Outlook Database |
| Inflation | Inflation, average consumer prices (year-on-year changes in %) | IMF World Economic Outlook Database |
| Current account/GDP | Current account balance in USD (% of nominal GDP) | IMF World Economic Outlook Database |
| Trade/GDP | External trade of the country in USD (% of nominal GDP) | World Bank |
| External debt | Gross external debt position in USD (% of nominal GDP) | World Bank/World Bank QEDS |
| Economic development | Dummy variable: 1 if a country is classified as advanced by IMF, zero otherwise | IMF |
| Default history | Dummy variable: 1 in the year of default and thereafter, zero otherwise | S&P, Fitch and Moody's sovereign default and recovery rates |
| Log of int. reserves | Natural logarithm of foreign currency reserves in million USD | IMF International Financial Statistics |
| Government debt/GDP | General government gross debt (% of nominal GDP) | IMF World Economic Outlook Database |
| Budget balance/GDP | General government net lending(+)/borrowing(−) - budget surplus or deficit balance in USD (% of nominal GDP) | IMF World Economic Outlook Database |
| **Institutional strength and political risk - Soft information** | | |
| Institutional quality | Composite indicator that includes indicators for law and order, bureaucracy quality, democratic accountability and corruption | International Country Risk Guide Table 3B |
| Governance | Composite indicator that includes indicators for government stability, socio-economic conditions, and investment profile | International Country Risk Guide Table 3B |
| **Economic and cultural proximity - Proxies for bias** | | |
| Trade proximity | Trade intensity of a country with the USA | OECD/WITS |
| Common language | Dummy variable: 1 if English is the common official language, zero otherwise | CEPII |
| Religious proximity | The probability that two randomly chosen individuals in the USA and particular country share the same religion | World Religion Data (Correlates of War) |
| Geographical distance | Physical distance (in km) based on latitude and longitude from Washington DC (U.S.) to the capital city of a country divided by 100 | CEPII |

*Table 4.* Summary statistics for the common and credit rating agency specific variables

| | *Obs* | *Mean* | *St.Dev.* | *Median* | *Min* | *Max* |
|---|---|---|---|---|---|---|
| GDP per capita | 1580 | 17.750 | 20.929 | 7.913 | 0.379 | 120.449 |
| Real GDP growth | 1580 | 0.034 | 0.036 | 0.034 | −0.151 | 0.345 |
| Inflation | 1580 | 0.051 | 0.065 | 0.033 | −0.037 | 0.857 |
| Current account/GDP | 1580 | −0.012 | 0.074 | −0.015 | −0.635 | 0.336 |
| Trade/GDP | 1580 | 0.915 | 0.632 | 0.762 | 0.207 | 4.426 |
| External debt/GDP | 1580 | 1.590 | 5.051 | 0.627 | 0.036 | 67.677 |
| Economic development | 1580 | 0.378 | 0.485 | 0.000 | 0.000 | 1.000 |
| Default history | 1580 | 0.164 | 0.370 | 0.000 | 0.000 | 1.000 |
| Log of int. reserves | 1580 | 9.698 | 1.809 | 9.756 | 4.967 | 15.169 |
| Government debt/GDP | 1580 | 0.539 | 0.352 | 0.453 | 0.001 | 2.371 |
| Budget balance/GDP | 1580 | −0.023 | 0.041 | −0.024 | −0.320 | 0.187 |
| Institutional quality | 1580 | 14.169 | 4.120 | 13.500 | 6.000 | 22.000 |
| Governance | 1580 | 23.319 | 4.240 | 23.083 | 12.375 | 34.000 |
| Trade proximity | 1580 | 0.011 | 0.027 | 0.002 | 0.000 | 0.201 |
| Common language | 1580 | 0.209 | 0.407 | 0.000 | 0.000 | 1.000 |
| Religious proximity | 1580 | 0.529 | 0.237 | 0.636 | 0.011 | 0.805 |
| Geographical distance | 1580 | 78.661 | 35.002 | 73.424 | 0.000 | 163.711 |
| **S&P** | | | | | | |
| Credit rating | 1382 | 13.241 | 5.143 | 13.000 | 1.000 | 21.000 |
| Downgrade | 1382 | 0.115 | 0.319 | 0.000 | 0.000 | 1.000 |
| Upgrade | 1382 | 0.132 | 0.338 | 0.000 | 0.000 | 1.000 |
| Outlook | 1382 | −0.064 | 0.510 | 0.000 | −1.000 | 1.000 |
| **Fitch** | | | | | | |
| Credit rating | 1433 | 13.373 | 5.169 | 13.000 | 1.000 | 21.000 |
| Downgrade | 1433 | 0.096 | 0.295 | 0.000 | 1.000 | 21.000 |
| Upgrade | 1433 | 0.117 | 0.322 | 0.000 | 0.000 | 1.000 |
| Outlook | 1399 | −0.042 | 0.499 | 0.000 | −1.000 | 1.000 |
| **Moody's** | | | | | | |
| Credit rating | 1580 | 13.123 | 5.260 | 12.000 | 1.000 | 21.000 |
| Downgrade | 1580 | 0.088 | 0.283 | 0.000 | 0.000 | 1.000 |
| Upgrade | 1580 | 0.092 | 0.289 | 0.000 | 0.000 | 1.000 |
| Outlook | 1344 | −0.042 | 0.510 | 0.000 | −1.000 | 1.000 |

Common variables: Country-year observations for 100 countries in the period from 1995 to 2018.
S&P: Country-year observations for 97 countries in the period from 2002 to 2018.
Fitch: Country-year observations for 98 countries in the period from 1999 to 2018.
Moody's: Country-year observations for 100 countries in the period from 1995 to 2018.

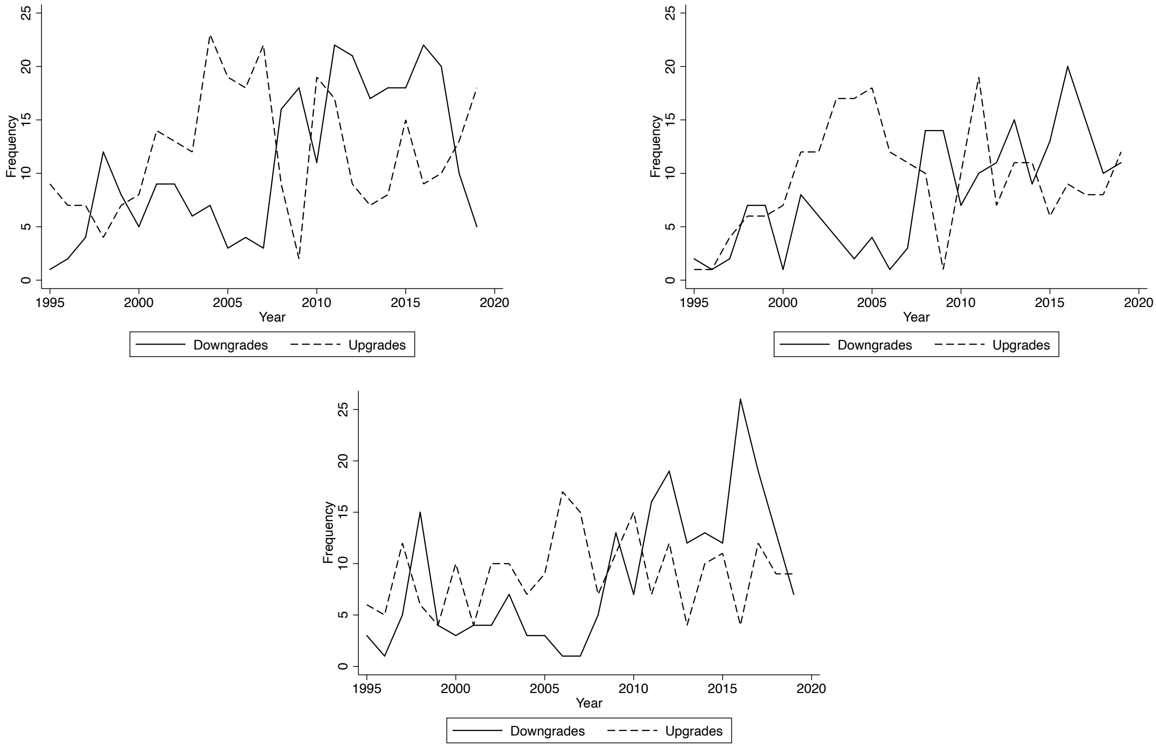
*Source:* Thomson Reuters Eikon, IMF, World Bank, Standard & Poor's, Fitch, Moody's, International Country Risk Guide, OECD/WITS, CEEPI, World Religion Data, own calculations

### 2.2.1 Downgrades and upgrades

Important variables in our models are changes in ratings, so we look at these in more detail. In Figure 2, we show the frequency of downgrades and upgrades per year since 1995. The frequency increases as we move towards the end of the period. This is because, in the 1990s, less (predominantly advanced) countries had a rating. Most of the emerging markets got their first time rating in the 2000s. In 1995, only 52 countries were rated by Standard & Poor's, 33 by Fitch and 53 by Moody's. These numbers increased to 121, 112 and 129 [10] in 2018. We observe a relatively similar pattern by all agencies. A detailed comparison of the timing of rating actions by Standard & Poor's, Fitch and Moody's is beyond the scope of this thesis.

[10]Note that we do not include all these countries in our analysis due to data limitations.

*Figure 2.* Downgrades and upgrades per year by Standard & Poor's (left), Fitch (right) and Moody's (bottom)




*Source:* Thomson Reuters Eikon, own calculations

In Table 5, we present the frequency of total downgrades and upgrades. We make a distinction between changes in ratings by one notch (+/- 1) or more than one notch. On average, in absolute terms, Moody's downgraded countries by more than one notch 59-times, while Standard & Poor's and Fitch only 43- and 44-times, respectively. However, in relative terms, these numbers are comparable. For example, downgrades of more than one notch happened to Greece in 2010. Similarly, for upgrades, there were 25 upgrades of more than one notch by Standard & Poor's, 29 by Fitch and 36 by Moody's. Overall, the samples are relatively balanced. We report 159 downgrades and 207 upgrades by Standard & Poor's, 138 and 197 by Fitch, and 139 and 181 by Moody's.

*Table 5.* Upgrades, downgrades and no-changes of credit ratings

| | Frequency | Percent | Cumulative |
|---|---|---|---|
| | S&P | | |
| Downgrade (>-1) | 43 | 3.11 | 11.51 |
| Downgrade (-1) | 116 | 8.39 | 8.39 |
| No change | 1016 | 73.52 | 85.02 |
| Upgrade (+1) | 182 | 13.17 | 98.19 |
| Upgrade (>1) | 25 | 1.81 | 100.00 |
| Total | 1382 | 100.00 | |
| | Fitch | | |
| Downgrade (>-1) | 44 | 3.07 | 9.63 |
| Downgrade (-1) | 94 | 6.56 | 6.56 |
| No change | 1098 | 76.62 | 86.25 |
| Upgrade (+1) | 168 | 11.72 | 97.98 |
| Upgrade (>1) | 29 | 2.02 | 100.00 |
| Total | 1433 | 100.00 | |
| | Moody's | | |
| Downgrade (>-1) | 59 | 3.73 | 8.80 |
| Downgrade (-1) | 80 | 5.06 | 5.06 |
| No change | 1260 | 79.75 | 88.54 |
| Upgrade (+1) | 145 | 9.18 | 97.72 |
| Upgrade (>1) | 36 | 2.28 | 100.00 |
| Total | 1580 | 100.00 | |

*Source:* Thomson Reuters Eikon, own calculations

### 2.2.2 Economic and cultural proximity

A fair body of existing literature argues that sovereign credit ratings are actually biased, as rating agencies favour their home countries or those close to them and disfavour emerging markets (De Moor et al., 2018; Fuchs & Gehring, 2017; Zheng, 2012). They claim this bias is substantial and mostly downward for emerging markets and upward for advanced countries. To control for potential bias, we include additional variables as proxies for economic and cultural proximity.

Concerning the economic proximity, Luitel et al. (2016) find evidence that US rating agencies favour countries, which have stronger geopolitical and trade ties with the US. We thus construct a variable reflecting the trade intensity of a country with the US in line with De Moor et al. (2018). The rationale behind this is that a more intense trade between the US and the respective country would lead to higher sovereign credit ratings of the latter. The measure is constructed as:

$$Trade\ proximity_{i,t} = \frac{Imports_{USA,i,t} + Exports_{i,USA,t}}{Total\ trade_{USA,t}} \tag{5}$$

where $Imports_{USA,i,t}$ denotes the imports from country $i$ to the USA in year $t$, $Exports_{i,USA,t}$

denotes the exports from the USA to country $i$ in year $t$, and $Total\ trade_{USA,t}$ denotes the total imports to USA and exports from USA in year $t$.

Studies by Fuchs and Gehring (2017); Gültekin-Karakaş et al. (2011); Özturk (2014) and De Moor et al. (2018) also find evidence of cultural proximity bias. We thus include three cultural proximity variables: a dummy variable that equals one if English is the common official language and zero otherwise, the probability that two randomly chosen individuals in the USA and particular country share the same religion, and geographical distance (in km) based on latitude and longitude from Washington DC (US) to the capital city of a particular country. For religious proximity, we take into account four major religious groups: Christianity, Islam, Judaism and others. The probability is then calculated using the following formula:

$$Religious\ proximity(r,s) = \sum_{w=1}^{4} p(r,w) \cdot p(s,w) \tag{6}$$

where $p(r,w)$ denotes the share of population in country $r$ that identifies as belonging to religion $w$ and $s = USA$.

## 2.3 Sovereign credit default swaps

In Chapter 6, we turn the attention to sovereign bond markets and their perception of textual sentiment or subjectivity in sovereign credit rating reports. Specifically, we analyse sovereign credit default swap spreads as opposed to sovereign bond yields because the former are typically more liquid, which leads to more accurate estimates of returns (Longstaff, Pan, Pedersen, & Singleton, 2011). Additionally, CDS spreads are a more direct measure of sovereign risk than sovereign bond spreads, since the latter are also affected by interest rates, changes in the supply of underlying assets and illiquidity effects in debt prices (Ang & Longstaff, 2013). A CDS is an insurance contract that provides protection in the event of default by the reference entity, particularly a sovereign entity. The periodic payment made by the CDS buyer to the CDS seller is expressed as a percentage (usually basis points) of the contract's notional value and is known as the CDS spread or the CDS premium (Ismailescu & Kazemi, 2010). We restrict the sample to 5-year CDS contracts and thus ensure that we include only the most liquid contracts among all maturities (Micu et al., 2006). We primarily use USD denominated contracts and euro denominated contracts, when USD denominated contracts are unavailable.

We collect daily sovereign CDS quotes of mid-premium (average between the bid and ask) from Thomson Reuters Datastream, ranging from December 14th 2007 to April 23rd 2020. The initial sample includes CDS data for 69 countries, specifically for 32 advanced and 37 emerging countries. Table 6 shows average CDS spreads by rating class and by agency. Average CDS spread for the highest rating class AAA/Aaa are relatively close, ranging from 34.07 basis points for Fitch to 35.86 for Moody's. The CDS spreads diverge for other rating classes, especially for the lower rating classes below B/B2, where the spreads range between 698.64 basis points for Moody's and 1055.24 basis points for Fitch.

*Table 6.* Average sovereign CDS spreads by rating class and credit rating agency, in basis points

| Rating | S&P | Fitch | Moody's | Average |
|---|---|---|---|---|
| AAA/Aaa | 35.46 | 34.07 | 35.86 | 35.13 |
| AA+/Aa1 | 38.64 | 51.72 | 48.01 | 46.12 |
| AA/Aa2 | 53.25 | 46.20 | 71.25 | 56.90 |
| AA-/Aa3 | 75.18 | 84.03 | 71.40 | 76.87 |
| A+/A1 | 74.86 | 91.03 | 93.89 | 86.59 |
| A/A2 | 132.76 | 109.64 | 125.13 | 122.51 |
| A-/A3 | 119.19 | 102.02 | 124.93 | 115.38 |
| BBB+/Baa1 | 154.52 | 150.00 | 174.97 | 159.83 |
| BBB/Baa2 | 170.03 | 156.85 | 151.34 | 159.41 |
| BBB-/Baa3 | 184.79 | 211.94 | 196.25 | 197.66 |
| BB+/Ba1 | 242.17 | 262.35 | 234.21 | 246.24 |
| BB/Ba2 | 251.70 | 265.53 | 239.10 | 252.11 |
| BB-/Ba3 | 372.73 | 322.03 | 296.40 | 330.38 |
| B+/B1 | 435.98 | 557.21 | 428.74 | 473.98 |
| B/B2 | 910.44 | 1055.24 | 698.64 | 888.10 |
| B-/B3 | 2321.05 | 1354.72 | 799.08 | 1491.62 |
| CCC+/Caa1 | 2799.12 | n.a. | 2541.15 | 2670.14 |
| CCC/Caa2 | 5932.25 | 7067.50 | 3262.71 | 5420.82 |
| CCC-/Caa3 | 7808.55 | n.a. | 5487.49 | 6648.02 |
| CC/Ca | 3884.83 | 4928.01 | 6407.62 | 5073.49 |
| C or lower | 3741.58 | 2207.14 | 14904.36 | 6951.03 |

*Source:* Thomson Reuters Datastream, own calculations

We shift the analysis from the low frequency of macroeconomic data to the high frequency of financial data. The former is useful for identifying the determinants of sovereign credit ratings and their understanding but is not particularly informative for sovereign credit spreads during a specific event (Blommestein, Eijffinger, & Qian, 2016). We thus use financial data as potential covariates of sovereign CDS spreads, which is in line with Blommestein et al. (2016); Dieckmann and Plank (2011); Fender, Hayo, and Neuenkirch (2012); Fontana and Scheicher (2016); Longstaff et al. (2011); Pan and Singleton (2008).

In addition to textual sentiment measures already defined in Table 3, we use a range of local and global financial factors, described in Table 7. Global financial indicators seem to have a substantial impact on sovereign CDS spreads (Blommestein et al., 2016; Longstaff et al., 2011). Specifically, Longstaff et al. (2011) find that the majority of sovereign credit risk can be linked to global factors. Their principal components analysis (PCA) shows that the first three principal components explain almost 80% of the variation in sovereign credit spreads, where the first principal component is linked to US stock returns and the second to changes in the VIX index. We thus include US excess returns and volatility risk premium in the set of variables.

*Table 7.* Analysis of credit default swaps: Definitions and sources of variables

| **Variable** | **Description** | **Source** |
|---|---|---|
| CDS spread | 5Y credit default swap spread, mid premium | Thomson Reuters Datastream |
| Local stock market return | Local stock market return in USD, calculated from the local MSCI index or, if unavailable, a local stock market index | Thomson Reuters Datastream |
| Exchange rate | The percentage change in the exchange rate of the local currency against the USD | Thomson Reuters Datastream |
| US excess return | The US stock market excess return, calculated as the return on all US stocks (MSCI USA) minus the one-month Treasury bill return | Thomson Reuters Eikon, FRED (Federal Reserve Bank of St. Louis) |
| Volatility risk premium | The change in the volatility risk premium, which is calculated as the difference between the VIX index and a measure of realised volatility for the S&P 100 index[a] | Thomson Reuters Eikon |
| 5Y CMT rate | The change in Treasury yields, based on the 5-year constant maturity treasury (CMT) rates | FRED (Federal Reserve Bank of St. Louis) |
| Investment to speculative grade dummy | Dummy variable: 1 if the rating changes from investment to speculative grade, zero otherwise | Thomson Reuters Eikon |
| Speculative to investment grade dummy | Dummy variable: 1 if the rating changes from speculative to investment grade, zero otherwise | Thomson Reuters Eikon |

[a] The measure of realised volatility for date t is based on the Garman and Klass (1980) open-high-low-close volatility estimator applied to the corresponding data for the S&P 100 index for the 20-day period from date $t-19$ to $t$.

# 3 Sentiment analysis and differences between credit rating agencies

## 3.1 Sentiment analysis

In this section, we apply the natural language processing methodology, described in section 2.1, to sovereign credit rating reports. We introduce six key sentiment and subjectivity measures, as potential proxies for the qualitative judgement of the rating committee, which we use or refer to throughout the dissertation.

Kearney and Liu (2014) define sentiment or tone as the degree of positivity or negativity in texts. They argue that sentiment can include both subjective judgement and objective reflection of economic conditions. The change in a credit rating is typically explained in an elaborate (text) report. Using textual analysis methods, one can analyse the reports and explore to what extent different sentiment measures relate to the ratings.

We collected Rating Action reports and Full Rating reports by Standard & Poor's, available between 2002 and 2018, Rating Action reports and Full Rating reports by Fitch, available between 1999 and 2018, and Rating Action reports by Moody's available between 1995 and 2018. These form the corpus for various textual analysis techniques, including sentiment analysis.

We use the dictionary-based approach using the LM financial dictionary by Loughran and McDonald (2011). Initially, most researchers used well-established dictionaries such as General Inquirer (GI) or DICTION. Kearney and Liu (2014) stress that these are general English language linguistic dictionaries rather than dictionaries that are specific to the finance domain. Loughran and McDonald (2011) find that almost three-quarters of negative words in GI/DICTION are typically not negative in the financial context. They conclude that the use of dictionaries derived outside the finance domain has the potential for errors that are not simply white noise. Consequently, researchers constructed finance specific dictionaries, such as LM, which led to more accurate and efficient sentiment scores. Additionally, like most studies, we apply proportional weighting of words, where every word is assumed to be equally important.

Our measure is the ratio (percentage) of the words in a given sentiment category to the total number of words in the text. We make two assumptions: (i) if more than one report is published in a calendar year, we take the sentiment from the last report in that year (similarly as with more than one sovereign credit rating per year); and (ii) if no reports are published in a calendar year, we assume there was no change in the prevailing sentiment/perception and take the value from the previous year.

We construct two different sentiment measures resulting from the dictionary approach. (Net) sentiment is the difference between positive and negative sentiment, where negative/positive sentiment is calculated as the ratio between the number of negative/positive words in the text and the total number of words. This is the most common measure in studies using the dictionary-based approach with the generic or custom dictionaries and proportional weighting (Kearney & Liu, 2014). Next, we define another relative measure in contrast to the absolute measure (i.e. raw percentage), namely polarity, as:

$$Polarity_{i,t} = \frac{pos_{i,t} - neg_{i,t}}{pos_{i,t} + neg_{i,t}} \tag{7}$$

where $pos_{i,t}$ is the count of positive words, and $neg_{i,t}$ is the count of negative words in the text.

*Table 8.* Sentiment analysis of sovereign credit rating reports: Definitions and sources of variables

| **Variable** | **Description** | **Source** |
|---|---|---|
| Net sentiment (W, dict) | Net textual sentiment/tone, measured as the difference between positive and negative sentiment, in % | S&P, Fitch & Moody's[a] |
| Negative sentiment | Negative textual sentiment/tone, measured as % of negative words in the credit rating report, in % | S&P, Fitch & Moody's[a] |
| Positive sentiment | Positive textual sentiment/tone, measured as % of positive words in the credit rating report, in % | S&P, Fitch & Moody's[a] |
| Polarity (W, dict) | Count of positive words minus the count of negative words, divided by the sum of positive and negative word counts, dictionary approach | S&P, Fitch & Moody's[a] |
| Polarity (S, ML) | Count of positive sentences minus the count of negative sentences, divided by the sum of positive and negative sentences counts, machine learning approach | S&P, Fitch & Moody's[a] |
| Subjectivity (W, dict) | Degree of subjectivity, measured as % of subjective words in the credit rating report, dictionary approach, in % | S&P, Fitch & Moody's[a] |
| Subjectivity (S, dict) | Degree of subjectivity, measured as % of subjective sentences in the credit rating report, dictionary approach | S&P, Fitch & Moody's[a] |
| Subjectivity (S, ML) | Degree of subjectivity, measured as % of subjective sentences in the credit rating report, machine learning approach | S&P, Fitch & Moody's[a] |

[a] Standard & Poor's Full Rating Reports and Rating Action reports, Fitch Full Rating Reports and Rating Action reports, Moody's Rating Action reports

The LM dictionary also includes categories for 'uncertainty' (terms expressing imprecision rather than exclusively focusing on risk, e.g. predictions or forecasts), 'strong modal' and 'weak modal' words (terms expressing levels of confidence). Therefore, as an alternative to polarity (positive/negative sentiment), we also introduce the subjectivity indicator. First, we define a new, broader category for 'subjectivity' that consists of the three before-mentioned categories. We then repeat the process described above using the newly constructed word list and apply the same assumptions. We obtain the subjectivity score, calculated as the ratio between the number of subjective words in the text and the total number of words. To ensure comparability to the machine learning approach, we also construct the subjectivity score at the sentence level, calculated as the ratio between the number of subjective sentences in the text and the total number of sentences. Subjective sentences are defined as sentences that contain at least one word from the 'subjectivity' category. Subjective sentences generally refer to personal opinion, emotion or judgement, whereas the objective refer to factual information. The motivation stems from the fact that qualitative judgement plays an important role in assigning sovereign credit ratings and could potentially be more efficiently detected by analysing subjectivity than simple negative/positive dichotomy. We define qualitative judgement of the rating committee as a subjective interpretation of soft information, which is unobservable and proxied by several indicators but may include potential bias as well. Cantor and Packer (1996) state that analysts may face several barriers when assessing country's political and economic status, which is, as Luitel et al. (2016) point out, especially true for emerging markets, where the data is usually limited and of questionable quality. This leads to analysts having to rely more on their qualitative judgement for such countries compared to the advanced markets. The increased use of qualitative judgement of the rating committee may thus be reflected in a higher subjectivity score and vice versa.

*Table 9.* Summary statistics for the sentiment and subjectivity measures, by credit rating agency

| | *Obs* | *Mean* | *St.Dev.* | *Median* | *Min* | *Max* |
|---|---|---|---|---|---|---|
| | | | **S&P** | | | |
| Net sentiment (W, dict) | 1382 | −1.330 | 2.029 | −0.985 | −8.730 | 4.260 |
| Negative sentiment | 1382 | 4.221 | 1.472 | 3.920 | 1.360 | 10.110 |
| Positive sentiment | 1382 | 2.891 | 1.133 | 2.810 | 0.000 | 6.910 |
| Polarity (W, dict) | 1382 | −0.181 | 0.266 | −0.152 | −1.000 | 0.611 |
| Polarity (S, ML) | 1382 | 0.419 | 0.288 | 0.460 | −1.000 | 1.000 |
| Subjectivity (W, dict) | 1382 | 2.703 | 0.929 | 2.640 | 0.250 | 6.530 |
| Subjectivity (S, dict) | 1382 | 34.233 | 10.040 | 33.330 | 4.170 | 71.150 |
| Subjectivity (S, ML) | 1382 | 34.547 | 9.365 | 33.330 | 9.380 | 67.860 |
| | | | **Fitch** | | | |
| Net sentiment (W, dict) | 1433 | −1.783 | 2.372 | −1.700 | −13.540 | 5.590 |
| Negative sentiment | 1433 | 4.922 | 1.784 | 4.740 | 0.330 | 14.950 |
| Positive sentiment | 1433 | 3.139 | 1.161 | 3.040 | 0.000 | 7.110 |
| Polarity (W, dict) | 1433 | −0.206 | 0.275 | −0.214 | −1.000 | 0.778 |
| Polarity (S, ML) | 1433 | 0.390 | 0.290 | 0.405 | −1.000 | 1.000 |
| Subjectivity (W, dict) | 1433 | 3.135 | 1.188 | 3.120 | 0.000 | 7.790 |
| Subjectivity (S, dict) | 1433 | 36.883 | 12.126 | 37.500 | 0.000 | 88.890 |
| Subjectivity (S, ML) | 1433 | 36.318 | 12.663 | 35.480 | 0.000 | 78.950 |
| | | | **Moody's** | | | |
| Net sentiment (W, dict) | 1580 | −1.156 | 2.526 | −0.870 | −14.290 | 8.760 |
| Negative sentiment | 1580 | 4.146 | 1.947 | 4.050 | 0.000 | 14.290 |
| Positive sentiment | 1580 | 2.990 | 1.381 | 2.900 | 0.000 | 8.760 |
| Polarity (W, dict) | 1580 | −0.138 | 0.349 | −0.143 | −1.000 | 1.000 |
| Polarity (S, ML) | 1580 | 0.286 | 0.457 | 0.363 | −1.000 | 1.000 |
| Subjectivity (W, dict) | 1580 | 2.446 | 1.161 | 2.400 | 0.000 | 7.600 |
| Subjectivity (S, dict) | 1580 | 30.293 | 12.220 | 30.430 | 0.000 | 69.230 |
| Subjectivity (S, ML) | 1580 | 43.504 | 15.295 | 42.110 | 0.000 | 100.000 |

S&P: Country-year observations for 97 countries in the period from 2002 to 2018.
Fitch: Country-year observations for 98 countries in the period from 1999 to 2018.
Moody's: Country-year observations for 100 countries in the period from 1995 to 2018.

*Source:* Standard & Poor's, Fitch, Moody's, own calculations

We obtain the final two sentiment and subjectivity measures using the machine learning approach. We define positive, negative, and neutral categories for sentiment and objective and subjective categories for subjectivity. We use Naïve Bayes algorithm for text classification, which we describe in subsection 2.1.2. After all the sentences in the complete corpus are classified, we construct sentiment and subjectivity measures using the initial classifications or combinations of them. The first is the polarity index, as defined above, where $pos_{i,t}$ is the count of positive sentences and $neg_{i,t}$ is the count of negative sentences in the text. The second is subjectivity, measured as the ratio between the number of subjective sentences in the text and the total number of sentences. Finally, these measures, together with other variables, are used for further analysis.

The summary statistics for textual analysis measures is provided in Table 9. We report the correlations between measures in Table 10. Sentiment/polarity are, on average, negatively correlated with subjectivity measures. We also report correlations with outlook. The correlations suggest that textual sentiment measures will have information value beyond the outlook variable.

*Table 10.* Pairwise correlation coefficients and corresponding significance levels between textual sentiment and subjectivity measures, and outlook, by credit rating agency

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| **Standard & Poor's** | | | | | | | |
| (1) Net sentiment (W, dict) | 1.0000 | | | | | | |
| (2) Polarity (W, dict) | 0.9552*** | 1.0000 | | | | | |
| (3) Polarity (S, ML) | 0.6883*** | 0.6731*** | 1.0000 | | | | |
| (4) Subjectivity (W, dict) | -0.1693*** | -0.1822*** | -0.2232*** | 1.0000 | | | |
| (5) Subjectivity (S, dict) | -0.1443*** | -0.1467*** | -0.2117*** | 0.8683*** | 1.0000 | | |
| (6) Subjectivity (S, ML) | -0.0972*** | -0.0555** | -0.1523*** | 0.2383*** | 0.3639*** | 1.0000 | |
| (7) Outlook | 0.5849*** | 0.5987*** | 0.4735*** | -0.1718*** | -0.1318*** | 0.0202 | 1.0000 |
| **Fitch** | | | | | | | |
| (1) Net sentiment (W, dict) | 1.0000 | | | | | | |
| (2) Polarity (W, dict) | 0.9503*** | 1.0000 | | | | | |
| (3) Polarity (S, ML) | 0.6034*** | 0.5869*** | 1.0000 | | | | |
| (4) Subjectivity (W, dict) | -0.1813*** | -0.1902*** | -0.2182*** | 1.0000 | | | |
| (5) Subjectivity (S, dict) | -0.1684*** | -0.1734*** | -0.2348*** | 0.8543*** | 1.0000 | | |
| (6) Subjectivity (S, ML) | -0.1174*** | -0.0670*** | -0.2152*** | 0.2638*** | 0.3822*** | 1.0000 | |
| (7) Outlook | 0.5003*** | 0.4950*** | 0.3718*** | -0.0521* | -0.0346 | -0.0184 | 1.0000 |
| **Moody's** | | | | | | | |
| (1) Net sentiment (W, dict) | 1.0000 | | | | | | |
| (2) Polarity (W, dict) | 0.9226*** | 1.0000 | | | | | |
| (3) Polarity (S, ML) | 0.5822*** | 0.5909*** | 1.0000 | | | | |
| (4) Subjectivity (W, dict) | -0.2191*** | -0.2132*** | -0.3021*** | 1.0000 | | | |
| (5) Subjectivity (S, dict) | -0.1708*** | -0.1891*** | -0.2636*** | 0.8493*** | 1.0000 | | |
| (6) Subjectivity (S, ML) | 0.0220 | 0.0389 | 0.0246 | -0.0031 | 0.1367*** | 1.0000 | |
| (7) Outlook | 0.5467*** | 0.5067*** | 0.4389*** | -0.1623*** | -0.1529*** | 0.0523* | 1.0000 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Standard & Poor's, Fitch, Moody's, own calculations

The dictionary approach and the machine learning approach have some advantages and disadvantages. Loughran and McDonald (2016) list several important advantages of a dictionary-based approach. By selecting a dictionary, researchers' subjectivity is avoided. Usually, large samples are generated since computer programs tabulate the frequency counts of words. Considering that most dictionaries are publicly available, the replication of other studies is straightforward. As Kearney and Liu (2014) argue, the dictionary approach is likely the easiest for economists and financiers to employ. However, as mentioned above, researchers should use finance-specific dictionaries. By doing so, the main issue is then the choice of a suitable weighting scheme. Additionally, the dictionary-based approach will, on average be less time-consuming and less costly than the machine learning approach, since the text in the 'training set' has to be manually categorised. However, as Li (2010) argues, it is highly likely that there is no existing dictionary for a particular type of text at hand, such as the case of credit rating reports. Even if such a dictionary exists, the dictionary-based approach does not take into consideration the context of a sentence or text. Additionally, the accuracy rate of machine learning is usually higher than the dictionary-based approach. Loughran and McDonald (2016) focus on Naïve Bayes, but their arguments can be generalized. Since machines process the text, large corpora can be included in the analysis. After the classification rules are established, the measuring of sentiment will not be exposed to any additional subjectivity of the researcher. However, they see the decreased transparency of the approach as a weakness because it will be difficult for others to replicate the results.

## 3.2 Differences between credit rating agencies

We look in detail at the differences between agencies in three important periods: at the beginning (1995) and end (2018) of our sample period, and right before the global financial crisis (2007). In 1995, there were 31 countries rated by all three agencies, out of which 14 had differences in ratings between at least two credit rating agencies. The differences by country and agency are presented in Table 11. Five countries were rated higher and five lower by Standard & Poor's than by Fitch. Eight sovereign credit ratings were higher and four lower by Standard & Poor's than by Moody's. Eight countries were assigned a higher rating and three a lower one by Fitch than by Moody's. The analysis shows that Standard & Poor's and Fitch, assigned relatively similar ratings on average, while both assigned slightly higher sovereign credit ratings than Moody's. Overall, the differences in sovereign credit ratings were relatively comparable with similar agreement rates and the maximum differences being +/- 2 rating notches.

Next, we analyse assigned sovereign credit ratings in 2007, right before the global

*Table 11.* Sovereign credit ratings for countries with differences in assigned ratings by the three rating agencies in 1995

| Country | S&P | Fitch | Moody's | S&P-Fitch | S&P-Moody's | Fitch-Moody's |
|---|---|---|---|---|---|---|
| Canada | AA+ | AA | Aa2 | 1 | 1 | 0 |
| Chile | A- | A- | Baa1 | 0 | 1 | 1 |
| Colombia | BBB- | BBB | Baa3 | -1 | 0 | 1 |
| Czech Republic | A | A- | Baa1 | 1 | 2 | 1 |
| Finland | AA- | AA- | Aa2 | 0 | -1 | -1 |
| Hong Kong | A | A+ | A3 | -1 | 1 | 2 |
| Ireland | AA | AA+ | Aa2 | -1 | 0 | 1 |
| Italy | AA | AA- | A1 | 1 | 2 | 1 |
| Norway | AAA | AAA | Aa1 | 0 | 1 | 1 |
| Poland | BB | BB+ | Baa3 | -1 | -2 | -1 |
| Portugal | AA- | AA- | A1 | 0 | 1 | 1 |
| South Africa | BB+ | BB | Baa3 | 1 | -1 | -2 |
| Sweden | AA+ | AA- | Aa3 | 2 | 2 | 0 |
| Turkey | B+ | BB- | Ba3 | -1 | -1 | 0 |
| Average | | | | 0.07 | 0.43 | 0.36 |

*Source:* Thomson Reuters Eikon, own calculations

financial crisis, and before credit rating agencies were heavily criticised and accused of unjustifiably inflating the ratings. Out of 79 countries rated by all three credit rating agencies, 46 had different levels assigned by at least two of the agencies. Table 12 shows that, similarly as in 1995, the differences in ratings between Standard & Poor's and Fitch were still relatively small. Standard & Poor's rated 14 sovereigns higher and 14 lower than Fitch. However, differently than in 1995, where discrepancies occurred, Standard & Poor's and Fitch on average assigned lower ratings than Moody's. Both assigned ratings higher than Moody's in 15 and 14 cases, respectively, but both rated lower in 20 cases. Standard & Poor's and Fitch agree 39% of the time, when Moody's does not, while Moody's agrees with them for approximately 1 in 4 cases when they do not. In total, Standard & Poor's and Fitch agree in 65% of cases, while Moody's agrees with them approximately 56% of the time. Most of the differences among the agencies range between -2 and 2 rating notches, while we detect much higher maximum differences. The obvious outlier is Argentina's sovereign credit rating by Fitch, which is lower by 7 notches than by Standard & Poor's and by 5 notches than by Moody's. This is because Argentina defaulted in 2001, and while Standard & Poor's and Moody's raised the ratings in 2003, Fitch stayed cautious until 2010. Another outlier is Moody's rating of Iceland, which is higher than that of Standard & Poor's and Fitch by 4 rating notches. This discrepancy is present in the previous years as well. The difference dropped to 2 notches in 2008.

Finally, Table 13 shows the differences in sovereign credit ratings between credit rating agencies in 2018, when there were 100 countries rated by all three agencies and 66 had different ratings assigned by at least one agency.

*Table 12.* Sovereign credit ratings for countries with differences in assigned ratings by the three rating agencies in 2007

| Country | S&P | Fitch | Moody's | S&P-Fitch | S&P-Moody's | Fitch-Moody's |
|---|---|---|---|---|---|---|
| Argentina | B+ | C | B3 | 7 | 2 | -5 |
| Australia | AAA | AA+ | Aaa | 1 | 0 | -1 |
| Bulgaria | BBB+ | BBB | Baa3 | 1 | 2 | 1 |
| Chile | A+ | A | A2 | 1 | 1 | 0 |
| China | A | A+ | A1 | -1 | -1 | 0 |
| Colombia | BB+ | BB+ | Ba2 | 0 | 1 | 1 |
| Costa Rica | BB | BB | Ba1 | 0 | -1 | -1 |
| Croatia | BBB | BBB- | Baa3 | 1 | 1 | 0 |
| Cyprus | A | AA- | A1 | -2 | -1 | 1 |
| Czech Republic | A | A | A1 | 0 | -1 | -1 |
| Dominican Republic | B+ | B | B2 | 1 | 1 | 0 |
| Ecuador | B- | CCC | Caa2 | 2 | 2 | 0 |
| El Salvador | BB+ | BB+ | Baa3 | 0 | -1 | -1 |
| Estonia | A | A | A1 | 0 | -1 | - |
| Greece | A | A | A1 | 0 | -1 | -1 |
| Guatemala | BB | BB+ | Ba2 | -1 | 0 | 1 |
| Hungary | BBB+ | BBB+ | A2 | 0 | -2 | -2 |
| Iceland | A+ | A+ | Aaa | 0 | -4 | -4 |
| Israel | A | A- | A2 | 1 | 0 | -1 |
| Italy | A+ | AA- | Aa2 | -1 | -2 | -1 |
| Jamaica | B | B+ | B1 | -1 | -1 | 0 |
| Japan | AA | AA | Aaa | 0 | -2 | -2 |
| Kazakhstan | BBB- | BBB | Baa2 | -1 | -1 | 0 |
| Korea | A | A+ | A2 | -1 | 0 | 1 |
| Kuwait | AA- | AA- | Aa2 | 0 | -1 | -1 |
| Latvia | BBB+ | BBB+ | A2 | 0 | -2 | -2 |
| Malta | A | A+ | A2 | -1 | 0 | 1 |
| Mongolia | BB- | B+ | B1 | 1 | 1 | 0 |
| Morocco | BB+ | BBB- | Ba1 | -1 | 0 | 1 |
| New Zealand | AA+ | AA+ | Aaa | 0 | -1 | -1 |
| Panama | BB | BB+ | Ba1 | -1 | -1 | 0 |
| Papua New Guinea | B+ | B | B1 | 1 | 0 | -1 |
| Peru | BB+ | BB+ | Ba2 | 0 | 1 | 1 |
| Philippines | BB- | BB | B1 | -1 | 1 | 2 |
| Poland | A- | A- | A2 | 0 | -1 | -1 |
| Portugal | AA- | AA | Aa2 | -1 | -1 | 0 |
| Romania | BBB- | BBB | Baa3 | -1 | 0 | 1 |
| Russia | BBB+ | BBB+ | Baa2 | 0 | 1 | 1 |
| Saudi Arabia | AA- | A+ | A1 | 1 | 1 | 0 |
| Slovakia | A | A | A1 | 0 | -1 | -1 |
| Suriname | B+ | B | B1 | 1 | 0 | -1 |
| Taiwan | AA- | A+ | Aa3 | 1 | 0 | -1 |
| Ukraine | BB- | BB- | B1 | 0 | 1 | 1 |
| Uruguay | B+ | BB- | B1 | -1 | 0 | 1 |
| Venezuela | BB- | BB- | B2 | 0 | 2 | 2 |
| Vietnam | BB | BB- | Ba3 | 1 | 1 | 0 |
| Average | | | | 0.13 | -0.17 | -0.30 |

*Source:* Thomson Reuters Eikon, own calculations

*Table 13.* Sovereign credit ratings for countries with differences in assigned ratings by the three rating agencies in 2018

| Country | S&P | Fitch | Moody's | S&P-Fitch | S&P-Moody's | Fitch-Moody's |
|---|---|---|---|---|---|---|
| Angola | B- | B | B3 | -1 | 0 | 1 |
| Azerbaijan | BB+ | BB+ | Ba2 | 0 | 1 | 1 |
| Bahrain | B+ | BB- | B2 | -1 | 1 | 2 |
| Belarus | B | B | B3 | 0 | 1 | 1 |
| Belgium | AA | AA- | Aa3 | 1 | 1 | 0 |
| Brazil | BB- | BB- | Ba2 | 0 | -1 | -1 |
| Bulgaria | BBB- | BBB | Baa2 | -1 | -1 | 0 |
| Chile | A+ | A | A1 | 1 | 0 | -1 |
| Colombia | BBB- | BBB | Baa2 | -1 | -1 | 0 |
| Congo, Republic of | B- | CC | Caa2 | 4 | 2 | -2 |
| Costa Rica | B+ | BB | B1 | -2 | 0 | 2 |
| Croatia | BB+ | BB+ | Ba2 | 0 | 1 | 1 |
| Cyprus | BBB- | BBB- | Ba2 | 0 | 2 | 2 |
| Czech Republic | AA- | AA- | A1 | 0 | 1 | 1 |
| Egypt | B | B | B3 | 0 | 1 | 1 |
| Estonia | AA- | AA- | A1 | 0 | 1 | 1 |
| Ethiopia | B | B | B1 | 0 | -1 | -1 |
| Georgia | BB- | BB- | Ba2 | 0 | -1 | -1 |
| Ghana | B | B | B3 | 0 | 1 | 1 |
| Greece | B+ | BB- | B3 | -1 | 2 | 3 |
| Guatemala | BB- | BB | Ba1 | -1 | -2 | -1 |
| Hong Kong | AA+ | AA+ | Aa2 | 0 | 1 | 1 |
| Iceland | A | A | A3 | 0 | 1 | 1 |
| India | BBB- | BBB- | Baa2 | 0 | -1 | -1 |
| Indonesia | BBB- | BBB | Baa2 | -1 | -1 | 0 |
| Iraq | B- | B- | Caa1 | 0 | 1 | 1 |
| Ireland | A+ | A+ | A2 | 0 | 1 | 1 |
| Israel | AA- | A+ | A1 | 1 | 1 | 0 |
| Italy | BBB | BBB | Baa3 | 0 | 1 | 1 |
| Jamaica | B | B | B3 | 0 | 1 | 1 |
| Japan | A+ | A | A1 | 1 | 0 | -1 |
| Kazakhstan | BBB- | BBB | Baa3 | -1 | 0 | 1 |
| Kenya | B+ | B+ | B2 | 0 | 1 | 1 |
| Korea | AA | AA- | Aa2 | 1 | 0 | -1 |
| Latvia | A | A- | A3 | 1 | 1 | 0 |
| Lithuania | A | A- | A3 | 1 | 1 | 0 |
| Malta | A- | A+ | A3 | -2 | 0 | 2 |
| Mexico | BBB+ | BBB+ | A3 | 0 | -1 | -1 |
| Mongolia | B | B | B3 | 0 | 1 | 1 |
| Morocco | BBB- | BBB- | Ba1 | 0 | 1 | 1 |
| Mozambique | C | C | Caa3 | 0 | -2 | -2 |
| New Zealand | AA | AA | Aaa | 0 | -2 | -2 |
| Nicaragua | B- | B- | B2 | 0 | -1 | -1 |
| Nigeria | B | B+ | B2 | -1 | 0 | 1 |
| Oman | BB | BB+ | Baa3 | -1 | -2 | -1 |
| Pakistan | B | B- | B3 | 1 | 1 | 0 |
| Paraguay | BB | BB+ | Ba1 | -1 | -1 | 0 |
| Peru | BBB+ | BBB+ | A3 | 0 | -1 | -1 |
| Poland | A- | A- | A2 | 0 | -1 | -1 |
| Portugal | BBB- | BBB | Baa3 | -1 | 0 | 1 |
| Russia | BBB- | BBB- | Ba1 | 0 | 1 | 1 |
| Rwanda | B | B+ | B2 | -1 | 0 | 1 |
| Saudi Arabia | A- | A+ | A1 | -2 | -2 | 0 |
| Serbia | BB | BB | Ba3 | 0 | 1 | 1 |
| Slovakia | A+ | A+ | A2 | 0 | 1 | 1 |
| Slovenia | A+ | A- | Baa1 | 2 | 3 | 1 |
| South Africa | BB | BB+ | Baa3 | -1 | -2 | -1 |
| Spain | A- | A- | Baa1 | 0 | 1 | 1 |
| Suriname | B | B- | B2 | 1 | 0 | -1 |
| Turkey | B+ | BB | Ba3 | -2 | -1 | 1 |
| Uganda | B | B+ | B2 | -1 | 0 | 1 |
| Ukraine | B- | B- | Caa1 | 0 | 1 | 1 |
| United States | AA+ | AAA | Aaa | -1 | -1 | 0 |
| Uruguay | BBB | BBB- | Baa2 | 1 | 0 | -1 |
| Vietnam | BB- | BB | Ba3 | -1 | 0 | 1 |
| Zambia | B- | B- | Caa1 | 0 | 1 | 1 |
| Average | | | | -0.14 | 0.17 | 0.30 |

*Source:* Thomson Reuters Eikon, own calculations

The dynamics between agencies in 2018 is more diverse compared to 1995 and 2007. Among the countries with different sovereign credit ratings, Standard & Poor's, on average, assigned lower ratings than Fitch, with only 12 examples of higher ratings and 21 of lower ratings. On the other hand, they were higher than Moody's, at 32 cases with higher ratings versus 20 cases with lower ratings. Fitch assigned higher ratings than Moody's, where 36 sovereign credit ratings were higher and 19 lower than by Moody's. Among the examples, where at least one agency assigned different ratings, Standard & Poor's and Fitch have the highest agreement rate, that is in half the cases, while Moody's agrees with them in only roughly one-fifth of the cases. Overall, the total agreement between Standard & Poor's and Fitch is also higher at 67%, while the agreement rate of Moody's with the two agencies is 48% and 45%, respectively. Similarly, as before, the differences in sovereign credit ratings vary between -2 and 2 rating notches. One exception is The Republic of Congo, rated 4 rating notches higher by Standard & Poor's than Fitch. Congo defaulted the previous year and Fitch, similarly as in the previous Argentinian example, maintained a lower rating longer than Standard & Poor's. Another exception is Greece, which is rated 3 rating notches higher by Fitch than Moody's. The discrepancies originated in 2012, when both Standard & Poor's and Fitch upgraded Greece after its default much faster than Moody's, and dropped to 1 in 2019.

To sum up, we observe a constant rate of agreement between Standard & Poor's and Fitch throughout the years, while the agreement rate between Moody's and the two agencies appears to be diminishing over the years. This may be explained by different approaches, as Standard & Poor's and Fitch estimate probability of default, whereas Moody's estimates expected loss. Alternatively, as Kiff et al. (2010) note, Standard & Poor's attach a relatively higher weight to the willingness to repay, while Moody's focuses more on the ability to repay. There is some evidence that Fitch is more reluctant to increase ratings after a default compared to Standard & Poor's and Fitch. These findings are consistent with Chen, Matousek, Stewart, and Webb (2019), who analyse herding behaviour of credit rating agencies and find that it generally exists towards Standard & Poor's. They argue that this is to be expected since Standard & Poor's is the most established credit rating agency. They also note that herding is more common towards Fitch than Moody's. Furthermore, Alsakka and ap Gwilym (2010) find that Standard & Poor's and Fitch have the lowest frequency of disagreement. Finally, we find that when credit rating agencies disagree, the disagreement is usually within one or two notches, which is in line with Hill et al. (2010). Kiff et al. (2010) note that disagreements across credit rating agencies are mainly due to the use of different factors and weights assigned to these factors, but also stress the importance of qualitative judgement.

Next, we examine whether the differences in ratings between agencies are significant by running a simple OLS regression. Additionally, we also check the responsiveness to each other's ratings. We estimate the following model:

$$Y_t = \alpha + \beta X_t + u_t \tag{8}$$

where $Y_t$ is the dependent variable and $X_t$ the explanatory variable. $\alpha$ and $\beta$ are the parameters to be estimated, with the former representing a difference in ratings between two agencies and the latter the responsiveness of Y's ratings to X's ratings. We test the null hypothesis $H_0 : \alpha = 0$ and $\beta = 1$, which means that there is no fundamental difference in the level of ratings of the agencies and that change in the level of rating by agency X leads to an equivalent change by agency Y. The alternative hypothesis $H_1$ implies a two-sided test.

*Table 14.* Estimation results for ordinary least squares: testing differences in credit ratings between agencies

| | S&P Credit rating | S&P Credit rating | Fitch Credit rating |
|---|---|---|---|
| Fitch Credit rating | 0.993*** | | |
| | (0.014) | | |
| Moody's Credit rating | | 0.969*** | 0.963*** |
| | | (0.010) | (0.013) |
| Constant | 0.036 | 0.598** | 0.785** |
| | (0.289) | (0.260) | (0.344) |
| Observations | 1930 | 1930 | 1930 |
| $R^2$ | 0.975 | 0.963 | 0.964 |
| $H_0 : \beta = 1$, F-statistics | 0.240 | 9.550 | 8.060 |
| $H_0 : \beta = 1$, p-value | 0.627 | 0.003 | 0.005 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

We run three regressions: (i) $Y_t$ = Standard & Poor's ratings and $X_t$ = Fitch ratings, (ii) $Y_t$ = Standard & Poor's ratings and $X_t$ = Moody's ratings, and (iii) $Y_t$ = Fitch ratings and $X_t$ = Moody's ratings. The results are presented in Table 14. Standard & Poor's and Fitch are the only pair that does not have a statistically significant difference in levels of ratings, while both have statistically significant differences in the level of ratings with Moody's. This is consistent with our previous findings. We then test the hypothesis that $\beta = 1$. The responsiveness of S&P ratings to Fitch ratings is not statistically significantly different from 1, indicating that a change in the ratings by one credit rating agency leads to an equivalent change by the other. On the other hand, the tests imply that changes in Moody's ratings do not lead to an equivalent change by S&P or Fitch. Overall, the results suggest there is a somewhat stronger relationship between Standard & Poor's and Fitch than between them and Moody's.

Next, we examine whether these relationships between agencies' credit ratings also exist between the credit rating reports, specifically between the six sentiment and subjectivity measures defined in the previous subsection. We again estimate Equation 8: (i) $Y_t$ = Standard & Poor's sentiment or subjectivity measure and $X_t$ = Fitch sentiment or subjectivity measure, (ii) $Y_t$ = Standard & Poor's sentiment or subjectivity measure and $X_t$ = Moody's sentiment or subjectivity measure, and (iii) $Y_t$ = Fitch sentiment or subjectivity measure and $X_t$ = Moody's sentiment or subjectivity measure. We test the null hypothesis $H_0 : \alpha = 0$ and $\beta = 1$, which means that there is no basic difference in the level of sentiment or subjectivity measures of the agencies and that change in the level of sentiment or subjectivity measure by agency X leads to an equivalent change by agency Y. The results for each measure are presented in Tables 15 through 20. They show that all credit rating agencies have statistically significant differences in levels of all six sentiment and subjectivity measures. Additionally, it seems that a change in any of the sentiment and subjectivity measures by one agency does not lead to an equivalent change by the other credit rating agency. However, the $\beta$ coefficients and $R^2$ for the three sentiment measures suggest that there exists a somewhat stronger relationship between S&P and Fitch, compared to the relationship with Moody's, which is in line with our previous results.

*Table 15.* Estimation results for ordinary least squares: testing differences in net sentiment between agencies

| | S&P Net sentiment (W, dict) | S&P Net sentiment (W, dict) | Fitch Net sentiment (W, dict) |
|---|---|---|---|
| Fitch Net sentiment | 0.407*** | | |
| | (0.029) | | |
| Moody's Net sentiment | | 0.281*** | 0.351*** |
| | | (0.026) | (0.029) |
| Constant | -1.417*** | -1.819*** | -1.309*** |
| | (0.302) | (0.304) | (0.384) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.261 | 0.172 | 0.209 |
| $H_0 : \beta = 1$, F-statistics | 416.660 | 749.360 | 504.81 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

*Table 16.* Estimation results for ordinary least squares: testing differences in polarity (W, dict) between agencies

| | S&P Polarity (W, dict) | S&P Polarity (W, dict) | Fitch Polarity (W, dict) |
|---|---|---|---|
| Fitch Polarity (dict) | 0.448*** | | |
| | (0.031) | | |
| Moody's Polarity (dict) | | 0.246*** | 0.252*** |
| | | (0.032) | (0.027) |
| Constant | -0.179*** | -0.243*** | -0.167*** |
| | (0.038) | (0.039) | (0.045) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.231 | 0.141 | 0.181 |
| $H_0 : \beta = 1$, F-statistics | 313.840 | 558.290 | 749.310 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

*Table 17.* Estimation results for ordinary least squares: testing differences in polarity (S, ML) between agencies

| | S&P Polarity (S, ML) | S&P Polarity (S, ML) | Fitch Polarity (S, ML) |
|---|---|---|---|
| Fitch Polarity (ML) | 0.449*** | | |
| | (0.031) | | |
| Moody's Polarity (ML) | | 0.242*** | 0.266*** |
| | | (0.033) | (0.033) |
| Constant | 0.329*** | 0.392*** | 0.262*** |
| | (0.043) | (0.047) | (0.045) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.258 | 0.189 | 0.206 |
| $H_0 : \beta = 1$, F-statistics | 317.220 | 512.860 | 491.090 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

*Table 18.* Estimation results for ordinary least squares: testing differences in subjectivity (W, dict) between agencies

| | S&P Subjectivity (W, dict) | S&P Subjectivity (W, dict) | Fitch Subjectivity (W, dict) |
|---|---|---|---|
| Fitch Subjectivity | 0.093*** | | |
| | (0.028) | | |
| Moody's Subjectivity | | 0.140*** | 0.087*** |
| | | (0.025) | (0.032) |
| Constant | 1.639*** | 1.596*** | 2.273*** |
| | (0.140) | (0.130) | (0.191) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.177 | 0.197 | 0.302 |
| $H_0 : \beta = 1$, F-statistics | 1051.01 | 1163.94 | 789.46 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

*Table 19.* Estimation results for ordinary least squares: testing differences in subjectivity (S, dict) between agencies

| | S&P Subjectivity (S, dict) | S&P Subjectivity (S, dict) | Fitch Subjectivity (S, dict) |
|---|---|---|---|
| Fitch Subjectivity (LM) | 0.086*** | | |
| | (0.026) | | |
| Moody's Subjectivity (LM) | | 0.114*** | 0.152*** |
| | | (0.032) | (0.040) |
| Constant | 22.590*** | 22.118*** | 26.630*** |
| | (1.794) | (1.809) | (2.185) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.194 | 0.202 | 0.155 |
| $H_0 : \beta = 1$, F-statistics | 1253.790 | 765.380 | 448.900 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

*Table 20.* Estimation results for ordinary least squares: testing differences in subjectivity (S, ML) between agencies

| | S&P Subjectivity (S, ML) | S&P Subjectivity (S, ML) | Fitch Subjectivity (S, ML) |
|---|---|---|---|
| Fitch Subjectivity (ML) | 0.135*** | | |
| | (0.027) | | |
| Moody's Subjectivity (ML) | | 0.104*** | 0.064 |
| | | (0.028) | (0.046) |
| Constant | 29.078*** | 27.555*** | 27.458*** |
| | (1.530) | (2.035) | (2.747) |
| Observations | 1505 | 1505 | 1505 |
| $R^2$ | 0.199 | 0.186 | 0.046 |
| $H_0 : \beta = 1$, F-statistics | 997.680 | 1039.350 | 410.670 |
| $H_0 : \beta = 1$, p-value | 0.000 | 0.000 | 0.000 |

Clustered standard errors in parentheses
Year dummies are included but the estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Source:* Own calculations

# 4 Understanding sovereign credit ratings: text-based evidence from the credit rating reports

## 4.1 Introduction

Sovereign credit ratings are important for a country since they imply its (credit) risk and thus impact government's cost of financing when accessing international financial markets, as well as costs of financing for individuals, residing in a particular country, and firms exposed to a particular country's sovereign risk. Therefore, one would expect that a sovereign credit rating is assigned based on thoroughly developed criteria and supported by data. However, as many studies note, this represents only one part of the rating (Afonso, 2003; Afonso et al., 2009; Butler & Fauver, 2006; Cantor & Packer, 1996; Özturk, 2014). The other part of the rating reflects the qualitative judgement or interpretation of the rating committee. The leading credit rating agencies themselves (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017) note that their sovereign ratings are merely an opinion and are assigned based on various quantitative factors and some qualitative reasoning.

Whereas the majority of studies tried to identify the determinants of sovereign credit ratings using (hard) macroeconomic explanatory variables (e.g. Afonso, 2003; Cantor & Packer, 1996), some prior studies also focused on the soft part of the rating by including explanatory variables that gauge political risk (e.g. Özturk, 2014). Some note that sovereign credit ratings are actually biased, as rating agencies favour their home countries or those close to them and disfavour emerging markets (De Moor et al., 2018; Fuchs & Gehring, 2017; Zheng, 2012). This bias seems substantial and mostly downward for emerging markets and upward for advanced countries.

All prior findings are based on examining the sovereign credit ratings themselves. However, credit ratings are accompanied by a rating report or an elaborate explanation motivating the rating or the rating change. This raises a question, whether it is possible to extract sentiment or tone reflecting the above mentioned qualitative judgement from those reports. To answer this question, we propose a different approach, namely textual analysis, which is becoming more extensively used in corporate finance when it comes to analysing public corporate disclosures/filings, media articles and internet messages (Kearney & Liu, 2014; Loughran & McDonald, 2016).

The main contributions of this chapter are twofold: methodological and practical. First, we build on the existing literature on the identification of the determinants of sovereign credit ratings but extend the traditional approach further by exploiting the textual analysis angle. To our knowledge, only one study (Agarwal et al., 2019) applies a similar approach. We thus include two key groups of variables in our analysis, namely textual sentiment scores and subjectivity scores, which we acquire by applying textual analysis methods to Standard & Poor's Rating Action and Full Reports, Fitch Rating Action and Full Reports, and Moody's Rating Action reports. The former is a group of indicators that measure negativity or positivity in texts, while the latter measures the degree of subjectivity in texts. We speculate that the sentiment score captures the general current perception of a country, while the qualitative judgement of the rating committee manifests itself in the subjectivity indicator. The aim is to answer to what extent does sentiment/subjectivity affect sovereign credit ratings across countries and over time. We examine two main groups of countries, namely emerging markets and advanced economies, due to previously identified biases in the literature. We also explore the behaviour of our key variables before and after the 2008 global financial crisis, since an increased demand for transparency following the crisis may have impacted the manner of delivering information through the reports.

Second, having identified the relative importance of sovereign credit ratings and their effect on countries' cost of financing, we believe this research is important, as it significantly contributes to understanding sovereign credit ratings and their formation. It sheds new light on the characteristics of bond markets, especially for emerging markets, a field that has not been studied thoroughly before. It offers important insights for policymakers, regulators, private and professional investors and financial institutions. A thorough understanding of sovereign credit ratings is crucial for (i) investors to be able to make fully informed investment decisions, (ii) for policymakers to be able to adjust policy measures to obtain a more favourable rating for the country's debt issues, and (iii) for financial institutions who hold a substantial part of government bonds on their balance sheets.

The first objective of this chapter is thus to determine whether sentiment or subjectivity scores offer any additional information not captured by the previously identified determinants of sovereign credit ratings, using a sample of 97 countries for the period of 2002-2018 by Standard & Poor's, 98 countries for the period of 1999-2018 by Fitch, and 100 countries for the period of 1995-2018 by Moody's. We find that soft information proxies greatly enhance the predictability of sovereign credit ratings, which supports the existing findings of Özturk (2014), Mellios and Paget-Blanc (2006) and Haque, Mark, and Mathieson (1998), who argue that including political variables has a strong impact on modelled ratings. We find evidence of economic proximity bias, but no indication of cultural proximity bias. This is partly consistent with previous studies, including De Moor et al. (2018), Luitel et al. (2016) and Fuchs and Gehring

(2017), who detect both proximity and cultural bias.

After adding the indicators for textual sentiment and subjectivity, the results support our initial hypothesis, that textual sentiment mirrors the general opinion of a country and retains explanatory power after including relevant determinants for institutional and political risk, as well as a potential bias. Furthermore, we find evidence that the qualitative judgement of the rating committee is expressed in the credit rating reports and reflected in the subjectivity score, especially when political risk variables are not included in the models.

The second objective of this chapter is to investigate the existence of differences in sentiment or subjectivity measures between emerging and advanced markets. We find significant differences in sentiment for one of the agencies, which implies that there are differences in the general opinion of groups of countries, but not the remaining two agencies. We also find evidence of differences in the subjectivity scores between groups of countries, implying that the rating committee employs qualitative judgement of different magnitude. We also run separate regressions for advanced and emerging markets. Even though the results are not statistically significant, there are substantial differences in coefficients for sentiment and subjectivity between advanced and emerging countries, supporting our previous finding. Additionally, we observe different determinants of sovereign credit ratings for both groups of countries, indicating the application of different weights to the qualitative judgement (Fuchs & Gehring, 2017; Zheng, 2012). The differences are in line with Bissoondoyal-Bheenick (2005) and Afonso (2003). The latter identifies GDP per capita as the most important explanatory variable for advanced countries and external debt for emerging markets. We also find evidence of an upward bias for advanced countries, similar to Gültekin-Karakaş et al. (2011), who argue that high-income (advanced) countries receive more favourable ratings compared to low-income (emerging) countries, holding all else constant.

The final objective is to explore the drivers behind sentiment and subjectivity scores. We find evidence supporting our previous findings since sentiment can be described by soft information, which reflects the general opinion of the country, and potential bias proxies, while subjectivity remains almost unexplained. This is in line with Vernazza and Nielsen (2015), who conclude that the subjective component in credit ratings is detrimental because it seems to be unrelated to the country's true credit risk.

The remainder of the chapter is structured as follows. In subsection 4.2, we describe the methodological framework. Next, in subsection 4.3 we discuss the overall results. More specifically, we examine the differences in sentiment and subjectivity measures between emerging and advanced economies, as well as before and after the onset of the global financial crisis. Furthermore, we investigate the potential determinants of sentiment and subjectivity indicators. Subsection 4.4 concludes.

## 4.2 The model

By definition, the sovereign credit ratings have an ordered structure. Early studies on the determinants of sovereign credit ratings used linear estimation techniques (OLS). This approach is problematic because it ignores the ordered structure of the ratings and assumes the distances between credit rating classes are equal (i.e. the transition from AAA/Aaa to AA+/Aa1 is treated equally as the transition from BBB-/Baa3 - investment grade to BB+/Ba1 - speculative grade). This problem can be avoided by using ordered response models (Afonso, Gomes, & Rother, 2011; Bissoondoyal-Bheenick, 2005; Mora, 2006; Reusens & Croux, 2017).

Subsequent studies predominantly applied either ordered response models (Bissoondoyal-Bheenick, 2005; Hu et al., 2002; Mellios & Paget-Blanc, 2006; Özturk, 2014), or both fixed and/or random effects estimation and ordered response models (Afonso et al., 2011; Erdem & Varli, 2014). We use an ordered logit with random effects[11] that takes into account both the panel structure of the dataset and the ordered nature of sovereign credit ratings (Afonso et al., 2009, 2011; Agresti & Natarajan, 2001; Erdem & Varli, 2014).

We estimate the following model:

$$y^*_{it} = \alpha_i + \beta' x_{it} + \varepsilon_{it} \tag{9}$$

where $y^*_{it}$ is the unobserved latent variable. The final rating is then given by several cut-off points:

$$y_{it} = \begin{cases} Aaa & \text{if} \quad y^*_{it} > c_{20} \\ Aa1 & \text{if} \quad c_{20} > y^*_{it} > c_{19} \\ \quad\vdots & \\ C & \text{if} \quad c_1 > y^*_{it} \end{cases} \tag{10}$$

[11]We chose random effects because no consistent estimator for an ordered logit (or probit) with fixed effects that can explicitly include individual fixed effects is available. Consequently, various estimation approaches were proposed in the literature but offered little guidance on when to use which estimator (Riedl & Geishecker, 2014).

Non-linearity is determined in the cut-off points. The parameters in equation (5) and cut-off points in (6) are estimated using maximum likelihood. By using the ordered logit with random effects, we assume both errors $\varepsilon_i$ and $\mu_{it}$ are normally distributed (Wooldridge, 2002).

The parameter estimates are interpreted the same as coefficients from a standard ordered logistic regression. We report the ordered log-odds (logit) regression coefficients, which are interpreted in the following way: for a one unit increase in the independent variable, the ordinal dependent variable changes by the regression coefficient in the ordered log-odds scale, holding all else constant. In order to get the proportional odds ratios, ordered logit coefficients have to be exponentiated. These can be interpreted in the following way: for a one unit increase in the independent variable, the odds for cases in a group that is greater than $k$ as opposed to less than or equal to $k$ are the proportional odds-times larger.

## 4.3 Results

We begin with a bivariate analysis of our key variables, namely net sentiment and subjectivity measures. The observations are pooled for all three agencies, where the value for each observation is averaged between the three agencies' credit action reports measures. We compare the means of advanced countries and emerging markets and test the differences in means. The results are reported in Panel A of Table 21. We find statistically significant differences in means for both sentiment and subjectivity measures between advanced and emerging economies. More specifically, the average net sentiment in sovereign credit reports for emerging markets is -1.63%, while the average net sentiment for advanced economies is -1.41%. Since the net sentiment is the difference between positive and negative sentiment, the results indicate that negative sentiment prevails over positive sentiment in both groups of countries, but more in emerging markets compared to advanced economies. Both polarity measures are also significantly higher for emerging markets. As expected, this indicates that general perception is more strongly expressed in the reports of emerging markets compared to advanced economies. On the other hand, we expected higher subjectivity scores for emerging markets due to data shortage and consequently, an increased emphasis on qualitative judgement. This is true only for the subjectivity measure from the machine learning approach.

We repeat the analysis separately for individual credit rating agencies. The results are presented in Appendix A. The individual results confirm the significant differences in means of sentiment measures, but the results for subjectivity measures are contradic-

*Table 21.* Bivariate pooled analysis: mean comparison of key variables for advanced economies (AE) and emerging markets (EME), and before and after the Global financial crisis (GFC)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | A: Emerging markets vs. advanced economies | | | | | |
| Mean (EME) | -1.634 | -0.200 | 0.306 | 2.605 | 32.528 | 39.204 |
| Mean (AE) | -1.141 | -0.143 | 0.426 | 2.775 | 33.550 | 37.480 |
| Diff. in means (EME-AE) | -0.493*** | -0.057*** | -0.120*** | -0.170*** | -1.022** | 1.724*** |
| | (0.091) | (0.012) | (0.015) | (0.044) | (0.451) | (0.473) |
| Observations (EME) | 1071 | Observations (Total) | | | 1669 | |
| Observations (AE) | 598 | | | | | |
| | B: Before vs. after the Global financial crisis (GFC) | | | | | |
| Mean (before GFC) | -1.273 | -0.150 | 0.393 | 2.154 | 28.923 | 41.315 |
| Mean (after GFC) | -1.577 | -0.199 | 0.321 | 2.997 | 35.466 | 36.819 |
| Diff. in means (before-after GFC) | 0.304*** | 0.049*** | 0.072*** | -0.843*** | -6.543*** | 4.496*** |
| | (0.101) | (0.013) | (0.016) | (0.039) | (0.467) | (0.523) |
| Observations (before GFC) | 656 | Observations (Total) | | | 1669 | |
| Observations (after GFC) | 1013 | | | | | |
| | C: Emerging markets before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.486 | -0.175 | 0.319 | 2.137 | 28.858 | 42.348 |
| Mean (after GFC) | -1.736 | -0.218 | 0.297 | 2.932 | 35.087 | 37.013 |
| Diff. in means (before-after GFC) | 0.250* | 0.043** | 0.022 | -0.795*** | -6.228*** | 5.335*** |
| | (0.133) | (0.017) | (0.021) | (0.050) | (0.603) | (0.686) |
| Observations (before GFC) | 440 | Observations (Total EME) | | | 1071 | |
| Observations (after GFC) | 631 | | | | | |
| | D: Advanced economies before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -0.837 | -0.099 | 0.543 | 2.191 | 29.055 | 39.213 |
| Mean (after GFC) | -1.313 | -0.168 | 0.360 | 3.105 | 36.092 | 36.500 |
| Diff. in means (before-after GFC) | 0.475*** | 0.069*** | 0.182*** | -0.914*** | -7.037*** | 2.713*** |
| | (0.138) | (0.018) | (0.022) | (0.062) | (0.711) | (0.737) |
| Observations (before GFC) | 216 | Observations (Total AE) | | | 598 | |
| Observations (after GFC) | 382 | | | | | |
| | E: Emerging vs. advanced markets before the Global financial crisis | | | | | |
| Mean (EME) | -1.486 | -0.175 | 0.319 | 2.137 | 28.858 | 42.348 |
| Mean (AE) | -0.837 | -0.099 | 0.543 | 2.191 | 29.055 | 39.213 |
| Diff. in means (EME-AE) | -0.649*** | -0.076*** | -0.223*** | -0.054 | -0.196 | 3.134*** |
| | (0.158) | (0.020) | (0.024) | (0.065) | (0.818) | (0.866) |
| Observations (EME) | 440 | Observations (Total before GFC) | | | 656 | |
| Observations (AE) | 216 | | | | | |
| | F: Emerging vs. advanced markets after the Global financial crisis | | | | | |
| Mean (EME) | -1.736 | -0.218 | 0.297 | 2.932 | 35.087 | 37.013 |
| Mean (AE) | -1.313 | -0.168 | 0.360 | 3.105 | 36.092 | 36.500 |
| Diff. in means (EME-AE) | -0.424*** | -0.050*** | -0.063*** | -0.173*** | -1.006** | 0.513 |
| | (0.108) | (0.014) | (0.018) | (0.046) | (0.447) | (0.515) |
| Observations (EME) | 631 | Observations (Total after GFC) | | | 1013 | |
| Observations (AE) | 382 | | | | | |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

tory. None of the differences in means of Standard & Poor's subjectivity measures are significant. Only the differences in means of sentence level subjectivity measures are significant for Fitch, and the difference in means of word level subjectivity measure for Moody's. We explore the differences between groups of countries further in subsection 4.3.2.

We also compare the means of key variables before and after the 2008 Global financial crisis and report the results in Panel B of Table 21. As expected, we detect a significant difference in sentiment and subjectivity scores. Sentiment measures were higher before the crisis. Specifically, the average net sentiment before the crisis was higher (-1.27%) than after the crisis (-1.58%). This corresponds to a lower average negative sentiment and a higher average positive sentiment before the crisis compared to after the crisis. The average subjectivity scores from the dictionary-based approach are higher after the crisis, e.g. at 3.00% (word level), compared to before the crisis, at 2.15% (word level), reflecting the increased demand for transparency of credit rating agencies' methodologies after the crisis. Focusing on individual agencies, the results for Fitch and Moody's confirm the pooled results, whereas the results for Standard & Poor's are either statistically insignificant or negligible for sentiment measures. We additionally investigate the relationships before and after the crisis in subsection 4.3.3.

Finally, we combine both approaches and compare means of emerging markets and advanced economies before and after the crisis. The results are shown in Panels C, D, E and F of Table 21 and are comparable to the overall sample results in Panels A and B. The most notable result is that textual sentiment measures for emerging markets changed less after the global financial crisis compared to advanced economies (Panels C and D). Another interesting find is that the difference in average net sentiment between emerging and advanced economies decreased after the crisis (Panels E and F). Similarly, the difference in average subjectivity from the dictionary-based approach between emerging and advanced economies is not statistically significant before the crisis but is significant after the crisis (Panels E and F). We observe the opposite for subjectivity from the machine learning approach. The latter also supports our argument of the increased transparency of credit rating agencies after the crisis, who were forced to present a more realistic picture of the advanced economies. For example, Gaillard (2012) argues that, before the European debt crisis, credit rating agencies attached too much weight to both the advanced economy status as well as eurozone membership of Greece. We continue with the overall sample analysis in the next subsection, before taking a closer look at the two country subgroups and subperiods in the subsequent two subsections.

### 4.3.1 Overall sample results

We extend the bivariate comparison to a multivariate analysis by exploring three baseline models. The first model (Model 1) contains only the variables that are considered the hard information, i.e. macroeconomic and fiscal strength variables defined in Table 3. Taking the methodologies (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017) and previous findings (Butler & Fauver, 2006; Özturk, 2014) into account it is evident that soft information plays an important role when assigning sovereign credit ratings. Model 2 is, therefore, an extension of Model 1 with proxies for institutional strength and political risk. To control for a potential bias identified in prior literature (Fuchs & Gehring, 2017; Luitel et al., 2016; Zheng, 2012), we include proxies for cultural and economic proximity in Model 3[12]. The results are presented in Table 22.

In Model 1, GDP per capita, inflation, current account, economic development, default history and government debt seem to have a significant effect on sovereign credit ratings for all three agencies. Additionally, external debt significantly explains part of the variability of Fitch's credit sovereign ratings, while international reserves play a significant role for ratings by Fitch and Moody's. The signs are as expected and found in earlier research. We find that sovereign credit ratings can, to some extent, be described by just a handful of variables. This is in line with previous research, including Cantor and Packer (1996), Afonso (2003) and Hill et al. (2010). After adding proxies for soft information in Model 2, inflation becomes insignificant for Moody's, whereas international reserves now have explanatory power for all three agencies. Out of newly added variables, namely institutional quality and governance, only the latter adds to the explanation of variability in sovereign credit ratings. When controlling for potential bias in Model 3, none but trade proximity are significant. Thus, there does not seem to be any evidence of cultural proximity bias, but the results suggest there is some economic proximity bias present in sovereign credit ratings. Contradictorily, De Moor et al. (2018); Luitel et al. (2016) and Fuchs and Gehring (2017) detect the presence of both types of bias.

[12] We also estimate Model 1 extended with proxies for bias. The results are comparable to the results of Model 1 and are available upon request.

*Table 22.* Estimation results of the ordered logit with random effects for the determinants of sovereign credit ratings for the three model specifications, by credit rating agency

| | Sovereign credit ratings | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| | S&P | Fitch | Moody's | S&P | Fitch | Moody's | S&P | Fitch | Moody's |
| GDP per capita | 0.150*** | 0.159*** | 0.121** | 0.171*** | 0.188*** | 0.150*** | 0.182*** | 0.196*** | 0.161*** |
| | (0.054) | (0.051) | (0.047) | (0.053) | (0.054) | (0.049) | (0.054) | (0.055) | (0.051) |
| Real GDP growth | 3.879 | -0.099 | 0.198 | 0.677 | -2.478 | -1.683 | 0.517 | -2.729 | -1.894 |
| | (3.704) | (2.818) | (2.389) | (3.718) | (3.168) | (2.247) | (3.760) | (3.206) | (2.250) |
| Inflation | -7.559*** | -7.173*** | -4.482* | -7.376*** | -5.668** | -2.783 | -7.163*** | -5.809** | -2.950 |
| | (2.624) | (2.433) | (2.296) | (2.054) | (2.412) | (2.382) | (2.093) | (2.371) | (2.435) |
| Current account/GDP | -5.899*** | -4.348** | -6.067*** | -4.819** | -3.740*** | -4.552** | -4.877** | -3.777*** | -4.753** |
| | (2.226) | (2.054) | (2.170) | (2.163) | (1.437) | (2.097) | (2.068) | (1.415) | (2.046) |
| Trade/GDP | 0.420 | 0.316 | -0.448 | 1.083 | 0.878 | -0.120 | 1.222* | 0.987 | -0.134 |
| | (0.853) | (0.722) | (0.676) | (0.764) | (0.701) | (0.598) | (0.730) | (0.679) | (0.610) |
| External debt/GDP | -0.019 | -0.191** | -0.138 | 0.002 | -0.180** | -0.124 | -0.008 | -0.184** | -0.130* |
| | (0.130) | (0.080) | (0.088) | (0.139) | (0.076) | (0.076) | (0.148) | (0.073) | (0.074) |
| Economic development | 10.199*** | 10.498*** | 8.772*** | 6.289*** | 7.512*** | 5.609*** | 5.684*** | 7.100*** | 5.487*** |
| | (2.065) | (1.643) | (1.678) | (1.936) | (1.768) | (1.665) | (1.944) | (1.788) | (1.689) |
| Default history | -3.881*** | -5.917*** | -6.187*** | -3.292*** | -5.248*** | -5.654*** | -3.102*** | -5.226*** | -5.554*** |
| | (1.198) | (1.106) | (1.216) | (0.992) | (1.017) | (1.228) | (0.953) | (1.007) | (1.225) |
| Log of int. reserves | 0.444 | 0.468* | 0.392* | 0.795** | 0.661** | 0.598*** | 0.615* | 0.555** | 0.461** |
| | (0.315) | (0.253) | (0.208) | (0.350) | (0.265) | (0.198) | (0.355) | (0.255) | (0.194) |
| Government debt/GDP | -9.921*** | -10.017*** | -8.691*** | -7.923*** | -8.276*** | -7.118*** | -8.142*** | -8.407*** | -7.360*** |
| | (1.504) | (1.236) | (1.284) | (1.211) | (1.079) | (1.133) | (1.267) | (1.079) | (1.157) |
| Budget balance/GDP | -0.107 | -0.875 | -0.675 | -3.896 | -5.031 | -5.380 | -4.385 | -5.359 | -5.439 |
| | (3.609) | (3.586) | (3.712) | (3.353) | (3.292) | (3.364) | (3.320) | (3.273) | (3.360) |
| Institutional quality | | | | 0.069 | -0.025 | -0.001 | 0.110 | -0.023 | -0.012 |
| | | | | (0.160) | (0.134) | (0.102) | (0.160) | (0.131) | (0.103) |
| Governance | | | | 0.469*** | 0.445*** | 0.430*** | 0.468*** | 0.452*** | 0.427*** |
| | | | | (0.063) | (0.061) | (0.051) | (0.061) | (0.060) | (0.052) |
| Trade proximity | | | | | | | 68.756*** | 31.837*** | 31.604*** |
| | | | | | | | (22.090) | (10.155) | (8.951) |
| Common language | | | | | | | -0.654 | -1.150 | -0.360 |
| | | | | | | | (0.812) | (0.845) | (0.686) |
| Religious proximity | | | | | | | 1.460 | -0.892 | 0.745 |
| | | | | | | | (1.713) | (1.574) | (1.394) |
| Geographical distance | | | | | | | 0.011 | -0.009 | 0.012 |
| | | | | | | | (0.010) | (0.012) | (0.010) |
| Observations | 1382 | 1433 | 1580 | 1382 | 1433 | 1580 | 1382 | 1433 | 1580 |

Clustered standard errors in parentheses. Cut-off estimates are not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we repeat the analysis after separately[13] adding indicators for (textual) sentiment and subjectivity to each of the three models. The results are shown in Table 23. Overall, sentiment and subjectivity measures have explanatory power in Model 1. If polarity from the machine learning approach in rating action reports increases by one, the ordered log-odds of being in a higher sovereign credit rating category will increase by 1.36 for Standard & Poor's, by 1.31 for Fitch, and by 0.71 for Moody's, holding all other variables constant. This is consistent with Agarwal et al. (2019), who also include sentiment in the analysis of CDS spreads and future downgrades, but do not introduce proxies for soft information and bias, and conclude that negative sentiment contains important information beyond credit ratings alone.

While sentiment measures from the dictionary-based approach become insignificant in Models 2 and 3, sentiment from machine learning approach does not. This result suggests that soft information is (to some extent) reflected in credit action reports via (textual) sentiment, which helps explain sovereign credit ratings if proxies for soft information are not included in the model. Additionally, the proxy for economic proximity bias remains significant in Model 3, and the coefficients are comparable (see Appendix B). Subjectivity is statistically significant (albeit marginally) in all three models for Moody's, which suggests there is some evidence of the credit rating committee's qualitative judgement being expressed in the reports and having explanatory power beyond the included proxies for soft information and potential bias. However, results for Standard & Poor's and Fitch do not confirm these findings. As a robustness check, we also estimate Model 3 using fixed effects, random effects, pooled OLS and ordinal logit, which are standard estimation techniques in prior literature. The results are reported in Appendix C and generally support our main conclusions.

We then examine the accuracy of correct predictions and predictions within one, two and three notches for the three baseline models and their extensions with sentiment and subjectivity measures. The results are presented in Table 24. Overall, Moody's correct predictions outperform correct predictions for Standard & Poor's and Fitch. While predictions within one notch are comparable between all three agencies, predictions within two and three notches for Standard & Poor's and Fitch exceed those of Moody's. By comparing the three baseline models, it is evident that adding soft information variables, specifically political risk and institutional strength variables, greatly improves the accuracy of predictions. The change in correct predictions is comparable for the three agencies, with increases ranging between 7.12 percentage points for Fitch and 7.89 percentage points for Standard & Poor's. Predictions within three notches

[13] We also include indicators for sentiment and subjectivity indicators simultaneously in all models. The results are comparable and available upon request.

*Table 23.* Estimation results of the ordered logit with random effects for the three model specifications for the determinants of sovereign credit ratings, that include sentiment and subjectivity scores, by credit rating agency and type of reports (Rating Action or Full Reports)

| | Sovereign credit ratings | | | | |
|---|---|---|---|---|---|
| | S&P | S&P (Full Reports) | Fitch | Fitch (Full Reports) | Moody's |
| | Model 1 | | | | |
| Net sentiment (W, dict) | 0.052 | 0.313*** | 0.080* | 0.280** | 0.121*** |
| | (0.042) | (0.108) | (0.041) | (0.128) | (0.037) |
| Polarity (W, dict) | 0.760*** | 2.663*** | 0.525 | 1.280* | 0.799*** |
| | (0.293) | (0.739) | (0.349) | (0.706) | (0.279) |
| Polarity (S, ML) | 1.362*** | 5.632*** | 1.311*** | 2.962*** | 0.711*** |
| | (0.293) | (1.063) | (0.333) | (0.787) | (0.192) |
| Subjectivity (W, dict) | -0.152* | -0.242 | -0.098 | -0.206 | -0.327*** |
| | (0.080) | (0.275) | (0.093) | (0.189) | (0.088) |
| Subjectivity (S, dict) | -1.218 | -3.265 | -1.253* | -2.516* | -2.305*** |
| | (0.747) | (2.661) | (0.759) | (1.438) | (0.758) |
| Subjectivity (S, ML) | 1.513* | -2.727 | -0.422 | -1.178 | -0.251 |
| | (0.816) | (2.514) | (0.576) | (2.007) | (0.631) |
| | Model 2 | | | | |
| Net sentiment (W, dict) | 0.033 | 0.266** | 0.025 | 0.119 | 0.044 |
| | (0.048) | (0.104) | (0.042) | (0.129) | (0.037) |
| Polarity (W, dict) | 0.492 | 2.081*** | 0.055 | 0.349 | 0.214 |
| | (0.344) | (0.761) | (0.367) | (0.673) | (0.284) |
| Polarity (S, ML) | 1.042*** | 4.764*** | 1.004*** | 1.963*** | 0.355* |
| | (0.323) | (0.998) | (0.311) | (0.743) | (0.210) |
| Subjectivity (W, dict) | -0.042 | 0.013 | 0.050 | -0.003 | -0.192** |
| | (0.085) | (0.268) | (0.087) | (0.192) | (0.090) |
| Subjectivity (S, dict) | -0.537 | -1.561 | -0.550 | -1.589 | -1.309* |
| | (0.735) | (2.572) | (0.688) | (1.365) | (0.756) |
| Subjectivity (S, ML) | 0.707 | -2.891 | -0.233 | -0.984 | -0.739 |
| | (0.720) | (2.246) | (0.597) | (2.089) | (0.665) |
| | Model 3 | | | | |
| Net sentiment (W, dict) | 0.033 | 0.232** | 0.021 | 0.094 | 0.043 |
| | (0.048) | (0.105) | (0.043) | (0.128) | (0.037) |
| Polarity (W, dict) | 0.458 | 1.828** | 0.042 | 0.212 | 0.214 |
| | (0.347) | (0.765) | (0.370) | (0.666) | (0.285) |
| Polarity (S, ML) | 1.024*** | 4.684*** | 1.010*** | 1.854** | 0.325 |
| | (0.324) | (1.016) | (0.310) | (0.767) | (0.207) |
| Subjectivity (W, dict) | -0.023 | 0.061 | 0.058 | 0.000 | -0.198** |
| | (0.087) | (0.261) | (0.086) | (0.193) | (0.091) |
| Subjectivity (S, dict) | -0.365 | -0.944 | -0.473 | -1.384 | -1.301* |
| | (0.733) | (2.476) | (0.686) | (1.379) | (0.774) |
| Subjectivity (S, ML) | 0.766 | -2.901 | -0.228 | -0.952 | -0.810 |
| | (0.720) | (2.251) | (0.602) | (2.128) | (0.673) |
| Observations | 1382 | 1422 | 1433 | 1232 | 1580 |

Clustered standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results are based on the analysis of the relationship between sovereign credit ratings and macroeconomic, political, economic and cultural proximity variables, depending on model specification, and sentiment and subjectivity measures, but only the coefficients for sentiment and subjectivity measures are reported. Full estimation results for the widest model specification (Model 3) are in Appendix B.

for Standard & Poor's, Fitch and Moody's increase by 8.90, 7.82 and 12.15 percentage points, respectively. Adding the proxies for potential bias only marginally improves the results, most notably correct predictions for Standard & Poor's, which increase to 32.20%. The results are comparable to Reusens and Croux (2017), who estimate a multi-year ordered probit without proxies for soft information and achieve 28% correctly predicted ratings. Özturk (2014) also applies ordered response models and can correctly predict 29.42% of ratings, while separately including proxies for governance to the model improves the predictions to between 31.61% and 42.47%.

*Table 24.* Predictions of estimated models (benchmark specifications and extensions with sentiment and subjectivity measures), by credit rating agency, in %

| | % predicted within n notches | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | S&P | | | | Fitch | | | | Moody's | | | |
| | n=0 | n=1 | n=2 | n=3 | n=0 | n=1 | n=2 | n=3 | n=0 | n=1 | n=2 | n=3 |
| | Model 1 | | | | | | | | | | | |
| No textual indicators | 22.65 | 48.91 | 72.43 | 85.60 | 22.82 | 47.66 | 70.83 | 85.62 | 28.42 | 49.30 | 66.46 | 78.48 |
| Net sentiment (W, dict) | 22.29 | 48.26 | 72.00 | 85.31 | 23.24 | 48.01 | 71.11 | 85.90 | 27.97 | 50.13 | 67.41 | 79.05 |
| Polarity (W, dict) | 22.29 | 47.97 | 71.85 | 85.17 | 23.03 | 48.22 | 71.11 | 85.97 | 27.78 | 49.81 | 67.47 | 78.92 |
| Polarity (S, ML) | 22.29 | 48.26 | 72.79 | 85.17 | 24.35 | 48.78 | 71.60 | 86.04 | 28.42 | 49.81 | 67.15 | 78.99 |
| Subjectivity (W, dict) | 22.50 | 49.06 | 72.94 | 86.03 | 23.73 | 48.50 | 71.60 | 86.11 | 29.30 | 50.76 | 68.54 | 81.08 |
| Subjectivity (S, dict) | 22.36 | 48.99 | 72.65 | 85.75 | 23.59 | 48.36 | 71.39 | 86.39 | 29.18 | 49.87 | 67.41 | 79.30 |
| Subjectivity (S, ML) | 23.23 | 49.06 | 72.87 | 86.11 | 22.89 | 47.52 | 70.76 | 85.55 | 28.35 | 49.11 | 66.39 | 78.42 |
| | Model 2 | | | | | | | | | | | |
| No textual indicators | 30.54 | 64.11 | 86.11 | 94.50 | 29.94 | 59.32 | 82.97 | 93.44 | 35.95 | 62.15 | 79.87 | 90.63 |
| Net sentiment (W, dict) | 30.75 | 63.60 | 85.96 | 94.50 | 29.94 | 59.04 | 82.69 | 93.37 | 35.25 | 61.71 | 79.75 | 90.57 |
| Polarity (W, dict) | 30.10 | 63.53 | 85.96 | 94.57 | 30.08 | 59.39 | 82.97 | 93.44 | 35.57 | 61.90 | 79.68 | 90.70 |
| Polarity (S, ML) | 31.04 | 62.81 | 85.75 | 94.43 | 29.24 | 59.53 | 82.83 | 93.09 | 35.89 | 61.65 | 79.81 | 90.82 |
| Subjectivity (W, dict) | 30.39 | 64.11 | 86.18 | 94.50 | 29.52 | 59.11 | 82.76 | 93.37 | 36.27 | 62.78 | 80.06 | 90.89 |
| Subjectivity (S, dict) | 30.10 | 63.97 | 86.25 | 94.50 | 29.73 | 59.39 | 82.97 | 93.51 | 35.51 | 62.22 | 79.43 | 90.82 |
| Subjectivity (S, ML) | 30.32 | 63.82 | 86.25 | 94.50 | 29.94 | 59.25 | 82.90 | 93.44 | 34.87 | 60.82 | 79.11 | 90.63 |
| | Model 3 | | | | | | | | | | | |
| No textual indicators | 32.20 | 64.83 | 86.32 | 94.65 | 30.15 | 60.50 | 83.53 | 94.28 | 34.75 | 62.28 | 80.13 | 91.33 |
| Net sentiment (W, dict) | 31.84 | 64.54 | 86.18 | 94.57 | 29.80 | 60.36 | 83.32 | 94.35 | 34.68 | 61.77 | 79.81 | 91.33 |
| Polarity (W, dict) | 31.62 | 64.47 | 85.89 | 94.65 | 30.29 | 60.43 | 83.46 | 94.28 | 34.62 | 62.03 | 79.94 | 91.27 |
| Polarity (S, ML) | 32.42 | 64.69 | 85.96 | 94.50 | 29.52 | 60.92 | 83.74 | 94.28 | 35.25 | 61.96 | 80.00 | 91.14 |
| Subjectivity (W, dict) | 32.20 | 64.69 | 86.40 | 94.57 | 29.87 | 60.50 | 83.46 | 94.21 | 35.38 | 61.71 | 80.95 | 91.71 |
| Subjectivity (S, dict) | 32.20 | 64.62 | 86.25 | 94.50 | 30.01 | 60.57 | 83.46 | 94.49 | 35.32 | 61.96 | 80.19 | 91.46 |
| Subjectivity (S, ML) | 32.13 | 64.91 | 86.25 | 94.72 | 30.29 | 60.50 | 83.32 | 94.14 | 35.25 | 61.96 | 79.56 | 91.27 |

Adding sentiment measures to the three models improves the accuracy of predictions to a certain extent for Model 1, most notably for Fitch, which increases to between 23.03% and 24.35%. This is in line with our previous finding that sentiment partly reflects soft information as expressed in the reports, but loses explanatory power when proxies for soft information are added to the model. Consequently, the predictions of Models 2 and 3 with sentiment are comparable to predictions without sentiment, which may be due to the correlation between soft information and sentiment. Similarly, subjectivity measures only slightly affect the accuracy of predictions. The most notable difference is in Model 1 for Fitch, where the per cent of accurately predicted sovereign credit ratings increases from 22.82% to 24.35% when including polarity at the sentence level. The best model in terms of correct predictions seems to be Model 3 with polarity from the machine learning approach for Standard & Poor's (32.42%), Model 3 with polarity from the dictionary-based approach and with subjectivity from the machine learning approach for Fitch (30.29%), and Model 2 with subjectivity from the dictionary-based approach at word level for Moody's (36.27%).

The highest accuracy within three notches is shared by Models 3 with and without polarity from the dictionary-based approach for Standard & Poor's (94.65%), Model 3 with subjectivity from the dictionary-based approach at sentence level for Fitch (94.49%), and Model 3 with subjectivity from the dictionary-based approach at word level for Moody's (91.71%).

### 4.3.2 Advanced economies vs. emerging markets

Given the existing evidence that credit rating agencies unjustifiably assign lower ratings to emerging countries compared to advanced markets (De Moor et al., 2018), we analyse the potential discrepancies in sentiment and subjectivity scores between these groups of countries. We introduce an interaction of economic development with sentiment and subjectivity measures, respectively[14]. The results are presented in Table 25.

The coefficients for sentiment and subjectivity measures correspond to emerging markets, while the interaction is the difference in coefficients between emerging and advanced markets. We focus on the rating action reports by the three agencies, which are directly comparable, but also show the results for full rating reports by Standard & Poor's and Fitch. Interaction terms with sentiment measures are predominantly insignificant, indicating there are negligible differences in measured textual sentiment between both groups of countries. We detect significant differences only for Moody's. These differences remain significant even after expanding the model with proxies for political risk and potential bias, suggesting there is additional information in textual sentiment regarding discrepancies between emerging and advanced markets. This could most likely mean the difference in the general perception of the two groups of countries, but could also mean either additional soft information not captured by the institutional quality and governance or bias we have not controlled for. Interaction with subjectivity measures is generally statistically insignificant for Standard & Poor's and Fitch, but statistically significant for Moody's, which is partially in contradiction with our initial results in Appendix A (Panels A). Interestingly, while there is little evidence of

[14]We also examine interaction with investment grade vs speculative grade, OECD member vs non-member, and previously defaulted vs never defaulted. The results are comparable and available upon request.

*Table 25.* Estimation results of the ordered logit with random effects for the three model specifications for the determinants of sovereign credit ratings, that include sentiment and subjectivity scores, as well as interaction with economic development, by credit rating agency and type of reports (Rating Action or Full Reports)

| | Sovereign credit ratings | | | | |
|---|---|---|---|---|---|
| | S&P | S&P (Full Reports) | Fitch | Fitch (Full Reports) | Moody's |
| | Model 1 | | | | |
| ED=1 × Net sentiment (W, dict) | 0.081 | 0.039 | 0.076 | 0.251 | 0.212** |
| | (0.123) | (0.382) | (0.082) | (0.284) | (0.082) |
| ED=1 × Polarity (W, dict) | 0.794 | 1.806 | 0.961 | 3.102** | 1.806*** |
| | (0.821) | (2.368) | (0.690) | (1.523) | (0.633) |
| ED=1 × Polarity (S, ML) | -0.169 | 0.402 | 0.818 | 3.921* | 1.229** |
| | (0.780) | (2.701) | (0.651) | (2.227) | (0.510) |
| ED=1 × Subjectivity (W, dict) | -0.350** | -1.228*** | -0.339* | -1.357*** | -0.466** |
| | (0.160) | (0.451) | (0.199) | (0.491) | (0.214) |
| ED=1 × Subjectivity (S, dict) | -2.874* | -16.106*** | -3.876** | -14.651*** | -3.696** |
| | (1.626) | (5.028) | (1.809) | (3.965) | (1.629) |
| ED=1 × Subjectivity (S, ML) | -0.542 | -16.787*** | 0.468 | -7.488 | 2.269 |
| | (2.060) | (5.215) | (1.331) | (5.600) | (1.474) |
| | Model 2 | | | | |
| ED=1 × Net sentiment (W, dict) | 0.074 | 0.156 | 0.032 | 0.023 | 0.190** |
| | (0.144) | (0.337) | (0.100) | (0.340) | (0.081) |
| ED=1 × Polarity (W, dict) | 0.738 | 2.079 | 0.492 | 1.781 | 1.543** |
| | (0.975) | (2.318) | (0.832) | (1.781) | (0.624) |
| ED=1 × Polarity (S, ML) | -0.397 | 0.543 | 0.407 | 2.214 | 1.441*** |
| | (0.884) | (2.534) | (0.622) | (2.132) | (0.475) |
| ED=1 × Subjectivity (W, dict) | -0.310* | -1.205*** | -0.231 | -1.539*** | -0.537*** |
| | (0.171) | (0.429) | (0.189) | (0.478) | (0.198) |
| ED=1 × Subjectivity (S, dict) | -1.707 | -15.203*** | -3.064* | -16.755*** | -4.471*** |
| | (1.554) | (4.525) | (1.590) | (3.701) | (1.533) |
| ED=1 × Subjectivity (S, ML) | 0.376 | -14.749*** | 0.744 | -5.381 | 2.770** |
| | (1.859) | (4.563) | (1.363) | (5.192) | (1.386) |
| | Model 3 | | | | |
| ED=1 × Net sentiment (W, dict) | 0.083 | 0.126 | 0.038 | 0.110 | 0.198** |
| | (0.144) | (0.352) | (0.100) | (0.335) | (0.082) |
| ED=1 × Polarity (W, dict) | 0.806 | 1.879 | 0.520 | 2.252 | 1.590** |
| | (0.975) | (2.423) | (0.836) | (1.751) | (0.625) |
| ED=1 × Polarity (S, ML) | -0.359 | 0.318 | 0.407 | 2.271 | 1.435*** |
| | (0.878) | (2.578) | (0.610) | (2.098) | (0.472) |
| ED=1 × Subjectivity (W, dict) | -0.304* | -1.177*** | -0.215 | -1.478*** | -0.516*** |
| | (0.172) | (0.431) | (0.191) | (0.465) | (0.199) |
| ED=1 × Subjectivity (S, dict) | -1.744 | -15.071*** | -3.030* | -16.384*** | -4.511*** |
| | (1.554) | (4.525) | (1.604) | (3.548) | (1.550) |
| ED=1 × Subjectivity (S, ML) | 0.165 | -14.180*** | 0.709 | -4.991 | 2.695* |
| | (1.868) | (4.618) | (1.379) | (5.085) | (1.394) |
| Observations | 1382 | 1422 | 1433 | 1232 | 1580 |

ED = Dummy variable for Economic development
Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the analysis of the relationship between sovereign credit ratings and macroeconomic, political, economic and cultural proximity variables, depending on model specification, and sentiment and subjectivity measures, but only the coefficients for sentiment and subjectivity measures are reported.

*Table 26.* Estimation results of the ordered logit with random effects for Model 3 for the determinants of sovereign credit ratings, estimated separately for advanced economies and emerging markets and by credit rating agency

| | Sovereign credit ratings | | | | | |
|---|---|---|---|---|---|---|
| | S&P | | Fitch | | Moody's | |
| | AE | EME | AE | EME | AE | EME |
| GDP per capita | 0.090 | 0.274*** | 0.149** | 0.331*** | 0.100 | 0.320*** |
| | (0.068) | (0.091) | (0.074) | (0.081) | (0.069) | (0.085) |
| Real GDP growth | -10.532 | 6.814* | -9.900 | 1.615 | -14.622*** | 0.510 |
| | (7.080) | (3.887) | (9.038) | (3.055) | (5.050) | (2.341) |
| Inflation | -25.371*** | -5.220*** | -26.991*** | -3.751* | -34.520*** | -2.016 |
| | (5.073) | (1.848) | (10.475) | (2.045) | (7.422) | (2.153) |
| Current account/GDP | -8.207** | -5.026* | -8.677*** | -4.855*** | -17.557*** | -2.358 |
| | (3.772) | (2.579) | (3.084) | (1.694) | (3.610) | (2.506) |
| Trade/GDP | 7.634*** | -0.230 | 4.967*** | -0.446 | 5.943*** | -0.518 |
| | (1.985) | (1.121) | (1.830) | (0.980) | (1.417) | (1.002) |
| External debt/GDP | -0.112 | 0.666 | -0.303** | 0.328 | -0.146 | 0.051 |
| | (0.184) | (1.432) | (0.138) | (1.047) | (0.208) | (1.092) |
| Default history | 0.008 | -3.438*** | -0.597 | -5.388*** | -1.246 | -5.429*** |
| | (1.373) | (1.042) | (1.699) | (1.081) | (1.977) | (1.085) |
| Log of int. reserves | -1.124*** | 1.482*** | -1.026*** | 1.272*** | -0.813** | 0.714*** |
| | (0.411) | (0.295) | (0.361) | (0.284) | (0.335) | (0.243) |
| Government debt/GDP | -15.542*** | -7.280*** | -12.844*** | -7.081*** | -15.584*** | -5.450*** |
| | (3.293) | (1.214) | (2.332) | (1.193) | (2.969) | (1.172) |
| Budget balance/GDP | -1.756 | -1.816 | -4.100 | -1.779 | -3.138 | -5.916 |
| | (5.465) | (4.088) | (11.023) | (3.724) | (6.705) | (4.331) |
| Institutional quality | 0.953* | -0.041 | -0.157 | -0.035 | -0.771** | 0.066 |
| | (0.500) | (0.194) | (0.329) | (0.177) | (0.388) | (0.122) |
| Governance | 0.489*** | 0.496*** | 0.476*** | 0.448*** | 0.672*** | 0.346*** |
| | (0.086) | (0.075) | (0.103) | (0.078) | (0.088) | (0.056) |
| Trade proximity | -117.444 | 67.723*** | -79.349* | 61.900** | -70.847 | 36.929** |
| | (127.048) | (24.970) | (46.060) | (30.532) | (53.024) | (16.482) |
| Common language | 6.737 | 0.302 | 1.259 | -0.973 | 5.472 | -0.150 |
| | (8.464) | (1.132) | (2.049) | (0.982) | (5.318) | (0.925) |
| Religious proximity | 18.466*** | 0.998 | 6.855 | -1.715 | 15.794** | -0.165 |
| | (6.681) | (1.819) | (6.806) | (1.551) | (6.399) | (1.334) |
| Geographical distance | -0.026 | 0.013 | -0.090** | 0.005 | -0.018 | 0.015 |
| | (0.087) | (0.013) | (0.040) | (0.012) | (0.020) | (0.010) |
| Observations | 551 | 831 | 591 | 842 | 598 | 982 |

AE = Advanced economies, EME = Emerging markets
Clustered standard errors in parentheses
Cut-off estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

differences in rating action reports by Standard & Poor's and Fitch, these differences are significant in the full credit rating reports. The detailed analysis suggests that the rating committees of all three agencies apply different degrees of qualitative judgement to advanced economies and emerging markets, which is in line with the initial hypothesis. The negative signs of the interaction terms suggest a stronger effect of subjectivity measures of advanced economies compared to emerging markets. Namely, a unit increase in Moody's subjectivity at word level in Model 3 would lead to a 0.52 unit decrease in the ordered log-odds of being in a higher credit rating category for emerging markets compared to advanced economies, holding all else equal.

*Table 27.* Estimation results of the ordered logit with random effects for Model 3 for the determinants of sovereign credit ratings, that include sentiment and subjectivity scores, estimated separately for advanced economies and emerging markets and by credit rating agency

| | Sovereign credit ratings | | | | | |
|---|---|---|---|---|---|---|
| | S&P | | Fitch | | Moody's | |
| | AE | EME | AE | EME | AE | EME |
| Net sentiment (W, dict) | 0.099 | -0.029 | 0.008 | 0.040 | 0.034 | 0.058 |
| | (0.338) | (0.050) | (0.067) | (0.057) | (0.091) | (0.049) |
| Polarity (W, dict) | 0.767 | 0.112 | 0.086 | 0.172 | 0.328 | 0.243 |
| | (0.648) | (0.392) | (0.642) | (0.504) | (0.588) | (0.371) |
| Polarity (S, ML) | 0.428 | 1.099*** | 0.612 | 1.176*** | 0.346 | 0.273 |
| | (0.547) | (0.375) | (0.461) | (0.429) | (0.404) | (0.269) |
| Subjectivity (W, dict) | 0.053 | 0.016 | 0.007 | -0.072 | 0.029 | -0.190* |
| | (0.158) | (0.122) | (0.142) | (0.100) | (0.181) | (0.109) |
| Subjectivity (S, dict) | 0.317 | -0.017 | -2.109 | -1.076 | -0.110 | -1.378 |
| | (1.041) | (0.961) | (1.325) | (0.902) | (1.629) | (0.985) |
| Subjectivity (S, ML) | 0.940 | 1.987** | 0.401 | -0.237 | 0.165 | -0.780 |
| | (1.514) | (0.883) | (1.340) | (0.795) | (1.098) | (0.847) |
| Observations | 551 | 831 | 591 | 842 | 598 | 982 |

AE = Advanced economies, EME = Emerging markets
Clustered standard errors in parentheses
Cut-off estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the analysis of the relationship between sovereign credit ratings and macroeconomic, political, economic and cultural proximity variables, depending on model specification, and sentiment and subjectivity measures, but only the coefficients for sentiment and subjectivity measures are reported.

We also run separate regressions for advanced economies and emerging markets of Model 3 with sentiment and subjectivity measures. The results without textual sentiment measures are reported in Table 26, while Table 27 shows the results for sentiment and subjectivity measures, respectively. Even though the coefficients are predominantly statistically insignificant, the differences in coefficients for sentiment and subjectivity measures between advanced and emerging markets are noticeable. Furthermore, different determinants seem to be influencing the sovereign credit ratings of advanced economies and emerging markets. Most notably, trade, religious proximity appear to have explanatory power for advanced economies. The latter, which has a positive sign, suggests there is some evidence that sovereign credit ratings are culturally biased upward for advanced markets. This is consistent with Gültekin-Karakaş et al. (2011), who note that high-income countries are rated higher than low-income countries, ceteris paribus. On the other hand, GDP per capita, default history and trade proximity significantly affect the sovereign credit ratings of emerging markets, where the latter points to a positive economic bias toward emerging markets. Additionally, we observe substantial differences in coefficients among the significant determinants for both groups of countries (e.g. inflation, government debt) or even opposite signs (e.g. international reserves). This may be evidence of different weighting schemes applied by the credit rating committee. According to Fuchs and Gehring (2017) and Zheng (2012), the degree of foreign bias in sovereign ratings varies across agencies due to different weights being applied to the qualitative judgement.

### 4.3.3 Global financial crisis

Additionally, we also examine the behaviour of sentiment and subjectivity measures before and after the 2008 global financial crisis. We introduce an interaction of the global financial crisis dummy (1 if the year of observation is 2008 or later, 0 otherwise) with sentiment and subjectivity measures[15]. The results are presented in Table 28. The coefficients for sentiment and subjectivity measures correspond to the period before the global crisis, while the interaction is the difference in coefficients between before and after the crisis. As expected, the difference in sentiment measures is significant in all three models, corresponding to the general negative financial climate after the crisis. However, the differences in subjectivity measures, apart from the measure from the machine learning approach, are not statistically significant, which is in contradiction with our initial results in Panel B of Table 21. A further analysis thus indicates that the global crisis did not disrupt the way the rating committee employs qualitative judgement.

[15]We also analyse interaction with systemic banking crisis dummy (Laeven & Valencia, 2018), and interaction with the crisis starting in 2010. The results are comparable and available upon request.

*Table 28.* Estimation results of the ordered logit with random effects for the three model specifications for the determinants of sovereign credit ratings, that include sentiment and subjectivity scores, as well as interaction iwith global crisis dummy, by credit rating agency and type of reports (Rating Action or Full Reports)

| | Sovereign credit ratings | | | | |
|---|---|---|---|---|---|
| | S&P | S&P (Full Reports) | Fitch | Fitch (Full Reports) | Moody's |
| | Model 1 | | | | |
| GFC=1 × Net sentiment (W, dict) | 0.187** | 0.491*** | 0.229*** | 0.367* | 0.258*** |
| | (0.084) | (0.174) | (0.053) | (0.218) | (0.054) |
| GFC=1 × Polarity (W, dict) | 1.510** | 4.049*** | 1.876*** | 1.772 | 1.791*** |
| | (0.693) | (1.439) | (0.445) | (1.251) | (0.417) |
| GFC=1 × Polarity (S, ML) | 2.327*** | 4.946*** | 2.091*** | 4.194*** | 1.884*** |
| | (0.608) | (1.535) | (0.486) | (1.443) | (0.405) |
| GFC=1 × Subjectivity (W, dict) | 0.230 | -0.069 | 0.018 | -0.249 | -0.125 |
| | (0.144) | (0.368) | (0.136) | (0.379) | (0.168) |
| GFC=1 × Subjectivity (S, dict) | 0.787 | -1.286 | 0.774 | 1.070 | -1.654 |
| | (1.350) | (3.656) | (1.284) | (3.235) | (1.623) |
| GFC=1 × Subjectivity (S, ML) | -2.982 | -7.426* | 2.441** | -0.926 | -0.310 |
| | (1.910) | (4.124) | (1.193) | (3.864) | (1.143) |
| | Model 2 | | | | |
| GFC=1 × Net sentiment (W, dict) | 0.187** | 0.652*** | 0.224*** | 0.353* | 0.210*** |
| | (0.078) | (0.161) | (0.050) | (0.214) | (0.056) |
| GFC=1 × Polarity (W, dict) | 1.475** | 4.883*** | 1.862*** | 1.738 | 1.426*** |
| | (0.629) | (1.380) | (0.426) | (1.242) | (0.438) |
| GFC=1 × Polarity (S, ML) | 2.496*** | 6.074*** | 1.768*** | 4.278*** | 1.792*** |
| | (0.619) | (1.334) | (0.499) | (1.489) | (0.390) |
| GFC=1 × Subjectivity (W, dict) | 0.250* | 0.151 | 0.161 | -0.168 | -0.168 |
| | (0.150) | (0.379) | (0.145) | (0.418) | (0.159) |
| GFC=1 × Subjectivity (S, dict) | 0.655 | 0.552 | 1.657 | 2.504 | -1.839 |
| | (1.364) | (3.861) | (1.313) | (3.665) | (1.551) |
| GFC=1 × Subjectivity (S, ML) | -3.768** | -5.940 | 2.916** | 0.965 | -0.806 |
| | (1.777) | (3.802) | (1.147) | (3.828) | (1.121) |
| | Model 3 | | | | |
| GFC=1 × Net sentiment (W, dict) | 0.168** | 0.611*** | 0.229*** | 0.361* | 0.197*** |
| | (0.076) | (0.156) | (0.052) | (0.210) | (0.055) |
| GFC=1 × Polarity (W, dict) | 1.310** | 4.598*** | 1.887*** | 1.816 | 1.343*** |
| | (0.605) | (1.342) | (0.441) | (1.215) | (0.436) |
| GFC=1 × Polarity (S, ML) | 2.374*** | 5.932*** | 1.772*** | 4.192*** | 1.737*** |
| | (0.619) | (1.290) | (0.506) | (1.483) | (0.379) |
| GFC=1 × Subjectivity (W, dict) | 0.228 | 0.067 | 0.144 | -0.186 | -0.195 |
| | (0.152) | (0.363) | (0.146) | (0.412) | (0.156) |
| GFC=1 × Subjectivity (S, dict) | 0.344 | 0.020 | 1.517 | 2.540 | -2.060 |
| | (1.335) | (3.797) | (1.317) | (3.626) | (1.550) |
| GFC=1 × Subjectivity (S, ML) | -3.990** | -6.955* | 2.635** | 1.638 | -0.897 |
| | (1.799) | (3.961) | (1.159) | (3.761) | (1.118) |
| Observations | 1382 | 1422 | 1433 | 1232 | 1580 |

GFC = Dummy variable for Global Financial Crisis
Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the analysis of the relationship between sovereign credit ratings and macroeconomic, political, economic and cultural proximity variables, depending on model specification, and sentiment and subjectivity measures, but only the coefficients for sentiment and subjectivity measures are reported.

### 4.3.4 Determinants of sentiment and subjectivity

As the last step, we regress sentiment and subjectivity measures on proxies for soft information and potential bias. We use random effects on a pooled sample. The results are presented in Table 29 and are in line with our previous findings.

*Table 29.* Estimation results of the random effects model with sentiment and subjectivity measures as dependent variables and institutional quality, governance and economic and cultural proximity as explanatory variables, pooled sample

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Institutional quality | -0.076*** | -0.012*** | -0.009** | 0.036** | 0.002 | -0.003 |
| | (0.026) | (0.003) | (0.004) | (0.017) | (0.002) | (0.002) |
| Governance | 0.187*** | 0.026*** | 0.032*** | -0.093*** | -0.007*** | 0.003** |
| | (0.026) | (0.003) | (0.004) | (0.014) | (0.001) | (0.001) |
| Trade proximity | 4.871** | 0.635** | 1.291*** | 3.553 | 0.157 | -0.341** |
| | (2.194) | (0.277) | (0.411) | (2.737) | (0.186) | (0.139) |
| Common language | -0.156 | -0.027 | -0.020 | -0.083 | -0.001 | -0.026** |
| | (0.192) | (0.027) | (0.038) | (0.143) | (0.013) | (0.013) |
| Religious proximity | 0.169 | 0.022 | 0.162* | -0.326 | -0.015 | 0.056* |
| | (0.366) | (0.047) | (0.083) | (0.300) | (0.029) | (0.030) |
| Geographical distance | 0.005* | 0.001** | 0.001** | 0.001 | 0.000 | -0.000 |
| | (0.003) | (0.000) | (0.001) | (0.002) | (0.000) | (0.000) |
| Constant | -5.213*** | -0.696*** | -0.471*** | 4.394*** | 0.456*** | 0.323*** |
| | (0.605) | (0.077) | (0.108) | (0.377) | (0.037) | (0.035) |
| Observations | 1669 | 1669 | 1669 | 1669 | 1669 | 1669 |
| $R^2$ | 0.103 | 0.120 | 0.153 | 0.031 | 0.021 | 0.034 |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Clustered standard errors in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Sentiment measures are explained by institutional quality, governance, trade proximity and geographical distance. The $R^2$ are comparable and range between 0.10 and 0.15. On the other hand, we were unable to explain much of the variability in subjectivity measures using the same explanatory variables. Only governance is significant for all three measures, and $R^2$ are relatively low, ranging between 0.02 and 0.03. We can conclude that sentiment to a limited degree captures soft information and potential bias, while subjectivity reflects the qualitative judgement of the rating committee and thus offers some additional information not captured by the determinants of sovereign credit ratings.

Additionally, we also repeat the analysis by using the subcomponents of institutional quality and governance. Institutional quality comprises of law and order, bureaucracy quality, democratic accountability and corruption, and governance includes government stability, socio-economic conditions and investment profile. The results are reported in

Table 30. Sentiment measures are predominantly explained by the subcomponents of governance, while only the bureaucracy quality and law and order subcomponents of institutional quality are statistically significant. Similarly, the subcomponents of governance and law and order are statistically significant when trying to explain subjectivity measures.

*Table 30.* Estimation results of the random effects model with sentiment and subjectivity measures as dependent variables and economic and cultural proximity, and subcomponents of governance and institutional quality as explanatory variables, pooled sample

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Trade proximity | 5.943*** | 0.813*** | 1.638*** | 0.327 | -0.126 | -0.217** |
| | (2.208) | (0.278) | (0.432) | (1.854) | (0.146) | (0.105) |
| Common language | -0.138 | -0.027 | -0.022 | -0.005 | 0.008 | -0.030** |
| | (0.206) | (0.028) | (0.039) | (0.117) | (0.011) | (0.013) |
| Religious proximity | -0.147 | -0.027 | 0.105 | -0.130 | -0.000 | 0.040 |
| | (0.403) | (0.050) | (0.088) | (0.253) | (0.026) | (0.030) |
| Geographical distance | 0.006** | 0.001** | 0.001*** | 0.002 | 0.000 | -0.000 |
| | (0.003) | (0.000) | (0.001) | (0.002) | (0.000) | (0.000) |
| Government stability | 0.121** | 0.020*** | 0.033*** | -0.199*** | -0.015*** | 0.012*** |
| | (0.058) | (0.007) | (0.009) | (0.023) | (0.003) | (0.002) |
| Socioeconomic conditions | 0.187*** | 0.024*** | 0.020* | 0.113*** | 0.012*** | -0.005 |
| | (0.066) | (0.007) | (0.010) | (0.033) | (0.004) | (0.004) |
| Investment profile | 0.272*** | 0.038*** | 0.044*** | -0.132*** | -0.012*** | 0.002 |
| | (0.057) | (0.007) | (0.009) | (0.028) | (0.003) | (0.003) |
| Corruption | 0.201 | 0.015 | 0.009 | -0.054 | -0.009 | -0.011 |
| | (0.129) | (0.017) | (0.021) | (0.052) | (0.005) | (0.007) |
| Law and order | -0.211** | -0.028** | -0.014 | -0.121** | -0.012** | 0.001 |
| | (0.105) | (0.013) | (0.016) | (0.051) | (0.005) | (0.006) |
| Democratic accountability | -0.022 | 0.001 | 0.009 | -0.013 | 0.001 | 0.008 |
| | (0.075) | (0.009) | (0.017) | (0.053) | (0.005) | (0.006) |
| Bureaucracy quality | -0.474** | -0.058*** | -0.054 | 0.148 | 0.013 | 0.005 |
| | (0.188) | (0.021) | (0.033) | (0.115) | (0.011) | (0.010) |
| Constant | -4.879*** | -0.679*** | -0.493*** | 5.020*** | 0.510*** | 0.271*** |
| | (0.721) | (0.088) | (0.128) | (0.343) | (0.034) | (0.039) |
| Observations | 1669 | 1669 | 1669 | 1669 | 1669 | 1669 |
| $R^2$ | | | | | | |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Clustered standard errors in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

## 4.4 Conclusion

This chapter offers a novel approach to analysing the determinants of sovereign credit ratings. We build on the existing methodology on the identification of determinants of sovereign credit ratings by extending the traditional regression with new measures (i.e. sentiment and subjectivity indicators), obtained by applying new approaches (i.e. textual analysis). There is substantial evidence regarding the impact of sentiment or tone (qualitative information) in corporate credit rating reports on corporate equity valuation (Agarwal et al., 2016; Kearney & Liu, 2014; Loughran & McDonald, 2016), but the evidence on the impact of sovereign credit rating reports on sovereign debt market is practically non-existent. Sovereign credit ratings can have economically more important consequences than firm-level credit ratings since they can affect the efficiency and stability of capital markets within and across countries. To our knowledge, only one study (Agarwal et al., 2019) applies textual analysis methods to these reports.

We examine the relationship between the sovereign credit ratings and the reports that accompany them. We achieve this by employing textual (sentiment) analysis methods to Standard & Poor's, Fitch and Moody's Rating Action reports, which produces six key indicators: net sentiment, polarity (dictionary-based and machine learning approach), and subjectivity (dictionary-based approach at the word and sentence level, and machine learning approach). We partially relate textual sentiment measures to proxies for soft information and to the general country perception, and subjectivity measures to the qualitative (expert) knowledge of the rating committee or their interpretation of soft information. We find that soft information plays an important role when assigning sovereign credit ratings, since adding the variables for political risk and institutional strength significantly improves the predictability of sovereign credit ratings. We can conclude that improving institutional strength and governance could lead to higher assigned ratings. Furthermore, we find evidence of economic proximity bias, which means that credit rating agencies (that are US-based) assign higher credit ratings to countries that have strong trade ties with the US. Results suggest that textual sentiment provides additional information not captured by traditional determinants of sovereign credit ratings, especially if soft information and bias proxies are not taken into account. On the other hand, our analysis indicates that the qualitative judgement of the rating committee manifests itself in the subjectivity score and is robust to model expansions for one of three credit rating agencies. Additionally, we find differences in sentiment between emerging and advanced markets for one of the agencies, which is most likely due to the difference in the general perception of these groups of countries and perceived political risk. We detect a significant difference in subjectivity scores for emerging markets compared to advanced economies, indicating that the rating committee employs different levels of qualitative judgement for these groups of countries. Potential data shortage or low-quality data for emerging markets, and consequently, an increased need for qualitative judgement may explain these differences.

We identify different determinants describing sovereign credit ratings of advanced and emerging countries and some indication of, on the one hand, an upward cultural bias towards the developed world, and, on the other hand, an upward economic bias towards the emerging markets. We also detect a change in sentiment after the 2008 global financial crisis, which may be due to the generally negative tone in the economy, but we do not register any difference in subjectivity. Finally, we find that sentiment can partially be explained by soft information and bias proxies, while subjectivity cannot, supporting our initial hypothesis.

The main finding of this chapter is thus that qualitative (expert) judgement of the rating committee can to some extent, be reflected in the subjectivity indicator obtained from the sovereign credit rating reports. This opens the door to further research opportunities, both in terms of data and methodology.

# 5 On the information content of sovereign credit rating reports: improving the predictability of rating transitions

## 5.1 Introduction

The formation of sovereign credit ratings has puzzled researchers for more than two decades. Cantor and Packer (1996) were among the first to delve into the determinants of sovereign credit ratings, with many following their footsteps (e.g. Afonso, 2003; Afonso et al., 2009; Butler & Fauver, 2006; Özturk, 2014). Nevertheless, a relatively large part of ratings has always remained unexplained by hard (macroeconomic) data (e.g. De Moor et al., 2018; Özturk, 2014). Credit rating agencies (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017) claim that part of the rating represents the qualitative knowledge of the rating committee.

Despite being aware of this, countries or government issuers rely on credit rating agencies and sovereign credit ratings, as they give a clear and relatively reliable signal to the international capital markets of their creditworthiness. Higher ratings mean lower borrowing costs and vice versa. Consequently, changes in sovereign credit ratings are equally, if not even more important, as they lead to deterioration or improvement of their borrowing costs in the future (Alsakka & ap Gwilym, 2013; Eijffinger, 2012). Of particular importance is the transition between investment and speculative grade. Kiff et al. (2010) find evidence of this transition breakpoint having significant effects on CDS spreads. Being able to predict rating transitions more accurately would have meaningful benefits for issuers and investors, who could better prepare for such changes. A few studies try to estimate sovereign transition matrices (e.g. Hill et al., 2010; Hu et al., 2002), but face limitations mainly due to data shortage. However, the underlying problem remains, namely that the qualitative judgment of the rating committee is not captured by the traditional determinants of sovereign credit ratings. As long as the qualitative part of the credit rating is left unexplained, attempting to better predict rating transitions is futile.

The main objective of this chapter is thus to further explore sentiment and subjectivity measures related to sovereign credit ratings further. We build on the approach of identifying the qualitative component, as proposed in the previous chapter. As already established, credit rating agencies normally issue sovereign credit rating reports along with the ratings themselves. We hypothesise that qualitative judgement or interpretation of the rating committee is expressed in the reports, which have been largely neglected in the past. We use dictionary-based textual sentiment methods to extract sentiment and subjectivity scores from the reports and find the latter helps in additionally explaining sovereign credit ratings even after the political risk, institutional strength and potential bias are controlled for.

We aim to determine whether textual sentiment and subjectivity measures improve performance of models predicting rating transitions, using a sample of 97 countries rated by Standard & Poor's in the period from 2002 to 2018, 98 countries rated by Fitch in the period from 1999 to 2018, and 100 countries rated by Moody's in the period from 1995 to 2018. Based on the use of the logistic regression, we focus on the classification accuracy of downgrades and upgrades. We define our binary dependent variable as equal to one if country $i$ is downgraded/upgraded at time $t$, and zero otherwise.

We use six different textual sentiment and subjectivity measures, using both dictionary-based and machine learning approaches on sovereign credit rating reports issued by Standard & Poor's, Fitch and Moody's. Sentiment measures, namely net sentiment and polarity, are based on detecting negative and positive words or sentences, while subjectivity measures focus on detecting opinion. We compare these measures by separately and simultaneously including them in models and explore which is the most informative (e.g. significantly affects upgrades/downgrades of the ratings).

We find that, on average, sentiment measures perform better than subjectivity measures. Specifically, we observe that correct classification of true positives, i.e. sensitivity, increases when we include sentiment measures in the models. This improvement is more pronounced for downgrades than upgrades. The increase in performance is more distinct for Moody's compared to Standard & Poor's and Fitch.

On the other hand, subjectivity scores, on average, offer relatively poor results compared to sentiment scores. There is also no clear winner for the textual analysis approach. With sentiment measures, the dictionary-based techniques appear to outperform machine learning. However, the machine learning approach surpasses dictionary-based applications for subjectivity measures.

Finally, credit rating agencies can comply with two different rating philosophies: with through-the-cycle rating philosophy, they take a longer horizon into account, while with the point-in-time approach, they focus on current information (Basel Committee on Banking Supervision, 2005). Credit rating agencies generally employ the through-the-cycle approach but have been criticised in the past on their failure to do so (Ferri et al., 1999; Kaminsky & Schmukler, 2002; Kiff et al., 2012). We thus estimate our models in both frameworks. Our findings suggest that, while credit rating agencies may, from time to time, fall off the wagon, they, on average, follow the through-the-cycle rating philosophy since taking into account past and future values leads to better model performance.

We believe that this chapter provides a significant methodological contribution. Although textual sentiment analysis is relatively widely used in corporate finance (e.g. Kearney & Liu, 2014; Loughran & McDonald, 2016), the application to sovereigns is limited, especially in the field of sovereign credit ratings. To our knowledge, only one study examines them. Agarwal et al. (2019) use a machine learning approach and find evidence of sentiment or tone having additional explanatory value. However, to the best of our knowledge, no study has explored more than one approach, and no comparative analysis exists on the performance of such measures. This thesis thus far offers the most comprehensive analysis of textual sentiment measures and their effect on sovereign credit ratings.

Additionally, practical implications are considerable, since sovereign credit ratings and changes in ratings have significant effects on both the international debt markets and governments' borrowing costs. The ability to more accurately predict changes in ratings would allow issuers to revise the timeline of debt financing by pursuing favourable lending conditions and allow investors to optimise their portfolios.

The rest of the chapter is structured as follows. In subsection 5.2, we describe the methodology. Next, we comment on the results: in subsection 5.3.1 we focus on the point-in-time analysis while we examine the though-the-cycle philosophy in subsection 5.3.2. In subsection 5.3.3, we perform robustness checks. We conclude in subsection 5.4.

## 5.2 The framework

We define the dependent variable as:

$$Y_{it} = \begin{cases} 1 & \text{if country } i \text{ is downgraded (upgraded) in time } t \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The probability that the binary dependent variable $Y_{it}$ equals one given the covariates is modelled using the following specification:

$$P(Y_{it} = 1|X_{it}) = \Lambda(\alpha + \beta X_{ti}) = \frac{e^{\alpha+\beta X_{ti}}}{1 + e^{\alpha+\beta X_{ti}}} \tag{12}$$

where $\alpha$ and $\beta$ are parameters to be estimated, and $X_{it}$ is a vector of a country-specific time-varying and time-invariant explanatory variables described in Table 3, depending on model specification. $\Lambda$ is the logistic distribution function, corresponding to the logit model.

We again explore three baseline models. The first model (Model 1) contains only macroeconomic and fiscal strength variables defined in Table 3. This corresponds to early studies of the determinants of sovereign credit ratings (e.g. Cantor & Packer, 1996). Next, we extend Model 1 with proxies for institutional strength and political risk. This is based on previous findings arguing that the credit rating committee takes soft information into account when assigning sovereign credit ratings (e.g. Butler & Fauver, 2006; Özturk, 2014). Finally, we include proxies for cultural and economic proximity in Model 3 to control for a potential bias identified in existing literature (Fuchs & Gehring, 2017; Luitel et al., 2016; Zheng, 2012). Next, we separately and simultaneously add the indicators for (textual) sentiment and subjectivity to each of the three models. In order to examine whether credit rating agencies employ the through-the-cycle or point-in-time approach, we extend the models by adding the one-period lags and leads of time-varying explanatory variables. Finally, since one could argue that textual sentiment is just a proxy for outlook, we include outlook in all our models as a robustness check.

Our aim is to compare the classification accuracy of all models. We focus on sensitivity, i.e. correct classification of true positives in percent (true positive rate, TPR), defined as:

$$Sensitivity = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{13}$$

where TP is the number of true positives, P is the number of actual positives in the sample, and FN is the number of false negatives. For example, in the case of downgrades, TP is the number of correctly classified (predicted) downgrades, FN is the number of downgrades identified as not a downgrade (upgrade or no change), and P is the number of actual downgrades in the sample.

We also report the overall classification accuracy and area under the ROC. The overall classification accuracy is the per cent of correctly classified (predicted) in the sample:

$$Overall = \frac{TP + TN}{P + N} \tag{14}$$

where TN is the number of true negatives, and N is the number of actual negatives in the sample. In the case of downgrades, TN is the number of correctly classified (predicted) non-downgrades (upgrade or no change), and N is the number of actual non-downgrades in the sample.

A ROC curve or receiver operating characteristic curve is a plot showing the relationship between sensitivity (y-axis) and false positive rate ($FPR = FP/N$), where FP is the number of false positives, e.g. the number of incorrectly classified downgrades, that are actually non-downgrades (x-axis). It shows the performance of the classification model at all classification thresholds. The area under the ROC curve (AUC) measures the total area under the curve and thus provides an aggregate measure of performance across all possible classification thresholds. It can be interpreted as the probability that the model ranks a random positive case more highly than a random negative case.

## 5.3 Results

### 5.3.1 Point-in-time approach

We begin our main analysis with baseline models assessed with current values, corresponding to the point-in-time rating philosophy. Table 31 shows that Model 1 correctly classifies 20.1% of downgrades by Standard & Poor's, 22.5% of downgrades by Fitch and 20.9% of downgrades by Moody's. The overall classification accuracy of agencies is also comparable, at 89.0%, 91.6% and 92.0%, respectively. Once we extend the specification to Models 2 and 3, we observe an improvement in the case of Moody's, specifically a rise to 21.6% and 24.5%, respectively. However, classification accuracy for downgrades deteriorates, albeit marginally, for Fitch, as it decreases to 21.0% for both models. The results for Standard & Poor's are puzzling, as the classification accuracy first increases to 22.0% in Model 2, but then decreases to 21.4% in Model 3.

Moving on to upgrades, we observe that both the accuracy of classifying upgrades vs downgrades and areas under ROC curves are significantly worse (Figure 3). This may be due to the credit rating agencies' reluctance to upgrade sovereigns after they have failed to downgrade them during a downturn. Ferri et al. (1999) give an example of the East Asian crisis and argue that credit rating agencies are motivated to remain cautious after failing to predict or even detect crises or downturns in order to restore their credibility. Since using the cut-off of 0.5 gives negligible results (see Appendix D), we lower the cut-off to 0.3. Table 32 shows that, differently from downgrades, the classification accuracy of upgrades varies across different credit rating agencies. The classification accuracy of Standard & Poor's is the highest out of the three, at 27.5%,

*Table 31.* Classification accuracy of downgrades, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable, following the point-in-time approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 20.126 | 89.001 | 0.839 | 22.464 | 91.626 | 0.844 | 20.863 | 91.899 | 0.861 |
| Net sentiment (W, dict) | 33.962 | 90.304 | 0.894 | 30.435 | 91.835 | 0.878 | 38.849 | 92.785 | 0.927 |
| Polarity (W, dict) | 39.623 | 90.955 | 0.902 | 25.362 | 91.486 | 0.876 | 33.094 | 92.089 | 0.917 |
| Polarity (S, ML) | 37.107 | 91.100 | 0.890 | 21.014 | 91.207 | 0.866 | 31.655 | 92.278 | 0.911 |
| Subjectivity (W, dict) | 21.384 | 89.363 | 0.840 | 23.188 | 91.766 | 0.846 | 17.266 | 91.646 | 0.873 |
| Subjectivity (S, dict) | 20.126 | 89.219 | 0.839 | 23.188 | 91.696 | 0.846 | 17.986 | 91.772 | 0.871 |
| Subjectivity (S, ML) | 20.755 | 89.219 | 0.839 | 21.014 | 91.626 | 0.843 | 20.144 | 91.962 | 0.862 |
| All textual indicators | 45.912 | 91.679 | 0.907 | 32.609 | 92.045 | 0.883 | 46.763 | 93.671 | 0.934 |
| | | | | Model 2 | | | | | |
| No textual indicators | 22.013 | 89.074 | 0.840 | 21.014 | 91.207 | 0.847 | 21.583 | 92.025 | 0.870 |
| Net sentiment (W, dict) | 35.849 | 90.738 | 0.894 | 30.435 | 91.975 | 0.882 | 37.410 | 92.911 | 0.928 |
| Polarity (W, dict) | 38.994 | 90.883 | 0.901 | 26.812 | 91.626 | 0.880 | 34.532 | 92.785 | 0.919 |
| Polarity (S, ML) | 35.849 | 90.955 | 0.892 | 23.913 | 91.486 | 0.868 | 32.374 | 92.278 | 0.912 |
| Subjectivity (W, dict) | 20.755 | 88.929 | 0.841 | 21.739 | 91.556 | 0.848 | 20.863 | 91.772 | 0.880 |
| Subjectivity (S, dict) | 19.497 | 88.857 | 0.841 | 20.290 | 91.277 | 0.848 | 20.144 | 91.835 | 0.878 |
| Subjectivity (S, ML) | 22.013 | 88.929 | 0.840 | 20.290 | 91.417 | 0.846 | 21.583 | 91.962 | 0.870 |
| All textual indicators | 44.654 | 91.534 | 0.907 | 33.333 | 91.975 | 0.885 | 46.043 | 93.544 | 0.935 |
| | | | | Model 3 | | | | | |
| No textual indicators | 21.384 | 89.219 | 0.845 | 21.014 | 91.207 | 0.846 | 24.460 | 92.278 | 0.872 |
| Net sentiment (W, dict) | 37.107 | 90.738 | 0.897 | 31.884 | 92.045 | 0.881 | 37.410 | 92.911 | 0.932 |
| Polarity (W, dict) | 39.623 | 91.027 | 0.904 | 26.812 | 91.696 | 0.880 | 35.252 | 92.848 | 0.922 |
| Polarity (S, ML) | 37.107 | 91.389 | 0.896 | 25.362 | 91.626 | 0.867 | 32.374 | 92.278 | 0.917 |
| Subjectivity (W, dict) | 24.528 | 89.508 | 0.848 | 22.464 | 91.486 | 0.848 | 20.863 | 92.025 | 0.882 |
| Subjectivity (S, dict) | 23.270 | 89.291 | 0.847 | 21.014 | 91.207 | 0.847 | 22.302 | 92.089 | 0.879 |
| Subjectivity (S, ML) | 21.384 | 89.001 | 0.845 | 21.739 | 91.486 | 0.846 | 23.022 | 92.215 | 0.872 |
| All textual indicators | 45.912 | 92.041 | 0.909 | 32.609 | 92.045 | 0.885 | 45.324 | 93.608 | 0.938 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

28.6% and 24.2% for the three models, respectively. The sensitivity of Fitch is slightly worse as it sits somewhere in the middle, at 15.5%, 16.1% and 18.5%. The classification accuracy of Moody's upgrades is relatively low at 6.9% for the three models. This discrepancy between Moody's and the other two agencies may be explained by Alsakka and ap Gwilym (2010), who find that Moody's on average upgrades sovereigns before the other agencies do. This means that Moody's potentially upgrades sovereigns when the rationale for upgrade is not yet reflected in the macroeconomic and political environment. Fitch lags behind and upgrades issuers when such action is supported by data, resulting in higher classification accuracy for upgrades in our models.

Next, we add the textual sentiment and subjectivity measures. The results for downgrades are reported in Table 31. They show that including textual sentiment measures, namely sentiment and dictionary-based polarity, significantly improves the classification of downgrades by the three agencies, while polarity from the machine learning approach positively affects only the results for two of the agencies. For Standard & Poor's, depending on the model, classification accuracy increases to between 34.0% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 39.7% (Models 1 and 3 with polarity from the dictionary-based approach, at word level). Similarly, for Moody's, classification accuracy increases to between 31.7% (Model 1 with polarity from the machine learning approach, at the sentence level) and 38.9% (Model 1 with the net sentiment from the dictionary-based approach, at word level). Sensitivity increases for Fitch as well, but the improvement is not that substantial. In some cases, sensitivity actually deteriorates (e.g. Model 1 with polarity from the machine learning approach, at the sentence level). It increases to between 23.9% (Model 2 with polarity from the machine learning approach, at the sentence level) and 31.9% (Model 3 with the net sentiment from the dictionary-based approach, at word level). On the other hand, the inclusion of subjectivity measures, on average, either lowers the classification accuracy of downgrades (most notably for Moody's) or has negligible effects on sensitivity. The most notable improvement is in the case of Standard & Poor's, Model 3 with subjectivity from the dictionary-based approach, where sensitivity increases from 21.4% without subjectivity to 24.5% with subjectivity. Overall, the single measure with the highest performance appears to be sentiment (dictionary-based approach). We also include all six measures simultaneously in the models and observe even higher sensitivity scores. These are higher for Standard & Poor's, at 45.9%, resulting from Models 1 and 3 with all measures, and for Moody's, at 46.8%, resulting from Model 1 with all measures, but lower for Fitch, at 33.3%, resulting from Model 2 with all measures. The observed discrepancy between both agencies when taking textual sentiment measures into account suggests that Standard & Poor's and Moody's express their opinion or qualitative judgement on downgrades in sovereign credit rating reports more efficiently than Fitch.

The results for upgrades are presented in Table 32. Similarly, as in the case of downgrades, we detect significant improvements in classification accuracy of upgrades when taking sentiment measures into account for all three agencies. Including sentiment measures, in most cases, more than doubles sensitivity for Moody's and significantly increases sensitivity for Standard & Poor's and Fitch in all three models. In the case of Moody's, classification accuracy rises to between 12.4% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 17.2% (Model 3 with polarity from the machine learning approach, at the sentence level). In the case of Standard & Poor's, which has the highest sensitivity levels among the three, sensitivity improves to between 30.2% (Model 2 with polarity from the machine learning approach, at the sentence level) and 36.8% (Model 3 with polarity from the machine learning approach, at the sentence level). In the case of Fitch, sensitivity improves to between 22.0% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 28.0% (Model 3 with polarity from the machine learning approach, at the sentence level).

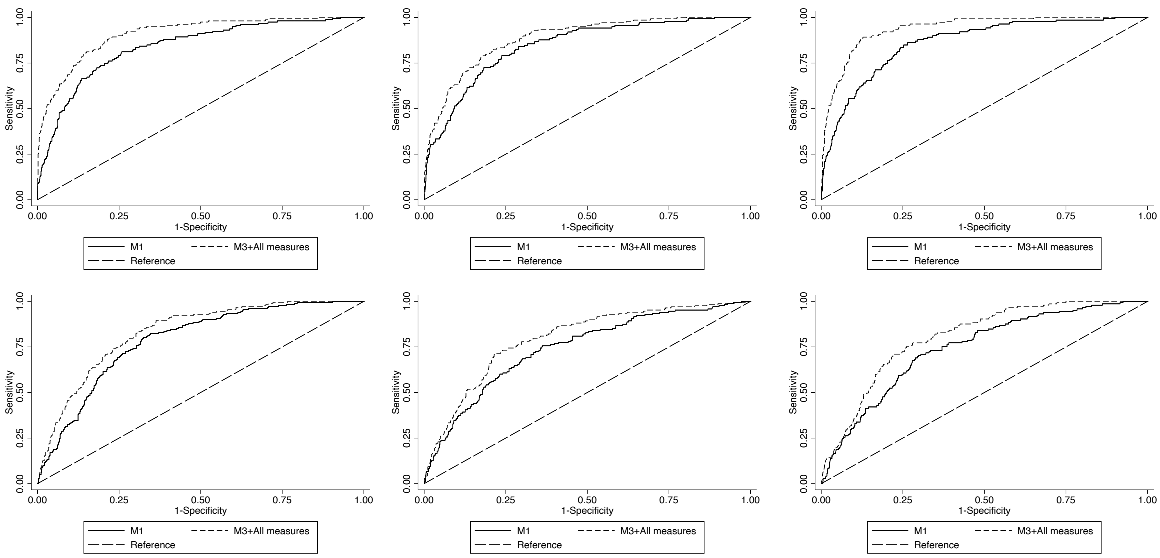
*Table 32.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, following the point-in-time approach, using the 0.3 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 27.473 | 84.370 | 0.785 | 15.476 | 87.439 | 0.742 | 6.897 | 89.747 | 0.742 |
| Net sentiment (W, dict) | 35.165 | 84.081 | 0.801 | 22.024 | 87.090 | 0.769 | 12.414 | 89.620 | 0.776 |
| Polarity (W, dict) | 35.165 | 84.153 | 0.799 | 22.619 | 87.299 | 0.766 | 13.793 | 89.747 | 0.774 |
| Polarity (S, ML) | 31.319 | 84.370 | 0.798 | 24.405 | 86.602 | 0.769 | 13.793 | 89.494 | 0.774 |
| Subjectivity (W, dict) | 30.220 | 84.515 | 0.800 | 15.476 | 87.439 | 0.743 | 8.276 | 89.937 | 0.742 |
| Subjectivity (S, dict) | 27.473 | 84.370 | 0.792 | 15.476 | 87.439 | 0.742 | 7.586 | 89.810 | 0.742 |
| Subjectivity (S, ML) | 26.923 | 84.226 | 0.785 | 17.262 | 87.369 | 0.745 | 7.586 | 89.620 | 0.752 |
| All textual indicators | 37.912 | 84.732 | 0.814 | 26.786 | 86.462 | 0.784 | 17.241 | 89.367 | 0.795 |
| | | | | Model 2 | | | | | |
| No textual indicators | 28.571 | 83.936 | 0.790 | 16.071 | 87.090 | 0.744 | 6.897 | 89.557 | 0.746 |
| Net sentiment (W, dict) | 31.319 | 83.502 | 0.806 | 22.619 | 87.090 | 0.770 | 13.793 | 89.747 | 0.779 |
| Polarity (W, dict) | 32.418 | 83.647 | 0.804 | 23.214 | 87.090 | 0.768 | 15.862 | 89.747 | 0.777 |
| Polarity (S, ML) | 30.220 | 83.719 | 0.802 | 24.405 | 86.950 | 0.770 | 13.793 | 89.367 | 0.776 |
| Subjectivity (W, dict) | 32.418 | 84.949 | 0.803 | 16.667 | 87.299 | 0.745 | 8.276 | 89.810 | 0.746 |
| Subjectivity (S, dict) | 26.923 | 83.792 | 0.797 | 16.071 | 87.090 | 0.744 | 7.586 | 89.747 | 0.747 |
| Subjectivity (S, ML) | 28.571 | 83.936 | 0.790 | 16.667 | 87.579 | 0.747 | 8.966 | 89.367 | 0.761 |
| All textual indicators | 36.813 | 84.588 | 0.817 | 26.786 | 86.462 | 0.785 | 18.621 | 89.051 | 0.800 |
| | | | | Model 3 | | | | | |
| No textual indicators | 24.176 | 83.068 | 0.803 | 18.452 | 86.741 | 0.762 | 6.897 | 89.241 | 0.764 |
| Net sentiment (W, dict) | 34.066 | 84.009 | 0.820 | 23.810 | 87.020 | 0.783 | 13.103 | 89.430 | 0.787 |
| Polarity (W, dict) | 34.066 | 83.792 | 0.818 | 23.810 | 87.230 | 0.781 | 15.172 | 89.430 | 0.785 |
| Polarity (S, ML) | 36.813 | 85.311 | 0.815 | 27.976 | 86.950 | 0.784 | 17.241 | 89.114 | 0.785 |
| Subjectivity (W, dict) | 31.868 | 84.009 | 0.814 | 19.643 | 86.811 | 0.763 | 6.897 | 89.430 | 0.764 |
| Subjectivity (S, dict) | 30.220 | 84.081 | 0.808 | 20.238 | 86.950 | 0.763 | 8.276 | 89.494 | 0.763 |
| Subjectivity (S, ML) | 24.176 | 83.068 | 0.803 | 17.857 | 87.090 | 0.763 | 9.655 | 89.557 | 0.777 |
| All textual indicators | 41.209 | 84.949 | 0.830 | 26.786 | 86.253 | 0.795 | 20.000 | 88.291 | 0.805 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

As mentioned above, Moody's generally upgrades sovereigns before the motivation for such action is reflected in the data. Therefore the credit rating committee has to justify their decision more extensively in the reports. This is reflected in sentiment scores and leads to higher model performance increases compared to downgrades and also compared to Standard & Poor's and Fitch. Regarding subjectivity measures, we notice relatively minor improvements compared to sentiment measures. The increase in classification accuracy is almost negligible for Fitch and Moody's. The obvious outlier is subjectivity from the dictionary-based approach at the word level in the Standard & Poor's sample. There seems to be no clear winner in the best performer category among subjectivity measures. Overall, the underlying models delivering the highest sensitivity on average are models with polarity scores from the machine learning approach, at the sentence level, followed by polarity from the dictionary-based approach, at the word level.

*Figure 3.* ROC curve comparison of the worst and best-performing models, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable (top), and with the upgrade dummy (equal to 1 if an upgrade occurs, 0 otherwise) as the dependent variable (bottom), following the point-in-time approach, by credit rating agency, i.e. Standard & Poor's (left), Fitch (centre) and Moody's (right)




*Source:* Own calculations

### 5.3.2 Through-the-cycle approach

Next, we continue with models assessed with lagged, current and forward values, corresponding to the through-the-cycle rating philosophy. Compared to models, following the point-in-time philosophy, we notice a substantial improvement in classification accuracy of downgrades, especially for Models 2 and 3, where sensitivity almost doubles, as evident from Table 33. Sensitivity for Moody's rises to 40.0% and 44.9%, respectively. Analysis for Standard & Poor's and Fitch yields slightly lower performance, at 37.4% and 38.3% for Model 2, and at 39.2% and 37.3% for Model 3. This is evidence supporting the notion that credit agencies follow the through-the-cycle rating philosophy, instead of the point-in-time philosophy. The overall performance of the model and area under the ROC curve increase as well, although the rise is not as high.

The improvement in sensitivity is even more pronounced for upgrades (see Table 34 and Appendix D). This is especially true for Moody's, where sensitivity in Models 1, 2, and 3 quadruples to 24.8%, 27.1% and 29.4%, respectively. This growth is extensive and supports the argument that Moody's is more forward-looking than Standard & Poor's and Fitch when it comes to upgrades. The performance of Fitch improves as well, as classification accuracy for upgrades more than doubles to 34.9% in Model 1, 34.2% in Model 2, and 40.1% in Model 3. The best performance is again detected in the Standard & Poor's sample, where sensitivities in the three baseline models stand at 46.1%, 50.3% and 54.8%.

*Table 33.* Classification accuracy of downgrades, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable, following the through-the-cycle approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | | Model 1 | | | | |
| No textual indicators | 28.571 | 90.263 | 0.858 | 32.031 | 91.981 | 0.870 | 28.462 | 92.126 | 0.872 |
| Net sentiment (W, dict) | 40.141 | 91.007 | 0.903 | 40.000 | 92.726 | 0.906 | 47.692 | 93.185 | 0.942 |
| Polarity (W, dict) | 43.662 | 91.341 | 0.912 | 33.600 | 91.687 | 0.908 | 40.000 | 93.042 | 0.928 |
| Polarity (S, ML) | 41.549 | 91.757 | 0.904 | 37.600 | 92.566 | 0.903 | 41.538 | 93.042 | 0.918 |
| Subjectivity (W, dict) | 31.690 | 90.425 | 0.859 | 32.800 | 92.406 | 0.882 | 34.615 | 92.539 | 0.894 |
| Subjectivity (S, dict) | 29.577 | 90.341 | 0.859 | 31.200 | 91.926 | 0.883 | 31.538 | 92.324 | 0.888 |
| Subjectivity (S, ML) | 29.577 | 90.591 | 0.861 | 32.000 | 92.006 | 0.875 | 30.000 | 92.037 | 0.879 |
| All textual indicators | 47.887 | 92.090 | 0.926 | 41.600 | 92.566 | 0.920 | 58.462 | 94.476 | 0.958 |
| | | | | | Model 2 | | | | |
| No textual indicators | 37.415 | 91.700 | 0.880 | 38.281 | 92.767 | 0.892 | 40.000 | 93.414 | 0.911 |
| Net sentiment (W, dict) | 43.662 | 91.424 | 0.920 | 40.800 | 92.966 | 0.919 | 54.615 | 93.902 | 0.956 |
| Polarity (W, dict) | 45.775 | 91.674 | 0.926 | 41.600 | 92.966 | 0.919 | 50.000 | 93.759 | 0.945 |
| Polarity (S, ML) | 46.479 | 92.340 | 0.919 | 44.000 | 92.886 | 0.915 | 53.846 | 94.046 | 0.937 |
| Subjectivity (W, dict) | 35.915 | 91.257 | 0.877 | 38.400 | 92.806 | 0.895 | 43.846 | 93.544 | 0.923 |
| Subjectivity (S, dict) | 35.915 | 91.174 | 0.878 | 35.200 | 92.406 | 0.897 | 40.769 | 93.329 | 0.919 |
| Subjectivity (S, ML) | 35.211 | 91.091 | 0.880 | 36.800 | 92.646 | 0.894 | 42.308 | 93.400 | 0.916 |
| All textual indicators | 50.000 | 92.590 | 0.935 | 42.400 | 92.726 | 0.928 | 59.231 | 94.261 | 0.966 |
| | | | | | Model 3 | | | | |
| No textual indicators | 39.161 | 91.761 | 0.888 | 37.302 | 92.693 | 0.894 | 44.882 | 93.478 | 0.913 |
| Net sentiment (W, dict) | 44.928 | 91.653 | 0.923 | 42.276 | 92.973 | 0.921 | 57.480 | 94.263 | 0.961 |
| Polarity (W, dict) | 48.551 | 92.243 | 0.928 | 41.463 | 92.811 | 0.922 | 51.181 | 93.755 | 0.949 |
| Polarity (S, ML) | 47.101 | 92.664 | 0.923 | 43.089 | 92.811 | 0.917 | 55.118 | 94.190 | 0.941 |
| Subjectivity (W, dict) | 40.580 | 91.821 | 0.885 | 39.837 | 92.973 | 0.898 | 44.882 | 93.755 | 0.925 |
| Subjectivity (S, dict) | 39.855 | 91.568 | 0.886 | 39.024 | 92.892 | 0.899 | 45.669 | 93.537 | 0.920 |
| Subjectivity (S, ML) | 38.406 | 91.400 | 0.890 | 39.024 | 92.892 | 0.897 | 48.819 | 93.755 | 0.917 |
| All textual indicators | 54.348 | 93.170 | 0.940 | 43.089 | 92.973 | 0.930 | 59.843 | 94.699 | 0.969 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

Similarly, as with the point-in-time approach, adding textual sentiment measures improves the results further (see Table 33). This is again more noticeable for sentiment measures compared to subjectivity measures, especially for Moody's. The classification accuracy for downgrades by Moody's climbs to between 40.0% (Model 1 with polarity from the dictionary-based approach, at word level) and 57.5% (Model 3 with the net sentiment from the dictionary-based approach, at word level). As before, the improvement is less profound and comparable for Standard & Poor's and Fitch, where sensitivity for the latter stands between 33.6% (Model 1 with polarity from the dictionary-based approach, at word level) and 44.0% (Model 2 with polarity from the machine learning approach, at the sentence level, and Model 3). Similarly for Standard & Poor's, the improvement after adding sentiment measures rises to between 40.1% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 48.6% (Model 3 with polarity from the dictionary-based approach, at word level). The addition of subjectivity measures only marginally helps to increase sensitivity, while in some cases, even decreases it.

Table 34 shows that sensitivity for upgrades when adding sentiment measures increases substantially for Standard & Poor's and Fitch, but a bit less for Moody's. Classification accuracy of upgrades for Moody's increases to between 28.7% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 35.7% (Model 3 with both polarity measures). On the other hand, the sensitivity of upgrades for Fitch, when including sentiment measures, ranges between 40.9% (Model 2 with polarity from the dictionary-based approach, at word level) and 53.7% (Model 3 with polarity from the machine learning approach, at the sentence level), while results for Standard & Poor's range between 50.0% (Model 2 with the net sentiment from the dictionary-based approach, at word level) and 64.6% (Model 3 with polarity from the machine learning approach, at the sentence level). This result is puzzling, as it contradicts our findings for downgrades and suggests that Standard & Poor's and Fitch, on average, explain their rationale for upgrades more clearly than Moody's. Extending the models with subjectivity measures, on average, offers a more noticeable rise than in previous cases. The predominantly leading models are models that include subjectivity measures from the machine learning approach, which are also the overall best performers for Moody's. In the case of Standard & Poor's and Fitch, sentiment measures, on average, outperform subjectivity measures.

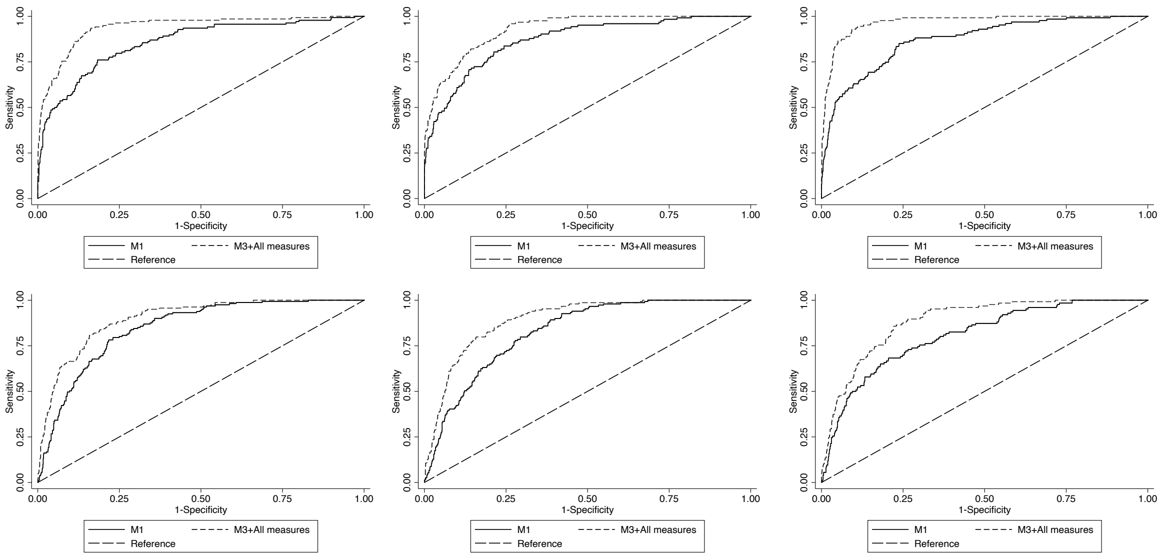
Overall, we find that sentiment measures perform better than subjectivity measures. Within the sentiment category, net sentiment appears to be a winner for downgrades, whereas polarity from the machine learning approach works best for upgrades. Within the subjectivity category, subjectivity from the dictionary-based approach outperforms machine learning approach for both downgrades and upgrades, albeit marginally for the latter.

*Table 34.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, following the through-the-cycle approach, using the 0.3 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | | Model 1 | | | | |
| No textual indicators | 46.108 | 85.475 | 0.845 | 34.868 | 86.635 | 0.820 | 24.806 | 89.907 | 0.808 |
| Net sentiment (W, dict) | 51.235 | 85.429 | 0.857 | 42.282 | 86.651 | 0.853 | 28.682 | 88.737 | 0.834 |
| Polarity (W, dict) | 52.469 | 85.762 | 0.858 | 41.611 | 86.091 | 0.853 | 31.783 | 89.527 | 0.830 |
| Polarity (S, ML) | 54.938 | 86.012 | 0.863 | 41.611 | 86.091 | 0.855 | 30.233 | 89.598 | 0.833 |
| Subjectivity (W, dict) | 58.642 | 86.928 | 0.864 | 34.899 | 86.491 | 0.828 | 23.256 | 89.311 | 0.811 |
| Subjectivity (S, dict) | 51.852 | 85.512 | 0.857 | 34.228 | 86.331 | 0.829 | 26.357 | 90.029 | 0.812 |
| Subjectivity (S, ML) | 45.062 | 84.929 | 0.845 | 38.926 | 87.290 | 0.827 | 33.333 | 90.316 | 0.818 |
| All textual indicators | 58.642 | 86.761 | 0.881 | 48.993 | 87.050 | 0.881 | 39.535 | 89.096 | 0.864 |
| | | | | | Model 2 | | | | |
| No textual indicators | 50.299 | 85.954 | 0.858 | 34.211 | 86.714 | 0.841 | 27.132 | 89.406 | 0.829 |
| Net sentiment (W, dict) | 50.000 | 85.096 | 0.869 | 42.282 | 86.251 | 0.860 | 34.109 | 88.881 | 0.848 |
| Polarity (W, dict) | 50.617 | 85.096 | 0.869 | 40.940 | 85.612 | 0.861 | 34.109 | 89.024 | 0.844 |
| Polarity (S, ML) | 54.938 | 85.595 | 0.872 | 47.651 | 87.050 | 0.867 | 31.008 | 89.168 | 0.845 |
| Subjectivity (W, dict) | 56.173 | 86.428 | 0.874 | 37.584 | 86.331 | 0.848 | 27.907 | 89.311 | 0.832 |
| Subjectivity (S, dict) | 54.321 | 86.095 | 0.867 | 36.242 | 86.491 | 0.848 | 31.008 | 89.742 | 0.831 |
| Subjectivity (S, ML) | 52.469 | 86.012 | 0.858 | 37.584 | 86.811 | 0.846 | 34.884 | 89.383 | 0.844 |
| All textual indicators | 60.494 | 87.261 | 0.888 | 54.362 | 88.249 | 0.892 | 46.512 | 89.527 | 0.878 |
| | | | | | Model 3 | | | | |
| No textual indicators | 54.819 | 86.511 | 0.868 | 40.132 | 86.180 | 0.848 | 29.365 | 88.623 | 0.841 |
| Net sentiment (W, dict) | 59.627 | 86.172 | 0.879 | 45.638 | 86.187 | 0.866 | 34.921 | 89.179 | 0.854 |
| Polarity (W, dict) | 60.870 | 86.341 | 0.880 | 45.638 | 86.187 | 0.868 | 35.714 | 89.397 | 0.850 |
| Polarity (S, ML) | 64.596 | 87.774 | 0.882 | 53.691 | 88.045 | 0.872 | 35.714 | 89.107 | 0.854 |
| Subjectivity (W, dict) | 57.143 | 87.099 | 0.880 | 43.624 | 86.995 | 0.856 | 33.333 | 89.397 | 0.843 |
| Subjectivity (S, dict) | 56.522 | 86.088 | 0.875 | 42.953 | 86.268 | 0.856 | 34.127 | 89.325 | 0.841 |
| Subjectivity (S, ML) | 56.522 | 86.425 | 0.867 | 40.940 | 86.349 | 0.854 | 38.889 | 90.051 | 0.857 |
| All textual indicators | 63.975 | 88.449 | 0.898 | 61.074 | 88.530 | 0.893 | 47.619 | 89.470 | 0.884 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Figure 4.* ROC curve comparison of the worst and best-performing models, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable (top), and with the upgrade dummy (equal to 1 if an upgrade occurs, 0 otherwise) as the dependent variable (bottom), following the through-the-cycle approach, by credit rating agency, i.e. Standard & Poor's (left), Fitch (centre) and Moody's (right)




*Source:* Own calculations

### 5.3.3 Robustness check

Credit rating agencies generally also publish outlook for issuers, which can be either stable, negative, or positive. Since outlook may be perceived as a substitute for textual sentiment and subjectivity, we perform a robustness check by including outlook in all models. Given that our previous findings suggest credit rating agencies follow the through-the-cycle rating philosophy, we only comment on the TTC results, while the PIT results are reported in Appendix D.

Starting with baseline models, we find that adding outlook further improves sensitivity (see Tables 35 and 36). The increase is highest for Moody's (51.8% for Model 1) and is comparable for Standard & Poor's (42.2% for Model 1) and Fitch (40.4% for Model 1). Furthermore, when adding sentiment and subjectivity measures, we do not observe any deterioration in improvements compared to models without outlook.

Specifically, when analysing downgrades and adding sentiment measures, sensitivity increases to between 62.5% (Model 1 with polarity from the machine learning approach, at the sentence level) and 74.3% (Model 3 with the net sentiment from the dictionary-based approach, at word level) for Moody's. As before, the increase is not as notable for Standard & Poor's and Fitch. For Fitch, it ranges between 48.7% (Model 1 with polarity from the machine learning approach, at the sentence level) and 58.4% (Model 2 with polarity from the machine learning approach, at the sentence level). Similarly, for Standard & Poor's, sensitivities range between 47.2% (Model 1 with the net sentiment from the dictionary-based approach, at word level) and 62.0% (Model 2 with polarity from the dictionary-based approach, at word level). The results for subjectivity measures support previous findings.

*Table 35.* Classification accuracy of downgrades, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the through-the-cycle approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | | Model 1 | | | | |
| No textual indicators | 42.177 | 91.460 | 0.903 | 40.351 | 92.454 | 0.930 | 51.786 | 93.414 | 0.950 |
| Net sentiment (W, dict) | 47.183 | 91.923 | 0.924 | 51.327 | 93.661 | 0.944 | 66.964 | 94.714 | 0.972 |
| Polarity (W, dict) | 52.817 | 92.590 | 0.929 | 49.558 | 92.994 | 0.945 | 63.393 | 94.714 | 0.965 |
| Polarity (S, ML) | 52.817 | 92.756 | 0.926 | 48.673 | 93.328 | 0.943 | 62.500 | 94.714 | 0.963 |
| Subjectivity (W, dict) | 40.141 | 91.007 | 0.900 | 51.327 | 93.495 | 0.936 | 58.036 | 94.107 | 0.953 |
| Subjectivity (S, dict) | 40.845 | 91.091 | 0.900 | 46.903 | 93.244 | 0.935 | 55.357 | 93.934 | 0.952 |
| Subjectivity (S, ML) | 38.732 | 90.674 | 0.902 | 43.363 | 92.827 | 0.932 | 52.679 | 93.761 | 0.952 |
| All textual indicators | 59.155 | 93.172 | 0.940 | 55.752 | 94.162 | 0.954 | 71.429 | 95.234 | 0.979 |
| | | | | | Model 2 | | | | |
| No textual indicators | 44.218 | 91.780 | 0.915 | 53.509 | 93.864 | 0.943 | 58.036 | 94.367 | 0.962 |
| Net sentiment (W, dict) | 55.634 | 93.089 | 0.937 | 56.637 | 94.662 | 0.953 | 73.214 | 95.581 | 0.976 |
| Polarity (W, dict) | 61.972 | 93.589 | 0.942 | 54.867 | 94.329 | 0.953 | 68.750 | 95.321 | 0.973 |
| Polarity (S, ML) | 54.225 | 92.923 | 0.939 | 58.407 | 94.746 | 0.951 | 69.643 | 95.754 | 0.969 |
| Subjectivity (W, dict) | 44.366 | 91.757 | 0.913 | 49.558 | 93.661 | 0.945 | 64.286 | 94.887 | 0.963 |
| Subjectivity (S, dict) | 43.662 | 91.590 | 0.913 | 51.327 | 93.828 | 0.944 | 60.714 | 94.627 | 0.963 |
| Subjectivity (S, ML) | 42.958 | 91.341 | 0.914 | 50.442 | 93.495 | 0.944 | 59.821 | 94.541 | 0.962 |
| All textual indicators | 60.563 | 93.422 | 0.951 | 58.407 | 94.662 | 0.958 | 75.893 | 96.014 | 0.982 |
| | | | | | Model 3 | | | | |
| No textual indicators | 46.853 | 92.326 | 0.919 | 50.442 | 93.808 | 0.945 | 59.633 | 94.561 | 0.964 |
| Net sentiment (W, dict) | 54.348 | 93.170 | 0.939 | 58.036 | 94.949 | 0.953 | 74.312 | 95.702 | 0.979 |
| Polarity (W, dict) | 59.420 | 93.761 | 0.944 | 57.143 | 94.276 | 0.954 | 66.972 | 95.439 | 0.975 |
| Polarity (S, ML) | 54.348 | 93.170 | 0.943 | 54.464 | 94.697 | 0.951 | 66.055 | 95.088 | 0.973 |
| Subjectivity (W, dict) | 48.551 | 92.243 | 0.918 | 50.893 | 93.939 | 0.946 | 64.220 | 94.825 | 0.967 |
| Subjectivity (S, dict) | 47.101 | 92.243 | 0.918 | 50.893 | 94.108 | 0.946 | 63.303 | 94.737 | 0.966 |
| Subjectivity (S, ML) | 46.377 | 92.327 | 0.919 | 52.679 | 94.192 | 0.945 | 59.633 | 94.649 | 0.965 |
| All textual indicators | 62.319 | 94.182 | 0.954 | 58.036 | 95.118 | 0.958 | 76.147 | 96.053 | 0.984 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

Finally, in the case of upgrades, the initial increase for all three credit rating agencies is substantial compared to results without outlook. The added value of including sentiment and subjectivity measures in the models is less pronounced as in the case of downgrades but far from negligible. Among sentiment measures, net sentiment from the dictionary-based approach appears to notably offer additional information not captured by Standard & Poor's and Fitch outlooks alone, while the same can be said for polarity from the machine learning approach and Moody's outlook. As with previous results, subjectivity measures, on average, add less or no significant value to the performance of the models.

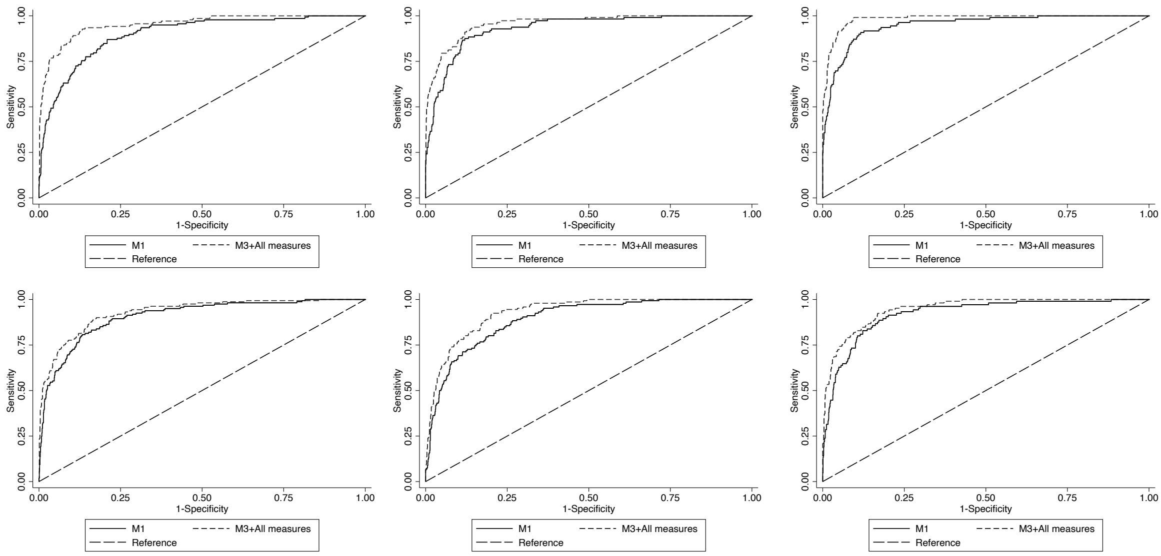
Overall, evidence suggests that sentiment and, to some extent, subjectivity measures offer important insights into the changes of sovereign credit ratings that go beyond the simple outlook of credit rating agencies.

*Table 36.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the through-the-cycle approach, using the 0.3 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 64.072 | 88.667 | 0.904 | 56.849 | 89.055 | 0.891 | 42.056 | 92.808 | 0.922 |
| Net sentiment (W, dict) | 66.049 | 89.842 | 0.910 | 63.699 | 89.741 | 0.900 | 46.729 | 93.068 | 0.925 |
| Polarity (W, dict) | 67.284 | 90.008 | 0.910 | 61.644 | 89.324 | 0.900 | 44.860 | 92.634 | 0.926 |
| Polarity (S, ML) | 66.049 | 89.675 | 0.911 | 61.644 | 89.741 | 0.907 | 48.598 | 93.241 | 0.930 |
| Subjectivity (W, dict) | 64.815 | 89.092 | 0.912 | 58.219 | 89.491 | 0.894 | 45.794 | 92.981 | 0.922 |
| Subjectivity (S, dict) | 63.580 | 89.092 | 0.909 | 59.589 | 89.575 | 0.895 | 42.991 | 92.981 | 0.922 |
| Subjectivity (S, ML) | 64.815 | 88.926 | 0.906 | 60.959 | 89.575 | 0.896 | 44.860 | 92.894 | 0.924 |
| All textual indicators | 66.049 | 89.675 | 0.918 | 65.753 | 90.075 | 0.921 | 50.467 | 93.674 | 0.936 |
| | | | | Model 2 | | | | | |
| No textual indicators | 66.467 | 88.667 | 0.911 | 60.274 | 89.386 | 0.900 | 43.925 | 92.808 | 0.928 |
| Net sentiment (W, dict) | 69.136 | 90.008 | 0.917 | 63.014 | 89.741 | 0.906 | 45.794 | 93.068 | 0.930 |
| Polarity (W, dict) | 67.901 | 89.675 | 0.917 | 60.274 | 88.991 | 0.905 | 48.598 | 93.241 | 0.930 |
| Polarity (S, ML) | 65.432 | 89.259 | 0.917 | 61.644 | 89.491 | 0.914 | 57.009 | 94.281 | 0.935 |
| Subjectivity (W, dict) | 67.284 | 89.425 | 0.918 | 62.329 | 89.324 | 0.904 | 45.794 | 92.894 | 0.928 |
| Subjectivity (S, dict) | 66.667 | 88.759 | 0.915 | 60.274 | 89.158 | 0.904 | 44.860 | 92.808 | 0.928 |
| Subjectivity (S, ML) | 66.667 | 89.009 | 0.914 | 63.014 | 89.908 | 0.904 | 48.598 | 93.501 | 0.932 |
| All textual indicators | 69.136 | 89.925 | 0.924 | 67.123 | 90.242 | 0.928 | 51.402 | 93.934 | 0.944 |
| | | | | Model 3 | | | | | |
| No textual indicators | 69.880 | 89.418 | 0.917 | 63.014 | 89.707 | 0.907 | 48.571 | 93.158 | 0.932 |
| Net sentiment (W, dict) | 73.913 | 90.809 | 0.924 | 65.753 | 90.152 | 0.912 | 51.429 | 93.947 | 0.933 |
| Polarity (W, dict) | 73.913 | 90.978 | 0.924 | 63.699 | 89.815 | 0.912 | 51.429 | 93.509 | 0.933 |
| Polarity (S, ML) | 72.050 | 90.388 | 0.923 | 65.753 | 89.983 | 0.919 | 56.190 | 94.123 | 0.938 |
| Subjectivity (W, dict) | 70.186 | 90.219 | 0.924 | 63.014 | 89.562 | 0.910 | 47.619 | 93.246 | 0.932 |
| Subjectivity (S, dict) | 69.565 | 90.051 | 0.921 | 63.014 | 89.731 | 0.910 | 45.714 | 92.807 | 0.932 |
| Subjectivity (S, ML) | 69.565 | 89.966 | 0.920 | 61.644 | 89.646 | 0.910 | 52.381 | 93.684 | 0.936 |
| All textual indicators | 73.292 | 90.388 | 0.930 | 67.123 | 89.899 | 0.930 | 54.286 | 94.035 | 0.949 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Figure 5.* ROC curve comparison of the worst and best-performing models, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable (top), and with the upgrade dummy (equal to 1 if an upgrade occurs, 0 otherwise) as the dependent variable (bottom), including the outlook as one of the explanatory variables and following the through-the-cycle approach, by credit rating agency, i.e. Standard & Poor's (left), Fitch (centre) and Moody's (right)




## 5.4 Conclusion

This chapter offers unique insights into sovereign rating transitions. We utilise textual analysis techniques to analyse sovereign credit rating reports in order to extract sentiment and subjectivity scores. To the best of our knowledge, this has not been done on such a comprehensive scale before. We apply both dictionary-based and machine learning methods and construct six different measures. We compare them in terms of the classification accuracy of both downgrades and upgrades. We find that, on average, sentiment measures ensure higher sensitivity than subjectivity measures. We additionally observe that downgrades are more easily predicted than upgrades. Relative to the textual analysis approach, dictionary-based methods seem to work best for sentiment measures, while machine learning techniques lead the race for subjectivity measures. Additionally, we estimate all our models with two rating philosophies in mind, namely point-in-time and through-the-cycle. Our results show that credit rating agencies follow the latter when assigning sovereign credit ratings. Finally, we perform robustness checks by adding an outlook to all models, which confirms our previous results.

We acknowledge the potential drawbacks of our approach, namely the neglect of different probabilities of rating changes for particular rating classes. We believe our approach is a necessary step in order to identify the most informative measures and best models, but mostly to highlight the importance of credit rating reports in the first place.

# 6 Sovereign credit rating announcements, markets and the role of textual sentiment in credit rating reports

## 6.1 Introduction

Credit rating agencies make various periodic or ad hoc rating announcements in order to reflect the latest available information about the sovereigns' creditworthiness, including rating downgrades or upgrade, rating outlooks and rating watches. The existing literature agrees that such rating announcements result in significant bond market reactions, either in changes in bond yields or in CDS spreads (Afonso et al., 2012; Drago & Gallo, 2016; Gande & Parsley, 2005; Ismailescu & Kazemi, 2010; Kiff et al., 2012). This leads us to question whether markets also react to additional information hiding in sovereign credit rating reports. According to Agarwal et al. (2019), there is some evidence of negative tone or sentiment impacting market returns.

CDS spreads represent the market price of creditworthiness, but they are affected by multiple factors that change constantly. On the other hand, sovereign credit ratings are perceived as relatively stable. Nevertheless, there is an important connection between the two. Drago and Gallo (2016) study euro area CDS markets and find significant market reactions to downgrades and upgrades, but not other forms of rating announcements. Ismailescu and Kazemi (2010) focus on emerging markets, where they identify meaningful effects of positive events on sovereign CDS spreads. Gande and Parsley (2005) claim that the market response to rating announcements is asymmetric, with notable increases in spreads following a negative event and trivial reaction following a positive event. Kiff et al. (2012) stress the importance of the shift from investment to speculative grade rating categories, and vice versa, for market returns.

The main objective of this chapter is to shift focus from rating announcements alone, to credit rating reports that accompany them. Building on the traditional event study, we examine how market participants perceive the credit rating reports. Specifically, we test the relationship between the six textual sentiment and subjectivity measures defined in subsection 3.1 and cumulative abnormal returns in CDS spreads during a three-day event window surrounding a rating announcement. The dataset covers daily CDS spreads for 69 countries from December 14th 2007 do April 23rd 2020. We take into account different types of rating announcements, namely downgrades and upgrades, or changes in outlook and watch status. We thus define four distinct events. With negative and positive events, we consider both changes in ratings and changes in outlooks/watches, whereas, with downgrades and upgrades, we focus solely on rating changes. We base our findings on four different samples. The focus is on the 'first mover' sample following Michaelides, Milidonis, and Nishiotis (2019), which is a pooled sample of rating announcements that are not foreshadowed by other rating announcements in the 20 trading days prior the observed event. The remaining three samples correspond to the three global credit rating agencies, Standard & Poor's, Fitch and Moody's.

We first aim to relate to the previous findings. Specifically, we test whether cumulative abnormal returns in CDS spreads immediately post the public rating announcement are statistically significantly different from zero. We estimate normal returns using the market model. Our results that both positive and negative rating announcements cause substantial disruption in global CDS markets are consistent with previous findings (Drago & Gallo, 2016; Ismailescu & Kazemi, 2010). Furthermore, market participants appear to value Standard & Poor's opinion above the opinions of Fitch or Moody's. We also survey market reactions to rating changes of advanced economies and emerging markets and find no discrepancies.

In the next step, we investigate the determinants of the cumulative abnormal returns. We add other financial variables that could potentially influence CDS returns, such as local stock market returns, US excess return and volatility risk premium. The 5-year constant maturity treasury rate seems to significantly affect cumulative abnormal CDS returns during negative events and downgrades, while US excess return impacts returns during positive events and upgrades. This supports existing findings, that global financial factors play a significant role in sovereign CDS spreads (Blommestein et al., 2016; Longstaff et al., 2011). We also control for the potentially economically important transition between speculative and investment grade categories. The upgrade from speculative to investment grade rating categories substantially influences cumulative abnormal returns in the observed event window, but not the other way around.

Finally, we examine if and how the six textual sentiment and subjectivity measures affect CDS returns during rating announcements. The general conclusion is that textual sentiment and subjectivity scores from sovereign credit rating reports do not have a meaningful effect on cumulative abnormal returns. However, there is limited evidence that subjectivity measures contribute to changes in CDS spreads during a positive event. This is an indication that market participants recognise the additional informational value in the credit rating agencies' opinion as expressed in the reports. Contrary to our expectations, the comparison of individual credit rating agencies' estimates indicates that markets assign a marginally superior meaning to Moody's opinion, as expressed in the reports, compared to Standard & Poor's or Fitch.

We believe this chapter contributes to existing literature both from a methodological and practical standpoint. First, to our knowledge, we perform the most comprehensive and detailed examination of the relationship between sovereign CDS or bond markets and textual analysis indicators, extracted from credit rating reports. Furthermore, similar studies are practically non-existent. To our knowledge, only one other study (Agarwal et al., 2019) tackles this question, but on a substantially smaller scale relative to our framework.

Second, the conclusions of our study have important implications for both market participants and borrowing countries. We provide them with novel insights into the importance of credit rating reports following credit rating announcements and their significance for market returns or borrowing costs. As already stressed in Chapter 4, any elimination of information asymmetries between credit rating agencies and market participants is essential for both private and professional investors, as well as borrowing governments.

The remainder of this chapter is structured as follows. In subsection 6.2, we outline the methodological framework for the analysis. We report the results in subsection 6.3, namely the results of the traditional event study, its extensions and, most importantly, the extensions with textual sentiment and subjectivity measures. We conclude in subsection 6.4.

## 6.2 The framework

We base our analysis on the event study framework. We define an event as a rating announcement, which can be either a downgrade, upgrade, or a change in the rating outlook/watch status. We thus define a comprehensive credit rating (CCR) in line with Gande and Parsley (2005) and Ismailescu and Kazemi (2010), which takes into account both the rating and the outlook. As Gande and Parsley (2005) argue, focusing only on rating changes could be too restrictive and could potentially eliminate valuable additional information provided by credit rating outlook and watch announcements. We construct the comprehensive credit rating by combining the numerical ratings already defined in Table 2 with additional values ranging from 0.5 for a positive outlook to -0.5 for a negative outlook. Stable outlook has a value of 0. Specifically, a change in the comprehensive credit rating from 21 to 20.5 means that a particular country's outlook has changed from stable to negative. A negative event is defined as a decrease in comprehensive credit rating, while a positive event is defined as an increase in comprehensive credit rating. As a robustness check, we also study downgrades and upgrades alone.

We begin with the pooled sample of 1459 observations, namely 553 by Standard & Poor's, 458 by Fitch, and 448 by Moody's. First, we exclude observations with multiple events on the same day, which reduces the sample to 1396 observations, with 748 negative and 648 positive events. Next, we construct the so-called 'first mover' sample in line with Michaelides et al. (2019). The objective is to control for contamination in the period before each rating announcement by removing any observations that undergo other changes in the comprehensive credit rating by any of the credit rating agencies in the 20 trading days preceding the event. Additionally, the sample includes only issuers for which CDS spreads are not constant in two days surrounding the event, i.e. not stale. This leaves us with a total of 1101 events, comprised of 590 negative and 511 positive events. If we define the event as downgrades or upgrades alone, the final sample consists of 579 observations, with 340 downgrades and 239 upgrades. The final number of events for different event definitions and by credit rating agency is presented in Table 37.

*Table 37.* Frequency of events by definition, where negative/positive event covers both the credit rating and outlook change, while downgrade/upgrade captures rating changes only, by credit rating agency

| | First mover | S&P | Fitch | Moody's |
|---|---|---|---|---|
| Negative event | 590 | 162 | 139 | 138 |
| Positive event | 511 | 175 | 130 | 139 |
| Total | 1101 | 337 | 269 | 277 |
| Downgrade | 340 | 97 | 76 | 76 |
| Upgrade | 239 | 87 | 63 | 61 |
| Total | 579 | 184 | 139 | 137 |

*Source:* Own calculations

Since the event study is only the basis for further analysis, we focus on the narrowest event window, namely the (-1,1) time interval. With the narrow definition of the event window and the 'first mover' sample restrictions, we avoid contamination by other events that could affect CDS spreads and bias the results of the analysis. We construct cumulative abnormal returns for the observed event window using the market model method in line with Drago and Gallo (2016) and Micu et al. (2006).

Following Micu et al. (2006), we calculate the CDS returns as:

$$R_{it}^{CDS} = \frac{S_{it}}{S_{it-1}} - 1 \tag{15}$$

where $S_{it}$ is the CDS spread for issuer $i$ on day $t$.

The alternative to this methodology is a calculation of absolute changes in CDS spreads, calculated as simple daily differences in basis points: $S_{it} - S_{it-1}$ (e.g. Finnerty, Miller, & Chen, 2013; Ismailescu & Kazemi, 2010). Micu et al. (2006) argue that the methodology used in this thesis is superior because it enables a comparison of returns across markets, and it adjusts for differences in the level of spreads across issuers.

We estimate normal returns using the market model, and calculate abnormal returns as the difference between realised and expected returns:

$$AR_{it}^{CDS} = R_{it}^{CDS} - \alpha_i^{CDS} - \beta_i^{CDS} R_{mt}^{CDS} \tag{16}$$

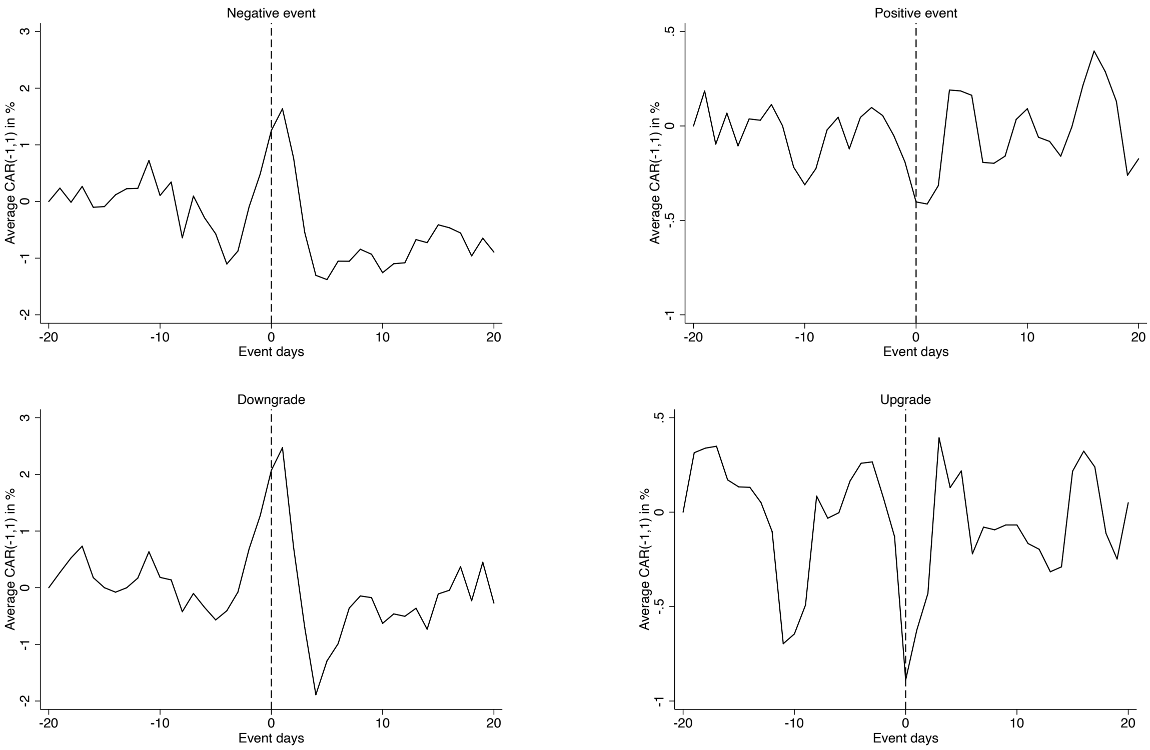
where $R_{it}^{CDS}$ is the daily return for issuer $i$ on day $t$, as calculated in (13), and $R_{mt}^{CDS}$ is the market return on day $t$. We construct an index based on CDS spreads included in our sample, and use is to calculate the market return. The index return is equal to the median CDS return for the total sample, which is consistent with Drago and Gallo (2016) and Micu et al. (2006). The parameters $\alpha_i^{CDS}$ and $\beta_i^{CDS}$ are estimated over a six-month period, beginning nine and ending three months before the event.

Instead of the market model approach, many studies apply the abnormal CDS spread changes (ASC) approach (e.g. Agarwal et al., 2019; Finnerty et al., 2013; Gande & Parsley, 2005; Hull, Predescu, & White, 2004; Ismailescu & Kazemi, 2010), where they calculate the simple difference between CDS spread changes and index changes. The advantage of using the market model over the latter approach is the ability to control for possible market-wide systemic factors that could affect CDS markets simultaneously.

Finally, the cumulative abnormal return is calculated as the sum of abnormal returns. Figure 6 shows the average CAR(-1,1) in the twenty trading days before and after each event for the 'first mover' sample. The average cumulative abnormal return suggests that the market reacts relatively strongly to a negative event or a downgrade. There is some evidence of anticipation of the event, but the majority of the reaction is at the time of the event. On the other hand, there is little evidence that the market reacts to a positive event or an upgrade on the day of the rating announcement. It appears that markets anticipate a positive rating announcement (either a positive outlook or an upgrade).

We then perform a traditional event study by employing a t-test to check whether the cumulative abnormal CDS returns are significantly different from zero. Specifically, we use a constant-only OLS model which enables us the use of robust standard errors.

*Figure 6.* Average cumulative abnormal return in the time interval (-1,1) in the 'first mover' sample for the four event definitions: negative event (upper left corner), positive event (upper right corner), downgrade (lower left corner) and upgrade (lower right corner), in %




*Source:* Thomson Reuters Datastream, own calculations

In the next step, we employ ordinary least squares to analyse the relationship between cumulative abnormal returns in the defined event window and financial variables listed in Table 7. Additionally, we also include the economic development dummy variable to control for potential differences between emerging and advanced countries, and a dummy controlling for the transition of in and out of investment grade. Finally, we include textual sentiment and subjectivity measures from credit rating action reports to check whether the market reacts to either textual sentiment or subjectivity expressed in the sovereign credit rating reports. We thus estimate the following model:

$$CAR_{it}^{CDS}(-1,1) = x'_{it}\beta + \varepsilon_{it} \tag{17}$$

with different specifications of $x'_{it}$ as discussed above.

As a robustness check, we also repeat the analysis for each of the three credit rating agencies separately, where we apply the same sample restrictions as above.

## 6.3 Results

### 6.3.1 Event study

We first perform a traditional event study for the four event definitions and four different samples, namely the 'first mover' sample and the Standard & Poor's, Fitch and Moody's sample (Model 1). The results are presented in Table 38. The cumulative abnormal returns are statistically significantly different from zero for all four definitions of events for the 'first mover' sample, suggesting that the market on average reacts to 'first' rating announcements, i.e. rating announcements not preceded by other rating announcements in the last month. Similarly, for Standard & Poor's, the cumulative abnormal returns are again statistically significantly different from zero for all four definitions of events. Additionally, the reactions are stronger than for the 'first mover' sample. As expected, the reactions are also stronger for downgrades/upgrades compared to changes in comprehensive credit rating. The signs of cumulative abnormal returns are in line with our expectations: positive for a negative event/downgrade, and negative for a positive event/upgrade. However, the cumulative abnormal returns are statistically significantly different from zero only for a negative event and downgrade for Moody's, and for a negative event for Fitch.

The conclusion that only negative rating announcements have significant impacts on CDS spreads is consistent with the existing literature (e.g. Afonso et al., 2012; Kiff

*Table 38.* Event study results for the cumulative abnormal returns in the time interval (-1,1) during rating announcements for the 'first mover' sample and S&P, Fitch and Moody's

| | First mover | S&P | Fitch | Moody's |
|---|---|---|---|---|
| | Negative event | | | |
| Constant | 0.020*** | 0.028*** | 0.011* | 0.017** |
| | (0.004) | (0.007) | (0.006) | (0.008) |
| Observations | 590 | 162 | 139 | 138 |
| | Positive event | | | |
| Constant | -0.004** | -0.010*** | 0.000 | 0.001 |
| | (0.002) | (0.003) | (0.004) | (0.003) |
| Observations | 511 | 175 | 130 | 139 |
| | Downgrade | | | |
| Constant | 0.023*** | 0.035*** | 0.010 | 0.020* |
| | (0.006) | (0.012) | (0.011) | (0.012) |
| Observations | 340 | 97 | 76 | 76 |
| | Upgrade | | | |
| Constant | -0.009*** | -0.015*** | -0.007 | 0.001 |
| | (0.003) | (0.005) | (0.005) | (0.006) |
| Observations | 239 | 87 | 63 | 61 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

et al., 2012). Kiff et al. (2012) find that Moody's transmit most of the incremental information value through negative credit warnings, i.e. outlooks and watches, but find a minimal market reaction to upgrades. The initial results suggest that the markets predominantly react to the rating information provided by Standard & Poor's. However, Kiff et al. (2012) point out, that Moody's issues credit rating outlooks or watches before rating changes more often than Standard & Poor's and Fitch, which impacts the information value of credit changes. Additionally, their analysis shows that Moody's and Fitch lag behind Standard & Poor's more often than Standard & Poor's lags behind others, especially regarding negative rating actions. This means that Fitch's or Moody's credit actions have less or no significant impact on markets compared to Standard & Poor's, which is corroborated in our results.

On the other hand, Ismailescu and Kazemi (2010) and Drago and Gallo (2016) show that markets react to both positive and negative rating announcements. Their conclusion is based on the Standard & Poor's sample, given that previous studies (e.g. Gande & Parsley, 2005) argue that Standard & Poor's rating changes are more frequent, less anticipated by markets and not preceded by other rating agencies. Focusing on Standard & Poor's results alone, our results are consistent.

### 6.3.2 Model extensions

We now analyse model extensions with additional explanatory variables: In Model 2, we add only the economic development dummy to check whether there are differences in market reactions between advanced economies and emerging markets. Model 3 includes financial variables listed in Table 7. Finally, Model 4 combines Models 2 and 3 with financial variables and economic development dummy.

Table 39 shows the results for Model 2. The results suggest there are no statistically significant differences in how markets perceive credit announcements of advanced economies over emerging markets either based on the 'first mover' sample or by a specific credit rating agency.

*Table 39.* Analysis of the effect of economic development on cumulative abnormal returns in the time interval (-1,1) during rating announcements for the 'first mover' sample and S&P, Fitch and Moody's

| | First mover | S&P | Fitch | Moody's |
|---|---|---|---|---|
| | Negative event | | | |
| Economic development | 0.007 | -0.002 | 0.008 | 0.028 |
| | (0.007) | (0.015) | (0.013) | (0.017) |
| Constant | 0.017*** | 0.028*** | 0.008 | 0.005 |
| | (0.004) | (0.010) | (0.008) | (0.006) |
| Observations | 590 | 162 | 139 | 138 |
| | Positive event | | | |
| Economic development | 0.001 | 0.009 | 0.008 | -0.001 |
| | (0.004) | (0.006) | (0.008) | (0.008) |
| Constant | -0.004* | -0.014*** | -0.003 | 0.001 |
| | (0.002) | (0.004) | (0.004) | (0.004) |
| Observations | 511 | 175 | 130 | 139 |
| | Downgrade | | | |
| Economic development | 0.008 | -0.006 | 0.030 | 0.030 |
| | (0.011) | (0.023) | (0.023) | (0.026) |
| Constant | 0.020** | 0.038* | -0.003 | 0.007 |
| | (0.008) | (0.020) | (0.013) | (0.010) |
| Observations | 340 | 97 | 76 | 76 |
| | Upgrade | | | |
| Economic development | 0.002 | 0.015 | 0.008 | 0.007 |
| | (0.006) | (0.009) | (0.009) | (0.014) |
| Constant | -0.010*** | -0.022*** | -0.010* | -0.002 |
| | (0.003) | (0.007) | (0.006) | (0.006) |
| Observations | 239 | 87 | 63 | 61 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We present the 'first mover' sample results for Model 3 in Table 40. We separately report the dummy capturing either the transition from investment grade to speculative grade (IG to SG dummy) or from speculative grade to investment grade (SG to IG dummy), depending on the event definition. The results show that different variables are affecting the cumulative abnormal returns for particular event definitions. 5-year constant maturity treasury rate negatively affects cumulative abnormal returns in the event of a negative rating announcement or a downgrade. Additionally, the local stock market also negatively affects returns in the case of a negative event. The results for a positive event or an upgrade are less agreeing. While the US excess return is statistically significant in both cases, the change in the local exchange rate against the dollar and the volatility risk premium impact returns in the case of a positive event, but the local stock market return and the 5-year constant maturity treasury rate affect returns in the case of an upgrade. To a limited degree, our results support the findings of Longstaff et al. (2011) and Blommestein et al. (2016), who claim that global financial factors have a substantial impact on sovereign CDS spreads. Agarwal et al. (2019), who analyse cumulative abnormal returns for a pooled sample of downgrades and upgrades, find that CDS spread changes react to local stock market return and the volatility risk premium.

*Table 40.* Analysis of the effect of financial variables on cumulative abnormal returns in the time interval (-1,1) during rating announcements for the 'first mover' sample

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.434** | 0.055 | -0.379 | -0.514** |
| | (0.199) | (0.100) | (0.281) | (0.211) |
| Exchange rate | 0.969 | 0.776* | 0.307 | -0.575 |
| | (0.608) | (0.409) | (0.685) | (0.583) |
| US excess return | 0.058 | -0.174*** | 0.088 | -0.154* |
| | (0.059) | (0.057) | (0.067) | (0.080) |
| 5Y CMT rate | -0.180* | -0.050 | -0.291** | -0.140* |
| | (0.107) | (0.063) | (0.136) | (0.074) |
| Volatility risk premium | 0.012 | 0.060*** | 0.049 | 0.040 |
| | (0.031) | (0.022) | (0.044) | (0.024) |
| IG to SG dummy | 0.007 | | 0.011 | |
| | (0.014) | | (0.014) | |
| SG to IG dummy | | -0.028*** | | -0.025*** |
| | | (0.008) | | (0.008) |
| Constant | 0.019*** | -0.002 | 0.022*** | -0.005 |
| | (0.004) | (0.002) | (0.006) | (0.003) |
| Observations | 535 | 461 | 312 | 209 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The transition from investment to speculative grade is not statistically significant. This result is surprising since most institutional investors are required to hold only investment grade securities in their portfolio and have to sell the assets with speculative grade credit rating. A possible explanation is that investors anticipate such an event and sell off the bonds beforehand. Micu et al. (2006), analysing corporate CDS spreads, conclude that market participants anticipate rating announcements. Additionally, Covitz and Harrison (2003) estimate that nearly three-quarters of the change in bond spreads takes place sometime in the six months before a rating downgrade. On the other hand, the transition from speculative to investment grade is statistically significant for both a positive event and an upgrade. Finnerty et al. (2013) argue, that this may be because a positive crossover is more difficult to accomplish than a negative crossover, and is thus more likely to surprise markets. The results are partially consistent with Kiff et al. (2012) and Agarwal et al. (2019), who find that both upgrades and downgrades in and out of the investment grade have a significant impact on CDS spreads.

The results by agency are reported in Tables 41, 42 and 43. Interestingly, the impact of explanatory variables on the cumulative abnormal return in the case of a negative rating announcement by Standard & Poor's is relatively limited, apart from the local stock market return, for a negative event, while the constant implies a significant market reaction. This may be evidence that the market participants base their reaction mostly on the credit rating announcement, indicating a somewhat higher degree of confidence in Standard & Poor's opinion. While the explanatory power of financial and dummy variables for a negative rating announcement by Fitch or Moody's is also almost nonexistent, the constant, apart from a negative event by Moody's, is also statistically insignificant. This suggests that the markets do not put as much weight to Fitch's or Moody's credit opinion as they do to Standard & Poor's in the case of negative rating announcements. This is in line with our previous findings.

In the case of a positive rating announcement by Standard & Poor's, the abnormal returns are mostly impacted by the local stock market return, US excess return and volatility risk premium. Additionally, the transition from speculative to investment grade significantly reduces the returns, corresponding to a reduced risk of default. The results for Fitch and Moody's are less straightforward. Fitch's positive event returns are impacted by the local stock market, change in the exchange rate and 5-year constant maturity treasury rate, while the change in the exchange rate explains most of the variability in the case of an upgrade. Nor the transition from speculative to investment grade, nor the constant are statistically significant, again indicating that markets valuation of Fitch's opinion is relatively low. Finally, apart from the transition from speculative to investment grade after an upgrade, nothing is statistically significant for Moody's positive rating announcements. The general conclusion is fairly similar as in the case of a negative rating announcement.

*Table 41.* Analysis of the effect of financial variables on cumulative abnormal returns in the time interval (-1,1) during rating announcements for S&P

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.604* | 0.127*** | -0.542 | -0.629** |
| | (0.308) | (0.043) | (0.423) | (0.314) |
| Exchange rate | -0.056 | 0.069 | -0.786 | -0.570 |
| | (0.834) | (0.637) | (0.880) | (0.866) |
| US excess return | 0.038 | -0.229** | 0.114 | -0.278** |
| | (0.132) | (0.091) | (0.127) | (0.126) |
| 5Y CMT rate | -0.220 | -0.004 | -0.269 | -0.112 |
| | (0.171) | (0.095) | (0.256) | (0.095) |
| Volatility risk premium | 0.001 | 0.060** | 0.023 | 0.054** |
| | (0.062) | (0.026) | (0.066) | (0.024) |
| IG to SG dummy | 0.007 | | 0.005 | |
| | (0.029) | | (0.026) | |
| SG to IG dummy | | -0.045*** | | -0.052*** |
| | | (0.016) | | (0.018) |
| Constant | 0.028*** | -0.005 | 0.036** | -0.009** |
| | (0.008) | (0.003) | (0.015) | (0.004) |
| Observations | 148 | 162 | 91 | 79 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table 42.* Analysis of the effect of financial variables on cumulative abnormal returns in the time interval (-1,1) during rating announcements for Fitch

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.437 | -0.758** | -1.151 | -0.378 |
| | (0.676) | (0.292) | (0.753) | (0.257) |
| Exchange rate | -0.347 | 1.210* | 0.267 | 1.621** |
| | (1.068) | (0.727) | (1.835) | (0.666) |
| US excess return | 0.088 | -0.111 | 0.165 | -0.182 |
| | (0.102) | (0.155) | (0.183) | (0.121) |
| 5Y CMT rate | 0.026 | -0.257* | -0.419 | -0.123 |
| | (0.174) | (0.133) | (0.401) | (0.128) |
| Volatility risk premium | -0.022 | 0.032 | 0.069 | -0.042 |
| | (0.053) | (0.041) | (0.092) | (0.046) |
| IG to SG dummy | 0.018 | | 0.033 | |
| | (0.032) | | (0.032) | |
| SG to IG dummy | | -0.005 | | -0.004 |
| | | (0.010) | | (0.010) |
| Constant | 0.012 | 0.002 | 0.006 | -0.003 |
| | (0.008) | (0.004) | (0.013) | (0.005) |
| Observations | 122 | 119 | 67 | 55 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table 43.* Analysis of the effect of financial variables on cumulative abnormal returns in the time interval (-1,1) during rating announcements for Moody's

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | 0.049 | -0.319 | 0.215 | -0.326 |
| | (0.584) | (0.442) | (1.024) | (0.809) |
| Exchange rate | 3.093** | 0.485 | 6.145 | -0.248 |
| | (1.512) | (0.748) | (3.777) | (1.565) |
| US excess return | -0.214 | 0.013 | -0.268 | 0.026 |
| | (0.194) | (0.109) | (0.264) | (0.204) |
| 5Y CMT rate | -0.007 | 0.017 | 0.254 | -0.264 |
| | (0.319) | (0.121) | (0.534) | (0.287) |
| Volatility risk premium | 0.072 | 0.051 | 0.175 | 0.009 |
| | (0.078) | (0.049) | (0.147) | (0.098) |
| IG to SG dummy | 0.048 | | 0.035 | |
| | (0.041) | | (0.042) | |
| SG to IG dummy | | -0.017 | | -0.026* |
| | | (0.014) | | (0.015) |
| Constant | 0.014* | 0.002 | 0.016 | 0.006 |
| | (0.008) | (0.004) | (0.014) | (0.008) |
| Observations | 125 | 120 | 69 | 50 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Overall, the results suggest that market participants value Standard & Poor's opinion above the opinion of the others, followed by Moody's and lastly Fitch. This is in line with previous findings (e.g. Chen et al., 2019; Kiff et al., 2012).

The results for Model 4 are presented in Appendix E and are fairly similar to results in Models 2 and 3.

### 6.3.3 Textual sentiment and subjectivity measures

Finally, we are interested to know whether markets react to textual sentiment or subjectivity expressed in sovereign credit rating reports. We separately add the six textual sentiment and subjectivity measures to Model 4, which controls for movements in financial markets and economic development. We report the coefficients for sentiment and subjectivity measures for the 'first mover' sample in Table 44. The results indicate that the cumulative abnormal returns are not affected by textual sentiment or subjectivity in the case of a negative rating announcement. However, there is some evidence that subjectivity measures at sentence level have a positive effect on cumulative abnormal returns in the event of an upward change in the comprehensive credit rating. This suggests that market participants attribute additional information to opinions expressed in the credit rating reports after an upgrade or a positive outlook.

*Table 44.* Relationship between cumulative abnormal returns in the time interval (-1,1) during rating announcements and sentiment and subjectivity measures for the 'first mover' sample

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Net sentiment (W, dict) | -0.001 | 0.001 | -0.001 | 0.001 |
| | (0.002) | (0.001) | (0.002) | (0.003) |
| Polarity (W, dict) | -0.016 | 0.005 | -0.024 | 0.002 |
| | (0.018) | (0.012) | (0.030) | (0.021) |
| Polarity (S, ML) | -0.001 | -0.003 | -0.012 | -0.014 |
| | (0.010) | (0.009) | (0.014) | (0.016) |
| Subjectivity (W, dict) | 0.002 | 0.003 | -0.001 | -0.000 |
| | (0.004) | (0.002) | (0.006) | (0.003) |
| Subjectivity (S, dict) | -0.011 | 0.054** | -0.051 | 0.039 |
| | (0.044) | (0.022) | (0.072) | (0.033) |
| Subjectivity (S, ML) | -0.007 | 0.038* | -0.025 | 0.031 |
| | (0.033) | (0.021) | (0.051) | (0.037) |
| Observations | 505 | 452 | 301 | 205 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the OLS analysis of the effect of financial variables, economic development and sentiment and subjectivity measures on cumulative abnormal returns, but only the coefficients for sentiment and subjectivity measures are reported.

The question remains, if market participants value additional information from the credit rating reports differently, depending on the credit rating agency providing the report. Results for the three agencies are reported in Tables 45, 46 and 47. Given that we have established that markets tend to trust Standard & Poor's opinion more over Fitch and Moody's, one would expect that sentiment and subjectivity measures for Standard & Poor's have a significant effect on returns more often than in the case of Fitch or Moody's. However, the estimates indicate that subjectivity measures extracted from Moody's rating action reports impact returns the most during a positive rating announcement. Specifically, the effect is statistically significant for both dictionary-based subjectivity measures in the case of a positive event, and dictionary-based subjectivity measure at word level in the case of an upgrade. The relationship between CDS returns and textual measures from Standard & Poor's credit action reports is a bit less pronounced, with statistically significant net sentiment during a negative event and subjectivity from the machine learning approach during a positive event. Fitch's credit rating reports appear to have no additional informational value for investors during positive rating announcements, but there is some evidence that polarity from the machine learning approach affects returns after a downgrade.

Overall, there is insufficient evidence that textual sentiment or subjectivity expressed in sovereign credit rating reports have meaningful effects on cumulative abnormal returns in the time interval (-1,1) during negative or positive rating announcements. This is in

*Table 45.* Relationship between cumulative abnormal returns in the time interval (-1,1) during rating announcements and sentiment and subjectivity measures for S&P

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Net sentiment (W, dict) | -0.007* | -0.001 | -0.003 | -0.006 |
| | (0.004) | (0.003) | (0.006) | (0.007) |
| Polarity (W, dict) | -0.047 | -0.014 | -0.003 | -0.042 |
| | (0.041) | (0.024) | (0.062) | (0.043) |
| Polarity (S, ML) | -0.018 | -0.018 | 0.004 | 0.008 |
| | (0.021) | (0.018) | (0.026) | (0.036) |
| Subjectivity (W, dict) | 0.001 | 0.003 | 0.009 | -0.007 |
| | (0.007) | (0.004) | (0.015) | (0.006) |
| Subjectivity (S, dict) | -0.026 | 0.024 | -0.082 | -0.062 |
| | (0.096) | (0.035) | (0.163) | (0.054) |
| Subjectivity (S, ML) | 0.071 | 0.070* | -0.024 | 0.078 |
| | (0.075) | (0.042) | (0.105) | (0.098) |
| Observations | 142 | 159 | 89 | 78 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the OLS analysis of the effect of financial variables, economic development and sentiment and subjectivity measures on cumulative abnormal returns, but only the coefficients for sentiment and subjectivity measures are reported.

*Table 46.* Relationship between cumulative abnormal returns in the time interval (-1,1) during rating announcements and sentiment and subjectivity measures for Fitch

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Net sentiment (W, dict) | 0.003 | -0.001 | -0.006 | 0.002 |
| | (0.003) | (0.002) | (0.006) | (0.003) |
| Polarity (W, dict) | 0.001 | -0.010 | -0.100 | 0.016 |
| | (0.045) | (0.021) | (0.086) | (0.024) |
| Polarity (S, ML) | -0.026 | 0.017 | -0.083** | 0.022 |
| | (0.035) | (0.015) | (0.041) | (0.026) |
| Subjectivity (W, dict) | -0.011 | 0.001 | -0.007 | 0.002 |
| | (0.010) | (0.004) | (0.011) | (0.004) |
| Subjectivity (S, dict) | -0.136 | 0.041 | 0.000 | 0.046 |
| | (0.102) | (0.031) | (0.159) | (0.042) |
| Subjectivity (S, ML) | -0.074 | -0.011 | 0.017 | 0.018 |
| | (0.059) | (0.033) | (0.098) | (0.037) |
| Observations | 109 | 116 | 64 | 54 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the OLS analysis of the effect of financial variables, economic development and sentiment and subjectivity measures on cumulative abnormal returns, but only the coefficients for sentiment and subjectivity measures are reported.

*Table 47.* Relationship between cumulative abnormal returns in the time interval (-1,1) during rating announcements and sentiment and subjectivity measures for Moody's

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Net sentiment (W, dict) | -0.004 | 0.003 | -0.005 | 0.004 |
| | (0.005) | (0.002) | (0.007) | (0.007) |
| Polarity (W, dict) | -0.007 | 0.022 | -0.034 | 0.023 |
| | (0.045) | (0.017) | (0.065) | (0.053) |
| Polarity (S, ML) | 0.001 | -0.006 | -0.008 | 0.003 |
| | (0.024) | (0.019) | (0.041) | (0.037) |
| Subjectivity (W, dict) | 0.010 | 0.009* | 0.012 | 0.014 |
| | (0.007) | (0.005) | (0.011) | (0.011) |
| Subjectivity (S, dict) | -0.073 | 0.131*** | -0.033 | 0.174* |
| | (0.078) | (0.048) | (0.143) | (0.091) |
| Subjectivity (S, ML) | -0.071 | 0.020 | -0.090 | -0.130 |
| | (0.061) | (0.043) | (0.103) | (0.104) |
| Observations | 123 | 119 | 67 | 50 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the OLS analysis of the effect of financial variables, economic development and sentiment and subjectivity measures on cumulative abnormal returns, but only the coefficients for sentiment and subjectivity measures are reported.

contradiction with Agarwal et al. (2019), who analyse Moody's credit action reports and find that a negative tone (sentiment) provides additional information beyond credit ratings alone. However, it is worth noting that they (i) treat sentiment separately as negative and positive tone, as opposed to net sentiment in our analysis, and (ii) base the conclusions on a pooled sample of both downgrades and upgrades.

## 6.4 Conclusion

The effects of sovereign credit rating changes on CDS and debt markets have been relatively thoroughly analysed by previous research (e.g. Drago & Gallo, 2016; Gande & Parsley, 2005; Ismailescu & Kazemi, 2010). However, like with the determinants of sovereign credit ratings, the effects of credit rating reports have been largely overlooked. The main objective of this chapter is thus to shed new light on the relationship between textual sentiment and subjectivity, extracted from sovereign credit rating reports, and the market response, namely CDS spread changes. Particularly, we focus on specific periods in the three days during a rating announcement by any of the three credit rating agencies, namely Standard & Poor's, Fitch and Moody's, that are normally accompanied by a credit rating action report. Specifically, we construct the so-called 'first mover' sample, where we remove observations with other rating announcements in the month preceding the event, but also look at credit rating agencies alone.

We first establish the event study framework, which is the basis for further analysis. The initial event study results indicate that the markets react to both positive and negative rating announcements when the rating announcement is not preceded by other rating announcements. This is also true for Standard & Poor's but not Fitch or Moody's, which suggests that markets put more emphasis on sovereign credit ratings assigned by Standard & Poor's as opposed to the remaining two credit rating agencies. We then extend the event study framework with additional explanatory variables for cumulative abnormal returns in the three-day window around rating announcements. We find no evidence that markets react differently to sovereign credit rating or outlook changes for advanced economies and emerging markets. Next, we include various financial variables, such as local market returns or US excess return. We observe different variables impacting the cumulative abnormal returns, depending on the event definition. The effect of 5-year constant maturity treasury rate is common to both negative announcements in general and downgrades in particular. The same is true for US excess return in the case of positive announcement and upgrades. We also examine the transitions between speculative and investment grade rating classes, which we expect to be economically more important. We find significant returns when ratings pass over from speculative to investment grade, but not vice versa.

Finally, we explore the possible connection between a market reaction and textual sentiment and subjectivity measures. Contrary to expectations, we conclude that the effect of textual sentiment and subjectivity measures is relatively limited. There is some indication that subjectivity measures positively affect cumulative abnormal returns after a positive credit rating announcement. Given that previous analysis indicates market participants place more weight to Standard & Poor's opinion compared to Fitch or Moody's, one would expect this to be reflected in the effect of textual sentiment and objectivity measures as well. However, the outcome implies that Moody's holds a slight edge over Standard & Poor's and Fitch in that department. Overall, we conclude that textual sentiment or subjectivity extracted from sovereign credit rating reports have no relevant effects on CDS market returns.

# Conclusion

Sovereign credit ratings, specifically the qualitative part of the ratings, have puzzled researchers for decades. Given that traditional approaches are relatively unsuccessful in determining the importance of the qualitative part, namely the credit rating committee's opinion, we adopt a different approach. We use sovereign credit rating reports as sources of additional information not taken into account by previous research that sheds new light on the qualitative judgement of the rating committee. The purpose of this dissertation is thus to bring attention to a relatively unexplored area of textual analysis, as it has been done in other fields of finance. We argue that textual sentiment and subjectivity extracted from the reports using a textual analysis approach helps in understanding sovereign credit ratings. In this chapter, we present the main findings of the doctoral dissertation, consisting of two parts. We first briefly summarise the main results, answering the three core research questions established in the Introduction. We then focus on outstanding issues and potential for future research.

## Main findings

*1. To what extent do textual sentiment and subjectivity help in explaining discrepancies between ratings of advanced economies and emerging markets, as well as before and after the global financial crisis?*

Overall, we find that textual sentiment significantly contributes to explaining sovereign credit ratings, which is especially true when soft information and bias proxies are not taken into account. Furthermore, evidence suggests that the qualitative judgement of the rating committee is captured by the subjectivity score for one of the three agencies.

There are meaningful discrepancies in textual sentiment between advanced economies and emerging markets for one of the three agencies, which we attribute to the difference in the general perception of two groups of countries. We also observe a significant difference in subjectivity measures for emerging markets compared to advanced economies, which indicates that the credit rating agencies attach different weights to the rating committee's qualitative judgement for the two groups of countries. This may be due to scarce and questionable data for emerging markets, leading to analysts having to rely more on qualitative factors.

We also analyse the behaviour of textual sentiment and subjectivity before and after the global financial crisis. We notice a change in sentiment after the crisis, which we ascribe to the general negative economic environment that lingered for several years.

However, we do not detect any difference in subjectivity, meaning that there has not been a disruption in the way the credit rating committee conveys its judgement due to the crisis.

Additionally, we notice that soft information, namely institutional strength and governance play an important role in explaining sovereign credit ratings. We also find evidence of economic proximity bias, suggesting that credit rating agencies assign higher ratings to countries that have strong trade ties with the US. By separately analysing advanced economies and emerging markets, we observe different determinants of sovereign credit ratings for each country group. The results also imply an upward cultural bias towards advanced economies and an upward economic bias towards emerging markets.

*2. How can we better model and predict rating transitions using additional textual information?*

The results show that, on average, sentiment measures perform better than subjectivity measures in predicting rating changes. The improvement is more notable for downgrades than upgrades. We register higher sensitivity scores for sentiment measures compared to subjectivity measures, on average. Dictionary-based methods appear to outperform machine learning algorithms for sentiment measures, while the reverse is true for subjectivity indicators. We examine model performance with two philosophical frameworks for assigning sovereign credit ratings in mind, namely the point-in-time and through-the-cycle approach. We confirm that credit rating agencies follow the through-the-cycle rating philosophy by taking a longer horizon into account. As a robustness check, we take rating outlook into account, which confirms our previous results.

*3. Do markets value additional information extracted from sovereign credit rating reports?*

We examine the relationship between textual sentiment and subjectivity measures and CDS markets. By focusing on the narrow window of three days surrounding the rating announcement, we conclude that the value market participants ascribe to sovereign credit rating reports is fairly limited. However, we do observe a positive effect of subjectivity measures on CDS spreads during a positive credit rating announcement. The results also indicate, that markets marginally value Moody's qualitative judgement expressed in the reports above the one expressed by the remaining two agencies.

Nevertheless, we detect a significant market reaction to both positive and negative rating announcements. The evidence also points to Standard & Poor's enjoying the highest reputation among the three agencies by the markets. We do not observe meaningful discrepancies in market reactions between advanced economies and emerging markets. When we extend the model with financial variables, we notice that different variables affect abnormal CDS returns among different event definitions. Finally, by examining the special role of the transition between investment and speculative grade rating classes, we register significant returns at the crossover from speculative to investment grade.

**Future research**

We believe this doctoral dissertation represents the most exhaustive and thorough examination of sovereign credit rating reports thus far. However, the results within the dissertation open the door to several potential future studies.

First, as already indicated in Chapter 5, the transition probabilities differ among specific rating classes, which we have not taken into account, as the focus is to identify the best performing textual sentiment and subjectivity measure. We thus do not estimate transition matrices, which is the natural continuation of our existing research. We note that this approach is problematic due to limited data availability for sovereign credit ratings, but a few existing studies manage to circumvent this issue Hill et al. (2010); Hu et al. (2002).

Second, due to unavailability of CDS spreads before December 2007, we are not able to examine the sovereign CDS market reactions to the informational value extracted from sovereign credit rating reports before the global financial crisis, and how the relationship evolved, if at all. The next step would thus be to acquire the missing CDS data and address this issue.

Third, we base our analysis on the traditional event study framework. We focus on the narrow three-day event window around the event as we concentrate on textual analysis measures, but generally, studies look at a wider array of event windows. The natural continuation of our research would thus be to extend the analysis by including additional time intervals. We hypothesise that textual sentiment and subjectivity indicators could potentially affect returns with a lag, given that investors may take some time to process the reports.

Regarding methodology, as Kearney and Liu (2014) point out, the textual analysis process can be improved. Although finance field dictionaries have been developed and their use is increasing, the construction and availability of more extensive field-specific and text type adjusted dictionaries are required for future studies. For instance, the LM dictionary, based on 10-K filings, may not be entirely appropriate for the analysis of credit rating reports. Furthermore, there is also room for improvements with term-weighting schemes, apart from equal weights applied in this dissertation. We barely scratch the surface of the machine learning algorithm assortment with the Naïve Bayes, as there are so many additional options, such as Support Vector Machines or neural networks. Applying an interdisciplinary approach by including linguists, psychologists, and computer scientists can result in developing other textual analysis approaches. All of this can lead to more accurate and efficient sentiment and subjectivity measures. Regarding the type of texts, there are sources that either have not been studied yet or have not been sufficiently studied, sovereign credit rating reports being just one of such examples.

The use of the textual analysis method in finance is still relatively new, and thus many approaches would be different in the context of what we know today. Some empirical results measuring sentiment in financial texts should be re-evaluated (Loughran & McDonald, 2016). Apart from that, in a digitalised world, where the available text databases are extensive, and computers are becoming more and more powerful, there is plenty to be done.

## References


Afonso, A. (2003). Understanding the determinants of sovereign debt ratings: Evidence for the two leading agencies. *Journal of Economics and Finance*, *27*(1), 56–74.

Afonso, A., Furceri, D., & Gomes, P. (2012). Sovereign credit ratings and financial markets linkages: Application to european data. *Journal of International Money and Finance*, *31*(3), 606 - 638.

Afonso, A., Gomes, P., & Rother, P. (2009). Ordered response models for sovereign debt ratings. *Applied Economics Letters*, *16*(8), 769–773.

Afonso, A., Gomes, P., & Rother, P. (2011). Short-and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, *16*(1), 1–15.

Afonso, A., & Jalles, J. T. (2019). Sovereign ratings and finance ministers' characteristics. *REM Working Paper 072-2019*.

Agarwal, S., Chen, V., Sim, G., & Zhang, W. (2019). An evaluation of moody's sovereign credit risk assessment. *Available at SSRN: https://dx.doi.org/10.2139/ssrn.2526648*.

Agarwal, S., Chen, V. Y., & Zhang, W. (2016). The information value of credit rating action reports: A textual analysis. *Management Science*, *62*(8), 2218–2240.

Agresti, A., & Natarajan, R. (2001). Modeling clustered ordered categorical data: A survey. *International Statistical Review*, *69*(3), 345–371.

Alsakka, R., & ap Gwilym, O. (2010). Leads and lags in sovereign credit ratings. *Journal of Banking & Finance*, *34*(11), 2614-2626.

Alsakka, R., & ap Gwilym, O. (2013). Rating agencies' signals during the european sovereign debt crisis: Market impact and spillovers. *Journal of Economic Behavior & Organization*, *85*, 144–162.

Amstad, M., & Packer, F. (2015). Sovereign ratings of advanced and emerging economies after the crisis. *BIS Quarterly Review December 2015*, 77–91.

Ang, A., & Longstaff, F. A. (2013). Systemic sovereign credit risk: Lessons from the u.s. and europe. *Journal of Monetary Economics*, *60*(5), 493 - 510.

Arezki, R., Candelon, B., & Sy, A. N. (2011). Sovereign rating news and financial markets spillovers: Evidence from the european debt crisis. *IMF Working Paper No. 11/68*.

Basel Committee on Banking Supervision. (2005). Studies on the validation of internal rating systems. *BIS Working Paper No. 14*.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit* (2nd ed.). Retrieved from: http://www.nltk.org/book/.

Bissoondoyal-Bheenick, E. (2005). An analysis of the determinants of sovereign ratings. *Global Finance Journal*, *15*(3), 251–280.

Blommestein, H., Eijffinger, S., & Qian, Z. (2016). Regime-dependent determinants of euro area sovereign cds spreads. *Journal of Financial Stability*, *22*, 10 - 21.

Boot, A. W., Milbourn, T. T., & Schmeits, A. (2006). Credit ratings as coordination mechanisms. *The Review of Financial Studies*, *19*(1), 81-118.

Brooks, R., Faff, R. W., Hillier, D., & Hillier, J. (2004). The national market impact of sovereign rating changes. *Journal of banking & finance*, *28*(1), 233–250.

Bruner, C. M., & Abdelal, R. (2005). To judge leviathan: sovereign credit ratings, national law, and the world economy. *Journal of Public Policy*, *25*(2), 191–217.

Butler, A. W., & Fauver, L. (2006). Institutional environment and sovereign credit ratings. *Financial Management*, *35*(3), 53–79.

Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic policy review*, *2*(2), 37–54.

Chen, Z., Matousek, R., Stewart, C., & Webb, R. (2019). Do rating agencies exhibit herding behaviour? evidence from sovereign ratings. *International Review of Financial Analysis*, *64*, 57 - 70.

Covitz, D. M., & Harrison, P. (2003). Testing conflicts of interest at bond rating agencies with market anticipation: Evidence that reputation incentives dominate. *Federal Reserve Board, Finance and Economics Discussion Series, No. 2003-68*.

Das, S. R. (2014). Text and context: Language analytics in finance. *Foundations and Trends® in Finance*, *8*(3), 145–261.

de Haan, J., & Amtenbrink, F. (2011). Credit rating agencies. *De Nederlandsche Bank Working Paper No. 278*.

De Moor, L., Luitel, P., Sercu, P., & Vanpée, R. (2018). Subjectivity in sovereign credit ratings. *Journal of Banking & Finance*, *88*, 366–392.

Dieckmann, S., & Plank, T. (2011, 07). Default Risk of Advanced Economies: An Empirical Analysis of Credit Default Swaps during the Financial Crisis*. *Review of Finance*, *16*(4), 903-934.

Drago, D., & Gallo, R. (2016). The impact and the spillover effect of a sovereign rating announcement on the euro area cds market. *Journal of International Money and Finance*, *67*, 264 - 286.

Eijffinger, S. C. (2012). Rating agencies: Role and influence of their sovereign credit risk assessment in the eurozone. *JCMS: Journal of Common Market Studies*, *50*(6), 912-921.

Erdem, O., & Varli, Y. (2014). Understanding the sovereign credit ratings of emerging markets. *Emerging Markets Review*, *20*, 42–57.

Fender, I., Hayo, B., & Neuenkirch, M. (2012). Daily pricing of emerging market sovereign cds before and during the global financial crisis. *Journal of Banking & Finance*, *36*(10), 2786 - 2794.

Ferri, G., Liu, L.-G., & Stiglitz, J. E. (1999). The procyclical role of rating agencies: Evidence from the east asian crisis. *Economic Notes*, *28*(3), 335-355.

Finnerty, J. D., Miller, C. D., & Chen, R.-R. (2013). The impact of credit rating announcements on credit default swap spreads. *Journal of Banking & Finance*, *37*(6), 2011 - 2030.

Fitch. (2017). *Sovereign rating criteria* (Tech. Rep.). Fitch Ratings.

Fontana, A., & Scheicher, M. (2016). An analysis of euro area sovereign cds and their relation with government bonds. *Journal of Banking & Finance*, *62*, 126 - 140.

Fuchs, A., & Gehring, K. (2017). The home bias in sovereign ratings. *Journal of the European Economic Association*, *15*(6), 1386–1423.

Fuertes, A.-M., & Kalotychou, E. (2007). On sovereign credit migration: A study of alternative estimators and rating dynamics. *Computational Statistics & Data Analysis*, *51*(7), 3448 - 3469.

Gaillard, N. (2012). *A century of sovereign ratings.* New York, NY: Springer.

Gande, A., & Parsley, D. C. (2005). News spillovers in the sovereign debt market. *Journal of Financial Economics*, *75*(3), 691–734.

Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *The Journal of Business*, *53*(1), 67-78.

Gültekin-Karakaş, D., Hisarcıklılar, M., & Öztürk, H. (2011). Sovereign risk ratings: biased toward developed countries? *Emerging Markets Finance and Trade*, *47*(2), 69–87.

Güttler, A., & Wahrenburg, M. (2007). The adjustment of credit ratings in advance of defaults. *Journal of Banking & Finance*, *31*(3), 751–767.

Haque, N. U., Mark, N. C., & Mathieson, D. J. (1998). The relative importance of political and economic variables in creditworthiness ratings. *IMF Working Paper No. 98/46*.

Hill, P., Brooks, R., & Faff, R. (2010). Variations in sovereign credit quality assessments across rating agencies. *Journal of Banking & Finance*, *34*(6), 1327–1343.

Hu, Y.-T., Kiesel, R., & Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings. *Journal of Banking & Finance*, *26*(7), 1383–1406.

Hull, J., Predescu, M., & White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking & Finance*, *28*(11), 2789 - 2811.

Ismailescu, I., & Kazemi, H. (2010). The reaction of emerging market credit default swap spreads to sovereign credit rating changes. *Journal of Banking & Finance*, *34*(12), 2861 - 2873. (International Financial Integration)

Kaminsky, G., & Schmukler, S. (2002). Emerging market instability: Do sovereign ratings affect country risk and stock returns? *World Bank Policy Research Papers 2678*.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185.

Kiesel, F. (2016). Capital market reactions and the content of credit rating reports: A textual analysis. *Retrieved from https://dx.doi.org/10.2139/ssrn.2781991*.

Kiff, J., Holland, A., Kisser, M., Nowak, S., Saab, S., Schumacher, L., ... Westin, A. (2010). The uses and abuses of sovereign credit ratings. In *Global financial stability report, october 2010: Sovereigns, funding, and systemic liquidity* (Vol. 10, pp. 85–122). Washington D.C., VA: International Monetary Fund.

Kiff, J., Kisser, M., & Schumacher, L. (2013). Rating through-the-cycle: what does the concept imply for rating stability and accuracy? *IMF Working Paper No. 13/64*.

Kiff, J., Novak, S. B., & Schumacher, L. (2012). Are rating agencies powerful? an investigation into the impact and accuracy of sovereign ratings. *IMF Working Paper No. 12/23*.

Klusak, P., Thornton, J., & Uymaz, Y. (2019). Do personal connections improve sovereign credit ratings? *Finance Research Letters*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1544612319302661

Laeven, M. L., & Valencia, M. F. (2018). Systemic banking crises revisited. *IMF Working Paper No. 18/206*.

Li, F. (2010). The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102.

Liu, S. (2014). The impact of textual sentiment on sovereign bond yield spreads: Evidence from the eurozone crisis. *Multinational Finance Journal*, *18*(3/4), 215–248.

Löffler, G., Norden, L., & Rieber, A. (2018). Salient news and the stock market impact of tone in rating reports. *Available at SSRN: https://dx.doi.org/10.2139/ssrn.2802570*.

Longstaff, F. A., Pan, J., Pedersen, L. H., & Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, *3*(2), 75–103.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, *66*(1), 35–65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Luitel, P., Vanpée, R., & De Moor, L. (2016). Pernicious effects: How the credit rating agencies disadvantage emerging markets. *Research in International Business and Finance*, *38*, 286–298.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* Cambridge, MA: MIT Press.

Mellios, C., & Paget-Blanc, E. (2006). Which factors determine sovereign credit ratings? *The European Journal of Finance*, *12*(4), 361–377.

Michaelides, A., Milidonis, A., & Nishiotis, G. P. (2019). Private information in currency markets. *Journal of Financial Economics*, *131*(3), 643 - 665.

Micu, M., Remolona, E. M., & Wooldridge, P. D. (2006). The Price Impact of Rating Announcements: Which Announcements Matter? *BIS Working paper No. 207*.

Moody's. (2016). *Sovereign bond ratings* (Tech. Rep.). Moody's Investor Service.

Mora, N. (2006). Sovereign credit ratings: Guilty beyond reasonable doubt? *Journal of Banking & Finance*, *30*(7), 2041–2062.

Özturk, H. (2014). The origin of bias in sovereign credit ratings: Reconciling agency views with institutional quality. *The Journal of Developing Areas*, *48*(4), 161–188.

Pagano, M., & Volpin, P. (2010). Credit ratings failures and policy options. *Economic Policy*, *25*(62), 401–431.

Pan, J., & Singleton, K. J. (2008). Default and recovery implicit in the term structure of sovereign cds spreads. *The Journal of Finance*, *63*(5), 2345-2384.

Purda, L. D. (2007). Stock market reaction to anticipated versus surprise rating changes. *Journal of Financial Research*, *30*(2), 301–320.

Reusens, P., & Croux, C. (2017). Sovereign credit rating determinants: A comparison before and after the european debt crisis. *Journal of Banking & Finance*, *77*, 108–121.

Riedl, M., & Geishecker, I. (2014). Keep it simple: estimation strategies for ordered response models with fixed effects. *Journal of Applied Statistics*, *41*(11), 2358-2374.

Rodríguez, I. M., Dandapani, K., & Lawrence, E. R. (2019). Measuring Sovereign Risk: Are CDS Spreads Better than Sovereign Credit Ratings? *Financial Management*, *48*(1), 229-256.

Sangiorgi, F., & Spatt, C. (2017). The economics of credit rating agencies. *Foundations and Trends® in Finance*, *12*(1), 1–116.

Sarkar, D. (2016). *Text analytics with python - a practical real-world approach to gaining actionable insights from your data.* New York, NY: Springer.

Standard & Poor's. (2017). *Sovereign rating methodology* (Tech. Rep.). S&P Global

Ratings.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *The general inquirer: A computer approach to content analysis. 1966*. Cambridge, MA: MIT Press.

Sy, A. N. (2002). Emerging market bond spreads and sovereign credit ratings: reconciling market views with economic fundamentals. *Emerging Markets Review*, *3*(4), 380–408.

Vernazza, D. R., & Nielsen, E. F. (2015). The damaging bias of sovereign ratings. *Economic Notes: Review of Banking, Finance and Monetary Economics*, *44*(2), 361–408.

White, L. J. (2010). Markets: The credit rating agencies. *Journal of Economic Perspectives*, *24*(2), 211–26.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Zheng, L. (2012). Are sovereign credit ratings objective? a tale of two agencies. *Journal of Applied Finance and Banking*, *2*(5), 43-61.

# APPENDICES

# Appendix A: Daljši povzetek v slovenskem jeziku

Bonitetne ocene so bistvenega pomena za državo pri dostopu do mednarodnih finančnih trgov, saj odražajo kreditno tveganje države in tako vplivajo na njene stroške financiranja. Poleg tega posredno vplivajo tudi na stroške financiranja rezidentov, tako fizičnih kot pravnih oseb. Pretežno jih dodeljujejo tri glavne bonitetne agencije, in sicer Standard & Poor's, Fitch in Moody's, ki pri tem upoštevajo tako kvantitativne kot tudi kvalitativne dejavnike. Agencije še posebej poudarjajo, da so njihove bonitetne ocene držav zgolj mnenje (Fitch, 2017; Moody's, 2016; Standard & Poor's, 2017).

Obstoječe študije, ki preučujejo bonitetne ocene držav, lahko pretežno ločimo v tri skupine. Prva poskuša identificirati dejavnike bonitetnih ocen držav (npr. Afonso, 2003; Afonso et al., 2009; Cantor & Packer, 1996; Özturk, 2014). Trdijo, da so ocene sestavljene iz trdih in mehkih informacij. Trdi del ocene večinoma temelji na makroekonomskih podatkih, npr. raven BDP na prebivalca, realna rast BDP, zunanji dolg, javni dolg in javnofinančni saldo. Mehki del ocene je težje izmeriti, vendar razni indikatorji za politično tveganje in institucionalno kakovost učinkovito zajamejo prevladujoče vzdušje v določeni državi. Pomemben del teh študij trdi, da bonitetne agencije razvijajočim se trgom neupravičeno dodeljujejo nižje bonitetne ocene glede na njim primerljiva razvita gospodarstva, kar posledično vodi do pristranskih bonitetnih ocen (De Moor et al., 2018; Fuchs & Gehring, 2017; Zheng, 2012). Druga skupina preučuje prehodne matrike bonitetnih ocen in poskuša izboljšati njihovo predvidljivost (npr. Hill et al., 2010; Hu et al., 2002). Tretja skupina se osredotoča na vpliv bonitetnih ocen držav na gibanje donosnosti državnih obveznic in CDS razmikov (angl. credit default swap, CDS spreads) (Afonso et al., 2012; Drago & Gallo, 2016; Gande & Parsley, 2005; Ismailescu & Kazemi, 2010; Kiff et al., 2012). Posledično je ta disertacija sestavljena iz treh glavnih delov, ki ustrezajo vsaki skupini, in sicer iz poglavij 4, 5 in 6.

Prvič, medtem ko so predhodne raziskave poskušale identificirati dejavnike bonitetnih ocen držav z uporabo različnih klasičnih tehnik ocenjevanja, je velik del bonitetnih ocen še vedno ostal nepojasnjen, odstotki pravilno napovedanih bonitetnih ocen držav pa so bili relativno nizki. Menimo, da zadnji (manjkajoči) del bonitetne ocene predstavlja kvalitativno presojo bonitetne komisije, ki je izražena v bonitetnih poročilih držav. Zato predlagamo alternativni pristop. Zaradi pomanjkanja predhodnih dokazov o informacijski vrednosti poročil o bonitetnih ocenah držav je cilj izkoristiti ta premalo izkoriščen vir kvalitativnih podatkov, da bi bolje razumeli različne vidike oblikovanja bonitetnih ocen. Analiziramo poročila bonitetnih agencij Standard & Poor's, Fitch in Moody's, ter z uporabo metod besedilne analize raziskujemo, v kolikšni meri se v ocenah odražajo različne mere besedilnega sentimenta in subjektivnosti. Ob upoštevanju prej omenjenih študij o pristranskosti bonitetnih ocen držav se osredotočamo na primerjavo dveh skupin držav: razvijajočih se trgov in razvitih gospodarstev. Predvidevamo, da bodo imeli razvijajoči se trgi višje mere sentimenta in subjektivnosti kot razvita gospodarstva, ker so podatki omejeni in potencialno nezanesljivi, kar vodi k večji vlogi kvalitativne presoje bonitetne komisije. Poleg tega domnevamo, da se bodo mere sentimenta in subjektivnosti verjetno spremenile po svetovni finančni krizi leta 2008 zaradi povečanih zahtev po transparentnosti bonitetnega procesa, pa tudi zaradi nedavnih kritik bonitetnih agencij, da so napihnile bonitetne ocene določenih držav (Agarwal et al., 2019; Gaillard, 2012), kar vodi k bolj realistični oceni državnega tveganja.

Drugič, spremembe bonitetnih ocen držav imajo pomembne ekonomske posledice, saj vplivajo na stroške financiranja (Alsakka & ap Gwilym, 2013; Eijffinger, 2012). Še več težav nastane zaradi čezmejnega prelivanja oz. učinkov na druge trge (Alsakka & ap Gwilym, 2013). Zato je napovedovanje zvišanj ali znižanj bonitetnih ocen pomembno za državo, pa tudi za njene ekonomske in finančne partnerje. To je še posebej pomembno za razvijajoče se trge, ki imajo na splošno razmeroma nizke bonitetne ocene in zato višje stroške zadolževanja. Tako posvečamo več pozornosti spremembam bonitetnih ocen držav oz. specifično zvišanjem in znižanjem bonitetnih ocen. Kot smo že ugotovili, so tradicionalni pristopi k analizi bonitetnih ocen držav omejeni, saj ne zaznajo kvalitativne presoje bonitetne komisije. Ker je sprememba bonitetne ocene razložena v poročilu o bonitetni oceni države, domnevamo, da bodo mere sentimenta in subjektivnosti lahko zajele prevladujoče mnenje in bodo tako pripomogle k večji predvidljivosti sprememb v bonitetnih ocenah.

Tretjič, poleg sprememb bonitetnih ocen držav bonitetne agencije izdajajo tudi bonitetne obete oz. napovedi (angl. outlooks, watches), ki prav tako lahko vplivajo na donosnost obveznic in CDS razmike. Številne študije ugotavljajo, da se trgi odzivajo na vse vrste objav bonitetnih agencij, zlasti pa kadar bonitetne ocene prečkajo mejo med naložbenimi in špekulativnimi bonitetnimi razredi (Drago & Gallo, 2016; Ismailescu & Kazemi, 2010; Kiff et al., 2012). V disertaciji poskušamo ugotoviti, ali se trgi odzivajo tudi na poročila o bonitetnih ocenah držav, ki vsebujejo dodatne kvalitativne informacije. To je področje, ki je bilo do sedaj slabo raziskano.

Glavni cilj disertacije je ponuditi izčrpno analizo poročil o bonitetnih ocenah držav glede na besedilni sentiment in subjektivnost, ter njihovo vlogo pri določanju bonitetnih ocen in sprememb bonitetnih ocen držav, pa tudi povezav z dolžniškimi trgi.

## Raziskovalna vprašanja

Naš cilj je obravnavati naslednji raziskovalni problem: *V kolikšni meri sentiment in subjektivnost vplivata na bonitetne ocene držav in trge po državah in skozi čas?*

Da bi lažje naslovili ta problem, se bomo osredotočili na tri glavna raziskovalna vprašanja, ki so v osnovi vsa povezana s sentimentom in mnenjem (subjektivnostjo), pridobljenim iz poročil o bonitetnih ocenah držav:

- *V kolikšni meri besedilni sentiment in subjektivnost pojasnjujeta razlike med implicitnimi in dejanskimi bonitetnimi ocenami razvitih gospodarstev in razvijajočih se trgov na splošno, pa tudi pred in po svetovni finančni krizi?*
- *Kako lahko z dodatnimi besedilnimi informacijami bolje modeliramo in napovedujemo prehode med ocenami?* in
- *Ali trgi cenijo dodatne informacije, pridobljene iz poročil o bonitetnih ocenah držav?*

## Znanstveni prispevek

Menimo, da ta disertacija pomembno prispeva k znanosti, tako z metodološkega kot praktičnega vidika. Prvič, kolikor nam je znano, izvajamo doslej najobsežnejšo analizo poročil o bonitetnih ocenah držav z uporabo novih in naprednih pristopov, tj. besedilne analize sentimenta. Čeprav se te metode sorazmerno pogosto uporabljajo v poslovnih financah, podobnih študij v mednarodnih financah, zlasti v zvezi z bonitetnimi ocenami držav, praktično ni. Tako zapolnjujemo pomembno praznino v literaturi. Poleg mer besedilnega sentimenta, ki so že prisotne v literaturi besedilnih analiz, uvajamo tudi mere besedilne subjektivnosti. Kolikor vemo, smo prvi, ki uporabljamo to definicijo. Menimo, da je to edinstven pristop k zaznavanju mnenja v bonitetnih ocenah držav.

Drugič, menimo, da ta disertacija pomembno prispeva k razumevanju bonitetnih ocen držav. To ima praktične posledice za vse udeležence na trgu, od zasebnih in institucionalnih vlagateljev do posojilojemalcev. Bonitetne ocene držav in spremembe leteh lahko bistveno vplivajo na donose in stroške financiranja, kar še posebej velja za nižje bonitetne razrede. Natančneje, poglobljeno razumevanje bonitetnih ocen držav in načina oblikovanja je pomembno za (i) vlagatelje, katerim omogoča zmanjšanje asimetrije informacij, (ii) države dolžnice, katerim je omogočeno izvajanje reform in ukrepov za zagotovitev ugodnejše bonitetne ocene, s čimer si zmanjšajo stroške financiranja in (iii) finančne in druge institucije, ki držijo državne obveznice, in imajo običajno stroge omejitve, odvisne od bonitetnih ocen držav. Z uporabo mer besedilnega sentimenta in subjektivnosti lahko izboljšamo predvidljivost sprememb bonitetnih ocen držav, kar izdajateljem omogoča, da prilagodijo časovnico financiranja dolga in zmanjšajo stroške, hkrati pa vlagateljem omogoča, da optimizirajo svoje portfelje in iščejo višje donose.

### Podatki

Glavni koncepti, obravnavani v tej disertaciji, temeljijo na uporabi pristopa besedilne analize. Natančneje, uporabljamo metode z namenom pridobivanja besedilnega sentimenta iz poročil o bonitetnih ocenah držav in tako oblikujemo tri mere besedilnega sentimenta in tri mere subjektivnosti, ki so podlaga za obravnavo vseh treh raziskovalnih vprašanj. Na poročilih Standard & Poor's, Fitch in Moody's uporabljamo tako metode, ki temeljijo na slovarju, kot tudi pristope strojnega učenja.

Bonitetne ocene držav pridobimo na portalu Thomson Reuters Eikon, pri čemer se osredotočamo na dolgoročne bonitetne ocene držav v tuji valuti, ki jih Standard & Poor's, Fitch in Moody's dodeljujejo številnim državam, tako razvitim gospodarstvom kot razvijajočim se trgom. Države so razvrščene tako v naložbene kot špekulativne bonitetne razrede.

Poleg tega sestavimo izčrpen nabor podatkov s spremenljivkami makroekonomske in fiskalne narave (IMF World Economic Outlook Database, World Bank), spremenljivkami institucionalne moči in političnega tveganja (International Country Risk Guide) ter spremenljivkami ekonomske in kulturne bližine (OECD, CEPII, World Religion Data) na letni ravni za prvi del disertacije (poglavji 4 in 5). Za zadnji del disertacije (poglavje 6) pridobimo tudi dnevne CDS razmike in druge visokofrekvenčne finančne spremenljivke (Thomson Reuters Datastream, Federal Reserve Bank of St. Louis).

### Analiza sentimenta

Kot že rečeno, uporabljamo metodologijo besedilne analize oz. obdelave naravnega jezika (angl. natural language processing) za analizo poročil o bonitetnih ocenah držav. Predstavljamo šest ključnih mer sentimenta in subjektivnosti kot potencialne cenilke kvalitativne presoje bonitetne komisije, ki jih uporabljamo ali se nanje sklicujemo v celotni disertaciji.

Kearney and Liu (2014) opredelita sentiment ali ton kot stopnjo pozitivnosti ali negativnosti v besedilih. Trdita, da lahko sentiment vključuje tako subjektivno presojo kot objektivni odraz gospodarskih razmer. Sprememba bonitetne ocene je običajno pojasnjena v izčrpnem (besedilnem) poročilu. Z metodami besedilne analize lahko analiziramo poročila in raziščemo, v kolikšni meri so različne mere sentimenta povezane z ocenami.

Zbirali smo poročila različnih dolžin o bonitetnih ocenah agencije Standard & Poor's, ki so na voljo med letoma 2002 in 2018, poročila o bonitetnih ocenah agencije Fitch, ki so na voljo med letoma 1999 in 2018, ter poročila o bonitetnih ocenah agencije Moody's, ki so na voljo med letoma 1995 in 2018. Ta oblikujejo korpus za različne tehnike besedilne analize, vključno z analizo sentimenta.

Uporabljamo slovarski pristop, ki temelji na finančnem slovarju LM avtorjev Loughran and McDonald (2011). Sprva je večina raziskovalcev uporabljala dobro uveljavljene slovarje, kot sta General Inquirer (GI) ali DICTION. Kearney and Liu (2014) poudarja, da gre za splošne jezikovne slovarje angleškega jezika in ne za slovarje, ki so specifično ustvarjeni za področje financ. Loughran and McDonald (2011) ugotovita, da skoraj tri četrtine negativnih besed v GI/DICTION v finančnem kontekstu običajno ni negativnih. Ugotavljata, da uporaba slovarjev, razvitih izven finančnega področja, lahko povzroči napake, ki niso zgolj beli šum. Posledično so raziskovalci oblikovali finančne slovarje, na primer slovar LM, kar je privedlo do natančnejših in učinkovitejših ocen sentimenta. Poleg tega, kot večina študij, uporabljamo sorazmerno ponderiranje besed, pri čemer velja, da je vsaka beseda enako pomembna.

Naša mera je razmerje (odstotek) besed v določeni kategoriji sentimenta in skupnega števila besed v besedilu. Predpostavljamo naslednje: (i) če je v koledarskem letu objavljenih več poročil, vzamemo sentiment iz zadnjega poročila v tem letu (podobno npr. naredimo tudi pri več kot eni bonitetni oceni na leto); in (ii) če v koledarskem letu ni bilo objavljeno nobeno poročilo, domnevamo, da ni prišlo do spremembe v prevladujočem sentimentu/percepciji in upoštevamo vrednost iz prejšnjega leta.

S pomočjo slovarskega pristopa ustvarimo dve različni meri sentimenta. (Neto) sentiment je razlika med pozitivnim in negativnim sentimentom, pri čemer se negativen/pozitiven sentiment izračuna kot razmerje med številom negativnih/pozitivnih besed v besedilu in skupnim številom besed. To je najpogostejša mera v študijah, ki uporabljajo slovarski pristop, s splošnimi ali prilagojenimi slovarji in sorazmernim ponderiranjem (Kearney & Liu, 2014). Nato določimo še eno relativno mero v nasprotju z absolutno mero (tj. surovi odstotek), in sicer polarnost (angl. polarity), kot:

$$Polarity_{i,t} = \frac{pos_{i,t} - neg_{i,t}}{pos_{i,t} + neg_{i,t}} \tag{18}$$

kjer je $pos_{i,t}$ število pozitivnih besed in $neg_{i,t}$ število negativnih besed v besedilu.

Slovar LM vključuje tudi kategorije za 'negotovost' (izrazi, ki izražajo nenatančnost in se izključno ne osredotočajo na tveganje), 'močne modalne' in 'šibke modalne' besede (izrazi, ki izražajo stopnjo prepričanja). Zato kot alternativo polarnosti (pozitiven/negativen sentiment) uvedemo tudi mero subjektivnosti. Najprej določimo novo, širšo kategorijo za 'subjektivnost', ki jo sestavljajo tri prej omenjene kategorije. Nato ponovimo zgoraj opisani postopek z novo zgrajenim seznamom besed in uporabimo enake predpostavke. Dobimo mero subjektivnosti, izračunano kot razmerje med številom subjektivnih besed v besedilu in skupnim številom besed. Da bi zagotovili primerljivost s pristopom strojnega učenja, izdelamo tudi oceno subjektivnosti na ravni stavkov, izračunano kot razmerje med številom subjektivnih stavkov v besedilu in skupnim številom stavkov. Subjektivni stavki so opredeljeni kot stavki, ki vsebujejo vsaj eno besedo iz kategorije 'subjektivnost'. Subjektivni stavki se običajno nanašajo na osebno mnenje, čustva ali presojo, medtem ko se objektivni nanašajo na dejanske informacije. Motivacija izhaja iz dejstva, da kvalitativna presoja igra pomembno vlogo pri dodeljevanju bonitetnih ocen držav in bi jo bilo morda učinkoviteje zaznati z analizo subjektivnosti kot pa s preprosto negativno/pozitivno dihotomijo. Kvalitativno presojo bonitetne komisije opredeljujemo kot subjektivno razlago mehkih informacij, ki je ni mogoče izmeriti in je podprta z več kazalniki, lahko pa vključuje tudi potencialno pristranskost. Cantor and Packer (1996) navajata, da se analitiki pri ocenjevanju političnega in ekonomskega stanja države lahko soočajo z več ovirami, kar, kot poudarja Luitel et al. (2016), še posebej velja za razvijajoče se trge, kjer so podatki običajno omejeni in/ali vprašljive kakovosti. Zaradi tega se morajo analitiki pri takih državah bolj zanašati na svojo kakovostno presojo kot pri razvitih gospodarstvih. Večja uporaba kvalitativne presoje bonitetne komisije se tako lahko odraža v višji meri subjektivnosti in obratno.

Z uporabo strojnega učenja dobimo zadnji dve meri sentimenta in subjektivnosti. Za sentiment definiramo pozitivno, negativno in nevtralno kategorijo, za subjektivnost pa objektivno in subjektivno kategorijo. Za razvrščanje besedila uporabljamo naïvni Bayesov algoritem. Ko so razvrščeni vsi stavki v celotnem korpusu, z osnovnimi klasifikacijami ali njihovimi kombinacijami sestavimo mere sentimenta in subjektivnosti. Prva je polarnost, kot je opredeljena zgoraj, kjer je $pos_{i,t}$ število pozitivnih stavkov in $neg_{i,t}$ število negativnih stavkov v besedilu. Druga je subjektivnost, merjena kot razmerje med številom subjektivnih stavkov v besedilu in skupnim številom stavkov. Na koncu se te mere skupaj z drugimi spremenljivkami uporabljajo za nadaljnjo analizo. Sentiment/polarnost sta v povprečju negativno povezana z merili subjektivnosti.

Slovarski pristop in pristop strojnega učenja imata nekaj prednosti in slabosti. Loughran and McDonald (2016) naštevata več pomembnih prednosti slovarskega pristopa. Z izbiro slovarja se izognemo subjektivnosti raziskovalcev. Običajno lahko ustvarimo velike vzorce, saj računalniški programi tabelirajo pogostost besed. Glede na to, da je večina slovarjev javno dostopnih, je poustvarjanje drugih študij bolj preprosto. Kot trdita Kearney and Liu (2014), je slovarski pristop verjetno najbolj enostaven za ekonomiste in finančnike. Kot je bilo omenjeno zgoraj, pa bi morali raziskovalci uporabljati slovarje, specifične za področje financ. S tem je glavno vprašanje potem izbira primerne utežne sheme. Poleg tega bo pristop, ki temelji na slovarju, v povprečju manj zamuden in cenejši od pristopa strojnega učenja, saj je treba besedilo v 'naboru za učenje' (angl. training set) ročno kategorizirati. Kot trdi Li (2010), je zelo verjetno, da slovar za določeno vrsto besedila sploh ne obstaja, kot na primer za poročila o bonitetnih ocenah. Tudi če tak slovar obstaja, pristop, ki temelji na slovarju, ne upošteva konteksta stavka ali besedila. Poleg tega je stopnja natančnosti strojnega učenja običajno višja od slovarskega pristopa. Loughran and McDonald (2016) se osredotočata na naïvni Bayes, vendar je njihove argumente mogoče posplošiti. Ker stroji obdelujejo besedilo, je v analizo mogoče vključiti veliko število besedil. Ko so pravila za razvrščanje določena, merjenje sentimenta ne bo izpostavljeno nobeni dodatni subjektivnosti raziskovalca. Vendar menita, da je zmanjšana preglednost pristopa šibka točka, ker bodo drugi težko ponovili rezultate.

### Glavne ugotovitve

*1. V kolikšni meri besedilni sentiment in subjektivnost pojasnjujeta razlike med implicitnimi in dejanskimi bonitetnimi ocenami razvitih gospodarstev in razvijajočih se trgov na splošno, pa tudi pred in po svetovni finančni krizi?*

Na splošno ugotavljamo, da besedilni sentiment bistveno prispeva k razlagi bonitetnih ocen držav, kar še posebej drži kadar ne upoštevamo kazalnikov mehkih informacij in približkov za pristranskost. Poleg tega dokazi kažejo, da mere subjektivnosti zajemajo kakovostno presojo bonitetne komisije ene izmed agencij.

Med razvitimi gospodarstvi in razvijajočimi se trgi obstajajo pomembne razlike v besedilnem sentimentu ene izmed agencij, kar pripisujemo razliki v splošnem zaznavanju teh skupin držav. Opažamo tudi znatno razliko v merah subjektivnosti za razvijajoče se trge v primerjavi z razvitimi gospodarstvi, kar kaže na to, da bonitetne agencije kvalitativni presoji bonitetne komisije za obe skupini držav pripisujejo različne uteži. Razlog za to so lahko manjkajoči in vprašljivi podatki za razvijajoče se trge, zaradi česar se morajo analitiki bolj zanašati na kvalitativne dejavnike.

Analiziramo vedenje besedilnega sentimenta in subjektivnosti pred in po svetovni finančni krizi. Po krizi opažamo spremembo v sentimentu, ki jo pripisujemo splošnemu negativnemu gospodarskemu okolju, ki je trajalo več let. Vendar ne zaznavamo nobene razlike v subjektivnosti, kar pomeni, da kriza ni spremenila načina oblikovanja kvalitativne presoje bonitetne komisije.

Poleg tega opažamo, da imajo mehke informacije, in sicer moč institucij (angl. institutional strength) in upravljanje (angl. governance), pomembno vlogo pri razlagi bonitetnih ocen držav. Najdemo tudi dokaze o pristranskosti glede na gospodarsko bližino, kar kaže na to, da bonitetne agencije državam, ki imajo močne trgovinske vezi z ZDA, dodeljujejo višje ocene. Pri ločeni analizi razvitih gospodarstev in razvijajočih se trgov opažamo različne dejavnike bonitetnih ocen za posamezno skupino držav. Rezultati kažejo tudi na pozitivno kulturno pristranskost do razvitih gospodarstev in pozitivno ekonomsko pristranskost do razvijajočih se trgov.

*2. Kako lahko z dodatnimi besedilnimi informacijami bolje modeliramo in napovedujemo prehode med ocenami?*

Rezultati kažejo, da se pri napovedovanju sprememb bonitetnih ocen v povprečju mere sentimenta odrežejo bolje kot mere subjektivnosti. Izboljšanje je opaznejše pri znižanju kot pri zvišanju. V povprečju zabeležimo višjo natančnost napovedi pri merah sentimenta v primerjavi z merami subjektivnosti. Zdi se, da slovarske metode prekašajo algoritme strojnega učenja za mere sentimenta, medtem ko velja obratno za mere subjektivnosti. Uspešnost modela preučujemo v dveh filozofskih okvirih za določanje bonitetnih ocen držav, in sicer pristop točkovne ocene in skozi cikel. Potrjujemo, da bonitetne agencije zasledujejo filozofijo ocenjevanja skozi celoten cikel z upoštevanjem daljšega obdobja. Pri preverjanju zanesljivosti modela upoštevamo tudi obete za prihodnjo bonitetno oceno (angl. outlook), kar potrjuje naše prejšnje rezultate.

*3. Ali trgi cenijo dodatne informacije, pridobljene iz poročil o bonitetnih ocenah držav?*

Preučujemo tudi povezavo med merami besedilnega sentimenta in subjektivnosti ter trgi s posli kreditnih zamenjav državnih vrednostnih papirjev (angl. credit default swaps, CDS). Če se osredotočimo na ozko okno treh dni okoli objave posamezne bonitetne agencije, ugotavljamo, da tržni udeleženci poročilom o bonitetni oceni držav pripisujejo precej omejeno vlogo. Vendar pa opažamo pozitiven učinek mer subjektivnosti na CDS razmike po pozitivni boniteti objavi. Rezultati tudi kažejo, da trgi nekoliko višje vrednotijo kakovostno presojo agencije Moody's, kot je izražena v poročilih, v primerjavi s presojo, izraženo v poročilih preostalih dveh agencij.

Kljub temu zaznamo pomemben odziv trga tako na pozitivne kot negativne objave bonitetnih ocen. Dokazi kažejo tudi na to, da Standard & Poor's na trgu uživa največji ugled med tremi agencijami. Ne opažamo pomembnih razlik v tržnih reakcijah med razvitimi gospodarstvi in razvijajočimi se trgi. Ko razširimo model s finančnimi spremenljivkami, opazimo, da različne spremenljivke vplivajo na nenormalne CDS donose glede na različne definicije dogodkov. S preučitvijo potencialno pomembnega prehoda med naložbenimi in špekulativnimi bonitetnimi razredi zaznamo statistično značilne donose pri prehodu iz špekulativnega v naložbeni razred.

**Prihodnje raziskave**

Menimo, da ta doktorska disertacija predstavlja doslej najbolj izčrpno in temeljito analizo poročil o bonitetnih ocenah držav. Vendar pa rezultati v disertaciji odpirajo vrata številnim potencialnim študijam v prihodnosti.

Prvič, kot navajamo v poglavju 5, se verjetnosti prehodov med posameznimi bonitetnimi razredi razlikujejo, česar nismo upoštevali, saj je v disertaciji poudarek na identifikaciji najučinkovitejših mer besedilnega sentimenta in subjektivnosti. Tako zaenkrat ne ocenjujemo prehodnih matrik, kar je sicer naravno nadaljevanje naših obstoječih raziskav. Ugotavljamo, da je ta pristop problematičen zaradi omejene razpoložljivosti podatkov o bonitetnih ocenah držav, vendar pa je nekaj obstoječim raziskavam uspelo to težavo zaobiti Hill et al. (2010); Hu et al. (2002).

Drugič, zaradi nedostopnosti podatkov o CDS razmikih pred decembrom 2007 nismo mogli preučiti reakcij trga državnih CDS razmikov na informacijsko vrednost, pridobljeno iz poročil o bonitetnih ocenah držav pred svetovno finančno krizo, in kako se je to razmerje razvijalo, če sploh. Da bomo lahko odgovorili na to vprašanje, je naslednji korak torej pridobitev manjkajočih podatkov.

Tretjič, analiza temelji na tradicionalnem okviru študije dogodkov (angl. event study). Osredotočamo se na ozko tridnevno okno okoli dogodka, saj se osredotočamo na mere, pridobljene iz besedilne analize, vendar na splošno študije obravnavajo širšo paleto oken okoli dogodkov. Tako bi bilo naravno nadaljevanje naše raziskave razširitev analize z vključitvijo dodatnih časovnih intervalov. Predvidevamo, da bi lahko mere besedilnega sentimenta in subjektivnosti potencialno vplivale na donos z zaostankom, saj lahko vlagatelji potrebujejo nekaj časa za obdelavo poročil.

Kar zadeva metodologijo, kot poudarjata Kearney and Liu (2014), je postopek besedilne analize mogoče izboljšati. Čeprav so bili razviti slovarji s področja financ in se njihova uporaba povečuje, je za prihodnje študije potrebna izdelava in razpoložljivost obsežnejših slovarjev, prilagojenih raziskovalnemu področju in tipu besedil. Na primer, slovar LM, ki temelji na letnih poročilih podjetij v ZDA, morda ni povsem primeren za analizo poročil o bonitetnih ocenah. Poleg tega je poleg enakih uteži, uporabljenih v tej disertaciji, prostor za izboljšave tudi s shemami ponderiranja. Z Naïvnim Bayesom se komajda dotaknemo gladine bazena algoritmov strojnega učenja, saj obstaja ogromno dodatnih, naprednih možnosti, kot so metoda podpornih vektorjev (angl. support vector machines) ali nevronske mreže (angl. neural networks). Z uporabo interdisciplinarnega pristopa z vključitvijo lingvistov, psihologov in podatkovnih znanstvenikov lahko razvijemo druge pristope k besedilni analizi. Vse to lahko privede do natančnejših in učinkovitejših mer sentimenta in subjektivnosti. Glede tipov besedil obstajajo viri, ki še niso bili preučeni ali niso bili dovolj proučeni, kjer so poročila o bonitetnih ocenah držav le eden izmed takih primerov.

Uporaba metode besedilne analize v financah je še vedno razmeroma nova, zato bi bili številni pristopi drugačni v okviru tega, kar poznamo danes. Nekatere empirične rezultate, ki merijo sentiment v finančnih besedilih, je treba ponovno oceniti (Loughran & McDonald, 2016). Poleg tega je v digitaliziranem svetu, kjer so nam na voljo obsežne baze besedilnih podatkov, računalniki pa so vse močnejši, mogoče narediti še veliko.

## Appendix B: Chapter 4 - Binary analysis results by credit rating agency

*Table B1.* Bivariate analysis of Standard & Poor's: mean comparison of key variables for advanced economies (AE) and emerging markets (EME), and before and after the Global financial crisis (GFC)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | A: Emerging markets vs. advanced economies | | | | | |
| Mean (EME) | -1.542 | -0.212 | 0.383 | 2.671 | 34.129 | 34.712 |
| Mean (AE) | -1.009 | -0.134 | 0.474 | 2.752 | 34.391 | 34.298 |
| Diff. in means (EME-AE) | -0.533*** | -0.077*** | -0.092*** | -0.081 | -0.262 | 0.414 |
| | (0.109) | (0.014) | (0.016) | (0.052) | (0.559) | (0.516) |
| Observations (EME) | 831 | Observations (Total) | | | 1382 | |
| Observations (AE) | 551 | | | | | |
| | B: Before vs. after the Global financial crisis (GFC) | | | | | |
| Mean (before GFC) | -1.394 | -0.167 | 0.461 | 2.290 | 31.420 | 35.404 |
| Mean (after GFC) | -1.302 | -0.187 | 0.401 | 2.881 | 35.445 | 34.178 |
| Diff. in means (before-after GFC) | -0.092 | 0.020 | 0.060*** | -0.591*** | -4.025*** | 1.226** |
| | (0.125) | (0.015) | (0.017) | (0.054) | (0.613) | (0.576) |
| Observations (before GFC) | 416 | Observations (Total) | | | 1382 | |
| Observations (after GFC) | 966 | | | | | |
| | C: Emerging markets before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.655 | -0.199 | 0.391 | 2.319 | 32.256 | 35.381 |
| Mean (after GFC) | -1.494 | -0.217 | 0.379 | 2.820 | 34.921 | 34.430 |
| Diff. in means (before-after GFC) | -0.161 | 0.018 | 0.012 | -0.500*** | -2.666*** | 0.952 |
| | (0.166) | (0.020) | (0.021) | (0.066) | (0.775) | (0.747) |
| Observations (before GFC) | 247 | Observations (Total EME) | | | 831 | |
| Observations (after GFC) | 584 | | | | | |
| | D: Advanced economies before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.011 | -0.120 | 0.563 | 2.247 | 30.200 | 35.438 |
| Mean (after GFC) | -1.008 | -0.141 | 0.435 | 2.975 | 36.245 | 33.794 |
| Diff. in means (before-after GFC) | -0.003 | 0.021 | 0.128*** | -0.727*** | -6.046*** | 1.643* |
| | (0.187) | (0.024) | (0.028) | (0.090) | (0.992) | (0.906) |
| Observations (before GFC) | 169 | Observations (Total AE) | | | 551 | |
| Observations (after GFC) | 382 | | | | | |
| | E: Emerging vs. advanced markets before the Global financial crisis | | | | | |
| Mean (EME) | -1.655 | -0.199 | 0.391 | 2.319 | 32.256 | 35.381 |
| Mean (AE) | -1.011 | -0.120 | 0.563 | 2.247 | 30.200 | 35.438 |
| Diff. in means (EME-AE) | -0.644*** | -0.079*** | -0.172*** | 0.072 | 2.056* | -0.056 |
| | (0.217) | (0.026) | (0.029) | (0.095) | (1.096) | (1.013) |
| Observations (EME) | 247 | Observations (Total before GFC) | | | 416 | |
| Observations (AE) | 169 | | | | | |
| | F: Emerging vs. advanced markets after the Global financial crisis | | | | | |
| Mean (EME) | -1.494 | -0.217 | 0.379 | 2.820 | 34.921 | 34.430 |
| Mean (AE) | -1.008 | -0.141 | 0.435 | 2.975 | 36.245 | 33.794 |
| Diff. in means (EME-AE) | -0.486*** | -0.076*** | -0.056*** | -0.155*** | -1.324** | 0.635 |
| | (0.124) | (0.017) | (0.019) | (0.058) | (0.619) | (0.594) |
| Observations (EME) | 584 | Observations (Total after GFC) | | | 966 | |
| Observations (AE) | 382 | | | | | |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table B2.* Bivariate analysis of Fitch: mean comparison of key variables for advanced economies (AE) and emerging markets (EME), and before and after the Global financial crisis (GFC)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | A: Emerging markets vs. advanced economies | | | | | |
| Mean (EME) | -1.905 | -0.218 | 0.358 | 3.147 | 37.512 | 38.475 |
| Mean (AE) | -1.610 | -0.190 | 0.434 | 3.118 | 35.986 | 33.245 |
| Diff. in means (EME-AE) | -0.295** | -0.028* | -0.076*** | 0.029 | 1.526** | 5.230*** |
| | (0.125) | (0.015) | (0.016) | (0.063) | (0.638) | (0.657) |
| Observations (EME) | 842 | Observations (Total) | | | 1433 | |
| Observations (AE) | 591 | | | | | |
| | B: Before vs. after the Global financial crisis (GFC) | | | | | |
| Mean (before GFC) | -1.219 | -0.137 | 0.453 | 2.507 | 32.293 | 35.298 |
| Mean (after GFC) | -2.094 | -0.244 | 0.355 | 3.481 | 39.411 | 36.880 |
| Diff. in means (before-after GFC) | 0.874*** | 0.107*** | 0.099*** | -0.974*** | -7.118*** | -1.581** |
| | (0.135) | (0.016) | (0.016) | (0.060) | (0.664) | (0.726) |
| Observations (before GFC) | 509 | Observations (Total) | | | 1433 | |
| Observations (after GFC) | 924 | | | | | |
| | C: Emerging markets before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.294 | -0.147 | 0.403 | 2.551 | 33.323 | 38.711 |
| Mean (after GFC) | -2.244 | -0.257 | 0.334 | 3.477 | 39.831 | 38.344 |
| Diff. in means (before-after GFC) | 0.950*** | 0.110*** | 0.069*** | -0.926*** | -6.508*** | 0.367 |
| | (0.180) | (0.021) | (0.021) | (0.082) | (0.931) | (0.955) |
| Observations (before GFC) | 300 | Observations (Total EME) | | | 842 | |
| Observations (after GFC) | 542 | | | | | |
| | D: Advanced economies before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.113 | -0.123 | 0.526 | 2.444 | 30.814 | 30.399 |
| Mean (after GFC) | -1.882 | -0.226 | 0.384 | 3.487 | 38.815 | 34.802 |
| Diff. in means (before-after GFC) | 0.769*** | 0.103*** | 0.142*** | -1.043*** | -8.001*** | -4.402*** |
| | (0.203) | (0.025) | (0.024) | (0.085) | (0.902) | (1.020) |
| Observations (before GFC) | 209 | Observations (Total AE) | | | 591 | |
| Observations (after GFC) | 382 | | | | | |
| | E: Emerging vs. advanced markets before the Global financial crisis | | | | | |
| Mean (EME) | -1.294 | -0.147 | 0.403 | 2.551 | 33.323 | 38.711 |
| Mean (AE) | -1.113 | -0.123 | 0.526 | 2.444 | 30.814 | 30.399 |
| Diff. in means (EME-AE) | -0.181 | -0.024 | -0.123*** | 0.107 | 2.509** | 8.312*** |
| | (0.230) | (0.029) | (0.026) | (0.092) | (1.064) | (1.156) |
| Observations (EME) | 300 | Observations (Total before GFC) | | | 509 | |
| Observations (AE) | 209 | | | | | |
| | F: Emerging vs. advanced markets after the Global financial crisis | | | | | |
| Mean (EME) | -2.244 | -0.257 | 0.334 | 3.477 | 39.831 | 38.344 |
| Mean (AE) | -1.882 | -0.226 | 0.384 | 3.487 | 38.815 | 34.802 |
| Diff. in means (EME-AE) | -0.362** | -0.031* | -0.050*** | -0.010 | 1.015 | 3.543*** |
| | (0.144) | (0.016) | (0.019) | (0.074) | 0.742 | 0.784 |
| Observations (EME) | 542 | Observations (Total after GFC) | | | 924 | |
| Observations (AE) | 382 | | | | | |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table B3.* Bivariate analysis of Moody's: mean comparison of key variables for advanced economies (AE) and emerging markets (EME), and before and after the Global financial crisis (GFC)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | A: Emerging markets vs. advanced economies | | | | | |
| Mean (EME) | -1.379 | -0.160 | 0.239 | 2.404 | 29.939 | 43.014 |
| Mean (AE) | -0.791 | -0.101 | 0.362 | 2.516 | 30.874 | 44.308 |
| Diff. in means (EME-AE) | -0.588$^{***}$ | -0.059$^{***}$ | -0.122$^{***}$ | -0.113$^{*}$ | -0.935 | -1.294 |
| | (0.123) | (0.018) | (0.024) | (0.063) | (0.642) | (0.794) |
| Observations (EME) | 982 | Observations (Total) | | | 1580 | |
| Observations (AE) | 598 | | | | | |
| | B: Before vs. after the Global financial crisis (GFC) | | | | | |
| Mean (before GFC) | -0.907 | -0.104 | 0.377 | 2.015 | 26.977 | 49.874 |
| Mean (after GFC) | -1.313 | -0.159 | 0.229 | 2.718 | 32.378 | 39.498 |
| Diff. in means (before-after GFC) | 0.407$^{***}$ | 0.055$^{***}$ | 0.148$^{***}$ | -0.703$^{***}$ | -5.402$^{***}$ | 10.376$^{***}$ |
| | (0.134) | (0.018) | (0.024) | (0.059) | (0.672) | (0.799) |
| Observations (before GFC) | 769 | Observations (Total) | | | 1580 | |
| Observations (after GFC) | 811 | | | | | |
| | C: Emerging markets before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -1.220 | -0.139 | 0.288 | 2.066 | 27.112 | 49.624 |
| Mean (after GFC) | -1.485 | -0.174 | 0.207 | 2.630 | 31.834 | 38.585 |
| Diff. in means (before-after GFC) | 0.265 | 0.036 | 0.081$^{***}$ | -0.565$^{***}$ | -4.722$^{***}$ | 11.039$^{***}$ |
| | (0.183) | (0.024) | (0.030) | (0.072) | (0.833) | (1.005) |
| Observations (before GFC) | 485 | Observations (Total EME) | | | 982 | |
| Observations (after GFC) | 497 | | | | | |
| | D: Advanced economies before vs. after the Global financial crisis | | | | | |
| Mean (before GFC) | -0.335 | -0.040 | 0.538 | 1.921 | 26.731 | 50.330 |
| Mean (after GFC) | -1.048 | -0.135 | 0.262 | 2.853 | 33.217 | 40.904 |
| Diff. in means (before-after GFC) | 0.714$^{***}$ | 0.095$^{***}$ | 0.277$^{***}$ | -0.932$^{***}$ | -6.485$^{***}$ | 9.426$^{***}$ |
| | (0.174) | (0.028) | (0.036) | (0.099) | (1.132) | (1.299) |
| Observations (before GFC) | 284 | Observations (Total AE) | | | 598 | |
| Observations (after GFC) | 314 | | | | | |
| | E: Emerging vs. advanced markets before the Global financial crisis | | | | | |
| Mean (EME) | -1.220 | -0.139 | 0.288 | 2.066 | 27.112 | 49.624 |
| Mean (AE) | -0.335 | -0.040 | 0.538 | 1.921 | 26.731 | 50.330 |
| Diff. in means (EME-AE) | -0.885$^{***}$ | -0.098$^{***}$ | -0.250$^{***}$ | 0.145 | 0.380 | -0.705 |
| | (0.203) | (0.030) | (0.036) | (0.099) | (1.236) | (1.403) |
| Observations (EME) | 485 | Observations (Total before GFC) | | | 769 | |
| Observations (AE) | 284 | | | | | |
| | F: Emerging vs. advanced markets after the Global financial crisis | | | | | |
| Mean (EME) | -1.485 | -0.174 | 0.207 | 2.630 | 31.834 | 38.585 |
| Mean (AE) | -1.048 | -0.135 | 0.262 | 2.853 | 33.217 | 40.904 |
| Diff. in means (EME-AE) | -0.437$^{***}$ | -0.039$^{*}$ | -0.055$^{*}$ | -0.223$^{***}$ | -1.382$^{**}$ | -2.319$^{***}$ |
| | (0.151) | (0.022) | (0.030) | (0.073) | (0.669) | (0.855) |
| Observations (EME) | 497 | Observations (Total after GFC) | | | 811 | |
| Observations (AE) | 314 | | | | | |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Standard errors in parentheses
$^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

# Appendix C: Chapter 4 - Model 3 results by credit rating agency

*Table C1.* Standard & Poor's: Estimation results of the ordered logit with random effects for the determinants of sovereign credit ratings, including macroeconomic, political, economic and cultural proximity variables, and sentiment and subjectivity measures

| | Sovereign credit ratings | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| GDP per capita | 0.181*** | 0.181*** | 0.188*** | 0.183*** | 0.183*** | 0.183*** |
| | (0.054) | (0.054) | (0.055) | (0.054) | (0.054) | (0.054) |
| Real GDP growth | 0.129 | -0.129 | -0.510 | 0.490 | 0.476 | 0.613 |
| | (3.725) | (3.686) | (3.541) | (3.739) | (3.742) | (3.782) |
| Inflation | -7.031*** | -6.988*** | -6.585*** | -7.124*** | -7.070*** | -7.307*** |
| | (2.075) | (2.074) | (2.034) | (2.113) | (2.123) | (2.080) |
| Current account/GDP | -5.046** | -5.211** | -5.167** | -4.879** | -4.880** | -4.902** |
| | (2.059) | (2.046) | (2.032) | (2.063) | (2.066) | (2.055) |
| Trade/GDP | 1.202* | 1.179 | 1.239* | 1.231* | 1.236* | 1.218* |
| | (0.729) | (0.726) | (0.753) | (0.733) | (0.734) | (0.728) |
| External debt/GDP | -0.004 | -0.000 | 0.006 | -0.008 | -0.006 | -0.014 |
| | (0.148) | (0.148) | (0.152) | (0.148) | (0.149) | (0.147) |
| Economic development | 5.706*** | 5.737*** | 5.662*** | 5.677*** | 5.674*** | 5.693*** |
| | (1.941) | (1.936) | (1.969) | (1.944) | (1.945) | (1.939) |
| Default history | -3.146*** | -3.182*** | -3.325*** | -3.107*** | -3.111*** | -3.103*** |
| | (0.952) | (0.950) | (0.950) | (0.952) | (0.947) | (0.953) |
| Log of int. reserves | 0.600* | 0.597* | 0.569 | 0.620* | 0.620* | 0.617* |
| | (0.359) | (0.357) | (0.361) | (0.360) | (0.358) | (0.353) |
| Government debt/GDP | -8.147*** | -8.131*** | -8.335*** | -8.150*** | -8.174*** | -8.062*** |
| | (1.274) | (1.275) | (1.274) | (1.259) | (1.268) | (1.276) |
| Budget balance/GDP | -4.717 | -4.989 | -5.331 | -4.461 | -4.515 | -4.222 |
| | (3.391) | (3.375) | (3.415) | (3.336) | (3.360) | (3.328) |
| Institutional quality | 0.110 | 0.108 | 0.107 | 0.110 | 0.110 | 0.104 |
| | (0.160) | (0.160) | (0.161) | (0.160) | (0.160) | (0.160) |
| Governance | 0.467*** | 0.464*** | 0.454*** | 0.466*** | 0.466*** | 0.465*** |
| | (0.061) | (0.061) | (0.059) | (0.059) | (0.060) | (0.060) |
| Trade proximity | 69.019*** | 68.700*** | 69.597*** | 68.613*** | 68.552*** | 68.677*** |
| | (22.072) | (21.986) | (21.937) | (22.237) | (22.143) | (22.099) |
| Common language | -0.659 | -0.654 | -0.706 | -0.654 | -0.654 | -0.628 |
| | (0.815) | (0.814) | (0.841) | (0.813) | (0.815) | (0.809) |
| Religious proximity | 1.462 | 1.466 | 1.275 | 1.445 | 1.435 | 1.483 |
| | (1.725) | (1.729) | (1.777) | (1.702) | (1.706) | (1.701) |
| Geographical distance | 0.011 | 0.011 | 0.010 | 0.011 | 0.011 | 0.011 |
| | (0.010) | (0.010) | (0.011) | (0.010) | (0.010) | (0.010) |
| Textual analysis measure | 0.033 | 0.458 | 1.024*** | -0.023 | -0.365 | 0.766 |
| | (0.048) | (0.347) | (0.324) | (0.087) | (0.733) | (0.720) |
| Observations | 1382 | 1382 | 1382 | 1382 | 1382 | 1382 |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Clustered standard errors in parentheses
Cut-off estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table C2.* Fitch: Estimation results of the ordered logit with random effects for the determinants of sovereign credit ratings, including macroeconomic, political, economic and cultural proximity variables, and sentiment and subjectivity measures

| | Sovereign credit ratings | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| GDP per capita | 0.196*** | 0.196*** | 0.199*** | 0.194*** | 0.197*** | 0.196*** |
| | (0.055) | (0.055) | (0.056) | (0.055) | (0.056) | (0.055) |
| Real GDP growth | -3.012 | -2.789 | -3.841 | -2.599 | -2.794 | -2.691 |
| | (3.132) | (3.203) | (3.048) | (3.159) | (3.246) | (3.220) |
| Inflation | -5.740** | -5.795** | -5.386** | -5.699** | -5.868** | -5.790** |
| | (2.398) | (2.397) | (2.328) | (2.347) | (2.371) | (2.376) |
| Current account/GDP | -3.850*** | -3.791*** | -4.026*** | -3.784*** | -3.806*** | -3.797*** |
| | (1.395) | (1.392) | (1.438) | (1.415) | (1.414) | (1.412) |
| Trade/GDP | 0.980 | 0.986 | 1.081 | 0.943 | 1.018 | 0.999 |
| | (0.682) | (0.680) | (0.690) | (0.678) | (0.677) | (0.681) |
| External debt/GDP | -0.183** | -0.184** | -0.180** | -0.180** | -0.186** | -0.185** |
| | (0.074) | (0.073) | (0.076) | (0.072) | (0.073) | (0.073) |
| Economic development | 7.114*** | 7.100*** | 7.141*** | 7.142*** | 7.074*** | 7.101*** |
| | (1.787) | (1.788) | (1.801) | (1.805) | (1.795) | (1.789) |
| Default history | -5.234*** | -5.226*** | -5.195*** | -5.226*** | -5.237*** | -5.237*** |
| | (0.996) | (1.007) | (0.971) | (1.008) | (1.008) | (1.010) |
| Log of int. reserves | 0.553** | 0.555** | 0.550** | 0.531** | 0.570** | 0.553** |
| | (0.258) | (0.256) | (0.266) | (0.263) | (0.262) | (0.256) |
| Government debt/GDP | -8.426*** | -8.408*** | -8.453*** | -8.463*** | -8.371*** | -8.408*** |
| | (1.072) | (1.077) | (1.065) | (1.083) | (1.081) | (1.078) |
| Budget balance/GDP | -5.708* | -5.434 | -7.334** | -5.233 | -5.449* | -5.414 |
| | (3.392) | (3.401) | (3.358) | (3.277) | (3.283) | (3.293) |
| Institutional quality | -0.022 | -0.023 | -0.017 | -0.020 | -0.027 | -0.025 |
| | (0.132) | (0.132) | (0.133) | (0.131) | (0.130) | (0.130) |
| Governance | 0.448*** | 0.451*** | 0.436*** | 0.459*** | 0.448*** | 0.451*** |
| | (0.060) | (0.060) | (0.058) | (0.059) | (0.058) | (0.060) |
| Trade proximity | 31.668*** | 31.824*** | 32.179*** | 32.312*** | 31.433*** | 31.795*** |
| | (10.155) | (10.151) | (10.708) | (10.139) | (10.148) | (10.179) |
| Common language | -1.149 | -1.150 | -1.189 | -1.152 | -1.148 | -1.154 |
| | (0.848) | (0.845) | (0.855) | (0.849) | (0.844) | (0.846) |
| Religious proximity | -0.924 | -0.899 | -0.982 | -0.911 | -0.891 | -0.903 |
| | (1.576) | (1.570) | (1.597) | (1.580) | (1.574) | (1.573) |
| Geographical distance | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 | -0.009 |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) |
| Textual analysis measure | 0.021 | 0.042 | 1.010*** | 0.058 | -0.473 | -0.228 |
| | (0.043) | (0.370) | (0.310) | (0.086) | (0.686) | (0.602) |
| Observations | 1433 | 1433 | 1433 | 1433 | 1433 | 1433 |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Clustered standard errors in parentheses
Cut-off estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table C3.* Moody's: Estimation results of the ordered logit with random effects for the determinants of sovereign credit ratings, including macroeconomic, political, economic and cultural proximity variables, and sentiment and subjectivity measures

| | Sovereign credit ratings | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| GDP per capita | 0.162*** | 0.162*** | 0.164*** | 0.172*** | 0.167*** | 0.159*** |
| | (0.051) | (0.051) | (0.052) | (0.053) | (0.052) | (0.051) |
| Real GDP growth | -2.323 | -2.147 | -2.486 | -1.900 | -1.960 | -1.986 |
| | (2.132) | (2.133) | (2.224) | (2.268) | (2.231) | (2.226) |
| Inflation | -2.781 | -2.838 | -2.728 | -2.781 | -2.840 | -2.936 |
| | (2.442) | (2.459) | (2.406) | (2.451) | (2.434) | (2.441) |
| Current account/GDP | -4.783** | -4.779** | -4.720** | -4.703** | -4.745** | -4.740** |
| | (2.061) | (2.050) | (2.037) | (2.012) | (2.040) | (2.067) |
| Trade/GDP | -0.133 | -0.138 | -0.122 | -0.069 | -0.112 | -0.165 |
| | (0.607) | (0.609) | (0.608) | (0.605) | (0.605) | (0.626) |
| External debt/GDP | -0.130* | -0.130* | -0.133* | -0.139* | -0.134* | -0.127* |
| | (0.075) | (0.075) | (0.075) | (0.075) | (0.076) | (0.075) |
| Economic development | 5.507*** | 5.498*** | 5.481*** | 5.448*** | 5.501*** | 5.662*** |
| | (1.699) | (1.692) | (1.694) | (1.647) | (1.653) | (1.718) |
| Default history | -5.601*** | -5.564*** | -5.575*** | -5.572*** | -5.583*** | -5.603*** |
| | (1.211) | (1.215) | (1.213) | (1.207) | (1.215) | (1.242) |
| Log of int. reserves | 0.455** | 0.456** | 0.465** | 0.509*** | 0.484** | 0.430** |
| | (0.194) | (0.195) | (0.194) | (0.190) | (0.190) | (0.191) |
| Government debt/GDP | -7.309*** | -7.325*** | -7.325*** | -7.382*** | -7.372*** | -7.448*** |
| | (1.163) | (1.163) | (1.151) | (1.142) | (1.149) | (1.173) |
| Budget balance/GDP | -5.805* | -5.739* | -5.999* | -5.674* | -5.747* | -5.585* |
| | (3.456) | (3.482) | (3.470) | (3.401) | (3.487) | (3.385) |
| Institutional quality | -0.009 | -0.010 | -0.016 | -0.027 | -0.022 | -0.017 |
| | (0.103) | (0.103) | (0.103) | (0.099) | (0.100) | (0.102) |
| Governance | 0.415*** | 0.419*** | 0.416*** | 0.410*** | 0.417*** | 0.432*** |
| | (0.053) | (0.053) | (0.052) | (0.051) | (0.051) | (0.050) |
| Trade proximity | 31.516*** | 31.654*** | 31.070*** | 32.194*** | 31.521*** | 32.408*** |
| | (8.810) | (8.870) | (8.831) | (9.180) | (8.875) | (8.934) |
| Common language | -0.342 | -0.345 | -0.337 | -0.328 | -0.328 | -0.387 |
| | (0.690) | (0.690) | (0.690) | (0.693) | (0.691) | (0.700) |
| Religious proximity | 0.755 | 0.745 | 0.687 | 0.790 | 0.739 | 0.749 |
| | (1.393) | (1.394) | (1.392) | (1.374) | (1.373) | (1.411) |
| Geographical distance | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.012 |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) | (0.010) |
| Textual analysis measure | 0.043 | 0.214 | 0.325 | -0.198** | -1.301* | -0.810 |
| | (0.037) | (0.285) | (0.207) | (0.091) | (0.774) | (0.673) |
| Observations | 1580 | 1580 | 1580 | 1580 | 1580 | 1580 |

(1) Net sentiment (W, dict), (2) Polarity (W, dict), (3) Polarity (S, ML), (4) Subjectivity (W, dict), (5) Subjectivity (S, dict), (6) Subjectivity (S, ML)
Clustered standard errors in parentheses
Cut-off estimates are not reported.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

# Appendix D: Chapter 4 - Robustness check

*Table D1.* Estimation results of sentiment and subjectivity scores as determinants of sovereign credit ratings from the fixed effects, random effects, pooled OLS and ordered logit model

| | Sovereign credit ratings | | | |
|---|---|---|---|---|
| | Fixed Effects | Random Effects | Pooled OLS | Ordered logit |
| | S&P | | | |
| Net sentiment (W, dict) | 0.000 | 0.009 | 0.035 | 0.041 |
| | (0.024) | (0.026) | (0.032) | (0.038) |
| Polarity (W, dict) | 0.173 | 0.234 | 0.397 | 0.439 |
| | (0.176) | (0.188) | (0.251) | (0.300) |
| Polarity (S, ML) | 0.460** | 0.464** | 0.526** | 0.846** |
| | (0.181) | (0.195) | (0.251) | (0.330) |
| Subjectivity (W, dict) | 0.010 | 0.001 | -0.023 | -0.016 |
| | (0.048) | (0.051) | (0.089) | (0.086) |
| Subjectivity (S, dict) | 0.291 | 0.195 | 0.211 | 0.008 |
| | (0.356) | (0.382) | (0.665) | (0.619) |
| Subjectivity (S, ML) | 1.273*** | 1.092*** | 1.473* | 1.257 |
| | (0.402) | (0.421) | (0.853) | (0.850) |
| | Fitch | | | |
| Net sentiment (W, dict) | 0.025 | 0.018 | 0.033 | 0.009 |
| | (0.022) | (0.023) | (0.030) | (0.035) |
| Polarity (W, dict) | 0.079 | 0.048 | 0.367 | 0.167 |
| | (0.177) | (0.187) | (0.275) | (0.314) |
| Polarity (S, ML) | 0.538*** | 0.531*** | 0.712** | 0.695** |
| | (0.178) | (0.182) | (0.308) | (0.311) |
| Subjectivity (W, dict) | -0.008 | 0.007 | 0.082 | 0.046 |
| | (0.044) | (0.045) | (0.060) | (0.075) |
| Subjectivity (S, dict) | -0.380 | -0.231 | 0.316 | -0.140 |
| | (0.335) | (0.334) | (0.494) | (0.672) |
| Subjectivity (S, ML) | -0.097 | 0.056 | 0.157 | -0.224 |
| | (0.292) | (0.298) | (0.462) | (0.625) |
| | Moody's | | | |
| Net sentiment (W, dict) | 0.042* | 0.035 | -0.011 | 0.007 |
| | (0.024) | (0.023) | (0.027) | (0.033) |
| Polarity (W, dict) | 0.341* | 0.285 | -0.094 | -0.058 |
| | (0.182) | (0.174) | (0.211) | (0.261) |
| Polarity (S, ML) | 0.193* | 0.164 | -0.121 | 0.011 |
| | (0.112) | (0.115) | (0.173) | (0.227) |
| Subjectivity (W, dict) | -0.076 | -0.077 | -0.051 | -0.148* |
| | (0.056) | (0.058) | (0.088) | (0.088) |
| Subjectivity (S, dict) | -0.684 | -0.635 | -0.183 | -0.507 |
| | (0.522) | (0.512) | (0.742) | (0.762) |
| Subjectivity (S, ML) | -0.677 | -0.576 | -0.176 | 0.049 |
| | (0.419) | (0.416) | (0.550) | (0.552) |
| Observations | S&P/Moody's/Fitch | | 1382/1433/1580 | |

Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The results are based on the analysis of the relationship between sovereign credit ratings and macroeconomic, political, economic and cultural proximity variables, and sentiment and subjectivity measures, but only the coefficients for sentiment and subjectivity measures are reported.

# Appendix E: Chapter 5 - Classification accuracy of downgrades and upgrades

*Table E1.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, following the point-in-time approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | | Model 1 | | | | |
| No textual indicators | 6.044 | 86.541 | 0.785 | 2.976 | 88.276 | 0.742 | 0.000 | 90.759 | 0.742 |
| Net sentiment (W, dict) | 6.593 | 86.469 | 0.801 | 4.762 | 88.067 | 0.769 | 0.690 | 90.633 | 0.776 |
| Polarity (W, dict) | 6.593 | 86.614 | 0.799 | 2.976 | 87.858 | 0.766 | 1.379 | 90.696 | 0.774 |
| Polarity (S, ML) | 6.593 | 86.686 | 0.798 | 4.167 | 88.416 | 0.769 | 0.000 | 90.696 | 0.774 |
| Subjectivity (W, dict) | 7.143 | 86.252 | 0.800 | 2.976 | 88.276 | 0.743 | 0.000 | 90.696 | 0.742 |
| Subjectivity (S, dict) | 6.593 | 86.614 | 0.792 | 2.976 | 88.276 | 0.742 | 0.000 | 90.759 | 0.742 |
| Subjectivity (S, ML) | 6.044 | 86.541 | 0.785 | 2.976 | 88.207 | 0.745 | 0.000 | 90.759 | 0.752 |
| All textual indicators | 13.187 | 87.048 | 0.814 | 4.762 | 88.346 | 0.784 | 2.759 | 90.823 | 0.795 |
| | | | | | Model 2 | | | | |
| No textual indicators | 6.044 | 86.686 | 0.790 | 2.976 | 88.276 | 0.744 | 0.000 | 90.696 | 0.746 |
| Net sentiment (W, dict) | 6.044 | 86.324 | 0.806 | 4.167 | 87.997 | 0.770 | 0.690 | 90.696 | 0.779 |
| Polarity (W, dict) | 6.044 | 86.469 | 0.804 | 4.762 | 88.137 | 0.768 | 1.379 | 90.633 | 0.777 |
| Polarity (S, ML) | 8.242 | 86.686 | 0.802 | 4.167 | 88.276 | 0.770 | 0.000 | 90.759 | 0.776 |
| Subjectivity (W, dict) | 8.791 | 86.541 | 0.803 | 2.976 | 88.276 | 0.745 | 0.000 | 90.696 | 0.746 |
| Subjectivity (S, dict) | 7.692 | 86.903 | 0.797 | 2.976 | 88.276 | 0.744 | 0.000 | 90.759 | 0.747 |
| Subjectivity (S, ML) | 6.044 | 86.686 | 0.790 | 4.167 | 88.346 | 0.747 | 0.000 | 90.570 | 0.761 |
| All textual indicators | 13.736 | 86.975 | 0.817 | 4.762 | 88.416 | 0.785 | 2.069 | 90.759 | 0.800 |
| | | | | | Model 3 | | | | |
| No textual indicators | 7.692 | 86.614 | 0.803 | 3.571 | 88.276 | 0.762 | 0.690 | 90.570 | 0.764 |
| Net sentiment (W, dict) | 10.989 | 86.324 | 0.820 | 4.762 | 87.718 | 0.783 | 0.690 | 90.443 | 0.787 |
| Polarity (W, dict) | 10.989 | 86.541 | 0.818 | 3.571 | 87.579 | 0.781 | 0.690 | 90.506 | 0.785 |
| Polarity (S, ML) | 9.341 | 86.107 | 0.815 | 5.952 | 88.346 | 0.784 | 0.690 | 90.570 | 0.785 |
| Subjectivity (W, dict) | 9.890 | 86.397 | 0.814 | 3.571 | 88.346 | 0.763 | 0.690 | 90.570 | 0.764 |
| Subjectivity (S, dict) | 9.341 | 86.614 | 0.808 | 3.571 | 88.276 | 0.763 | 0.690 | 90.633 | 0.763 |
| Subjectivity (S, ML) | 7.692 | 86.614 | 0.803 | 3.571 | 88.276 | 0.763 | 0.690 | 90.506 | 0.777 |
| All textual indicators | 15.385 | 86.975 | 0.830 | 6.548 | 88.067 | 0.795 | 3.448 | 90.886 | 0.805 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Table E2.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, following the through-the-cycle approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 15.569 | 86.991 | 0.845 | 4.605 | 87.579 | 0.820 | 3.101 | 90.623 | 0.808 |
| Net sentiment (W, dict) | 22.840 | 87.677 | 0.857 | 14.765 | 88.090 | 0.853 | 9.302 | 91.320 | 0.834 |
| Polarity (W, dict) | 21.605 | 87.344 | 0.858 | 17.450 | 88.489 | 0.853 | 7.752 | 91.033 | 0.830 |
| Polarity (S, ML) | 23.457 | 87.261 | 0.863 | 14.094 | 88.090 | 0.855 | 6.977 | 90.961 | 0.833 |
| Subjectivity (W, dict) | 25.309 | 87.510 | 0.864 | 9.396 | 88.169 | 0.828 | 3.876 | 90.674 | 0.811 |
| Subjectivity (S, dict) | 19.753 | 86.511 | 0.857 | 8.725 | 88.090 | 0.829 | 3.876 | 90.674 | 0.812 |
| Subjectivity (S, ML) | 15.432 | 86.761 | 0.845 | 5.369 | 87.450 | 0.827 | 5.426 | 90.603 | 0.818 |
| All textual indicators | 29.630 | 87.594 | 0.881 | 26.846 | 88.649 | 0.881 | 14.729 | 91.392 | 0.864 |
| | | | | Model 2 | | | | | |
| No textual indicators | 18.563 | 86.512 | 0.858 | 9.211 | 87.893 | 0.841 | 4.651 | 90.551 | 0.829 |
| Net sentiment (W, dict) | 21.605 | 87.094 | 0.869 | 16.779 | 88.489 | 0.860 | 8.527 | 91.033 | 0.848 |
| Polarity (W, dict) | 22.222 | 87.261 | 0.869 | 19.463 | 88.649 | 0.861 | 7.752 | 90.746 | 0.844 |
| Polarity (S, ML) | 26.543 | 87.510 | 0.872 | 15.436 | 87.930 | 0.867 | 7.752 | 90.961 | 0.845 |
| Subjectivity (W, dict) | 27.778 | 87.760 | 0.874 | 13.423 | 88.569 | 0.848 | 5.426 | 90.674 | 0.832 |
| Subjectivity (S, dict) | 24.074 | 86.928 | 0.867 | 14.765 | 88.489 | 0.848 | 6.202 | 90.746 | 0.831 |
| Subjectivity (S, ML) | 19.753 | 86.511 | 0.858 | 10.067 | 87.850 | 0.846 | 10.078 | 90.818 | 0.844 |
| All textual indicators | 36.420 | 89.092 | 0.888 | 28.188 | 88.729 | 0.892 | 14.729 | 90.818 | 0.878 |
| | | | | Model 3 | | | | | |
| No textual indicators | 23.494 | 86.914 | 0.868 | 11.184 | 87.609 | 0.848 | 5.556 | 90.725 | 0.841 |
| Net sentiment (W, dict) | 26.708 | 87.605 | 0.879 | 19.463 | 88.368 | 0.866 | 9.524 | 91.068 | 0.854 |
| Polarity (W, dict) | 27.329 | 87.774 | 0.880 | 18.792 | 88.288 | 0.868 | 8.730 | 90.850 | 0.850 |
| Polarity (S, ML) | 32.919 | 88.027 | 0.882 | 17.450 | 88.045 | 0.872 | 11.111 | 91.285 | 0.854 |
| Subjectivity (W, dict) | 32.298 | 88.196 | 0.880 | 13.423 | 88.126 | 0.856 | 5.556 | 90.559 | 0.843 |
| Subjectivity (S, dict) | 27.329 | 87.521 | 0.875 | 14.765 | 88.288 | 0.856 | 5.556 | 90.632 | 0.841 |
| Subjectivity (S, ML) | 22.360 | 86.256 | 0.867 | 12.752 | 87.884 | 0.854 | 12.698 | 90.777 | 0.857 |
| All textual indicators | 38.509 | 88.702 | 0.898 | 26.846 | 88.611 | 0.893 | 15.079 | 90.850 | 0.884 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Table E3.* Classification accuracy of downgrades, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the point-in-time approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 27.673 | 90.159 | 0.856 | 23.077 | 91.923 | 0.848 | 31.200 | 92.039 | 0.896 |
| Net sentiment (W, dict) | 34.591 | 90.376 | 0.894 | 28.462 | 91.994 | 0.877 | 47.200 | 93.304 | 0.935 |
| Polarity (W, dict) | 40.252 | 91.100 | 0.903 | 24.615 | 91.637 | 0.874 | 39.200 | 92.262 | 0.925 |
| Polarity (S, ML) | 40.252 | 91.534 | 0.890 | 23.077 | 91.494 | 0.867 | 38.400 | 92.262 | 0.922 |
| Subjectivity (W, dict) | 29.560 | 90.376 | 0.855 | 22.308 | 91.422 | 0.853 | 30.400 | 91.890 | 0.898 |
| Subjectivity (S, dict) | 28.931 | 90.449 | 0.855 | 21.538 | 91.637 | 0.852 | 32.000 | 92.039 | 0.900 |
| Subjectivity (S, ML) | 28.302 | 90.232 | 0.856 | 22.308 | 91.851 | 0.847 | 31.200 | 92.039 | 0.896 |
| All textual indicators | 46.541 | 91.751 | 0.908 | 31.538 | 92.137 | 0.880 | 54.400 | 93.676 | 0.940 |
| | | | | Model 2 | | | | | |
| No textual indicators | 30.189 | 90.449 | 0.856 | 23.846 | 91.708 | 0.850 | 32.000 | 91.890 | 0.900 |
| Net sentiment (W, dict) | 34.591 | 90.593 | 0.894 | 30.769 | 92.495 | 0.880 | 46.400 | 93.006 | 0.936 |
| Polarity (W, dict) | 41.509 | 91.100 | 0.903 | 26.154 | 91.851 | 0.878 | 36.800 | 91.890 | 0.927 |
| Polarity (S, ML) | 38.365 | 91.245 | 0.892 | 23.846 | 91.637 | 0.868 | 39.200 | 92.336 | 0.924 |
| Subjectivity (W, dict) | 30.818 | 90.593 | 0.856 | 23.077 | 91.422 | 0.853 | 29.600 | 91.667 | 0.903 |
| Subjectivity (S, dict) | 32.704 | 90.738 | 0.856 | 23.077 | 91.637 | 0.853 | 30.400 | 91.815 | 0.903 |
| Subjectivity (S, ML) | 30.818 | 90.521 | 0.856 | 24.615 | 91.994 | 0.850 | 32.000 | 91.890 | 0.900 |
| All textual indicators | 45.912 | 91.823 | 0.909 | 32.308 | 92.280 | 0.883 | 54.400 | 93.899 | 0.941 |
| | | | | Model 3 | | | | | |
| No textual indicators | 32.704 | 90.810 | 0.863 | 23.077 | 91.637 | 0.850 | 31.200 | 92.113 | 0.903 |
| Net sentiment (W, dict) | 37.107 | 90.738 | 0.897 | 30.769 | 92.280 | 0.879 | 48.000 | 93.452 | 0.940 |
| Polarity (W, dict) | 40.252 | 91.389 | 0.905 | 26.923 | 91.994 | 0.878 | 39.200 | 91.964 | 0.931 |
| Polarity (S, ML) | 40.252 | 91.389 | 0.896 | 23.846 | 91.494 | 0.867 | 42.400 | 92.857 | 0.929 |
| Subjectivity (W, dict) | 33.333 | 90.593 | 0.863 | 23.846 | 91.637 | 0.854 | 31.200 | 92.336 | 0.905 |
| Subjectivity (S, dict) | 32.704 | 90.738 | 0.863 | 22.308 | 91.494 | 0.854 | 31.200 | 92.188 | 0.906 |
| Subjectivity (S, ML) | 32.704 | 90.738 | 0.862 | 25.385 | 91.923 | 0.849 | 32.000 | 92.411 | 0.903 |
| All textual indicators | 45.283 | 91.896 | 0.911 | 32.308 | 92.352 | 0.882 | 53.600 | 93.973 | 0.945 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Table E4.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the point-in-time approach, using the 0.3 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| **Model 1** | | | | | | | | | |
| No textual indicators | 25.824 | 83.647 | 0.788 | 16.970 | 86.919 | 0.749 | 11.905 | 89.435 | 0.783 |
| Net sentiment (W, dict) | 35.165 | 84.226 | 0.801 | 21.818 | 86.991 | 0.773 | 15.873 | 89.062 | 0.800 |
| Polarity (W, dict) | 35.165 | 84.081 | 0.799 | 23.636 | 87.134 | 0.770 | 16.667 | 88.914 | 0.799 |
| Polarity (S, ML) | 31.319 | 84.370 | 0.798 | 26.061 | 86.776 | 0.773 | 15.873 | 88.988 | 0.800 |
| Subjectivity (W, dict) | 30.220 | 84.153 | 0.802 | 16.970 | 86.848 | 0.749 | 13.492 | 89.360 | 0.783 |
| Subjectivity (S, dict) | 28.022 | 83.575 | 0.795 | 16.364 | 86.848 | 0.749 | 13.492 | 89.509 | 0.782 |
| Subjectivity (S, ML) | 25.275 | 83.502 | 0.788 | 18.788 | 87.062 | 0.752 | 15.873 | 88.467 | 0.789 |
| All textual indicators | 39.560 | 84.877 | 0.813 | 27.273 | 86.347 | 0.788 | 22.222 | 89.360 | 0.812 |
| **Model 2** | | | | | | | | | |
| No textual indicators | 26.923 | 83.502 | 0.793 | 18.788 | 86.776 | 0.751 | 11.905 | 88.988 | 0.786 |
| Net sentiment (W, dict) | 32.418 | 83.792 | 0.805 | 23.030 | 86.919 | 0.774 | 14.286 | 88.690 | 0.802 |
| Polarity (W, dict) | 32.967 | 83.719 | 0.804 | 23.636 | 86.776 | 0.772 | 15.873 | 89.062 | 0.801 |
| Polarity (S, ML) | 30.220 | 83.647 | 0.802 | 26.667 | 87.062 | 0.775 | 16.667 | 88.914 | 0.801 |
| Subjectivity (W, dict) | 30.769 | 84.370 | 0.806 | 18.788 | 86.848 | 0.751 | 11.111 | 88.988 | 0.786 |
| Subjectivity (S, dict) | 27.473 | 83.285 | 0.799 | 18.788 | 86.776 | 0.751 | 11.905 | 88.988 | 0.786 |
| Subjectivity (S, ML) | 26.923 | 83.502 | 0.793 | 18.182 | 87.062 | 0.754 | 15.079 | 88.467 | 0.795 |
| All textual indicators | 37.912 | 84.732 | 0.817 | 29.697 | 86.490 | 0.790 | 19.841 | 88.170 | 0.816 |
| **Model 3** | | | | | | | | | |
| No textual indicators | 26.923 | 83.575 | 0.804 | 19.394 | 86.562 | 0.770 | 9.524 | 88.467 | 0.794 |
| Net sentiment (W, dict) | 34.615 | 84.153 | 0.819 | 24.242 | 86.919 | 0.788 | 17.460 | 88.988 | 0.808 |
| Polarity (W, dict) | 34.066 | 84.153 | 0.818 | 24.242 | 86.919 | 0.786 | 19.841 | 89.062 | 0.807 |
| Polarity (S, ML) | 36.813 | 85.384 | 0.815 | 30.909 | 86.848 | 0.790 | 19.048 | 88.690 | 0.807 |
| Subjectivity (W, dict) | 32.967 | 83.936 | 0.815 | 21.212 | 86.633 | 0.771 | 10.317 | 88.542 | 0.794 |
| Subjectivity (S, dict) | 29.670 | 83.647 | 0.809 | 20.000 | 86.490 | 0.771 | 10.317 | 88.467 | 0.794 |
| Subjectivity (S, ML) | 26.374 | 83.430 | 0.804 | 21.818 | 86.848 | 0.771 | 13.492 | 88.244 | 0.806 |
| All textual indicators | 41.758 | 84.949 | 0.830 | 29.091 | 85.776 | 0.800 | 23.810 | 88.467 | 0.823 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Table E5.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the point-in-time approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | | Model 1 | | | | |
| No textual indicators | 7.143 | 86.469 | 0.788 | 3.030 | 88.206 | 0.749 | 0.794 | 90.402 | 0.783 |
| Net sentiment (W, dict) | 6.593 | 86.541 | 0.801 | 4.848 | 87.920 | 0.773 | 1.587 | 90.402 | 0.800 |
| Polarity (W, dict) | 6.593 | 86.758 | 0.799 | 3.636 | 87.848 | 0.770 | 0.794 | 90.179 | 0.799 |
| Polarity (S, ML) | 5.495 | 86.541 | 0.798 | 4.848 | 88.277 | 0.773 | 0.794 | 90.030 | 0.800 |
| Subjectivity (W, dict) | 8.791 | 86.614 | 0.802 | 3.030 | 88.134 | 0.749 | 0.794 | 90.476 | 0.783 |
| Subjectivity (S, dict) | 8.791 | 86.397 | 0.795 | 3.030 | 88.206 | 0.749 | 0.794 | 90.551 | 0.782 |
| Subjectivity (S, ML) | 7.143 | 86.469 | 0.788 | 3.030 | 88.134 | 0.752 | 0.794 | 90.402 | 0.789 |
| All textual indicators | 12.637 | 86.975 | 0.813 | 6.061 | 88.277 | 0.788 | 2.381 | 90.253 | 0.812 |
| | | | | | Model 2 | | | | |
| No textual indicators | 7.143 | 86.614 | 0.793 | 3.030 | 88.134 | 0.751 | 0.794 | 90.253 | 0.786 |
| Net sentiment (W, dict) | 7.143 | 86.397 | 0.805 | 4.242 | 87.920 | 0.774 | 1.587 | 90.253 | 0.802 |
| Polarity (W, dict) | 6.044 | 86.469 | 0.804 | 3.636 | 87.920 | 0.772 | 0.794 | 90.253 | 0.801 |
| Polarity (S, ML) | 7.692 | 86.614 | 0.802 | 4.242 | 88.206 | 0.775 | 1.587 | 90.179 | 0.801 |
| Subjectivity (W, dict) | 9.890 | 86.903 | 0.806 | 3.030 | 88.134 | 0.751 | 0.794 | 90.253 | 0.786 |
| Subjectivity (S, dict) | 7.692 | 86.397 | 0.799 | 3.030 | 88.134 | 0.751 | 0.794 | 90.253 | 0.786 |
| Subjectivity (S, ML) | 7.143 | 86.614 | 0.793 | 4.242 | 88.349 | 0.754 | 0.000 | 90.253 | 0.795 |
| All textual indicators | 13.187 | 86.975 | 0.817 | 6.061 | 88.420 | 0.790 | 2.381 | 90.253 | 0.816 |
| | | | | | Model 3 | | | | |
| No textual indicators | 8.242 | 86.686 | 0.804 | 3.636 | 88.063 | 0.770 | 0.794 | 90.327 | 0.794 |
| Net sentiment (W, dict) | 10.989 | 86.614 | 0.819 | 5.455 | 87.706 | 0.788 | 0.794 | 90.030 | 0.808 |
| Polarity (W, dict) | 10.440 | 86.469 | 0.818 | 3.636 | 87.420 | 0.786 | 0.794 | 89.955 | 0.807 |
| Polarity (S, ML) | 9.341 | 86.107 | 0.815 | 7.273 | 88.420 | 0.790 | 3.175 | 90.402 | 0.807 |
| Subjectivity (W, dict) | 12.088 | 87.048 | 0.815 | 3.636 | 87.920 | 0.771 | 0.794 | 90.327 | 0.794 |
| Subjectivity (S, dict) | 9.341 | 86.614 | 0.809 | 3.636 | 87.920 | 0.771 | 0.794 | 90.402 | 0.794 |
| Subjectivity (S, ML) | 8.242 | 86.686 | 0.804 | 3.636 | 88.063 | 0.771 | 1.587 | 90.253 | 0.806 |
| All textual indicators | 14.286 | 86.758 | 0.830 | 7.879 | 88.420 | 0.800 | 6.349 | 90.774 | 0.823 |

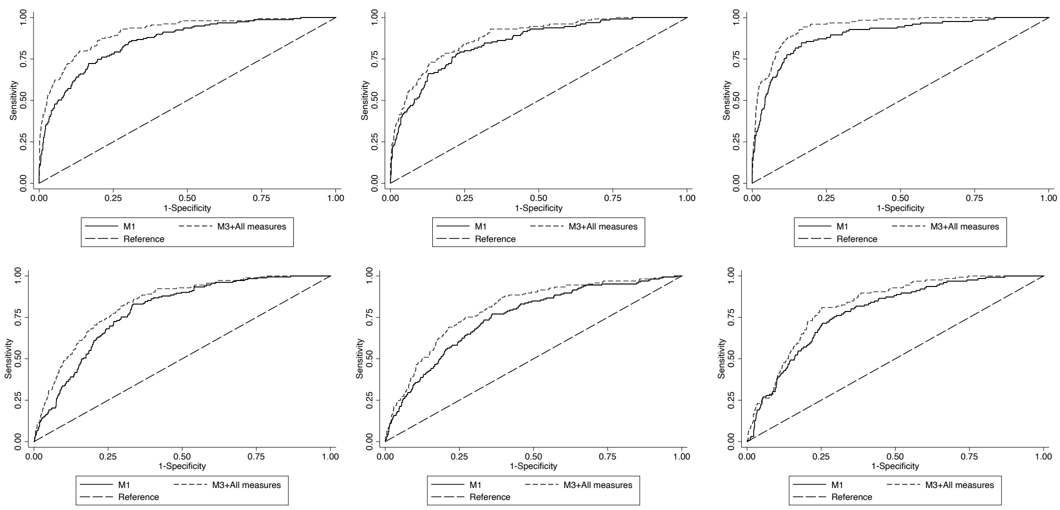
Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Table E6.* Classification accuracy of upgrades, based on the logit model with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable, including the outlook as one of the explanatory variables and following the through-the-cycle approach, using the 0.5 cutoff, by credit rating agency

| | S&P | | | Fitch | | | Moody's | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC | Sensitivity (%) | Overall (%) | AUC ROC |
| | | | | Model 1 | | | | | |
| No textual indicators | 50.898 | 91.301 | 0.904 | 34.247 | 90.216 | 0.891 | 61.682 | 92.201 | 0.922 |
| Net sentiment (W, dict) | 52.469 | 90.758 | 0.910 | 39.726 | 90.492 | 0.900 | 62.617 | 92.114 | 0.925 |
| Polarity (W, dict) | 53.086 | 90.758 | 0.910 | 37.671 | 90.158 | 0.900 | 63.551 | 92.288 | 0.926 |
| Polarity (S, ML) | 53.704 | 91.923 | 0.911 | 41.096 | 90.575 | 0.907 | 67.290 | 93.154 | 0.930 |
| Subjectivity (W, dict) | 50.617 | 91.174 | 0.912 | 35.616 | 89.992 | 0.894 | 65.421 | 92.721 | 0.922 |
| Subjectivity (S, dict) | 53.086 | 91.424 | 0.909 | 34.932 | 90.075 | 0.895 | 60.748 | 92.201 | 0.922 |
| Subjectivity (S, ML) | 51.852 | 91.257 | 0.906 | 33.562 | 89.908 | 0.896 | 63.551 | 92.114 | 0.924 |
| All textual indicators | 55.556 | 91.590 | 0.918 | 47.260 | 91.326 | 0.921 | 71.028 | 92.894 | 0.936 |
| | | | | Model 2 | | | | | |
| No textual indicators | 48.503 | 90.742 | 0.911 | 37.671 | 90.464 | 0.900 | 65.421 | 92.721 | 0.928 |
| Net sentiment (W, dict) | 53.086 | 91.257 | 0.917 | 39.726 | 90.409 | 0.906 | 67.290 | 92.721 | 0.930 |
| Polarity (W, dict) | 50.617 | 91.007 | 0.917 | 39.041 | 90.325 | 0.905 | 66.355 | 92.808 | 0.930 |
| Polarity (S, ML) | 53.704 | 91.424 | 0.917 | 41.781 | 90.742 | 0.914 | 68.224 | 92.808 | 0.935 |
| Subjectivity (W, dict) | 51.235 | 90.841 | 0.918 | 36.986 | 90.158 | 0.904 | 65.421 | 92.461 | 0.928 |
| Subjectivity (S, dict) | 52.469 | 91.007 | 0.915 | 37.671 | 90.075 | 0.904 | 65.421 | 92.634 | 0.928 |
| Subjectivity (S, ML) | 50.617 | 91.007 | 0.914 | 35.616 | 89.825 | 0.904 | 66.355 | 92.894 | 0.932 |
| All textual indicators | 59.259 | 91.757 | 0.924 | 49.315 | 91.576 | 0.928 | 71.963 | 93.414 | 0.944 |
| | | | | Model 3 | | | | | |
| No textual indicators | 53.012 | 91.357 | 0.917 | 37.671 | 89.791 | 0.907 | 62.857 | 92.281 | 0.932 |
| Net sentiment (W, dict) | 57.143 | 91.821 | 0.924 | 43.151 | 90.741 | 0.912 | 66.667 | 92.456 | 0.933 |
| Polarity (W, dict) | 55.280 | 91.653 | 0.924 | 41.781 | 90.657 | 0.912 | 63.810 | 92.544 | 0.933 |
| Polarity (S, ML) | 57.143 | 91.906 | 0.923 | 45.890 | 90.909 | 0.919 | 70.476 | 93.158 | 0.938 |
| Subjectivity (W, dict) | 52.795 | 90.978 | 0.924 | 41.781 | 90.741 | 0.910 | 63.810 | 92.018 | 0.932 |
| Subjectivity (S, dict) | 54.658 | 91.315 | 0.921 | 39.041 | 90.152 | 0.910 | 64.762 | 92.193 | 0.932 |
| Subjectivity (S, ML) | 53.416 | 91.400 | 0.920 | 37.671 | 89.899 | 0.910 | 66.667 | 92.895 | 0.936 |
| All textual indicators | 58.385 | 91.653 | 0.930 | 49.315 | 91.330 | 0.930 | 72.381 | 93.246 | 0.949 |

Sensitivity (%) = true positive rate, percent of correctly classified true positives
Overall (%) = overall classification accuracy, percent correctly classified
AUC ROC = area under ROC (receiver operating characteristic) curve

*Figure E1.* ROC curve comparison of the worst and best performing models, based on the logit model with the downgrade dummy (equal to 1 if a downgrade occurs, 0 otherwise) as the dependent variable (top), and with the upgrade dummy (equal to 1 if a upgrade occurs, 0 otherwise) as the dependent variable (bottom), including the outlook as one of the explanatory variables and following the point-in-time approach, by credit rating agency i.e. Standard & Poor's (left), Fitch (center) and Moody's (right)



## Appendix F: Chapter 6 - Model 4 results

*Table F1.* Analysis of the effect of financial variables and economic development on cumulative abnormal returns in the time interval (-1,1) during rating announcements for the ‘first mover’ sample

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.432** | 0.055 | -0.382 | -0.524** |
| | (0.199) | (0.101) | (0.278) | (0.213) |
| Exchange rate | 0.970 | 0.785* | 0.304 | -0.530 |
| | (0.607) | (0.405) | (0.682) | (0.585) |
| US excess return | 0.058 | -0.176*** | 0.089 | -0.165* |
| | (0.058 ) | (0.058) | (0.066) | (0.085) |
| 5Y CMT rate | -0.180* | -0.049 | -0.291** | -0.139* |
| | (0.107) | (0.063) | (0.136) | (0.073) |
| Volatility risk premium | 0.012 | 0.060*** | 0.049 | 0.040 |
| | (0.030) | (0.023) | (0.044) | (0.025) |
| IG to SG dummy | 0.007 | | 0.010 | |
| | (0.014) | | (0.015) | |
| SG to IG dummy | | -0.028*** | | -0.026*** |
| | | (0.008) | | (0.009) |
| Economic development | 0.000 | -0.002 | -0.001 | -0.004 |
| | (0.007) | (0.004) | (0.012) | (0.006) |
| Constant | 0.019*** | -0.001 | 0.022** | -0.002 |
| | (0.005) | (0.003) | (0.010) | (0.004) |
| Observations | 535 | 461 | 312 | 209 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table F2.* Analysis of the effect of financial variables and economic development on cumulative abnormal returns in the time interval (-1,1) during rating announcements for S&P

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.635** | 0.127*** | -0.663 | -0.615* |
| | (0.307) | (0.042) | (0.415) | (0.328) |
| Exchange rate | 0.002 | 0.061 | -0.643 | -0.572 |
| | (0.835) | (0.626) | (0.885) | (0.883) |
| US excess return | 0.061 | -0.226** | 0.151 | -0.272** |
| | (0.134) | (0.092) | (0.127) | (0.127) |
| 5Y CMT rate | -0.206 | -0.004 | -0.315 | -0.116 |
| | (0.172) | (0.097) | (0.250) | (0.101) |
| Volatility risk premium | 0.010 | 0.059** | 0.003 | 0.054** |
| | (0.061) | (0.026) | (0.073) | (0.024) |
| IG to SG dummy | 0.005 | | -0.003 | |
| | (0.028) | | (0.029) | |
| SG to IG dummy | | -0.045*** | | -0.051*** |
| | | (0.016) | | (0.019) |
| Economic development | -0.015 | 0.001 | -0.027 | 0.002 |
| | (0.015) | (0.006) | (0.028) | (0.009) |
| Constant | 0.035*** | -0.006 | 0.048* | -0.010 |
| | (0.012) | (0.004) | (0.025) | (0.007) |
| Observations | 148 | 162 | 91 | 79 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table F3.* Analysis of the effect of financial variables and economic development on cumulative abnormal returns in the time interval (-1,1) during rating announcements for Fitch

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | -0.424 | -0.758** | -1.068 | -0.377 |
| | (0.673) | (0.292) | (0.722) | (0.257) |
| Exchange rate | -0.317 | 1.208* | 1.010 | 1.608** |
| | (1.087) | (0.726) | (2.028) | (0.681) |
| US excess return | 0.087 | -0.111 | 0.112 | -0.179 |
| | (0.102) | (0.157) | (0.185) | (0.130) |
| 5Y CMT rate | 0.027 | -0.257* | -0.454 | -0.119 |
| | (0.175) | (0.135) | (0.422) | (0.129) |
| Volatility risk premium | -0.022 | 0.032 | 0.053 | -0.042 |
| | (0.053) | (0.041) | (0.089) | (0.048) |
| IG to SG dummy | 0.017 | | 0.030 | |
| | (0.032) | | (0.031) | |
| SG to IG dummy | | -0.005 | | -0.004 |
| | | (0.010) | | (0.009) |
| Economic development | 0.003 | 0.000 | 0.026 | 0.001 |
| | (0.013) | (0.007) | (0.025) | (0.009) |
| Constant | 0.011 | 0.002 | -0.007 | -0.003 |
| | (0.011) | (0.005) | (0.018) | (0.006) |
| Observations | 122 | 119 | 67 | 55 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Table F4.* Analysis of the effect of financial variables and economic development on cumulative abnormal returns in the time interval (-1,1) during rating announcements for Moody's

| | Negative event | Positive event | Downgrade | Upgrade |
|---|---|---|---|---|
| Local stock market | 0.162 | -0.314 | 0.272 | -0.322 |
| | (0.565) | (0.445) | (1.007) | (0.810) |
| Exchange rate | 3.126** | 0.485 | 6.456* | -0.253 |
| | (1.541) | (0.752) | (3.773) | (1.581) |
| US excess return | -0.252 | 0.013 | -0.268 | 0.029 |
| | (0.193) | (0.110) | (0.269) | (0.215) |
| 5Y CMT rate | -0.038 | 0.022 | 0.210 | -0.264 |
| | (0.317) | (0.120) | (0.539) | (0.292) |
| Volatility risk premium | 0.067 | 0.054 | 0.139 | 0.007 |
| | (0.076) | (0.047) | (0.144) | (0.096) |
| IG to SG dummy | 0.051 | | 0.040 | |
| | (0.041) | | (0.042) | |
| SG to IG dummy | | -0.017 | | -0.026 |
| | | (0.014) | | (0.016) |
| Economic development | 0.027* | -0.003 | 0.033 | 0.002 |
| | (0.015) | (0.008) | (0.021) | (0.015) |
| Constant | 0.002 | 0.003 | -0.001 | 0.005 |
| | (0.007) | (0.005) | (0.016) | (0.010) |
| Observations | 125 | 120 | 69 | 50 |

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$