Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology
Management

Bø, Jørgen Frost
Lian, Magnus Lysholm
Smeby, Karl Magnus

The Impact of Political Factors on Bank CDS Spreads

A Data-Driven Approach

Master's thesis in MTIOT
Supervisor: Maria Lavrutich & Stavros Zenios
June 2020

NTNU
Norwegian University of
Science and Technology

Bø, Jørgen Frost
Lian, Magnus Lysholm
Smeby, Karl Magnus

The Impact of Political Factors on Bank CDS Spreads

A Data-Driven Approach

Master's thesis in MTIOT
Supervisor: Maria Lavrutich & Stavros Zenios
June 2020

Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

NTNU
Norwegian University of
Science and Technology

Abstract

This thesis empirically analyzes the determinants of CDS spreads from a global sample of 46 listed banks over the 2005–2019 period. We use traditional accounting- and market-based variables, in addition to two novel political and policy variables as well as a news sentiment variable. We apply a data-driven approach to variable selection in order to identify redundancies in existing literature. Using a panel fixed effects approach, we find that (1) political stability and policy uncertainty are important drivers of bank credit risk, (2) news sentiment is found to be important *in addition* to political and policy variables, (3) market variables are overall more important in explaining bank CDS spreads than accounting variables, (4) variable selection methods show that there are redundancies in the set of traditional variables found to be significant in the existing literature, and (5) by using a data-driven approach to variable selection on all of the available variables, we obtain simpler models with higher explanatory power.

Sammendrag

I denne oppgaven analyserer vi empirisk hvilke variabler som er viktige for *CDSene* til et globalt utvalg av 46 børsnoterte banker i perioden fra 2005 til 2019. Vi studerer tradisjonelle regnskaps- og markedsvariabler, i tillegg til to originale variabler som måler politisk risiko og risiko ved politiske retningslinjer, og én nyhetssentimentvariabel. Vi bruker en datadrevet tilnærming til variabelseleksjon for å identifisere overflødige variabler i eksisterende litteratur. Ved å bruke et paneldata med fikserte enhetseffekter finner vi at; (1) politisk stabilitet og usikkerhet ved politiske retningslinjer er viktige drivere av bankers kredittrisiko, (2) nyhetssentiment er viktig *i tillegg til* de politiske variablene, (3) markedsbaserte variabler er viktigere i å forklare CDSer enn regnskapsvariabler, (4) variabelselsksjonsmetoder viser at det er overflødigheter i settet av tradisjonelle variabler funnet til å være viktig i eksisterende litteratur, og (5) ved å bruke en datadrevet tilnærming til variabelselsksjon på alle tilgjengelige variabler, oppnår vi enklere modeller med bedre forklaringskraft.

Preface

This thesis concludes our Master of Science degree in Industrial Economics and Technology Management at the Norwegian University of Science and Technology (NTNU). It is original and independent work by Jørgen Frost Bø, Magnus Lysholm Lian, and Karl Magnus Smeby, written during the spring of 2020.

We would like to thank our supervisors, Associate Professor Maria Lavrutich at the Department of Industrial Economics and Technology Management (NTNU) and Professor Stavros Zenios at the Accounting and Finance Department (University of Cyprus), for helpful guidance, inspiration and advice. Their interest in our work has been truly valuable during the completion of our master's thesis. We also appreciate the help from Professor Sjur Westgaard at the Department of Industrial Economics and Technology Management (NTNU) for contributing to laying the foundation for this thesis. A thankful note is also directed to Giovanni Pagliardi at the Finance Faculty at BI Norwegian Business School, for beneficial collaboration and feedback. Moreover, we would like to thank Morten Risstad in Sprebank 1 Markets for providing data, guidance and valuable insights on the banking industry. Our thesis and analyses have benefited considerably from his involvement, and we are grateful for his advice.

Jørgen Frost Bø
Magnus Lysholm Lian
Karl Magnus Smeby

Trondheim, June 11, 2020

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Both the global Financial Crisis in 2007-2008 and the subsequent European Debt Crisis resulted in reduced economic performance and activity across the entire economy, greatly affecting people's everyday life. Due to systemic banking crises remaining relatively frequent and costly, several studies examining bank credit risk have been encouraged (Drago et al., 2017).

According to GlobalData, the global banking industry was ranked the third largest industry by revenue and the largest industry by profits in 2019 (GlobalData, 2019). Banks are the main provider of credit and therefore facilitate new investments contributing to economic growth. They have important societal tasks, connecting savers with borrowers and facilitating payments and transactions, and are part of a larger financial system of groundskeepers, including governments, central banks, regulators, and supervisors. These institutions try to ensure that banks operate efficiently and safely, and in the interests of the broader society. Banks' pivotal role in society outlines our main motivation for studying credit risk in the global banking industry.

In this thesis, we study the determinants of banks' credit risk. We measure this credit risk through *Credit Default Swaps* (CDSs), which is a financial instrument that insures the buyer against a default or credit event in the underlying firm. CDSs are acknowledged as the preferred measure of credit risk in the literature (Augustin et al., 2015). Among others, Ericsson et al. (2009) argue that CDSs are more liquid and provide a purer measure of credit risk than alternatives.

The existing literature on determinants of bank CDS spreads primarily focuses on the impact of traditional financial variables, mainly accounting and market variables. However, there is no general consensus on the impact of several of these. In particular, results vary on the sign and impact of bank size, the yield curve and the Fama-French factors.

By including a comprehensive set of financial variables and using a data-driven approach to variable selection, we investigate if there are redundancies in the set of variables previously found to impact bank CDS spreads. Furthermore, using novel political factors, we analyze the impact of political (in)stability and policy uncertainty on bank CDS spreads.

In doing so, we contribute to the existing literature in two ways. First, we conduct a comprehensive literature review, identifying the variables shown to impact the CDS spreads in previous studies. The classical approach to selecting variables has been based on theory, paradigms or the researcher's own hypotheses. To our knowledge, no prior work uses a data-driven approach to selecting variables in research related to bank CDS spreads.

Secondly, a growing body of empirical research suggests that political and policy variables are important determinants of several financial variables in other asset classes (Pástor and Veronesi, 2013; Dai and Zhang, 2019). Periods that feature political (in)stability and policy uncertainty may increase investors' risk perception in the banking industry due to potential changes in the macro or regulatory environment. However, no existing research has aimed to capture the effect of political risk on *bank* credit risk, or made an explicit separation between political stability and policy uncertainty.

Political and policy topics are typically important in the coverage of financial news. Media may provide early warnings of a deteriorating credit situation, and news can influence the beliefs of market participants and induce investors to withdraw funds from financial markets (Hillert et al., 2012). We address this in our thesis by analyzing if news sentiment affects the impact of the political and policy variables.

We run a total of four models and two robustness tests. First, we build a baseline model consisting of the variables that existing literature has found important. Secondly, we apply data-driven variable selection techniques, the Least Absolute Shrinkage and Selection Operator (LASSO) and the Stepwise Forward Floating Selection (SFFS), to the baseline model in order to optimize the set of variables. Thirdly, we add the novel political and policy variables and study their impact on the bank CDS spreads. Fourthly, we add the news sentiment variable. Finally, we run two robustness tests, one by applying variable selection techniques to all variables, and one by testing the in-sample robustness.

Our main findings can be summarized as follows. First, political stability and policy uncertainty are important drivers of bank credit risk and capture additional variance of bank CDS spreads. This has, to the best of our knowledge, not yet been shown in existing studies. Secondly, news sentiment variables are found to be important *in addition* to the political and policy variables. Thirdly, we find that market variables are overall more important in explaining bank CDS spreads than accounting variables, implying that the general market conditions are very important to assess. The penultimate finding is that variable selection methods show that there are redundancies in the set of traditional variables found significant in the existing literature. Finally, using a data-driven approach to variable selection on all of the available variables, we remove these redundancies and obtain simpler models with higher explanatory power.

The rest of the thesis is structured as follows; Chapter 2 and Chapter 3 provide an introduction to the banking sector and the CDS market, respectively. The former focuses on banks' role in society, regulatory environment, particular asset structure, and distinct risk factors. All these are arguments to why CDS spreads on banks should be studied separately from other firms. The latter explains the dynamics of the CDS market and why CDS spreads are the preferred measure of credit risk.

A thorough literature review on the determinants of bank CDS spreads is presented in Chapter 4. Prior work on political, policy and news sentiment variables in relation to financial markets is also presented.

In Chapter 5 we present our data set. We describe how the variables are collected and constructed, as well as stating hypotheses on their impact on the CDS spreads.

Chapter 6 lays out our choice of model and modelling techniques. We explain the use of panel data with the fixed effects approach and outline how the chosen variable selection techniques work. Results and discussions of our models are presented in Chapter 7, while Chapter 8 concludes.

Chapter 2

Banks - Impact and Distinction From Other Firms

The focus of this thesis is the analysis of the determinants of bank CDS spreads. We study banks separately from other corporates because of their role in society, regulatory environment, particular asset structure, and distinct risk factors. In what follows, we discuss these particularities in more detail.

2.1 How Banks are Different

The importance of studying banks specifically stems from their fundamentally important role in society. Globally, banks are regulated in different ways depending on which financial system they comply to. Still, all banks have some fundamental similarities that are prominent no matter the regulations. They,

1. Connect savers with borrowers

2. Facilitate transactions

First, banks engage in financial intermediation and help society grow. By raising deposits from households and companies, banks turn these funds into credit by providing loans to customers. The major part of bank loans are provided to non-financial corporations and households. This way, banks have an important task in facilitating new investments which in turn contribute to economic growth. It is therefore essential that banks are able to operate and lend money in both economically stable and unstable times.

Second, banks facilitate transactions and make payments safe and swift. To be able to make everyday payments, easy access to savings is required. The lack of such system would have severe consequences, as trivial everyday activities and services become complicated to complete.

Another motivation for studying banks separately is related to the regulatory environment and exposure to distinct risk factors in their line of business. Three of the most important risk factors banks face are:

1. **Credit risk**. The risk that borrowers will not repay their debt

2. **Regulatory risk**. The risk that authorities will change the regulatory framework that banks must comply with

3. **Liquidity risk**. The risk that a bank is unable to meet its short-term financial obligations

Banks are more prone to the first two risk factors than other firms. Since banks are engaged in lending activities, they hold a credit risk on their customers. The credit risk arises from the possibility that the customer at a later point in time will default, resulting in a loan loss that reduces the value of the bank's assets. Loan losses may arise from unexpected economic developments that significantly reduce the financial strength of borrowers, or simply from poor credit risk management in the bank.

Due to their importance on the economy and society, banks meet stringent requirements from authorities. These requirements may be changed depending on the economic situation, creating a regulatory risk for banks. Among other things, there are requirements for how much capital banks should have. If a bank has significant loan losses, its capital ratio will decline and may fall below the regulatory requirement. Failing to meet regulatory requirements may result in fines or other sanctions from the authorities.

Liquidity risk is as important for banks as other companies. However, due to the important societal roles of banks, authorities are more concerned with bank liquidity risk than other firms' liquidity. Therefore banks, unlike other firms, receive liquidity requirements from authorities.

Since these risk factors may affect the bank's probability of going bankrupt, determinants of bank credit risk should be studied separately from other firms. Most research on firm's default risk determinants exclude banks from the empirical investigations. According to Sclip et al. (2019), the reason for this is that the asset structure of banks is very different from other corporations.

Also Raunig and Scheicher (2009) argue that banks differ in a number of characteristics from other firms. They highlight that the composition of banks' balance sheets, bank's central functions in the economy, and their regulatory environment set them apart from other firms. They also find empirical evidence that banks' credit risk behave differently than that of other firms, i.e. a different set of variables are significant in explaining them. Due to their central role in the economy, banks, unlike other corporates, have historically been bailed-out by authorities when approaching bankruptcy. This is done so they can withhold their day-to-day activities and facilitate economic activity. Recent examples of bank bail-outs include the Financial Crisis and the Euro Crisis.

2.2 Banking Crises: Two Examples From Recent History

The Financial Crisis of 2007-2008 and the subsequent European Debt Crisis stress the key role banks play for the financial system and global economy. The impact of these crises outlines the importance of understanding risks in the banking industry. In the following, we describe both crises in brief.

2.2.1 The Financial Crisis of 2007-2008

Deregulation of the financial markets had allowed large US financial institutions to issue mortgage-backed debt with poor collateral and low credit quality. Among the buyers of this debt were other US and European banks, who deemed these products attractive

(Ramskogler, 2015). In order to continue meeting the demand, more mortgages were needed, spurring the issuance of mortgages to subprime borrowers. It became evident that the loan packages suffered from very poor credit quality, i.e. the *credit risk* was high, and that the banks that owned these loans would not be repaid. In combination with high leverage levels among banks, confidence in the financial markets dropped, and it became difficult to obtain new financing, even for the most solid banks. With few or no sources of funding, some banks were unable to meet their obligations and therefore went bankrupt.

Prior to the Financial Crisis, European banks had been through a longer period of consolidation. Cross-border M&A activities among European banks were particularly high around year 2000, both within Europe, but also in the US. Lack of supervision and a consistent regulatory framework made it easy for European banks to take on increased leverage and expand their business. By 2008, European banks had become more global and interconnected. However, the high leverage meant they held less capital to deal with potential future loan losses.

By the time the Financial Crisis began in the US, the European banks were also heavily exposed. Employment and financial markets in developed countries globally dropped sharply and global productivity reduced significantly (Kouki et al., 2017; Eichhorst et al., 2010). In addition, bankruptcies spiked, also for non-financial firms (Blinder, 2013).

2.2.2 The European Debt Crisis of 2009-2014

While many American banks rebounded after the Financial Crisis, the European banking industry went into a long and deep debt crisis. The signing of the Maastricht treaty in 1992, obliged the EU countries to limit their budget deficits to 3% of GDP and hold low debt levels. However, according to Eurostat, the average budget deficit in Greece, Ireland, UK, Italy, Portugal and Spain increased dramatically to a 11.3% average in 2009, while their public debt to GDP averaged 86%, see Figure 2.1.

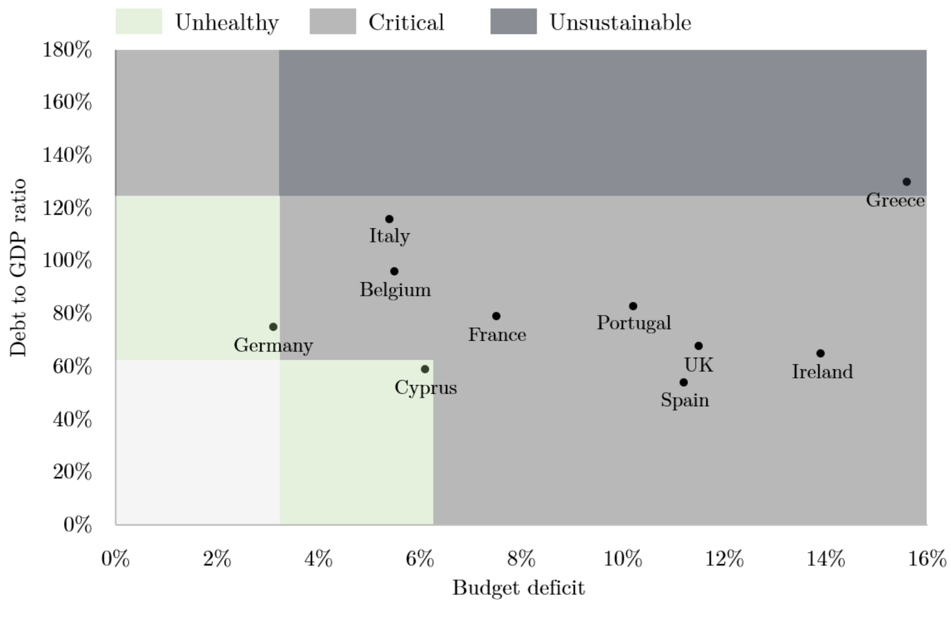

Figure 2.1: Overview of debt ratios and budget deficits, EU countries, year-end 2009 (Eurostat)

Moreover, in the aftermath of the Financial Crisis there were few prospects of economic growth in the EU area. This made it harder for governments to pay off, or refinance, their debt as public inflows were reduced. Therefore, 2010 saw the "Troika" (The ECB, IMF and European Commission) organizing a 100bn EUR bailout of Greece and a 85bn EUR bailout of Ireland, whilst in 2011 Portugal got a 78bn EUR bailout package. In 2012, Spain also received a relief package (Copelovitch et al., 2016).

Throughout these years, several European banks needed public aid to survive. One of the reasons banks lost capital in the period was that they were large owners of sovereign debt, which in many cases decreased in value during the European Debt Crisis period, as can be seen from Figure 2.2 (Evans et al., 2008). They also suffered loan losses in the aftermath of the Financial Crisis. Seeing as these banks were systematically important, many of them received help from government. A total of 114 European banks were aided by governments in the 2007-2014 period (Gerhardt and Vennet, 2017).

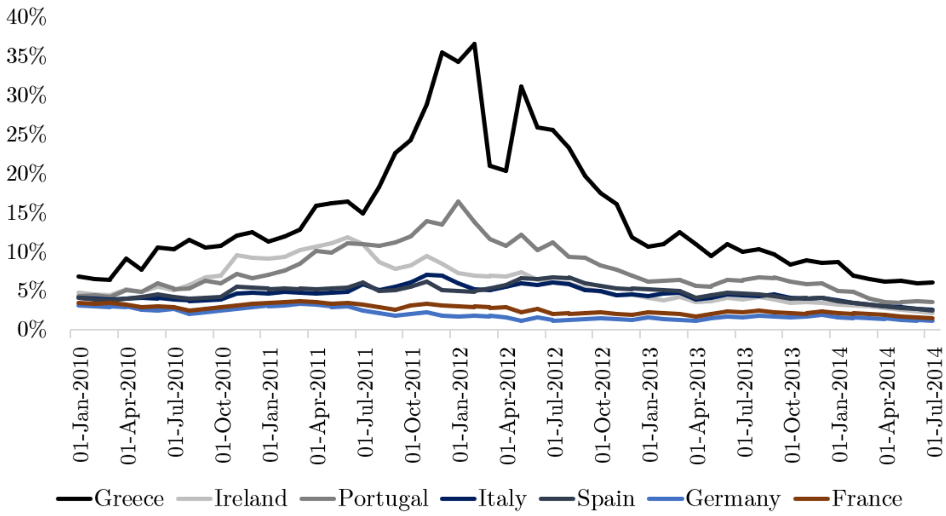

Figure 2.2: Overview of the yield on 10-year government bonds for selected European countries (Datastream)

Also the European Debt Crisis showed the impact a failing banking system has on the economy and society as a whole, as well as the costs needed to restore the system. To prevent banks from needing bailouts to the large extent seen during these crises, stricter regulations on banks were imposed by the regulators. The EU Basel III and the European Banking Union are examples of two resulting products. We refer to Appendix A for further details.

Chapter 3

Credit Default Swaps

In our thesis, we opt to use credit default swaps (CDSs) as our measure of credit risk, and in the following we argue why they are the preferred measure compared to other metrics. In addition, we give a brief introduction to the global CDS market.

3.1 Comparison of Credit Risk Measures

CDSs work by transferring credit risk between two parties, the insurance *buyer* and the insurance *seller*. The insurance buyer pays for insurance on some fixed amount. In order to incentivize the insurance seller to sell insurance on this amount, the buyer has to make regular payments to the seller. The size of this payment varies according to the credit risk of the underlying company. The annualized payment to the insured amount ratio is called the *CDS spread*. The CDS spread fluctuates as a result of changes in probability that the reference entity will experience a credit event. In case of a contractually defined credit event, the buyer will receive the insured amount as payoff. The CDS spread therefore reflects the credit risk of the entity in question. According to Augustin et al. (2015) CDSs are a widely used measure of credit risk in the literature.

An alternative metric of credit risk is *credit ratings*. Credit ratings provide information on the creditworthiness of the issuer and are made by credit rating agencies (CRAs). There are three main global CRAs; Moody's, Standard & Poor's (S&P) and Fitch. A disadvantage of using credit ratings compared to CDS spreads is that rating changes are rare and most often associated with the release of quarterly statements, whereas CDSs are traded daily and their prices are continuously updated (Ericsson et al., 2009). Moreover, credit ratings are often hard to obtain for many companies as the universe covered by CRAs is limited (Ericsson et al., 2009). Credit ratings have also received critique for not fully reflecting the true credit risks in underlying entities (Hilscher and Wilson, 2016). Lastly, CRAs receive fees from the companies they cover and are therefore incentivized to give more optimistic ratings (Morkoetter et al., 2017; Park and Lee, 2018).

A third measure of credit risk is *bond credit spreads*. Banks issue bonds as a source of funding. The yield on the bond is the risk premium debt investors require in order to buy the bond. Ericsson et al. (2009) identifies several advantages of using CDSs as a measure of credit risk, compared to bond yields. First, bond yields include factors not related to credit risk, such as systematic risk unrelated to default and illiquidity (Elton et al., 2001; Longstaff et al., 2005). Huang and Huang (2012) conclude that less than 25 percent of the credit spread of corporate bonds is attributable to credit risk. As a consequence of the inclusion of non-default components, the changes in underlying credit risk take longer time to be incorporated in the bond spreads (Ericsson et al., 2009). The faster speed of diffusion to CDS spreads compared to bond yields, is supported by Blanco et al. (2005).

Secondly, CDS markets are more liquid than that of bonds. CDSs usually trade on accessible platforms in contrast to bonds which typically trade through investment bank brokers. Trading in bonds is less frequent than that of comparable CDSs. As a consequence, studies focusing on CDS data often use daily frequency, whereas studies on bonds or credit ratings are mainly conducted on data of lower granularity. Moreover, the trading of CDSs does not require the underlying security to issue bonds, whereas using bond yields as a measure of credit risk necessarily involves the issuance of bonds by the underlying entity.

Thirdly, the bond market has higher trading friction, though brokerage costs and high bid-ask spreads, compared to the CDS market (Oehmke and Zawadowski, 2016). Also, CDSs have the advantage that they, to a greater extent, are standardized with constant maturity, whereas uniform bond yields can only be obtained by interpolating bond yields of different maturities (Avino et al., 2019; Blanco et al., 2005). Lastly, many corporate bonds have embedded options, further complicating the measurement of credit risk based on corporate bond yields (Yongjun Tang and Yan, 2008).

3.2 The CDS market

CDSs were first introduced in 1994 by the J.P. Morgan Inc. in order to transfer credit risk from their balance sheet. CDSs have gained widespread attention the last 10 years, largely because of their role during the financial crisis (Augustin et al., 2015).

CDSs are often bought for hedging purposes, most commonly a bond holder will buy a CDS on the same company to reduce the credit risk on the bond. If the insurance buyer does not own bonds in the company for which he buys the CDS, the CDS position is called *naked*.

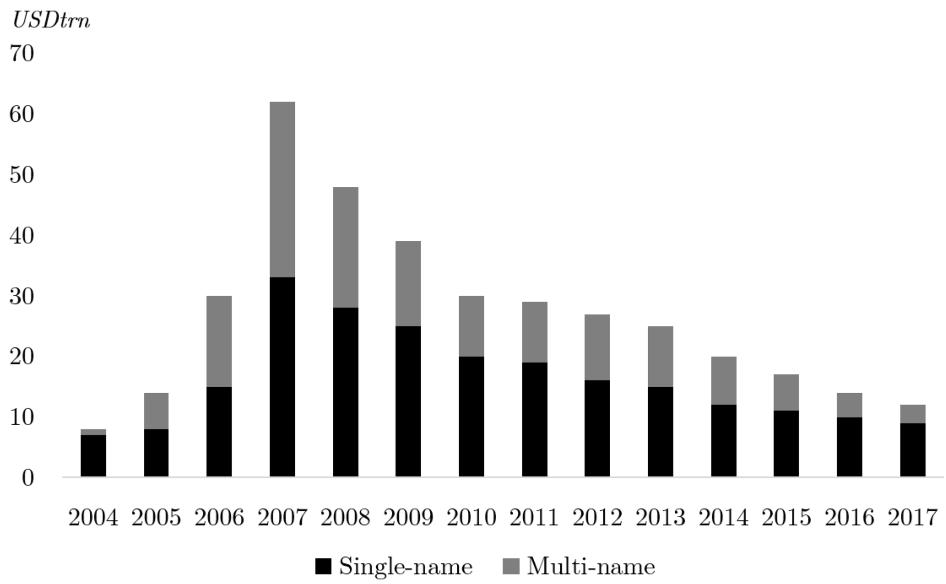

Figure 3.1: Total notional amounts outstanding, 2004-2017 (Bank for International Settlements, 2018).

The market size for CDSs reached its peak in 2007/2008, when the total notional amount outstanding was approximately USD 60 trillion. As can be seen from Figure 3.1, the market has been shrinking continuously since then and in 2017 the total notional amount outstanding was approximately USD 10 trillion. Over 70% of the total outstanding CDS amount is related to CDSs with 1-5 years maturity. The most common maturity is 5 years (Ericsson et al., 2009).

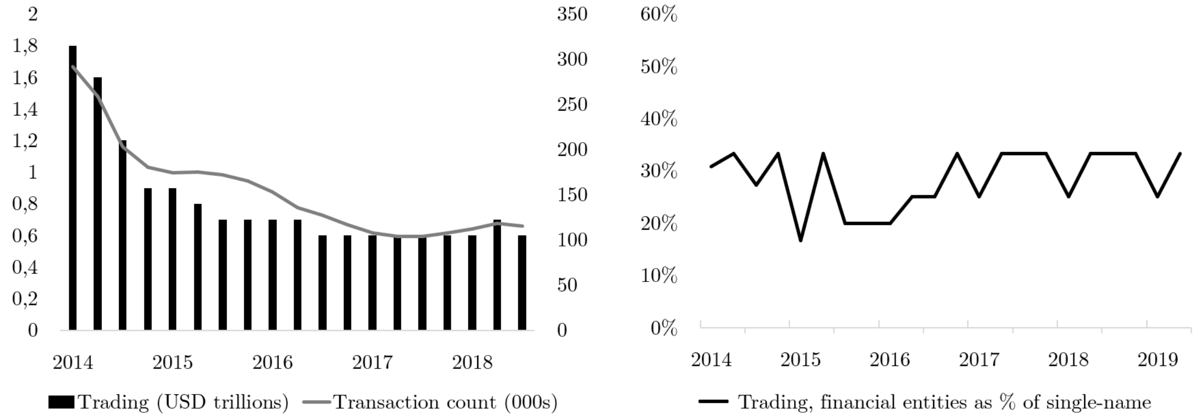

Figure 3.2: LHS: Global trading amount (left axis) and frequency (right axis) on single name entities. RHS: Global trading amount on financial entities as percentage of total trading in corporate single name entities (ISDA, 2019).

CDSs are either *single-name*, meaning that they insure a single entity (company or nation), or *index*, meaning that they insure a collection of different entities.

According to the International Swaps and Derivatives Association, the trading activity in single-name CDSs amounted to approximately USD 0.6 trillion during Q3 2018 (see Figure 3.2), spread across approximately 115 000 trades (ISDA, 2019). In comparison, NASDAQ trading volumes are at around USD 0.2 trillion *daily*. The global CDS market is therefore a lot less liquid than that of equities.

The total number of single-name entities for which there were recorded transactions in Q2 2019, was 799. The top 100 single-name entities account for 69% of total single-name volumes. From 2014 to 2019, 27 entities were consistently among the top 100 most

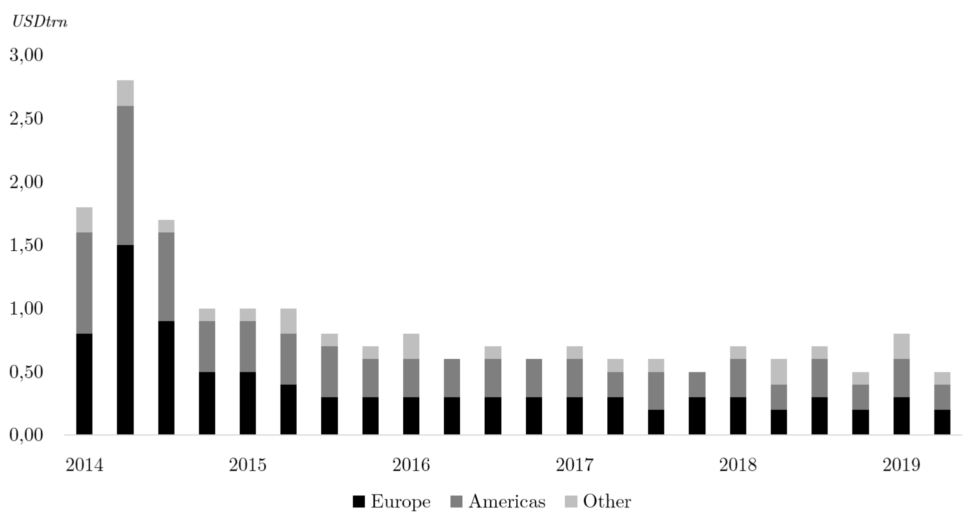

Figure 3.3: Trading amount on single name entities, split by geography of underlying entity (ISDA, 2019).

traded entities. Of these 27, 10 were banks and 13 were sovereign CDSs. As can be seen from Figure 3.2, of total trading in single-name entities, the financial sector accounts for approximately 30%, as of Q2 2019.

Furthermore, as of Q2 2019, Figure 3.3 shows that Europe and the Americas accounted for 80% of trading in single-name entities. According to Nikkei (2016), China opened for CDS trading in 2016, and therefore not many studies on determinants of bank CDS spreads have included Chinese banks.

Chapter 4

Literature Review

Our thesis is closely related to the growing body of research on the determinants of bank CDS spreads. In this chapter, we present a thorough review of the existing literature. Additionally, since we study the impact of political stability and policy uncertainty, previous work related to these variables' impact on financial markets is reviewed.

4.1 Previous Research on Bank CDS Spread Determinants

In total, we have identified 11 studies in the existing literature that investigate the determinants of bank CDS spreads. An overview of these studies is presented in Table 4.1.

Table 4.1: Overview of the previous literature on CDS determinants on banks studied in this thesis.

Author	**Raunig and Scheicher (2009)**	**Annaert et al (2013)**	**Chiaramonte and Casu (2013)**	**Hasan et al. (2016)**	**Samaniego-Medina et al. (2016)**	**Smales (2016)**
#Banks	41	32	57	161	45	10
Period	2003-2007	2004-2010	2005-2011	2001-2011	2004-2010	2004-2010
Frequency	Monthly	Weekly	Quarterly	Annual	Annual	Daily
Levels/returns	Levels	Returns	Levels	Levels	Levels	Returns
Geography	Europe/US	Europe	Global	Global	Europe	Europe/US
Model	Panel regression	Panel regression	Panel regression	Panel regression	Panel regression	Panel regression

Author	**Drago et al. (2017)**	**Benbouzid et al. (2017)**	**Koutmos (2018)**	**Guesmi et al. (2018)**	**Sclip et al. (2019)**	
#Banks	63	26	20	Single banking index	38	
Period	2007-2016	2004-2011	2002-2017	2007-2016	2005-2015	
Frequency	Quarterly	Annual	Daily	Weekly	Quarterly	
Levels/returns	Levels	Levels	Returns	Both	Returns	
Geography	Europe/US	Global	Europe/US	US	Europe	
Model	Panel regression	Panel regression	Quantile regression	NARDL	Panel regression	

The studies on CDS spreads have various geographical focus, however most include either European or US banks, or both. In our study, we expand the geographical focus to include banks from Europe, US, Canada and Australia. We use semi-annual CDS spreads over a 14 year period from 2005-2019. The frequency of our CDS data is in line with that of the existing literature, which ranges from daily to annual. In general, papers that exclude accounting variables tend to use higher frequency CDS data. This is due to the fact that market variables are accessible at higher frequency (typically daily), while accounting variables can only be collected quarterly.

The results of the previous studies on bank CDS spread determinants is presented in Table 4.2. Each row represents a variable[1]. In the table, "+" represents a significant positive variable coefficient, "-" a significant negative coefficient and "0" means the variable was not found significant in the article[2]. Finally, the rightmost column summarizes the findings of all the articles on each variable.

Table 4.2: Overview of variables included in literature. Each column represent one article. The variables are grouped by firm-specific variables (both accounting and market), market variables and Fama-French variables.

Variable	Raunig and Scheicher (2009)	Annaert et al. (2013)	Chiaramonte and Casu (2013)	Hasan et al. (2016)	Samaniego-Medina et al. (2016)	Smales (2016)	Drago et al. (2017)	Benbouzid et al. (2017)	Koutmos (2018)	Guesmi et al. (2018)	Sclip et al. (2019)	*Summary*
Firm-specific												
Accounting												
Leverage			+	+	-	0	+	0			+	-/+/0
Efficiency				0	0			+				+/0
Profitability			-	-	0			-			-	-/0
Size					+	0	-				-	-/+/0
Asset quality			-	-	-		-	-			-	-
Funding stability							0				0	-/0
Liquidity			-	-	-			0				-/0
Income diversification											+	+
Market												
Stock return		-			-	-			-			-
Stock volatility	+	0		+	+	+			0		+	+/0
CDS liquidity		-			-	0						-/0
Credit rating	0				-		-	0				-/0
Market												
Risk free rate	0	-		0	0	-	0			0		-/0
Yield curve	0	-				-	0				+	-/+/0
Market volatility	0	0			+	0	+		0	+		+/0
Market return		-			0		-			-	-	-/0
Financial crisis			+				+					+
Housing prices								0				0
Interbank risk					-				0			-/0
Stock skew									+			+
Stock kurtosis									0			0
Forex volatility									+			+
Commodities									0			0
Fama-French												
FF Mkt-Rf						+			-			-/+
FF SMB						+			0	0		+/0
FF HML						0			-	-		-/0

The literature on CDS determinants for banks has focused on traditional financial variables. In Table 4.2, these variables are grouped in three categories: Firm-specific variables, market variables and Fama-French variables. Most of the studies include market variables, around half include accounting variables, while only three of the examined studies include Fama-French variables.

To our knowledge, Raunig and Scheicher (2009) give the first contribution on the determinants of bank CDS spreads, using a data set of monthly CDS spreads on 41

[1]Different articles may use different proxies for these variables. For simplicity we have grouped them into an appropriate common variable. For an overview of the different proxies used by each article, we refer to Table B.1 in the Appendix.

[2]Each research article may contain several models and get different results for the same variable. We only look at the results in the model with most variables over the full sample period, since this model is most similar to our model.

banks and 162 non-banks in Europe and the US from January 2003 to December 2007. They run a panel fixed effects regression containing the risk free interest rate, the yield curve, stock volatility (both market and firm-specific) and the Moody's KMV empirical default probability. When studying banks only, on the full sample period, they find that only the firm-specific stock volatility is a significant factor explaining CDS spreads. In particular, higher volatility increases the spreads.

Building on the work of Raunig and Scheicher (2009), Annaert et al. (2013) perform a similar analysis on their data set consisting of weekly CDS spreads on European banks between 2004 and 2010, thereby also covering the period of the financial crisis. In addition to the variables used by Raunig and Scheicher (2009), Annaert et al. (2013) include the bid-ask spread on the CDS quotes (as a proxy for CDS liquidity) and stock returns (both firm-specific and market returns). Both stock returns and CDS liquidity are found to be significant determinants of CDS spreads, such that improving stock prices and increased CDS liquidity yield lower CDS spreads. In contrast to Raunig and Scheicher (2009), Annaert et al. (2013) find that market and firm-specific stock volatility are insignificant, while the yield curve and risk-free interest rate are significant determinants of CDS spreads (with negative sign).

Both Raunig and Scheicher (2009) and Annaert et al. (2013) make strong cases for the impact of market-based variables on bank CDS spreads, but they do not investigate the potential effect of accounting variables.

Chiaramonte and Casu (2013) are the first to include accounting variables in their regressions on bank quarterly CDS spreads. They study the impact of eight balance sheet ratios on CDS prices in Europe, the US, Australia and Japan from 2005-2011. The eight variables include two measures each for asset quality, leverage, profitability and liquidity. The results of their analysis suggest that bank balance sheet ratios are important determinants of bank CDS spreads. More specifically, they find that improving asset quality (as measured by a lower loan loss reserve to gross loans ratio), profitability and balance sheet liquidity results in lower CDS spreads, while higher leverage (lower equity/assets ratio) increase CDS spreads.

Hasan et al. (2016) give, to our knowledge, the first contribution which combines market-based variables and accounting variables when investigating determinants of bank CDS spreads. They use a data set of 161 global banks and look at annual CDS spreads from 2001 to 2011. Their results on the impact of accounting variables are largely in line with Chiaramonte and Casu (2013). However, they also include a variable for cost efficiency, which is found insignificant. Among the market-based variables, stock volatility is found significant, while the risk-free rate is found insignificant (similar to Raunig and Scheicher (2009), but in contrast to Annaert et al. (2013)).

Samaniego-Medina et al. (2016) study the determinants of CDS spreads over a data set of annual CDS spreads on 45 European banks from 2004 to 2010. The 15 independent variables studied were, like Hasan et al. (2016), related to accounting data and market data. They find that four accounting variables are significant determinants of bank CDS spreads: The non-performing loan ratio, the size, the leverage ratio and the liquidity ratio (net loans to total assets). Samaniego-Medina et al. (2016) are the first to include bank size (measured by total assets) as a determinant of CDS spreads. Interestingly, they find that CDS spreads increase with the size of the bank. As a possible explanation, they refer to De Jonghe (2010) who argues that larger banks tend to be riskier due to the *moral hazard problem*. According to De Jonghe (2010), the moral hazard problem stems from the fact that larger banks, especially system-important banks, are too big to fail and therefore have a tendency to receive rescue-packages from regulatory authorities when they are close to bankruptcy[3]. This may incentivize managers of such banks to take higher risks, since the downside is protected. Qu (2020) has recently found empirical evidence for this in the Chinese banking sector. Among the market variables, market volatility is found significant by Samaniego-Medina et al. (2016), while the market return and risk-free rate are found insignificant.

Smales (2016) adds to the literature by including three Fama-French factors (HML, SMB and MKT) as measures of market risk. In the previous literature, the connection between Fama-French factors and the CDS spread on firms have been claimed to be negative (Galil et al., 2013). The intuition behind this is that higher Fama-French factor levels indicate better economic conditions (higher assets value) and therefore lower credit spreads (Galil et al., 2013). However, this relationship had not been investigated on CDS spreads of banks, prior to Smales (2016). Smales (2016) finds that market based variables are more important than accounting-based variables. In particular, no accounting variables are found to be significant in the baseline model, whereas the majority of the market based variables are found significant. The latter include the risk-free rate and the yield curve with negative coefficients, and SMB and MKT with positive coefficients. Furthermore, Smales (2016) finds that firm stock return and volatility are significant with negative and positive coefficients respectively. This is consistent with the findings of previous literature.

Drago et al. (2017) examine CDS spreads across both Europe and the US, in the 2007-2016 time period. Their study includes both market-based and accounting-based variables in order to find determinants of one week ahead CDS spreads. They estimate a panel regression and conclude that the main market-based variables are stock return and volatility. Also, the accounting-based variables leverage, asset quality and bank size are found to be of importance. Drago et al. (2017) find a negative relationship between bank size and CDS spreads, in contrast to Samaniego-Medina et al. (2016). In that context, it should be noted that Drago et al. (2017) differ from Samaniego-Medina et al. (2016) by analyzing bank CDS spreads in a broader geographical context and in a larger time span, stretching further into the post-crisis era.

Similar to Drago et al. (2017), Benbouzid et al. (2017) study bank CDS spreads across different geographies using a wide array of accounting-based and market-based variables. The CDSs used were collected from the 2004-2011 time period. Using a panel approach, similar to Chiaramonte and Casu (2013) and Hasan et al. (2016), they find that the asset quality and bank profitability are helpful in determining bank CDS spread. Also, unlike Samaniego-Medina et al. (2016) and Hasan et al. (2016), Benbouzid et al. (2017) finds that financial efficiency (measured by overhead costs to total assets) is significant with a positive sign. On the other hand, leverage is not found to be significant, contrary to the findings of Chiaramonte and Casu (2013), Hasan et al. (2016), and Drago et al. (2017).

A recent contribution to the literature is Koutmos (2018) who studies 20 global system-important banks, 14 of which European, and six from the US. He does not rely on accounting variables and mostly uses a range of stock-market implied variables, in addition to interbank risk and the Fama-French factors. Using a quantile regression framework, he finds that HML and MKT are significant determinants of CDS spreads, both with negative sign, while SMB is not significant. This is conflicting to the findings of Smales (2016), however in accordance with Guesmi et al. (2018). A reason for the inconsistencies could be that Guesmi et al. (2018) do not study CDS spreads on indi-

[3]Examples of this were shown in Chapter 2 with the Financial Crisis and the Euro Crisis.

vidual banks, but on a bank CDS index. Interbank risk is not found to be a significant determinant of CDS spreads by Koutmos (2018).

Furthermore, Koutmos (2018) finds that market and firm-specific stock volatility are not significant in the middle quantile. This is similar to Annaert et al. (2013), but at the same time contradicts the conclusions of the majority of the preceding literature. Moreover, he finds that the skewness of market returns is significant in the middle quantile. Also volatility in the Forex markets is significant across all quantiles. Among the contributions discussed, Koutmos (2018) is the only one to include market skewness and forex volatility as determinants of CDS spreads.

Sclip et al. (2019) study a set of CDS spreads on 28 European banks during the 2005-2015 period. They include both accounting based and market based variables. Of the accounting variables, they largely rely on proxies as to how well the banks commit to the Basel III regulations. They find that the quality of bank assets, as measured by the NPL ratio, is a significant determinant of CDS spreads. This echoes the findings of Samaniego-Medina et al. (2016), who also purely study European banks. Similar to Benbouzid et al. (2017), Sclip et al. (2019) do not find leverage to be a determinant of CDS spreads. Of the market variables, Sclip et al. (2019), like Drago et al. (2017) and Guesmi et al. (2018) find market return to be of importance. Sclip et al. (2019) also deem the yield curve to be a determinant of CDS spreads, with a positive sign. This is in contrast with the findings of Annaert et al. (2013) and Smales (2016) who find this relationship to be negative.

Based on the above discussion, there are no general consensus on the impact of several financial variables. In particular, results vary a lot on the sign and impact of bank size, the yield curve and the Fama-French variables. A reason for this may be that the previous studies have large variations in their set of variables. In fact, no study includes all variables that have been shown to significantly determine bank CDS spreads. Our goal is to build on the work of the previous literature by including all these variables in a more comprehensive setting. Different variables may have similar explanatory power, and hence by including all variables we can identify any redundancies in the findings of the previous literature.

4.2 Previous Research on Political and Policy Uncertainty

In addition to traditional variables, we study the impact of three novel factors related to political stability, policy uncertainty, and news sentiment.

An increasing amount of literature focuses on the impact of political uncertainty on financial markets (Dai and Zhang, 2019). Pástor and Veronesi (2012) develop a general equilibrium model and prove analytically that the expected value of the stock return at the announcement of a policy change is negative. Building upon this, Pástor and Veronesi (2013) find that that the Economic Policy Uncertainty (EPU) index, introduced by Baker and Bloom (2016), is negatively associated with a wide range of economic conditions, such as the Chicago Fed National Activity Index, industrial production growth and the Shiller price–earnings ratio.

Despite the above findings, limited research has focused on how political uncertainty affects credit risk, and to our best knowledge, no research has focused on the impact of political/policy variables on *bank* CDS spreads. Among the few studies that investigate the impact of political variables in credit markets are Kaviani et al. (2017), Liu and Zhong (2017), and Wang et al. (2018).

Kaviani et al. (2017) investigate whether policy uncertainty affects credit risk using US bond data covering the period 2002–2015. They find that policy uncertainty is positively associated with corporate bond spreads, controlling for bond-issue, firm characteristics, firm and credit-rating fixed effects, as well as macroeconomic conditions and economic uncertainty.

Liu and Zhong (2017) and Wang et al. (2018) also focus on the link between political uncertainty and credit risk. Unlike Kaviani et al. (2017), who use bond yield spreads to measure credit risk, they measure an individual firm's credit risk by using CDS spreads. To our knowledge Liu and Zhong (2017) and Wang et al. (2018) are the only groups of researchers studying the influence of political uncertainty on CDS spreads. Their results motivate us to study the same relationships in the bank CDS market.

Using national elections as a proxy for political uncertainty and using a sample of firms with single-name CDS across 30 countries, Liu and Zhong (2017) find that elections cause an increase in CDS spreads.

Wang et al. (2018) employ the EPU index of Baker and Bloom (2016), and document evidence that increases in the uncertainty index lead to increases in the CDS spreads, and the impacts can persist for up to eight quarters.

We note that these modelling approaches have two limitations. First, they use a narrow approach to quantify political uncertainty. Liu and Zhong (2017) solely use a dummy variable reflecting elections, as a proxy for political uncertainty, whereas Wang et al. (2018) only use the EPU index, which aims to capture *policy* uncertainty. The EPU is a composite measure based on three componenets: The percentage of news articles related to policy uncertainty in large newspapers, the magnitude of federal tax code provisions set to expire, and the dispersion of economic forecasts of the consumer price index and purchases of goods and services by governments. The main downside of the index, according to critics, is that it is relying on newspaper coverage when it is clear that media reports exhibit considerable bias in favoring negative news (Čižmešija et al., 2017). Thus there exists a limitation which we aim to address by using variables constructed on the basis of expert opinions and a wider definition of political and policy uncertainty.

Secondly, neither Liu and Zhong (2017) nor Wang et al. (2018) include variables related to both political stability and policy uncertainty. Douglass C. North, co-recipient of the 1993 Nobel Memorial Prize in Economic Sciences, argues for separating the analysis of political rules from the economic policy choices (North, 1991). However, most empirical finance studies in this area make a latent assumption that politics matter because of the policies they usher in, and don't explicitly differentiate between the two. In general, confounding the two entails loss of information (Gala et al., 2018). We therefore perform a more rigorous study by including variables related both to political (in)stability and policy uncertainty.

Policy and political news are typically an important part of the news coverage of financial newspapers. It may therefore be the case that political and policy risk is captured by the news sentiment. Al-Maadid et al. (2020) study the impact of business and political news on stock market returns in the Gulf Cooperation Council (GCC) countries, while Rambaccussing and Kwiatkowski (2020) forecast macroeconomic variables based on economic policy news in UK newspapers. Hence, the borderlines between news sentiment and political and policy risk may be small. This is particularly the case if political and policy variables are created based on news articles, like the EPU (Baker and Bloom, 2016). To address this, we include a sentiment variable. This is particularly relevant as most studies on CDS spread determinants include only quantitative information from financial reports, securities markets or macroeconomic publications. Such models do not take into account potentially important qualitative information released directly from firms through corporate filings or from other sources such as news articles.

A large body of theoretical and empirical studies show that price movements in financial markets are influenced by financial news (Tetlock, 2007; Boudoukh et al., 2012; Calomiris and Mamaysky, 2019; Fang and Peress, 2009; Hillert et al., 2012). Their studies focus on the stock markets. However, the relationship between news sentiment and the credit markets, where institutional investors are dominant, is not well studied.

Among the few studies conducted on the influence of news sentiment in the CDS market are those of Smales (2016) and Tsai et al. (2016). Smales (2016), one of the 11 papers included in our literature study from Table 4.1, also includes variables for news sentiment in his panel regression on US and European bank CDS spreads[4]. He finds that there is a significant relationship between news sentiment and changes in bank CDS spreads. Tsai et al. (2016) investigate US corporate CDS spreads using Wall Street Journal news articles and US company public filings. Similar to Smales (2016), they find that a more negative news sentiment is associated with higher CDS spread.

To summarize, a growing body of empirical research suggests that political and policy uncertainty impact financial variables, yet no empirical research has made an explicit separation between the two or studied their impact on *bank* CDS spreads in particular. Moreover, recent publications suggest that news sentiment affects a wide array of financial markets. In order to differentiate between the impact of political stability and policy uncertainty, and news sentiment, we include both in our model.

[4]The news sentiment variable was not included in Table 4.2 because it is not regarded as a traditional financial variable.

Chapter 5

Data Description and Sign Hypotheses

5.1 Data Description and Construction

5.1.1 CDS spreads

Our CDS data is collected on 46 banks from developed countries (Europe, USA, Canada and New Zealand). As outlined in Section 3.2, these markets account for over 80% of total trading in single-name CDSs. A full list of the banks used in this study is given in Table 5.1.

Table 5.1: Banks for which CDSs are collected and analyzed in this study.

Europe (27)		USA (11)	Canada (4)	Australia (4)
Barclays	Commerzbank	Wells Fargo	Bank of Montreal	Commonwealth Bank Of Australia
Bank of Ireland	KBC Groep	Morgan Stanley	Bank of Nova Scotia	Westpac Banking Corp
BNP Paribas	SEB	Charles Schwab	Canadian Imperial Bank of Commerce	Australia and New Zealand Banking Group
Credit Agricole	Swedbank	American Express	Toronto-Dominion Bank	National Australia Bank
Deutsche Bank	Banco de Sabadell	Bank Of America		
Santander	Erste Group Bank	Citibank		
Societe Generale	Unione Di Banche Italiane	Fifth Third Bank Group		
Lloyds Banking Group	Svenska Handelsbanken	Goldman Sachs		
UniCredit Spa	Banca Comercial Portugues	JP Morgan		
Intesa Sanpaolo	UBS	Comerica		
Banco Bilbao	Credit Suisse	Bank of New York Mellon		
Nordea	DNB			
Standard Chartered	Royal Bank of Scotland			
Danske Bank				

We use the CDS spread of the last trading day in each half-year period, from the second half of 2005 until the second half of 2019, totalling 29 time periods. We therefore have 1 334 observations of CDS spreads. The spreads collected are quoted in basis points to the second decimal. The data is collected from Bloomberg, where IHS Markit is the data provider of CDS spreads. We also use Datastream, where IHS Markit also is the data provider, as a secondary source of CDS spreads in those cases where Bloomberg data have missing values. In line with literature, the logarithm of the CDS spreads is used as dependent variable. Figure 5.1 shows the average CDS spreads segmented by geography. Similar to Chiaramonte and Casu (2013) and Samaniego-Medina et al. (2016), among others, we use levels rather than differences in our equations for CDS spreads. This is because we are more interested in explaining the spread than in making predictions.

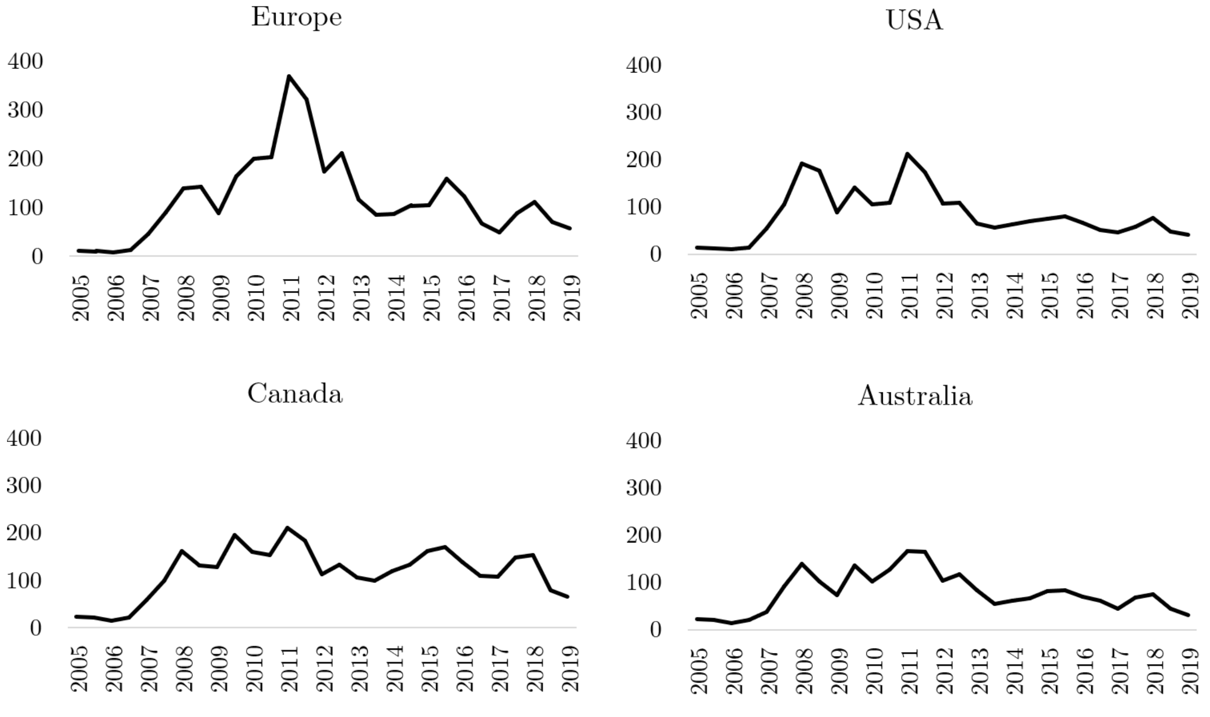

Figure 5.1: Average CDS spreads, segmented by geography.

We have included as many banks as possible from our selected geographies. Given that our data stretches over a 14 year time period, several banks have been excluded from the study as their CDS spreads do not exist over the whole time period. Moreover, the banks need to be listed for us to have access to their accounting data, and as such, our data set only contains listed banks.

5.1.2 Independent variables

In our analysis, we focus on 26 traditional financial variables, two novel political stability and policy uncertainty variables, and one news sentiment variable. The traditional financial variables are used by the articles covered in our study of the existing literature on CDS spreads in Chapter 4. Of these, twelve are firm-specific variables, eleven are market variables, and three are Fama-French variables, see Table 4.2. Based on what is most common in literature, we have chosen an appropriate proxy for each variable[5]. Several variables have been excluded from our analysis due to the lack of data and because existing literature has found them insignificant in explaining variations in bank CDS spreads[6].

[5]Different articles have used different proxies for the same variable. For an overview of the variable proxies used by each article, see Table B.1 in the Appendix

[6]We have not been able to collect data for all variables due to data access restrictions and low data quality. The following variables have therefore been left out of the study: asset quality, funding stability, firm liquidity, income diversification, CDS liquidity, credit rating, stock skew and forex volatility. Furthermore, house prices, interbank risk, stock kurtosis and commodities due to insignificance in existing literature.

Hence, we are left with 17 variables; 14 traditional financial variables from the existing literature, two political/policy variables, and one news sentiment variable.

Table 5.2: Overview of independent variables used in our model.

Name in model	Variable	Variable description	Sign (hypothesis)	Sign (literature)	Source
Firm-specific					
Accounting					
LEV	Leverage	Total liabilities/Total assets	+	-/+/0	Datastream
EFF	Efficiency	OPEX/Revenue	+	+/0	Datastream
PROF	Profitability	Net income / Total equity (ROE)	-	-/0	Datastream
SIZE	Size	ln(Total assets). Total assets in USD	-/+	-/+/0	Datastream
Market					
SVOL	Stock volatility	6-month historical volatility of stock, annualized	+	+/0	Inside calculation
SRET	Stock return	Stock price of bank	-	-	Inside calculation
Market					
RF	Risk free rate	Yield on US 3-month Treasury	-	-/0	FRED
YLD	Yield curve	US 10-year government bond yield less yield on 1-year Treasury bonds	-	-/+/0	FRED
MVOL	Market volatility	VSTOXX for European banks, VIX for non-European banks	+	+/0	Datastream
MRET	Market return	STOXX50 for European banks, VSTOXX for non-European banks	-	-/0	Datastream
CRIS	Crisis	Dummy: Financial Crisis (2007-2008) and European Debt Crisis	+	+	Inside calculation
Fama-French					
MKT	Mkt-Rf	Excess return on the market portfolio	-/+	-/+	K. French website
SMB	SMB	Small capitalization portfolio minus big capitalization portfolio return	-	+/0	K. French website
HML	HML	High book-to-market portfolio minus low book-to-market portfolio return	-	-/0	K. French website
Political					
POLT	Political (in)stability	The 6-month portfolio return of going long politically unstable countries and short politically stable countries	+	N/A	Gala et al. (2018)
POLC	Policy uncertainty	The 6-month portfolio return of going long countries with high policy uncertainty and short countries with low policy uncertainty	+	N/A	Gala et al. (2018)
Sentiment					
SENT	Sentiment	Single PCA component constructed using 7 IMF news-based sentiment indices (Fear, Crisis, Negative, Positive, Hedging, Risk, Opinion)	+	N/A	IMF

Table 5.2 gives a short description of how we measure each variable. For all accounting variables, we use second quarter (Q2) data points for the first half of the year (H1), and fourth quarter (Q4) data points for the second half of the year (H2). To obtain the return on equity, the *profitability* proxy, we annualize the Q2 and Q4 net income due to shortages of Q1 and Q3 P&Ls in Datastream. For all variables with daily data, we select the prices at the end of Q2 and Q4, respectively. This is true also for the *stock return*. However, for the *stock volatility*, we compute the annualized historical volatility based on the last 6 months of trading in the stock.

The market variables *risk free rate* and *yield curve* are collected in the same manner as *stock return* as they are available on a daily frequency. The *market volatility* is measured using the VSTOXX index for European banks and the VIX for non-European banks. These indices measure the option-implied volatility for the STOXX50 index and the S&P500 index, respectively, which we also use to measure *market return*. We also include a Crisis dummy variable. It is activated in the H2 2007 to H1 2009 and H1 2011 to H2 2013 time periods. The activation is made at these time periods to indicate the peaks of the Financial Crisis and the European Debt Crisis.

We follow the approach of Smales (2016), including the three factors from the original 1993 paper where Fama and French propose their model, *Mkt-Rf, SMB* and *HML* for developed countries (Fama and French, 1993).

Our political variables are sourced from Gala et al. (2018) who have constructed indices for policy and political uncertainty. The factors are constructed on the basis of the World Economic Survey (WES) conducted by the International Institute for Economic

Research (IFO) with funding from the European Commission. The results for politics and policy are published each year, both in May and November. The survey has been conducted by the same research center since 1992, and is answered by a panel of over 1 000 experts related to 42 countries. The experts satisfy professional requirements set by WES, and IFO controls conflict of interest, to increase the reliability of the survey (Gala et al., 2018).

The data is well suited for our work as it provides longitudinal data which allows for ”analysis of economic, financial, political and investment climate across countries and how it has changed over time” (Stangl, 2007). Moreover it allows us to separate political stability from policy uncertainty. The political and policy variables are constructed on the basis of two questions, outlined below in Table 5.3.

Table 5.3: Questions related to political and policy risk, asked in IFO survey.

Variable	Question
Policy uncertainty	“[A]ssess the importance of the following problems the economy of your country is facing at present: Lack of confidence in the government’s economic policy.”
Political (in)stability	“[A]ssess the importance of the following factors which influence the climate for foreign investors in this country: political instability is absent, low or high.”

The answers are given numerically, on a scale from 1-9 for political stability and 0-100 for policy uncertainty, respectively. Based on the answers, countries are rated from lowest to highest in both dimensions. The factors are then created by using factor-mimicking portfolios. As described by Gala et al. (2018), the portfolios are formed on the last day of the month of each WES announcement, and are rebalanced semi-annually. By construction, these portfolios maximize the spread in the politics and policy variables, so that differences in their returns can be more accurately attributed to differences between political and policy risk. The politics factor is the return of a portfolio going long on low stability countries and short on high stability countries, and the policy factor is the return of going long on low policy confidence countries and short on high policy confidence

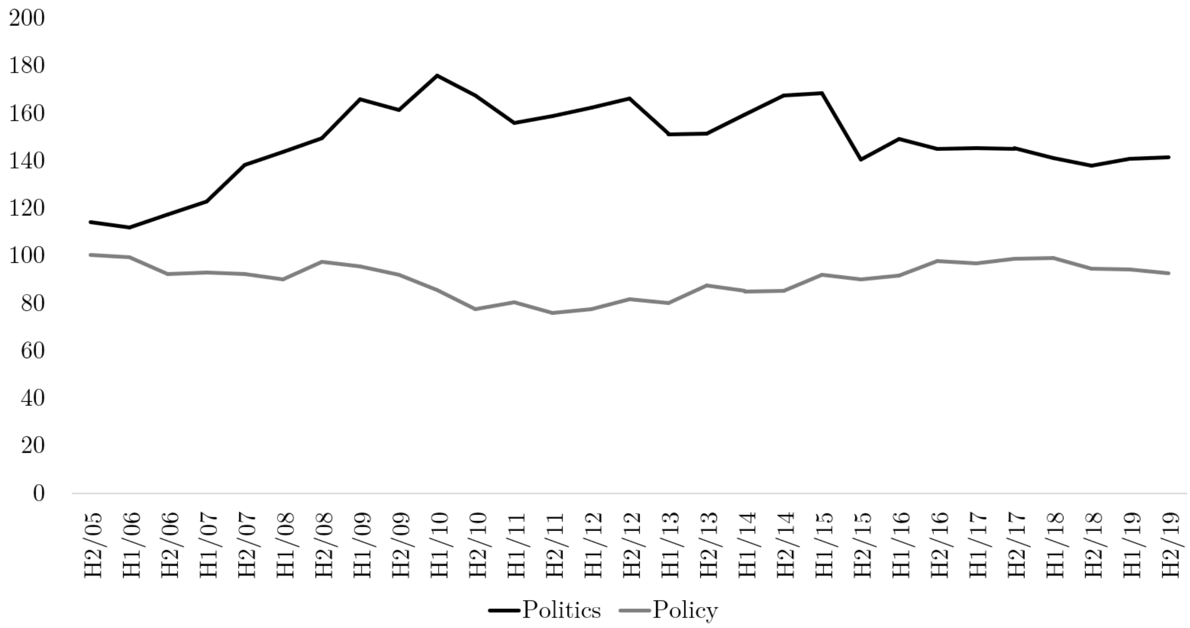

Figure 5.2: Political and policy variables plotted at levels, H2 2005 until H2 2019.

countries. In both cases the MSCI Investable Indices is used in order to gain exposure to country-level market returns. The two factors therefore give an investor exposure to the underlying political and policy risks. When the underlying risk increases, according to classical financial theory, the factor returns will increase. Plots of the variables are shown in Figure 5.2. A detailed description of this data set and its construction is given in Gala et al. (2018).

The news sentiment variable is included in order to control that our political factors are not captured by news sentiment. We use a novel data set published by the IMF in 2019 (Huang et al., 2019)[7]. It includes a set of 7 indices constructed on the back of a database containing over 3 million news articles from the Financial Times newspaper. The news articles cover business, finance and economic topics, and hence is an appropriate source of news to construct sentiment indicators (Huang et al., 2019). The indices are constructed through analyzing the frequency of semantically similar words to what the index reflects. The seven indices are Crisis, Fear, Hedging, Opinion, Negative, Positive and Risk. Examples for "Fear", "Crisis" and "Risk" are given in Table 5.4. For a substantial documentation of the indices, we point to the work of Huang et al. (2019).

Table 5.4: Examples of semantically similar words for different IMF indices.

Index	Semantically similar words
Fear	*Worry, concern, anxiety, suspicion, fearful, hope, etc*
Crisis	*Depression, financial crisis, etc*
Risk	*Warn, risk, threat, hazard, impact, terror, danger etc*

The indices are available for 20 countries[8]. A weakness with this data is that the selection of countries for which the indices are constructed have little overlap with the countries which are included in our study. In order to mitigate the problem of a small overlap, we take the average values for the indices corresponding to countries which

Table 5.5: Descriptive statistics for all variables.

	Lever-age	Efficiency	Profitability	Size	Stock volatility	Stock return	Risk free rate	Yield curve	Market volatility	Market return	Crisis	MKT	SMB	HML	Political (in)stability	Policy uncertainty	Senti-ment	CDS
count	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334	1334
mean	0.94	0.64	10.39	13.33	0.33	0.00	1.24	1.47	20.69	0.00	0.38	0.04	-0.00	-0.00	0.01	-0.00	-0.00	107.32
std	0.03	1.96	16.58	1.27	0.23	0.98	1.62	1.08	7.29	0.98	0.49	0.12	0.04	0.05	0.06	0.05	0.19	119.99
min	0.85	-21.95	-156.92	8.81	0.08	-2.27	0.03	-0.29	11.04	-1.83	0.00	-0.34	-0.09	-0.08	-0.17	-0.09	-0.31	4.00
25%	0.93	0.41	6.06	12.60	0.19	-0.67	0.06	0.52	15.27	-0.76	0.00	-0.02	-0.02	-0.04	-0.02	-0.02	-0.19	43.31
50%	0.94	0.58	11.60	13.61	0.26	-0.29	0.26	1.62	18.12	-0.04	0.00	0.06	-0.00	-0.01	0.02	-0.01	0.10	80.00
75%	0.96	0.73	15.95	14.24	0.37	0.62	1.96	2.37	24.06	0.53	1.00	0.10	0.02	0.02	0.05	0.02	0.16	132.23
max	0.99	58.79	179.43	15.35	2.15	3.15	4.98	3.38	43.87	2.30	1.00	0.25	0.08	0.13	0.14	0.09	0.23	1663.97
skew	-0.76	19.01	-1.73	-1.22	3.14	0.68	1.29	0.03	1.16	0.32	0.50	-1.07	-0.06	0.54	-0.41	0.28	-0.46	4.90
kurt	0.06	600.42	29.04	1.78	14.08	-0.18	0.35	-1.19	1.17	-0.55	-1.75	2.17	0.16	0.26	0.95	-0.29	-1.38	40.55

[7]To the best of our knowledge, few, if any, other open source historical sentiment indicators are available.

[8]A complete list of these countries is given in Figure B.1 in the Appendix.

are included in our study. Some of the indices are then reversed so that the expected relationship with CDS spreads is positive for all indices.

Descriptive statistics of the variables included in our analysis is shown in Table 5.5, and a correlation matrix of the variables is presented in Figure B.2 in the Appendix.

5.2 Coefficient Sign Hypotheses

In the following, our hypotheses on which sign the regression coefficients of the variables should have based on theory and empirical results in literature is presented. An overview is given in Table 5.2 on page 21.

5.2.1 Firm-specific variables

Leverage

In line with the majority of the studies outlined in Chapter 4, we include leverage as a variable in our model. According to classical asset pricing theory, higher leverage indicates a shorter distance to the default barrier and a higher probability of default (Merton, 1974). Therefore, we expect a positive relationship between leverage and CDS spreads.

Efficiency

In line with Samaniego-Medina et al. (2016), Benbouzid et al. (2017) and Hasan et al. (2016), we include a bank-specific efficiency variable in our model. When efficiency decreases, i.e. OPEX/Revenue increases, banks operate less efficiently and reduce their cash flow. Thus we expect this variable to have a positive relationship the CDS spreads.

Profitability

Following Samaniego-Medina et al. (2016), Sclip et al. (2019), Hasan et al. (2016), Benbouzid et al. (2017) and Chiaramonte and Casu (2013), we include a profitability variable in our model. Earnings reflect a bank's income-producing ability. It is essential for a bank to remain viable, fund growth, and sustain and increase capital. Therefore, a bank with higher return on its equity is more financially sound and has lower default risk (Hasan et al., 2016). In line with this thinking, we hypothesize that the profitability variable should have a negative relationship with the CDS spreads. The majority of results in existing literature confirm this relationship.

Size

Following Sclip et al. (2019), Drago et al. (2017), Smales (2016) and Samaniego-Medina et al. (2016) we use a variable for bank size. According to the moral hazard theory introduced by De Jonghe (2010), larger banks may be more attracted to increasing risk taking, reducing market discipline and creating competitive distortions because of their "too big to fail" mentality. Conversely, larger banks may be less prone to risk because of their managerial capacities and efficiencies (Baselga-Pascual et al., 2015). This argument has been reversed by certain researchers who argue that the historical government bailouts of large banks speaks in favour of lower default probability associated with large banks

(Sclip et al., 2019). The relationship between bank size and CDS spreads is therefore not obvious, and we argue that the relationship could be both positive or negative. The literature is also inconclusive, having found both a positive and negative relationship.

Stock volatility

As Raunig and Scheicher (2009), Annaert et al. (2013), Hasan et al. (2016), Samaniego-Medina et al. (2016), Smales (2016), and Koutmos (2018) we include a variable for the firm-specific stock volatility. Higher asset volatility theoretically leads to higher credit spreads because it increases the likelihood that the default threshold is hit (Annaert et al., 2013). Moreover, from an asset pricing point of view, increases in firm-specific volatility lead to higher cash flow betas and an increase in default probability. We therefore believe this variable should inhibit a positive relationship with CDS spreads, in line with empirical results presented in the literature.

Stock return

We follow Annaert et al. (2013), Samaniego-Medina et al. (2016), Sclip et al. (2019), Hasan et al. (2016), Koutmos (2018) and Smales (2016) and include a firm-specific stock return variable. If stock returns are negative, leverage measured in market values will increase, leading to higher CDS spreads (Annaert et al., 2013). A negative relation between stock returns and CDSs is thus expected. Also, since equity returns reflect a firms' future prospects, positive returns indicate lower default risk and may thus also lead to lower spreads. In line with these arguments we hypothesize that the stock return variable should have a negative sign. The literature agrees on the sign of this variable being negative.

5.2.2 Market variables

Risk free rate

Following the majority of the studies outlined in Chapter 4, we include a variable for the risk-free rate. Longstaff and Schwartz (1995) argue that a higher interest rate (spot rate) increases future value of firms. Collin-Dufresne et al. (2001) further argue that an increase in spot rate reduces the probability of default for corporates. As stated by Annaert et al. (2013), the risk-free interest rate constitutes the drift in the risk neutral world. The higher it is, the less likely default becomes. These arguments support a negative connection between spot rate and credit spreads. Longstaff and Schwartz (1995) empirically confirm the negative relationship. This relationship is further confirmed by Annaert et al. (2013) and Smales (2016) in our literature review. As a consequence, we hypothesize that this variable exhibits a negative relationship with the CDS spreads.

Yield curve

The term structure is widely regarded as a business cycle predictor (Estrella and Mishkin, 1997; Annaert et al., 2013). A high yield curve anticipates improved economic growth. Therefore, a negative relationship is expected with CDS spreads. Moreover, to the extent that the yield curve gives information about future interest rates, a negative relation with CDS spreads follows. An increase in the yield curve would indicate higher future interest rates which imply lower credit risk. The negative relationship is found by Annaert et al.

(2013) and Smales (2016) in our literature review, and we therefore believe this variable should have a negative relationship with the CDS spreads.

Market volatility

The larger the volatility, the higher the uncertainty about the economic prospects is the assumption. A positive relation with credit spreads therefore follows (Annaert et al., 2013). In line with the majority of the literature, we include a market volatility variable. In the vast majority of papers included in our literature study on bank CDS spread determinants, this variable has a significant positive sign. We therefore hypothesize it should have a positive relationship with the CDS spreads.

Market return

General business climate improvements, reflected in market returns, will decrease the probabilities of default and will also increase the recovery rates (Annaert et al., 2013). A negative relation with CDS spreads thus follows. We follow Annaert et al. (2013), Samaniego-Medina et al. (2016), Drago et al. (2017), Guesmi et al. (2018) and Sclip et al. (2019) by including a market wide stock index return. It could be argued that the individual stock return already captures the information contained in the indices. However, firm stock returns are quite noisy and the danger exists that the firm-specific returns swamp the economy-wide content of an index return (Annaert et al., 2013). We hypothesize a negative relationship with the CDS spreads for the market return variable.

Crisis

As Chiaramonte and Casu (2013) and Drago et al. (2017), a crisis variable is included as a dummy variable. It is no surprise that we expect to see higher CDS spreads during bank crisis, where economic conditions are weaker and more uncertain. Figure 5.1 has also confirmed this, showing that CDS spreads are higher both during the Financial Crisis and the European Debt Crisis. Therefore, we expect a positive relationship between the crisis variable and the CDS spreads, as was obtained by Chiaramonte and Casu (2013) and Drago et al. (2017).

5.2.3 Fama-French variables

Market excess return (Mkt-Rf)

This variable constitutes the excess return on the market over the risk free rate. No hypothesis is clearly stated for this variable in our literature review. However, Guesmi et al. (2018) reference Galil et al. (2013), who state that this variable should have a negative relationship to CDS spreads. It should be noted that Galil et al. (2013) study the determinants of CDS spreads for *corporates in general* rather than focusing purely on banks. They argue that higher values for this factor indicate higher asset values and therefore lower spreads. Based on this we hypothesize that the coefficient for this variable should be negative. However, we note that the empirical results for this variable is not conclusive, where the results of Koutmos (2018) give a negative coefficient, whereas Smales (2016) concludes on it being positive.

Small-Minus-Big (SMB)

In line with Samaniego-Medina et al. (2016), Benbouzid et al. (2017) and Guesmi et al. (2018), the SMB variable is included in our model. The SMB factor is the spread between the returns of a small capitalization portfolio and a big capitalization portfolio. Also here, no hypothesis is stated in our literature review. Again, references are made to Galil et al. (2013) who argue that this variable should have a negative sign using the argument that higher factor values for this variable indicate that investors lower their distress risk aversion, which should lead to lower CDS spreads. We therefore hypothesize that there should be a negative relationship between this variable and the CDS spreads. The literature is not conclusive on this factor, where Koutmos (2018) and Guesmi et al. (2018) find it insignificant, whereas Smales (2016) finds it positively related to bank CDS spreads.

High-Minus-Low (HML)

In line with Samaniego-Medina et al. (2016), Benbouzid et al. (2017) and Guesmi et al. (2018), the HML variable is included in our model. The HML factor incorporates the spread in return between a high book-to-market portfolio and a low book-to-market portfolio. In the literature, this is the only Fama-French factor to exhibit a significant and negative relationship with bank CDS spreads (Koutmos, 2018). Book-to-market ratios are according to Fama and French (1993) associated with leverage, where high book-to-market firms, often being value firms, typically carry high leverage, whereas low book-to-market firms, often being growth firms, carry lower leverage (Koutmos, 2018). When the spread between the two portfolios increase, this is associated with a decrease in investors' distress risk aversion, which should be associated with a decline in CDS spreads. Following this argument, we believe the factor should have a negative sign.

5.2.4 Political and policy variables

Countries are rated 0-100 for confidence in economic policy and 1-9 for political stability, with higher values corresponding to higher confidence and more stability

Political risk

The political risk factor gives exposure to political stability. An increase in this factor indicates greater political uncertainty. As discussed in Section 4.2, Wang et al. (2018) find that greater political risk gives higher CDS spreads for corporates. We believe this finding can be translated to banks as well, and therefore hypothesize that this factor has a positive relationship with the bank CDS spreads.

Policy risk

The policy risk factor gives exposure to policy uncertainty. An increase in such a factor has by Wang et al. (2018) been found to increase the CDS spreads of corporates. We believe this factor also increases the CDS spreads of banks, and a positive sign is therefore expected.

5.2.5 Sentiment variable

Sentiment variable

Our sentiment variable is the first PCA component representing 83% of the variance in the seven underlying sentiment factors. All underlying factors are manipulated in such a way that an increase in the factor value corresponds to *worse* sentiment. We therefore expect this variable to have a positive relationship with bank CDS spreads.

Chapter 6

Model and Modelling Techniques

This chapter lays out our choice of model and modelling techniques. We use fixed effects panel regression. We provide theoretical foundation for analysis of CDS spreads using panel regressions, as well as more advanced variable selection techniques.

6.1 Panel Data Regression

Given the data described in the previous chapter, we need a model to estimate data comprising both time series and cross sectional entities. We call this setup a *panel of data* or *longitudinal data*, henceforth *panel data* (Brooks, 2019). Such a setup will cover data in two dimensions, both across time and space. In our case the *time* dimension will be the semi-annual data from second half 2005 throughout 2019, whilst the *space* dimension consist of the 46 banks located in developed countries (Europe, US, Canada, and Australia).

As we saw in the literature review in Chapter 4, the vast majority of recent studies on CDSs in the banking industry use panel data as their data structure. A similarity across literature is the low degree of granularity in the time dimension, often annual or quarterly, compared to the historically more frequently used time series models with daily or even hourly granularity. In addition, the models in our literature study stretch over a shorter time period of approximately 7.5 years. Baltagi (2013) argue that panels of different sizes require different econometric treatment. With a long time series for panels with large N, issues of non-stationarity in the time series must be dealt with. In contrast, panels with shorter time series do not need to be concerned with these issues since T is short for each entity. This is the case in most of the literature we have studied, this thesis included.

Even though panel data has lower data granularity and sometimes shorter time periods compared to time series, Brooks (2019) outlines the following advantages of panel data:

1. More complex problems can be tackled and a broader range of issues can be addressed than with pure time series or cross sectional data alone.

2. The degrees of freedom is increased and thus the power of the test is improved.

3. The data structure is flexible, allowing us to possibly remove the impact of certain forms of omitted variables bias in regression results.

The baseline panel data model we will use as our starting point in this thesis can be written as follows in econometric terms,

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 EFF_{it} + \beta_3 PROF_{it} + \beta_4 SIZE_{it} \\ &+ \beta_5 SVOL_{it} + \beta_6 SRET_{it} \\ &+ \beta_7 RF_t + \beta_8 YLD_t + \beta_9 MVOL_t + \beta_{10} MRET_t + \beta_{11} CRIS_t \\ &+ \beta_{12} MKT_t + \beta_{13} SMB_t + \beta_{14} HML_t + u_{it} \end{aligned} \tag{6.1}$$

where the CDS spread for bank $i = 1, ..., N$ at time $t = 1, ..., T$ is the dependent variable, α is the intercept term, β_j is the time and bank invariant coefficient to be estimated for the $j = 1, ..., P$ independent variables, and u_{it} is the error term. Note that the market and Fama-French variables only varies across time, not banks.

6.1.1 The Fixed effects model

The simplest way to account for entity or time differences in behaviour is by letting the coefficient of the variables to vary in either dimension. When we allow these unknown but fixed coefficients to vary through time or space we have a *fixed effects model* (Mátyás and Sevestre, 2008).

In the entity fixed effects model, the intercept α is allowed to vary from bank to bank, while the coefficients of the independent variables are assumed to be constant in both dimensions[9]. Following Stock and Watson (2018), consider Equation 6.1 above, but altered into the following:

$$CDS_{it} = \alpha + \beta_1 LEV_{it} + \beta_z z_i + u_{it}, \tag{6.2}$$

where z_i represents the unobserved variables that vary across banks but *not* over time, e.g. the corporate governance policies or the organizational culture of a bank. For illustration purposes only, only the first variable, LEV, is included. If we can observe z_i we get an unbiased and consistent estimator of β_1, the parameter of interest. However, if z_i is unobserved, we only get this if the second condition of *omitted variables* bias, i.e. $Cov(x_{it}, z_i) = 0$, is not fulfilled[10]. The question of how to estimate z_i when it is unobserved and $Cov(x_{it}, z_i) \neq 0$ thus arises.

Equation 6.2 generalizes to panels with varying intercepts:

$$CDS_{it} = \alpha_i + \beta_1 LEV_{it} + u_{it}, \tag{6.3}$$

where $\alpha_i = \alpha + \beta_z z_i$ models the impact of omitted time-invariant variables on CDS_{it} and will thus represent the *entity fixed effects* of the regression. By *demeaning* CDS_{it} and LEV_{it}, i.e. subtracting the entity mean from Equation 6.3, we get:

$$CDS_{it} - \overline{CDS}_i = (\alpha_i + \beta_1 LEV_{it} + u_{it}) - (\alpha_i + \beta_1 \overline{LEV}_i + \overline{u}_i)$$

$$\widetilde{CDS}_{it} = \beta_1 \widetilde{LEV}_{it} + \widetilde{u}_{it} \tag{6.4}$$

where $\overline{CDS}_i$, $\overline{LEV}_i$, and $\overline{u}_i$ is the entity mean of the variables, and $\widetilde{CDS}_{it}$, $\widetilde{LEV}_{it}$, and $\widetilde{u}_{it}$ is the entity demeaned variables. In Equation 6.4, α_i has disappeared and the

[9] *Entity* fixed effects, and not *time* fixed effects, will solely be addressed as this is what we use in our methodology. The two approaches are however analogous.

[10] More on omitted variables bias can be found in Duffy and Smith (2020) amongst others.

OLS can thus be run on just one normal equation instead of two. This is often referred to the *within transformation* and is a preferred method of estimating the fixed effects model due to the computational ease it provides compared to the also common Least Squared Dummy Variables (LSDV) method (Mátyás and Sevestre, 2008; Stock and Watson, 2018).

In this thesis we use the entity fixed effects model estimated by the within transformation. This approach lets us utilize the powers of the panel data, while it offers computational ease compared to LSDV. It is important to note that both approaches produce the same result.

A drawback is the degrees of freedom we lose when working with transformed variables. The usual $NT - P$ is modified to $NT - N - P$, since in order to transform the variables the N banks' mean must be computed, resulting in the loss of N degrees of freedom. Another, simpler transformation, the *first-difference transformation*, can also be done. This approach does not require any adjustments of the degrees of freedom. However, it introduces serial correlation in the transformed version, and thus the OLS estimation is no longer BLUE as it affects the efficiency of the estimation leading to the conclusion that the parameters are more precise than they really are (Mátyás and Sevestre, 2008). The loss of degrees of freedom and using the within transformation thus seems more attractive for our purpose. A second disadvantage of the within transformation is that we lose the ability to determine the influences of all of the variables that affect CDS_{it} but do not vary over time (Brooks, 2019). However, in our case, to determine this influence is not of interest and will therefore not alter our decision to use a within estimation.

6.1.2 Model selection and Random effects consideration

An alternative approach when estimating a panel regression is to use the *random effects model*, sometimes referred to as the error components model. This model shares many traits with the fixed effects model, for example different intercept terms for each entity but constant over time. The major difference is that the intercepts are assumed to be drawn from a global, common intercept α plus a random variable ϵ_i that varies over time but is constant across entities (Brooks, 2019). The random effects regression thus becomes (using only LEV as in Equation 6.2):

$$CDS_{it} = \alpha + \beta_1 LEV_{it} + \omega_{it}, \quad \omega_{it} = \epsilon_i + u_{it} \tag{6.5}$$

Unlike the fixed effects method there are no dummy variables to capture the heterogeneity across the banks, instead this happens with the ϵ_i-term. This term is assumed to have zero mean, independent of the individual regressors' errors u_{it} and the independent variables, and has constant variance (Brooks, 2019).

Whether fixed or random effects is the better model varies based on the data. A rule of thumb suggested by Brooks (2019) states that random effects is more appropriate when the banks in the sample can be thought of as having been randomly selected from the population, but fixed effects is more appropriate when they effectively constitute the entire population. In our case, to the best of our knowledge, all listed banks with available CDS spreads over the whole time period in the selected geographies have been included. The tendency is that larger banks have CDS spreads, and thus we can argue that the most important banks in each geography are included. This is therefore an argument in favor of the fixed effects model.

Random effects offers some advantages compared to fixed effects. In contrary to fixed effects, random effects does not remove the independent, unobserved variables that do not vary over time, making it possible to calculate the impact on the dependent variable. Also, random effects requires fewer parameters to be estimated, as there are no transformations as in fixed effects, and thereby upholding the degrees of freedom discussed in the section above.

The major drawback with random effects follows from the important assumption about the composite error term ω_{it}. It has to be uncorrelated with *all* of the independent variables. This means that we require both the random bank error term ϵ_i *and* the individual errors u_{it} to be independent of all of the independent variables. This demand is stricter than that of fixed effects. Another way to put it, if the unobserved, omitted variables are correlated with the observed, independent variables, fixed effects should be used. Intuitively, there may be some kind of correlation between the unobserved and the observed variables, as there are several variables we were not able to include due to different limitations, e.g. data availability. As an example, Arosa et al. (2014) highlight the correlation between organizational culture (unobserved) and leverage ratio (observed). In general, this argument trumps the advantages of random effects, and gives an indication that fixed effects model is the best alternative.

A Hausman test could support this claim. It tests whether the assumption is valid for the random effects estimator. If it does not hold, the parameters will be biased and inconsistent. This follows from the fact that the estimator will ascribe all of the increase in the dependent variable to the independent variables when in reality some of it arises from the error term. The test thus looks to see if there is a correlation between the unique errors and the regressors in the model. Under the null hypothesis, that there is no correlation between the two, the random effects model is recommended. The alternative hypothesis suggests the fixed effects model (Brooks, 2019). The p-value on the Hausman test in all models we run in this thesis is 0.00. The interpretation of this is that we reject the null hypothesis and select the fixed effects model.

6.2 Variable Selection

In addition to the panel data with fixed effects approach described above, our methodology described in the next chapter makes use of one other concept, variable selection, that demands introduction. We use this approach to identify redundancies in the wide range of CDS spread determinants. In this section, we will address our motivation for utilizing variable selection techniques and present the LASSO and SFFS techniques used in our model.

We see three main arguments for using variable selection in our model. First, the classical approach to selecting variables has been based on theory, paradigms or the researcher's own hypotheses. To our knowledge, no prior work uses a data-driven approach to selecting variables in research related to bank CDS spreads. Secondly, there is no consensus in the literature as to which set of variables explain the bank CDS spreads. This may be caused by the vast array of variables proposed to explain banks' CDS spreads, see Table 4.2 in the Appendix, which could cause redundancies in the models. Thirdly, even though different researchers find different variables to be important for their data set, there is no guarantee that the same will be the case in our data set. Hence, we propose a more data-driven and flexible approach to selecting variables than what is common in existing literature, while maintaining or even improving the ability to explain the variance of the CDS spreads.

Variable selection, or more commonly *feature selection* in machine learning terminology, is the process of selecting a subset of relevant variables for use in a model. The working premise is that it is generally better to have fewer independent variables in a model. The goal of variable selection, as stated by Kuhn and Johnson (2019), is to reduce the number of variables without compromising predictive performance. In our case, the most reasonable interpretation of this goal would be to reduce the number of variables without compromising the ability to explain variance. Variable selection also has the benefit of simplifying models to make them easier to interpret by researchers, reduce training time of complex models and reduce overfitting.

There are mainly three types of variable selection techniques (Kuhn and Johnson, 2019):

- **Filter** methods are simple and fast. They conduct a superficial analysis of the variables to determine which are important and then only provide these to the model.
- **Wrappers** use an external search procedure to choose different subsets of the whole predictor set to evaluate in a model. It separates variable search from the model fitting process.
- **Embedded** methods do variable selection as part of their model fitting, and is a "catch all" category for the rest of the methods.

Filters are usually faster to implement and computationally faster than both wrappers and embedded methods. Wrappers use iterative search methods in order to find the best variable subset. In our case, we do not have a lot of data and a wrapper would skim through every possible subset (in a greedy manner) in a matter of seconds, implying that a filter method loses its advantage over wrappers. The wrapper methods have a disadvantage related to having the most potential of overfitting the variables and thus require external validation (i.e. a test set). An embedded method is a suitable mid-point between the two, including its variable selection while fitting the model. There is no rule on which variable selection technique to use in which case, thus we include two methods in order to be able to compare our results.

6.2.1 LASSO (embedded)

As variables become more correlated with each other, the estimated βs in the normal linear equation get inflated and become unstable. Also, if one variable is linearly dependent of one or more of the other variables, it is not possible to complete the inversion required to solve the minimization problem in OLS.

Tibshirani (1996) proposed a modification to the sum of squares regression in order to solve this, a *penalized* regression equation, called the least absolute shrinkage and selection operator (LASSO). Modified to our use, the LASSO can be expressed as

$$SSE = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(CDS_{it} - \widehat{CDS}_{it}\right)^2 + \lambda \sum_{j=1}^{P} |\beta_j| \tag{6.6}$$

where N is the number of banks, T is the number of time periods, and P is the number of variables. By modifying the penalization term λ, typically in the range 10^{-8} to 10^{-1}, coefficients are forced to zero. In practice, the LASSO selects variables during the course of the process so that it is optimal for the model in question. The main task of the LASSO is thus to eliminate variables[11].

The LASSO is an automatic method as it does not let researchers affect which variables are removed. This may be the biggest drawback with the method in general, yet we deem it suitable for our purpose as it ensures complete objectivity. The thought is to have a data-driven approach, and so the LASSO offers a suitable solution to this.

6.2.2 SFFS (wrapper)

In addition to LASSO, we apply another method to confirm the robustness of our results. The Stepwise Forward Floating Selection (SFFS) technique uses a sequence of steps to include one-at-a-time variable from the previous steps' subset based on a certain metric. Often p-value is used, but this has been highly criticized by a number of researchers (Kuhn and Johnson, 2019). Instead, we use mean squared error (MSE) as scoring metric. SFFS is a greedy algorithm, and will therefore only consider the best solution in the next step given the current set of variables. Therefore, it has a good chance to reach a *local* optimum. To be able to evaluate a larger number of variable subset combinations, and increase the probability of reaching a *global* optimum, a possibility to exclude a variable after the selection step, is added. This second step, the *floating* step, is performed conditional on yielding a better score, and will not be performed unless an improvement is present. The procedure goes on until the algorithm reaches the predefined number of variables. This number is optimized so that the fixed effects model yielding the best AIC is chosen[12]. An overview of the algorithm is shown in Figure C.1 in the Appendix.

The major drawback of SFFS compared to the LASSO is that it more easily can be stuck in a local optimum and never reach a global optimum. This follows from the fact that it is a greedy algorithm, constantly searching for the "best at the time" solution in order to show immediate benefit, instead of allowing for detours to get to an overall better solution. In addition, as touched upon above, a wrapper method has a higher probability of overfitting the data set.

[11] The closely related ridge regression, which preceded the LASSO and has squared β_j instead of an absolute term, combats collinearity in larger degree than selecting variables.

[12] Akaike Information Criterion (AIC) is described in Section 7.1

Chapter 7

Results and Discussion

Our methodology consists of five main steps. A short overview is given below in Table 7.1, before a detailed description of each step and the respective results follow.

Table 7.1: Overview of methodology.

Model	Methodology	Description
1	Replication	We use variables from existing literature to find which of these affect bank CDS spreads
2	Variable selection	We use a data-driven approach to variable selection, aiming to find redundancies in the set of variables included in model 1
3	Include political/ policy variables	We include political stability and policy uncertainty variables and analyze their effect on bank CDS spreads
4	Include sentiment variable	We include a sentiment variable to see if there is overlap with political uncertainty variables
5	Robustness tests	We perform robustness tests to analyze the robustness of our results

7.1 Model 1: Baseline Model - Replication of Existing Literature

The first model we run is a replication of existing literature. The goal is to find significant variables when including all variables from existing literature in the same model and on the same data set.

Table 5.2 summarizes the variables used in existing literature on the determinants of bank CDS spreads. All variables in the model are standardized in order to be able to compare coefficients in the regression results[13].

The fixed effects panel regression model is summarized in Equation 7.1, each line describing the firm-specific accounting variables, the firm-specific market variables, the

[13]Standardization is done by subtracting each variable's mean from the observed value, and dividing by the variable's standard deviation.

market variables, and the Fama-French variables respectively.

$$\begin{aligned}CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 EFF_{it} + \beta_3 PROF_{it} + \beta_4 SIZE_{it} \\ &+ \beta_5 SVOL_{it} + \beta_6 SRET_{it} \\ &+ \beta_7 RF_t + \beta_8 YLD_t + \beta_9 MVOL_t + \beta_{10} MRET_t + \beta_{11} CRIS_t \\ &+ \beta_{12} MKT_t + \beta_{13} SMB_t + \beta_{14} HML_t + u_{it}\end{aligned} \tag{7.1}$$

The results in Table 7.2 suggest that our baseline model in general is in accordance with theory and the empirical results presented in existing literature. In particular, the results on *size*, *stock volatility*, *stock return*, *risk-free rate*, *yield curve*, *market volatility*, *market return*, *crisis*, and *HML* are all in accordance with theory *and* empirical results of the existing literature.

Table 7.2: Results Model 1: Baseline model - replication of determinants from existing literature. *Sign (hypothesis)* indicates the sign hypotheses from Section 5.2, while *Sign (literature)* indicates the sign of the results from existing literature. "+" indicates significantly positive, "-" significantly negative, and "0" insignificant. Multiple symbols indicate disagreements. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

	Baseline model	Sign (hypothesis)	Sign (literature)
Firm-specific			
Accounting			
LEV	-0.040	-	-/+/0
	(0.036)		
EFF	-0.012	+	+/0
	(0.011)		
PROF	0.019	-	-/0
	(0.013)		
SIZE	0.599 ***	-/+	-/+/0
	(0.071)		
Market			
SVOL	0.045 ***	+	+/0
	(0.017)		
SRET	-0.167 ***	-	-
	(0.017)		
Market			
RF	-0.649 ***	-	-/0
	(0.029)		
YLD	-0.163 ***	-	-/+/0
	(0.027)		
MVOL	0.114 ***	+	+/0
	(0.023)		
MRET	-0.102 ***	-	-/0
	(0.020)		
CRIS	0.195 ***	+	+
	(0.028)		
Fama French			
MKT	0.027	-/+	-/+
	(0.019)		
SMB	-0.031 **	-	+/0
	(0.013)		
HML	-0.040 ***	-	-/0
	(0.012)		
Intercept	4.176 ***		
	(0.015)		
Hausmann test	0.00		
AIC	1303.8		
Adj. R^2	0.8150		

Four of our variables are insignificant in explaining the bank CDS spreads, namely *profitability*, *leverage*, *efficiency* and the Fama-French variable *market excess return*. Given that these variables have been found significant in certain previous studies, a brief discussion is warranted.

While the majority of studies in the existing literature find *leverage* positively linked to the bank CDS spreads, we find it insignificant in our baseline model. A possible explanation is that the information is captured by *stock return*. As discussed in Section 5.2, if the stock price decreases, the leverage measured in market value will increase. The *stock return* variable is significant and negative, as expected. Furthermore, the minimum leverage ratio in the banking industry is regulated by authorities. Since banks make money by using deposits to issue credit, they want to maximise their leverage. Therefore, bank leverage ratios are often just above regulatory requirements. This means that leverage in the banking industry, to a larger extent than for other industries, has a tendency to be similar across banks and over time (given that regulatory requirements are not changed). This is also supported by the descriptive statistics in Table 5.5, which shows that the leverage ratio has the lowest standard deviation of all variables. Moreover, the data of CDS spreads are collected at the last trading day of each six month period, whereas Q2 and Q4 results are posted later, meaning that the market might not have incorporated the leverage levels in the CDS spreads. It should be noted that both Benbouzid et al. (2017) and Smales (2016) also find this variable to be insignificant.

The second variable found insignificant in our baseline model is *efficiency*. Similar to the case of leverage, the OPEX/Revenue is an accounting ratio that is published at a later point in time than the corresponding date we have used to collect the CDS spreads. Of the three studies including the efficiency ratio, Hasan et al. (2016), Samaniego-Medina et al. (2016) and Benbouzid et al. (2017), only Benbouzid et al. (2017) find it significant.

Profitability is the third variable that is insignificant in the baseline model. The profitability variable is used by three studies in the existing literature. Both Chiaramonte and Casu (2013) and Hasan et al. (2016) find it negatively related to the CDS spreads, whereas Samaniego-Medina et al. (2016) find it insignificant. A possible explanation for this is that neither Chiaramonte and Casu (2013) nor Hasan et al. (2016) incorporate *stock return* as variable in their model whereas Samaniego-Medina et al. (2016) do. This is supported by Alagaam (2019) who finds that the ROE is reflected in the stock prices of banks.

Mkt-Rf is also insignificant in our model. There are only two studies in the existing literature which include this variable, namely Smales (2016) and Koutmos (2018). While Koutmos (2018) gets significant negative coefficients for this variable across all quantiles (in accordance with the hypothesis of the literature), Smales (2016) gets a positive and significant coefficient in his panel regression. Smales (2016), however, gets a negative and significant coefficient for the post financial crisis period. Given that there are only two papers with opposing results that study the impact of this variable, it is difficult to make any conclusions on the significance and sign of the *Mkt-Rf* factor based on the literature. The fact that it is insignificant in our model may be because our model includes variables that capture the effect of *Mkt-Rf*. For instance, neither Koutmos (2018) nor Smales (2016) include a proxy for *market return* in their models, whereas we do.

We find *SMB* significant with a negative sign, in line with the theory presented in the literature, but different from empirical findings in existing literature. Both Koutmos (2018) and Guesmi et al. (2018) find that SMB is insignificant. Smales (2016) finds a significant and positive relationship between the SMB factor and CDS spreads. We note that the study of Smales (2016) is not as rigorous as ours. He studies CDS spreads from 2006 to 2010, and the data set only consists of 10 banks, whereas the articles in our literature study analyze around 50 banks on average. Seeing as only three articles have studied the *SMB* factor in relation to bank CDS spreads, with varying results, we argue that there is not sufficient research done on the significance and sign of the *SMB* factor. Our result is the first to show a significant negative relationship between bank CDS spreads and the *SMB* factor. This is in line with the theory stated in Chapter 5.

Interestingly, Table 7.2 shows that while three of four accounting variables are insignificant, all market variables are strongly significant. In this context, it seems evident that the participants in the CDS market base their decisions on high-frequency financial variables such as interest rates and equity prices. These variables contain information in real time that reflects future economic trends. Accounting data on the other hand are available at lower frequencies (i.e. quarterly) and are more appropriate to describe the past economic performance.

The baseline model suggests that the *risk free rate* and the *size* are most important, weighing these variables most in the model. Also the *crisis* variable, the *stock return* and the *yield curve* are given high weights. Again, this strengthens the case of the market variables with *risk free rate*, *yield curve* and *crisis* all providing information about general market conditions. This suggests that weakening market conditions generally will increase bank CDS spreads, irrespective of differences or relative performance between banks.

In what follows, we use the Akaike Information Criterion (AIC) to compare our models. AIC is an estimate of relative quality of statistical models for a given set of data. More precisely, AIC estimates the amount of information lost by a model. In doing so, AIC deals with the trade-off between the goodness of fit of the model and the simplicity of the model. Thus, AIC provides a means for model selection. Burnham and Anderson (2002) state that AICs should only be compared by their difference, not the absolute value, and propose a rescaling: $\Delta_i = AIC_i - AIC_{min}$. Here AIC_i is the model under question, while AIC_{min} is the best model yet. They outline some simple rules of thumb with regards to model preference; Models having $\Delta_i < 2$ have substantial support, $4 < \Delta_i < 7$ have considerably less support, and models having $\Delta_i > 10$ have essentially no support at all. This model has $AIC = 1303.8$, and consequently takes its place as the best model yet. Thus, $AIC_{min} = 1303.8$. In what follows, we will calculate Δ_i's in order to compare our models.

We also report adjusted R^2. In this model, the $AdjR^2 = 0.8150$. The existing literature report R^2 between 0.20 and 0.87 in comparable models. Only Sclip et al. (2019) and Benbouzid et al. (2017) report higher R^2 than us, with 0.87 and 0.85 respectively, in their most comparable models. For these models they have 841 and 130 bank CDS spread observations, significantly less than our 1334 observations.

7.2 Model 2: Variable Selection on Baseline Model

We continue by applying a data-driven approach to selecting the variables from our baseline model described in Equation 7.1. For this, we use two variable selection techniques, the LASSO and the SFFS method. The LASSO excludes both the *efficiency* and *Mkt-Rf* variable, which, interestingly, were also deemed insignificant in the baseline model. Similar to LASSO, SFFS removes the *efficiency* variable, while it opts to keep the *Mkt-Rf*. SFFS also removes *leverage* and *profitability*. In doing so, the SFFS method further supports the importance of market variables, as opposed to accounting variables.

Equation 7.2 specifies the LASSO model.

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 PROF_{it} + \beta_3 SIZE_{it} \\ &+ \beta_4 SVOL_{it} + \beta_5 SRET_{it} \\ &+ \beta_6 RF_t + \beta_7 YLD_t + \beta_8 MVOL_t + \beta_9 MRET_t + \beta_{10} CRIS_t \\ &+ \beta_{11} SMB_t + \beta_{12} HML_t + u_{it} \end{aligned} \tag{7.2}$$

Table 7.3: Results Model 2: Variable selection on baseline model. N.S. refers to not selected. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

Model	LASSO		SFFS	
Firm-specific				
Accounting				
LEV	-0.043		N.S.	
	(0.036)			
EFF	N.S.		N.S.	
PROF	0.019		N.S.	
	(0.013)			
SIZE	0.609	***	0.564	***
	(0.071)		(0.066)	
Market				
SVOL	0.042	**	0.038	**
	(0.017)		(0.017)	
SRET	-0.169	***	-0.166	***
	(0.017)		(0.017)	
Market				
RF	-0.642	***	-0.662	***
	(0.029)		(0.024)	
YLD	-0.147	***	-0.169	***
	(0.024)		(0.026)	
MVOL	0.093	***	0.113	***
	(0.017)		(0.023)	
MRET	-0.104	***	-0.098	***
	(0.020)		(0.019)	
CRIS	0.189	***	0.191	***
	(0.027)		(0.027)	
Fama French				
MKT	N.S.		0.023	
			(0.019)	
SMB	-0.026	**	-0.029	**
	(0.013)		(0.013)	
HML	-0.040	***	-0.042	***
	(0.012)		(0.011)	
Intercept	4.178	***	4.177	***
	(0.015)		(0.015)	
Hausmann test	0.00		0.00	
AIC	1303.0		1302.5	
Adj. R^2	0.8148		0.8146	

Equation 7.3 specifies the SFFS model.

$$
\begin{aligned}
CDS_{it} = \alpha &+ \beta_1 SIZE_{it} \\
&+ \beta_2 SVOL_{it} + \beta_3 SRET_{it} \\
&+ \beta_4 RF_t + \beta_5 YLD_t + \beta_6 MVOL_t + \beta_7 MRET_t + \beta_8 CRIS_t \\
&+ \beta_9 MKTt + \beta_{10} SMB_t + \beta_{11} HML_t + u_{it}
\end{aligned}
\tag{7.3}
$$

The results from the regressions are shown in Table 7.3, and are very similar to our baseline model. In fact, the signs and significance are unchanged for all variables. We therefore refer to Section 7.1 for a discussion on these variables.

Further, we observe that the AIC decreases for both models, from $AIC_{min} = 1303.8$ to $AIC_{LASSO} = 1303.0$ and $AIC_{SFFS} = 1302.5$. This implies that $AIC_{min} = 1302.5$. These are small improvements and we can not say for certain that the new models are better than the previous.

Looking at the adjusted R^2, it decreases marginally for both models, from 0.8150 in the baseline model to $AdjR^2{}_{LASSO} = 0.8148$ and $AdjR^2{}_{SFFS} = 0.8146$.

7.3 Model 3: Adding Political and Policy Variables

We extend both the previous models by including the political and policy variables discussed in Chapter 5. Equation 7.4 specifies the extended LASSO model.

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 PROF_{it} + \beta_3 SIZE_{it} \\ &+ \beta_4 SVOL_{it} + \beta_5 SRET_{it} \\ &+ \beta_6 RF_t + \beta_7 YLD_t + \beta_8 MVOL_t + \beta_9 MRET_t + \beta_{10} CRIS_t \\ &+ \beta_{11} SMB_t + \beta_{12} HML_t \\ &+ \beta_{13} POLT_t + \beta_{14} POLC_t + u_{it} \end{aligned} \tag{7.4}$$

Table 7.4: Results Model 3: Adding political and policy variables. Baseline model including variable selection, and political and policy variables. Blanks indicate that the variable selection method excludes the variable from the model. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

Model	LASSO	SFFS
Firm-specific		
Accounting		
LEV	- 0.037	
	(0.036)	
EFF		
PROF	0.020	
	(0.013)	
SIZE	0.594 ***	0.548 ***
	(0.071)	(0.066)
Market		
SVOL	0.033 *	0.029 *
	(0.017)	(0.017)
SRET	- 0.167 ***	-0.163 ***
	(0.016)	(0.016)
Market		
RF	- 0.601 ***	-0.620 ***
	(0.030)	(0.026)
YLD	- 0.091 ***	-0.117 ***
	(0.027)	(0.029)
MVOL	0.086 ***	0.119 ***
	(0.017)	(0.022)
MRET	- 0.105 ***	-0.010 ***
	(0.020)	(0.019)
CRIS	0.152 ***	0.155 ***
	(0.028)	(0.028)
Fama French		
MKT		0.044 **
		(0.019)
SMB	- 0.059 ***	-0.067 ***
	(0.015)	(0.016)
HML	- 0.058 ***	-0.061 ***
	(0.012)	(0.013)
Political		
POLT	0.039 ***	0.041 ***
	(0.014)	(0.014)
POLC	0.054 ***	0.056 ***
	(0.013)	(0.014)
Intercept	4.192 ***	4.191 ***
	(0.015)	(0.015)
Hausmann test	0.00	0.00
AIC	1283.9	1280.1
Adj. R^2	0.8177	0.8199

Equation 7.5 specifies the extended SFFS model.

$$\begin{aligned} CDS_{it} = & \alpha + \beta_1 SIZE_{it} \\ & + \beta_2 SVOL_{it} + \beta_3 SRET_{it} \\ & + \beta_4 RF_t + \beta_5 YLD_t + \beta_6 MVOL_t + \beta_7 MRET_t + \beta_8 CRIS_t \\ & + \beta_9 MKTt + \beta_{10} SMB_t + \beta_{11} HML_t \\ & + \beta_{12} POLT_t + \beta_{13} POLC_t + u_{it} \end{aligned} \tag{7.5}$$

As seen from the results in Table 7.4, both *political stability* and *policy uncertainty* factors are significant, at the strongest level for both models, in explaining the CDS spreads. The results suggest that the political variables carry information relevant for bank CDS spreads, yet is not captured by variables studied in existing literature. Moreover, the signs of the coefficients for both the political and policy variables are positive, which is what we hypothesized in Section 5.2; Increasing political (in)stability and policy uncertainty generally increase the bank CDS spreads.

These results are in line with a growing body of research which has found political risk to impact different asset classes, such as equity markets, equity options market and firms' investment decisions (Dai and Zhang, 2019). As mentioned in Chapter 4, most empirical finance studies do not explicitly differentiate between the two. We perform a more rigorous study by including variables related both to political and policy uncertainty. Our results show that both political (in)stability *and* policy uncertainty have a significant impact on the CDS spreads, where an increase in the factors entails heightened credit risk.

Looking at the non-political variables, in the LASSO model all maintain their signs and significance from the preceding model. For the SFFS method, *Mkt-Rf* is now significant with a positive coefficient.

The AIC decreases for both models, from the old $AIC_{min} = 1302.5$ to $AIC_{POL\_LASSO} = 1283.9$ and $AIC_{POL\_SFFS} = 1280.1$. The new AIC_{min} is therefore $AIC_{min} = 1280.1$, obtained by adding political and policy variables to the optimal set of variables proposed by SFFS. This implies that $\Delta_{LASSO} = 19.1$ from the previous model, while $\Delta_{SFFS} = 22.4$. These are substantial improvements yielding no support for the previous models, and suggests that when including political and policy variables, the relative quality of the model increases significantly.

Also the adjusted R^2 improves, increasing from $AdjR^2{}_{LASSO} = 0.8148$ to $AdjR^2{}_{POL\_LASSO} = 0.8177$, and $AdjR^2{}_{SFFS} = 0.8146$ to $AdjR^2{}_{POL\_SFFS} = 0.8199$, suggesting the best fit among the models considered so far.

7.4 Model 4: Adding News Sentiment Variable

In Chapter 4, we emphasized that political stability and policy uncertainty may be closely related to news sentiment. Therefore, to investigate whether political stability and policy uncertainty have explanatory power that cannot be attributed to the effect of news sentiment, we extend our models by including a news sentiment variable.

The news sentiment data was introduced in Section 5.2, as seven different indices related to various types of news. In this thesis, we do not focus on the impact of distinct type of news. Instead, our main goal is to investigate the impact of a broader news sentiment, and the relation it has to the political and policy factors. Therefore, we perform a principal component analysis (PCA) on the news sentiment variables[14]. By performing a PCA, we can capture the variability in the news sentiment indices with fewer variables. In addition, the principal components represent a more general measure of news sentiment since they are a linear combination of all seven sentiment indices. Figure 7.1 shows the percentage of in-sample variance by each PCA component. We use the "elbow" method to select an appropriate number of principal components. The first principal component accounts for as much as ~83% of the in-sample variance, and we therefore opt to use one principal component in our model.

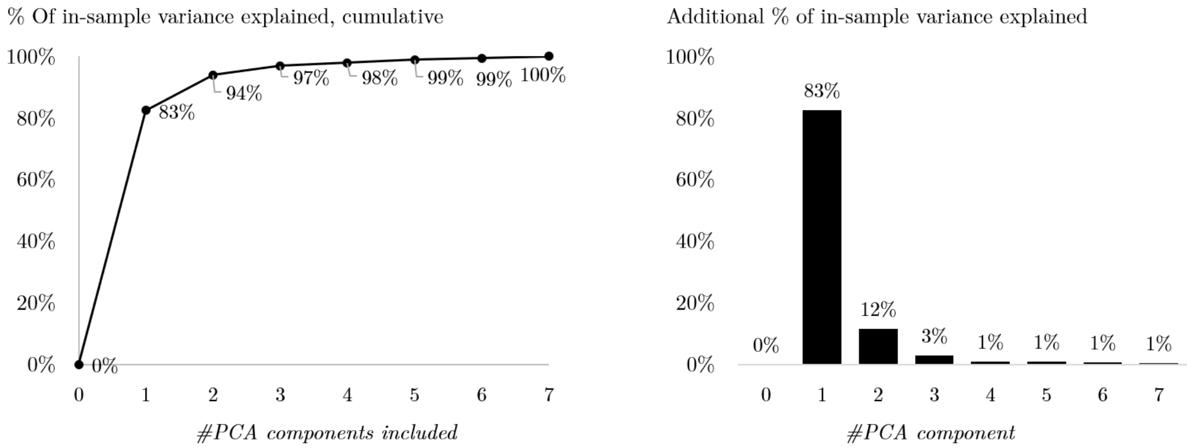

Figure 7.1: Percentage of in-sample variance explained by including PCA components.

For the LASSO, Equation 7.6 extends Equation 7.4 by including this component for measuring news sentiment.

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 PROF_{it} + \beta_3 SIZE_{it} \\ &+ \beta_4 SVOL_{it} + \beta_5 SRET_{it} \\ &+ \beta_6 RF_t + \beta_7 YLD_t + \beta_8 MVOL_t + \beta_9 MRET_t + \beta_{10} CRIS_t \\ &+ \beta_{11} SMB_t + \beta_{12} HML_t \\ &+ \beta_{13} POLT_t + \beta_{14} POLC_t \\ &+ \beta_{15} SENT_t + u_{it} \end{aligned} \tag{7.6}$$

[14]See Appendix D for a more thorough review of the theory behind PCA

The similar extension for the SFFS from Equation 7.5 is stated in Equation 7.7.

$$\begin{aligned} CDS_{it} = {} & \alpha + \beta_1 SIZE_{it} \\ & + \beta_2 SVOL_{it} + \beta_3 SRET_{it} \\ & + \beta_5 RF_t + \beta_4 YLD_t + \beta_6 MVOL_t + \beta_7 MRET_t + \beta_8 CRIS_t \\ & + \beta_9 MKTt + \beta_{10} SMB_t + \beta_{11} HML_t \\ & + \beta_{12} POLT_t + \beta_{13} POLC_t \\ & + \beta_{14} SENT_t + u_{it} \end{aligned} \tag{7.7}$$

Table 7.5: Results Model 4: Adding news sentiment variable. Baseline results including variable selection, political and policy variables, and news sentiment principal component. Blanks indicate that the variable selection method excludes the variable from the model. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

Model	LASSO	SFFS
Firm-specific		
Accounting		
LEV	0.003	
	(0.035)	
EFF		
PROF	0.015	
	(0.013)	
SIZE	0.532 ***	0.511 ***
	(0.069)	(0.064)
Market		
SVOL	0.074 ***	0.075 ***
	(0.017)	(0.017)
SRET	-0.165 ***	-0.160 ***
	(0.016)	(0.016)
Market		
RF	-0.415 ***	-0.415 ***
	(0.037)	(0.035)
YLD	0.033	0.010
	(0.031)	(0.032)
MVOL	0.110 ***	0.157 ***
	(0.017)	(0.022)
MRET	-0.078 ***	-0.075 ***
	(0.020)	(0.019)
CRIS	0.155 ***	0.164 ***
	(0.028)	(0.027)
Fama French		
MKT		0.059 **
		(0.019)
SMB	-0.068 ***	-0.079 ***
	(0.015)	(0.019)
HML	-0.044 ***	-0.046 ***
	(0.012)	(0.012)
Political		
POLT	0.048 ***	0.051 ***
	(0.013)	(0.013)
POLC	0.068 ***	0.075 ***
	(0.013)	(0.013)
Sentiment		
SENT	0.154 ***	0.160 ***
	(0.019)	(0.019)
Intercept	4.191 ***	4.187 ***
	(0.015)	(0.015)
Hausmann test	0.00	0.00
AIC	1217.2	1206.5
Adj. R^2	0.8268	0.8298

Table 7.5 shows the results of these models. First, we observe that both the polit-

ical and policy factors remain significant with positive sign in both cases. This is very interesting, since it suggests that the explanatory power of *political stability* and *policy uncertainty* cannot be attributed to the effect of *news sentiment*. Thereby, we are able to confirm that *political stability* and *policy uncertainty* have important and distinct explanatory power on bank CDS spreads.

Furthermore, the *news sentiment* variable is significant in both models, suggesting that also this is an important determinant of bank CDS spreads. The sign of the coefficient is positive, indicating that news indicating heightened risk yields higher CDS spreads.

Again, AIC decreases for both models from the old $AIC_{min} = 1280.1$ to $AIC_{SENT\_LASSO} = 1217.2$ and $AIC_{SENT\_SFFS} = 1206.5$. This implies that the new $AIC_{min} = 1206.5$, when adding a news sentiment variable to the variable set proposed by SFFS. $\Delta_{LASSO} = 66.7$ from the political model, while $\Delta_{SFFS} = 73.6$. These are large improvements suggesting that *news sentiment* has important additional explanatory power on bank CDS spreads, and that a model containing *political stability*, *policy uncertainty* and *news sentiment* is preferred.

This is also supported by the adjusted R^2 which increases from $AdjR^2{}_{POL\_LASSO} = 0.8177$ to $AdjR^2{}_{SENT\_LASSO} = 0.8268$, and $AdjR^2{}_{POL\_SFFS} = 0.8199$ to $AdjR^2{}_{POL\_SFFS} = 0.8298$.

Among the other traditional financial variables, the inclusion of news sentiment results in the yield curve becoming insignificant in both models. Therefore, information contained in the yield curve variable might be captured in the sentiment variable. This argument is in line with the work of Gotthelf and Uhl (2019), which finds that sentiments from news articles can explain and predict movements in the term structure of U.S. government bonds.

7.5 Model 5A and 5B: Robustness Tests

To test the robustness of our results, we perform two robustness tests. The first test investigates whether the political and policy variables are selected by LASSO and SFFS when they can select freely from all our variables. The idea of the second test is to show that the model results are consistent regardless of which of the 46 banks we use as input. In the following we present the tests and the results obtained from running them.

7.5.1 Model 5A: Variable selection test

The first test is done by running variable selection on *all* available variables. The LASSO method removes *efficiency*, *profitability*, *yield curve* and *Mkt-Rf*, while the SFFS method exclude *leverage*, *efficiency*, *profitability* and *yield curve*. Equation 7.8 states the LASSO model.

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 LEV_{it} + \beta_2 SIZE_{it} \\ &+ \beta_3 SVOL_{it} + \beta_4 SRET_{it} \\ &+ \beta_5 RF_t + \beta_6 MVOL_t + \beta_7 MRET_t + \beta_8 CRIS_t \\ &+ \beta_9 SMB_t + \beta_{10} HML_t \\ &+ \beta_{11} POLT_t + \beta_{12} POLC_t \\ &+ \beta_{13} SENT_t + u_{it} \end{aligned} \tag{7.8}$$

Equation 7.9 states the SFFS model.

$$\begin{aligned} CDS_{it} = \alpha &+ \beta_1 SIZE_{it} \\ &+ \beta_2 SVOL_{it} + \beta_3 SRET_{it} \\ &+ \beta_4 RF_t + \beta_5 MVOL_t + \beta_6 MRET_t + \beta_7 CRIS_t \\ &+ \beta_8 MKT_t + \beta_9 SMB_t + \beta_{10} HML_t \\ &+ \beta_{11} POLT_t + \beta_{12} POLC_t \\ &+ \beta_{13} SENT_t + u_{it} \end{aligned} \tag{7.9}$$

Table 7.6 shows that the political and policy variables are included by both variable selection methods. They are also significant with correct sign in the regression. Again, this underlines their importance in explaining bank CDS spreads, even when including *all* variables studied in existing literature and a sentiment variable. Moreover, the sentiment variable is also included and is significant in the regression. Again, this solidifies the argument that also news sentiment affects spreads.

The *yield curve*, *efficiency*, and *profitability* variables are removed by both the LASSO and the SFFS method. The former, as discussed in Section 7.4, is in line with the findings of Gotthelf and Uhl (2019) who show that news sentiment help explain the yield curve, meaning that information contained in the *yield curve* variable may be captured by our *sentiment* variable. Recall that *efficiency* was deemed insignificant in our first model presented in Section 7.1. There, we argued that the reason for this may be that the publication of accounting data lags behind the corresponding CDS spreads. Further, *Profitability* was also left out by SFFS in Equation 7.3, while it was deemed insignificant in the regression with the LASSO technique. Recall from Section 7.1 that none of the previous studies that included this variable also included a *stock return* variable, which has been shown to capture some of the similar effects.

Table 7.6: Results Model 5A: Variable selection test. Variable selection on all variables in this study. N.S. refers to not selected. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

Model	LASSO	SFFS
Firm-specific		
Accounting		
LEV	0.006	N.S.
	(0.035)	
EFF	N.S.	N.S.
PROF	N.S.	N.S.
SIZE	0.525 ***	0.511 ***
	(0.069)	(0.064)
Market		
SVOL	0.068 ***	0.075 ***
	(0.017)	(0.017)
SRET	- 0.161 ***	- 0.159 ***
	(0.016)	(0.016)
Market		
RF	- 0.443 ***	- 0.423 ***
	(0.021)	(0.019)
YLD	N.S.	N.S.
MVOL	0.108 ***	0.158 ***
	(0.017)	(0.022)
MRET	- 0.083 ***	- 0.077 ***
	(0.019)	(0.018)
CRIS	0.164 ***	0.166 ***
	(0.027)	(0.026)
Fama French		
MKT	N.S.	0.060 ***
		(0.018)
SMB	- 0.058 ***	- 0.078 ***
	(0.013)	(0.014)
HML	- 0.044 ***	- 0.046 ***
	(0.012)	(0.012)
Political		
POLT	0.042 ***	0.049 ***
	(0.012)	(0.012)
POLC	0.063 ***	0.074 ***
	(0.013)	(0.013)
Sentiment		
SENT	0.145 ***	0.158 ***
	(0.016)	(0.016)
Intercept	4.189 ***	4.187 ***
	(0.015)	(0.015)
Hausmann test	0.00	0.00
AIC	1215.9	1204.6
Adj. R^2	0.8267	0.8281

Like in Equation 7.2, we observe that the *Mkt-Rf* variable is excluded by LASSO. However, it is included by SFFS and has a significant and positive relationship with the CDS spreads, in contrast to what we found in Equation 7.1. As discussed in Section 7.1, Koutmos (2018) finds this variable to exhibit a significant and *negative* relationship with the CDS spreads, whereas Smales (2016) finds it positively related to the spreads.

Leverage is included by LASSO, but found insignificant in the regression, while SFFS does not select the variable. Recall from Section 7.1, that a possible explanation is that the information is captured by the *stock return* variable, as for the *profitability* variable. As discussed in Section 5.2, if the stock price decreases, the leverage measured in market value will increase.

Based on the AIC, both models are superior to other models presented in this thesis. $AIC_{LASSO} = 1215.9$ and $AIC_{SFFS} = AIC_{min} = 1204.6$, yielding $\Delta_{LASSO} = 1.3$ and $\Delta_{SFFS} = 12.6$. The last improvement provides substantial support that this is the best model in-sample. However, as discussed in Section 6.2, wrappers tend to overfit the training data, and it is a possibility that this model does so as well. In any way, these scores further support our previously mentioned arguments, namely that political and sentiment variables should be considered important determinants of bank CDS spreads, and that by using a data-driven approach to variable selection, one obtains simpler models with less redundancies and higher explanatory power. This is also supported by the adjusted R^2 which increases from $AdjR^2_{SFFS} = 0.8268$ to $AdjR^2_{SFFS} = 0.8281$, and is approximately unchanged for the LASSO.

7.5.2 Model 5B: In-sample robustness test

In order to show consistency within our results regardless of which banks are included in the models, we conduct an in-sample robustness test. The test is performed by randomly drawing 35 of the 46 banks and running the respective model on this subset instead of the full set[15]. Performing 100 iterations, the main goal here is to rule out that our results are driven by particularities of certain banks in our data set (i.e. outlier banks)[16].

The median parameter coefficients and corresponding standard errors of the 100 iterations are presented in Table 7.7. The test is performed on model 1 and model 2-4 with LASSO. Overall, these results show that the regressions results for each model are consistent. This implies that there are a substantial number of banks pushing the model to produce the same results, and not a few essential banks that must be included in order to obtain the results in the models.

The *stock volatility* variable is insignificant in the robustness test on the political/policy model, whilst it is significant when running on the full sample of banks. This implies that for a few banks there is enough evidence to reject the null hypothesis, and that these banks are essential in order to get a significant *stock volatility* variable. A possible explanation for this could be that the banks in question have very high volatility compared to the others, and that the corresponding CDS data for these banks is high when the volatility is high. These "outlier" banks will have impacted the standardization process by pushing the other, non-outlier banks' stock volatility towards the mean (i.e. 0). The larger part of the standard deviation would have been explained by the outlier banks, making the others less important, and consequently the *stock volatility* variable insignificant when the outlier banks are excluded.

It is important to note that this is an *in-sample test* and thus only test our selection of banks. It cannot immediately be extended to account for new banks not originally in the data set. However, Figure 3.3 shows that Europe and US alone account for 80% of the global CDS trading in single-name entities. We include the vast majority of the listed banks with CDSs in these markets.

[15]We draw 35 banks in order to keep N sufficiently larger than T, avoiding stationarity issues

[16]We do no more than 100 iterations in order to avoid convergence to the original data set, and no less in order to get a reasonable probability for including possible outliers.

Table 7.7: Results Model 5B: In-sample robustness test on the baseline model and the three LASSO models. 35 banks are drawn from the set of 46, and each model is run on the sub-sample. This process is done 100 times. Model 1-4 stands for the baseline model (1), baseline model with LASSO selection (2), political/policy model with LASSO (3), and the sentiment model with LASSO (4). N.S. = Not Selected. ***significant at 1%-level, **significant at 5%-level, *significant at 10%-level.

Model	1		2		3		4	
Firm-specific								
Accounting								
LEV	-0.056		-0.054		-0.051		-0.001	
	(0.044)		(0.044)		(0.043)		(0.043)	
EFF	-0.013		N.S.		N.S.		N.S.	
	(0.012)							
PROF	0.019		0.023		0.020		0.016	
	(0.015)		(0.015)		(0.015)		(0.015)	
SIZE	0.614	***	0.624	***	0.626	***	0.523	***
	(0.089)		(0.089)		(0.089)		(0.088)	
Market								
SVOL	0.037	**	0.038	*	0.025		0.072	***
	(0.020)		(0.020)		(0.020)		(0.020)	
SRET	-0.165	***	-0.167	***	-0.168	***	-0.163	***
	(0.020)		(0.020)		(0.020)		(0.020)	
Market								
RF	-0.638	***	-0.643	***	-0.582	***	-0.404	***
	(0.035)		(0.035)		(0.035)		(0.044)	
YLD	-0.165	***	-0.166	***	-0.094	***	0.035	
	(0.032)		(0.032)		(0.032)		(0.037)	
MVOL	0.117	***	0.118	***	0.089	***	0.110	***
	(0.027)		(0.026)		(0.020)		(0.020)	
MRET	-0.110	***	-0.110	***	-0.119	***	-0.083	***
	(0.024)		(0.024)		(0.024)		(0.024)	
CRIS	0.201	***	0.200	***	0.163	***	0.161	***
	(0.032)		(0.032)		(0.033)		(0.033)	
Fama French								
MKT	0.027		N.S.		N.S.		N.S.	
	(0.023)							
SMB	-0.028	**	-0.029	*	-0.055	***	-0.067	***
	(0.015)		(0.015)		(0.018)		(0.017)	
HML	-0.037	***	-0.038	***	-0.056	***	-0.041	***
	(0.013)		(0.013)		(0.014)		(0.014)	
Political								
POLT					0.036	**	0.049	***
					(0.016)		(0.016)	
POLC					0.054	***	0.068	***
					(0.016)		(0.015)	
Sentiment								
SENT							0.161	***
							(0.022)	
Intercept	4.190	***	4.180	***	4.201	***	4.200	***
	(0.018)		(0.018)		(0.018)		(0.018)	

Chapter 8

Conclusion

In this thesis, we study the determinants of CDS spreads in the banking industry. Using semi-annual data on 46 banks from 2005 to 2019, we analyze a comprehensive set of variables included in historical studies conducted on bank CDS spreads, in addition to three novel variables related to political stability, policy uncertainty, and news sentiment.

Our primary conclusion is that political stability and policy uncertainty are significant drivers of bank CDS spreads, where increased (in)stability and uncertainty increase the spreads. This has, to the best of our knowledge, not yet been shown in existing studies. Secondly, this finding holds also when including a news sentiment variable, underlining the political factors' importance. Thirdly, our data-driven approach to variable selection removes redundant variables previously found important determinants of bank CDS spreads in the literature. A fourth finding is that market variables in general are more important than accounting variables. Finally, we found that with our data-driven approach, we obtain simpler models with less redundancies and higher explanatory power measured by AIC and Adjusted R^2.

Our thesis offers several interesting directions for further research. First, it would be interesting to include banks from a wider geography. Our results are obtained using European, US, Canadian, and Australian banks, yet excludes banks from e.g. Asia. Employing a wider geography could yield interesting insights into the behaviour of bank CDS spreads in non-developed countries. For example, China opened for CDS trading in 2016 and may be an interesting area of research going forward.

A substantial body of research has found news sentiment to impact a wide array of financial variables. However, no previous studies consider its effect on CDS spreads using a broad set of banks over a long time period. We find that news sentiment significantly explains variation in bank CDS spreads, and as such, further investigation into its impact is prompted. In particular, a promising direction for further research is the impact of more specific news topics.

Predictive models incorporating political and policy risk factors also provide opportunity for further research. In this thesis, we have studied the effect of different variables on the CDS spreads, but it would be interesting to incorporate these novel findings in larger predictive models with more data. A more comprehensive study on CDS spread prediction could be valuable for banks in terms of foreseeing periods of high credit risk, giving them an opportunity to adjust their risk-exposure accordingly.

References

Al-Maadid, A., Caporale, G.M., Spagnolo, F., Spagnolo, N., 2020. The impact of business and political news on the GCC stock markets. Research in International Business and Finance .

Alagaam, A., 2019. The Relationship Between Profitability and Stock Prices: Evidence from the Saudi Banking Sector. Journal of Finance and Accounting .

Annaert, J., De Ceuster Marc, M., Van Roy, P., Vespro, C., 2013. What determines euro area bank CDS spreads? Journal of International Money and Finance .

Arosa, C.M.V., Richie, N., Schuhmann, P.W., 2014. The impact of culture on market timing in capital structure choices. Research in International Business and Finance .

Augustin, P., Subrahmanyam, M.G., Tang, D.Y., Wang, S.Q., 2015. Credit Default Swaps: Past, Present, and Future. SSRN Electronic Journal .

Avino, D.E., Conlon, T., Cotter, J., 2019. Credit default swaps as indicators of bank financial distress. Journal of International Money and Finance .

Baker, N., Bloom, S., 2016. Measuring economic policy uncertainty. The Quarterly Journal of Economics .

Baltagi, B.H., 2013. Econometric Analysis of Panel Data. 5th ed., John Wiley & Sons.

Bank for International Settlements, 2018. The credit default swap market: What a difference a decade makes. Technical Report. Bank for International Settlements.

Baselga-Pascual, L., Trujillo-Ponce, A., Cardone-Riportella, C., 2015. Factors influencing bank risk in Europe: Evidence from the financial crisis. North American Journal of Economics and Finance .

Benbouzid, N., Mallick, S.K., Sousa, R.M., 2017. Do country-level financial structures explain bank-level CDS spreads? Journal of International Financial Markets, Institutions and Money .

Blanco, R., Brennan, S., Marsh, I.W., 2005. American Finance Association An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps. The Journal of Finance .

Blinder, A.S., 2013. After the music stopped: the financial crisis, the response, and the work ahead. Penguin Press.

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2012. Which News Moves Stock Prices? A Textual Analysis. NBER Working Paper .

Brooks, C., 2019. Introductory Econometrics for Finance. Cambridge University Press.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference. Springer New York.

Calomiris, C.W., Mamaysky, H., 2019. Monetary Policy and Exchange Rate Returns: Time-Varying Risk Regimes. NBER Working Paper .

Cattell, R.B., 1966. The scree test for the number of factors. Multivariate Behavioral Research .

Chiaramonte, L., Casu, B., 2013. The determinants of bank CDS spreads: Evidence from the financial crisis. European Journal of Finance .

Čižmešija, M., Lolić, I., Sorić, P., 2017. Economic policy uncertainty index and economic activity: What causes what? Croatian Operational Research Review .

Collin-Dufresne, P., Goldstein, R.S., Martin, J.S., 2001. The Determinants of Credit Spread Changes. Journal of Finance .

Copelovitch, M., Frieden, J., Walter, S., 2016. The Political Economy of the Euro Crisis. Comparative Political Studies .

Dai, L., Zhang, B., 2019. Political Uncertainty and Finance: A Survey. Asia-Pacific Journal of Financial Studies .

De Jonghe, O., 2010. Back to the basics in banking? A micro-analysis of banking system stability. Journal of Financial Intermediation .

Drago, D., Tommaso, C.D., Thornton, J., 2017. What determines bank CDS spreads? Evidence from European and US banks. Finance Research Letters .

Duffy, S., Smith, J., 2020. Omitted-variable bias and other matters in the defense of the category adjustment model: A comment on Crawford (2019). Journal of Behavioral and Experimental Economics .

Eichhorst, W., Escudero, V., Marx, P., Tobin, S., 2010. The Impact of the Crisis on Employment and the Role of Labour Market Institutions. IZA Discussion Paper .

Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C., 2001. Explaining the Rate Spread on Corporate Bonds. The Journal of Finance .

Ericsson, J., Jacobs, K., Oviedo, R., 2009. The determinants of credit default swap premia. Journal of Financial and Quantitative Analysis .

Estrella, A., Mishkin, F.S., 1997. The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank. European Economic Review .

Eurostat, . Eurostat Online Database. URL: `https://ec.europa.eu/eurostat/web/government-finance-statistics/data/database`.

Evans, P., Hasan, I., Lozano-Vivas, A., 2008. Deregulation and convergence of Banking: The EU experience. Finnish Economic Papers .

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics .

Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. Journal of Finance .

Gala, V., Pagliardi, G., Zenios, S.A., 2018. Politics, Policy, and International Stock Returns. SSRN Electronic Journal .

Galil, K., Shapir, O.M., Amiram, D., Benzion, U., 2013. The Determinants of CDS Spreads. SSRN Electronic Journal .

Gerhardt, M., Vennet, R.V., 2017. Bank bailouts in Europe and bank performance. Finance Research Letters .

GlobalData, 2019. Oil and gas sector tops Forbes 2019 list of world's largest public firms in revenue generation, finds GlobalData. URL: `shorturl.at/wxIK3`.

Gotthelf, N., Uhl, M.W., 2019. News Sentiment: A New Yield Curve Factor. Journal of Behavioral Finance .

Guesmi, K., Dhaoui, A., Goutte, S., Abid, I., 2018. On the determinants of industry-CDS index spreads: Evidence from a nonlinear setting. Journal of International Financial Markets, Institutions and Money .

Hasan, I., Liu, L., Zhang, G., 2016. The Determinants of Global Bank Credit-Default-Swap Spreads. Journal of Financial Services Research .

Hillert, A., Jacobs, H., Müller, S., 2012. Media Makes Momentum. SSRN Electronic Journal .

Hilscher, J., Wilson, M., 2016. Credit ratings and credit risk: Is one measure enough? Management Science .

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology .

Huang, C., Simpson, S., Ulybina, D., Roitman, A., 2019. News-based Sentiment Indicators. International Monetary Fund .

Huang, J.Z., Huang, M., 2012. How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk? The Review of Asset Pricing Studies .

ISDA, 2019. Global Credit Default Swaps Market Study. Technical Report. ISDA.

Jolliffe, I.T., 2002. Principal Component Analysis. Encyclopedia of Statistics in Behavioral Science .

Kaviani, M.S., Kryzanowski, L., Maleki, H., Savor, P., 2017. Policy Uncertainty and Corporate Credit Spreads. Fox School of Business Research Paper .

Kouki, M., Belhadj, R., Chikhaoui, M., 2017. Impact of Financial Crisis on GDP Growth: The Case of Developed and Emerging Countries. International Journal of Economics and Financial Issues .

Koutmos, D., 2018. Asset pricing factors and bank CDS spreads. Journal of International Financial Markets, Institutions and Money .

Kuhn, M., Johnson, K., 2019. Feature Engineering and Selection: A Practical Approach for Predictive Models. Chapman and Hall/CRC.

Liu, J., Zhong, R., 2017. Political uncertainty and a firm's credit risk: Evidence from the international CDS market. Journal of Financial Stability .

Longstaff, F.A., Mithal, S., Neis, E., 2005. Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market. Journal of Finance .

Longstaff, F.A., Schwartz, E.S., 1995. A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. Journal of Finance .

Mátyás, L., Sevestre, P., 2008. The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Springer Berlin Heidelberg.

Merton, R.C., 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The Journal of Finance .

Morkoetter, S., Stebler, R., Westerfeld, S., 2017. Competition in the credit rating Industry: Benefits for investors and issuers. Journal of Banking and Finance .

Nikkei, 2016. China debuts first credit default swap market. URL: `https://asia.nikkei.com/Business/Finance/China-debuts-first-credit-default-swap-market`.

North, D.C., 1991. Institutions. Journal of Economic Perspectives .

Oehmke, M., Zawadowski, A., 2016. The Anatomy of the CDS Market. The Review of Financial Studies .

Park, G., Lee, H.Y., 2018. Opportunistic behaviors of credit rating agencies and bond issuers. Pacific Basin Finance Journal .

Pástor, , Veronesi, P., 2012. Uncertainty about Government Policy and Stock Prices. The Journal of Finance .

Pástor, , Veronesi, P., 2013. Political uncertainty and risk premia. Journal of Financial Economics .

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science .

Qu, Z., 2020. The Impact of Income Diversification on Chinese Banks: Bank Performance. Springer.

Rambaccussing, D., Kwiatkowski, A., 2020. Forecasting with news sentiment: Evidence with UK newspapers. International Journal of Forecasting .

Ramskogler, P., 2015. Tracing the origins of the financial crisis. OECD Journal: Financial Market Trends 2014, 47–61.

Raunig, B., Scheicher, M., 2009. Are Banks Different? Evidence from the CDS Market. Working Papers, Oesterreichische Nationalbank (Austrian Central Bank) .

Samaniego-Medina, R., Trujillo-Ponce, A., Parrado-Martínez, P., di Pietro, F., 2016. Determinants of bank CDS spreads in Europe. Journal of Economics and Business .

Sclip, A., Girardone, C., Miani, S., 2019. Large EU banks' capital and liquidity: Relationship and impact on credit default swap spreads. British Accounting Review .

Smales, L.A., 2016. News sentiment and bank credit risk. Journal of Empirical Finance .

Stangl, A., 2007. European Data Watch: Ifo World Economic Survey Micro Data. Journal of Applied Social Science Studies .

Stock, J.H., Watson, M.W., 2018. Introduction to Econometrics. 4th ed., Pearson.

Tetlock, P., 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. The Journal of Finance .

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society .

Tsai, F.T., Lu, H.M., Hung, M.W., 2016. The impact of news articles and corporate disclosure on credit risk valuation. Journal of Banking and Finance .

Wang, X., Xu, W., Zhong, Z.K., 2018. Economic policy uncertainty, CDS spreads, and CDS liquidity provision. Journal of Futures Markets .

Yongjun Tang, D., Yan, H., 2008. Liquidity and credit default swap spreads. AFA 2007 Chicago Meetings Paper .

Appendices

Appendix A

Regulatory Overview

An overview of the most important regulatory changes implemented following the Financial Crisis and the Euro Crisis, including the Basel III requirements and the European Banking Union, is given in Figure A.1.

Requirement	**Description**	**Implementation**
EU Basel III Capital Requirements	Requires banks to hold a higher amount of minimum capital ratio	Started implemented in 2014
EU Basel III Leverage Ratio	Hinders banks in taking on too much debt	Started implementation 2013
EU Basel III Liquidity Coverage ratio	Requires banks to hold high quality liquid assets to meet cash outflows for 30 days	At 100% coverage since 2019
EU Basel III Net Stable Funding Ratio (NSFR)	Requires banks to better match assets and liabilities. Requires stable funding over a 1 year period	Implemented by certain countries already
Banking Union (Single Supervisory Mechanism)	ECB takes over as single supervisor for largest Eurozone banks	Active from 2014
Minimum Requirement For Own Funds and Eligable Liabilities	Banks required to have loss-absorbing capacity to the degree that tax payers won't be exposed to losses	Currently ongoing
OTC Derivatives Reforms	Forces the most liquid OTC contracts onto electronic markets for more transparency	Largely implemented in the EU

Figure A.1: Overview of regulatory changes, post Euro-crisis

Appendix B

Data Description

B.1 Variable Proxies Used in the Existing Literature

Variable	Raunig and Scheicher (2009)	Annaert et al. (2013)	Chiaramonte and Casu (2013)	Hasan et al. (2015)	Samaniego-Medina et al. (2016)	Smales (2016)	Drago et al. (2017)	Benbouzid et al. (2017)	Koutmos (2018)	Guesmi et al. (2018)	Sclip (2019)
Asset quality			Loan loss reserves/gross loans, Unreserved Impaired Loans/Equity	Loan loss provision	Impaired/Gross loans		Ratio of non-performing loans to total assets, Ratio of loan loss provisions to total loans	Impaired loans to equity,			Loan loss provisions/total assets, Loan loss reserves to non-performing loans, non-performing loans to total loans
Capitalization			Tier 1 ratio, Equity/total assets	Leverage ratio	Eq/Total assets ratio	Leverage ratio	Tier 1 capital ratio, leverage/total assets	Tier 2 capital, Leverage ratio			Basel III leverage, Equity to Total Assets, CET1 ratio, T2 capital ratio, Market capital ratio, Loans/Total assets
Profitability			ROA, ROE	ROE, interest expense/total liabilities	ROA			EBITA/Assets			ROE, ROA
Efficiency				Cost efficiency	Cost/income (CIR)						
Size					Total assets	Ln(Total Assets), Ln(Market Cap)	Ln(total assets)				Ln(total assets)
Funding stability							Deposits/total liabilities				NSFR 10, NSFR 14
Liquidity			Net Loans/Deposits and Short-Term Funding (%), Liquid Assets/Deposits and Short-Term Funding (%)	liquid assets to total assets	Interbank ratio, Net loans/total assets			Liquidity ratio			
Diversification											Interest income over total income
Liquidity CDS		Change in bid-ask spread on underlying CDS			Bid-ask spread (level)	Amihud Illiquidity Measure					
Stock return		Bank stock return (2 week) historical			Equity return	Daily stock return			Stock return		
Stock volatility		Bank volatility, StdDev (1 week) historical		Historical volatility of bank stock	Equity volatility	Daily stock return volatility			Stock volatility		60 days standard deviation of bank stock returns
Credit risk/rating	Moodys risk database: KMV	Difference between Merrill Lynch 5yr BBBB and AAA corporate yields			Dummy var (AAA and A)		Index ranging from 1 (Moody's rating Ca) to 17 (Aaa) on bank and sovereign	Economic risk rating, Financial Risk rating			
Housing								House prices			
Financial crisis			Dummy				Dummy				
Interbank risk									LIBOR-OIS		
Spot interest rate (economic condition)	Risk-free interest rate (5 yr)	2 yr gov. Yield (Eikon benchmark)		10 yr gov. bond yield	10-year Treasury bond rate	3M T-bill	Yield on 5-year government bonds			5yr US Treasury	
Yield curve (economic condition)	10yr minus 3month bond yield	10yr-5yr yield (Eikon benchmark)				10yr-3M (term spread)	10-year government bond yield less yield on 2-year Treasury bonds				The difference between the 10-year government bond yield for each country
Market volatility	VIX and VSTOXX	VSTOXX			VSTOXX	VIX	VIX and VSTOXX		VSTOXX, VXO, MOVE	VIX	
Fama French Factors						HML, SMB, Mkt-Rf			HML, SMB, Mkt-Rf, RMW, CMA	HML, SMB	
Market return		Eikon Euro area stock market index			STOXX50		STOXX50 (for European) and S&P500 (for US)			Financial Industry stock returns US	The natural logarithm of the Eurostoxx 600 index
Stock skew									Market and stock skew		
Stock kurtosis									Market and stock kurtosis		
Forex volatility									Forex volatility index		
Commodities									Commodities Index (JPM)		

Table B.1: Overview of the variable proxies used in existing literature on bank CDS spreads

B.2 IMF News Sentiment Indices

The countries for which IMF has created the 7 news sentiment indices are shown in Figure B.1. We only use the indices for Denmark, Finland, Norway, Spain and Sweden, and compute an average based on these. 9 of the 46 banks included in our sample are from these countries.

Countries for which sentiment variables are made			
Argentina	Denmark	Mexico	Sweden
Bolivia	Finland	Norway	Thailand
Brazil	Indonesia	Peru	Turkey
Chile	Israel	Philippines	Uruguay
Colombia	Malaysia	Spain	Venezuela

Figure B.1: The 20 countries for which IMF has created news sentiment indices. Grey cells indicate overlap with our geography.

B.3 Correlation Matrix

Figure B.2 shows the correlation matrix of all the variables included in our models.

	Leverage	Efficiency	Profitability	Size	Stock volatility	Stock return	Risk free rate	Yield curve	Market volatility	Market return	Crisis	MKT	SMB	HML	GPZ Politics	GPZ Policy	Sentiment	CDS
Leverage	1,00	0,37	0,04	0,33	0,26	0,02	0,21	0,00	0,24	-0,17	0,30	-0,12	0,02	0,11	-0,20	0,14	-0,39	-0,07
Efficiency	0,37	1,00	-0,03	-0,03	0,33	-0,12	0,15	0,02	0,22	-0,23	0,27	-0,12	-0,01	0,11	-0,17	0,12	-0,36	-0,02
Profitability	0,04	-0,03	1,00	-0,13	-0,29	0,29	0,45	-0,29	-0,22	0,18	-0,06	0,10	0,05	0,04	0,00	-0,10	-0,24	-0,40
Size	0,33	-0,03	-0,13	1,00	0,07	0,06	-0,17	0,15	0,09	-0,03	0,13	-0,02	-0,05	-0,04	-0,03	0,03	0,07	-0,14
Stock volatility	0,26	0,33	-0,29	0,07	1,00	-0,50	-0,37	0,35	0,71	-0,57	0,41	-0,46	0,10	-0,08	-0,06	0,37	-0,13	0,50
Stock return	0,02	-0,12	0,29	0,06	-0,50	1,00	0,51	-0,42	-0,43	0,63	-0,25	0,21	-0,05	0,06	0,06	-0,22	-0,22	-0,61
Risk free rate	0,21	0,15	0,45	-0,17	-0,37	0,51	1,00	-0,81	-0,41	0,44	-0,17	0,07	-0,15	0,11	0,01	-0,16	-0,56	-0,83
Yield curve	0,00	0,02	-0,29	0,15	0,35	-0,42	-0,81	1,00	0,41	-0,58	0,28	0,11	0,34	0,03	-0,22	0,15	0,17	0,65
Market volatility	0,24	0,22	-0,22	0,09	0,71	-0,43	-0,41	0,41	1,00	-0,60	0,32	-0,65	0,04	-0,05	-0,12	0,29	-0,06	0,54
Market return	-0,17	-0,23	0,18	-0,03	-0,57	0,63	0,44	-0,58	-0,60	1,00	-0,42	0,24	-0,17	-0,10	0,26	-0,33	0,02	-0,55
Crisis	0,30	0,27	-0,06	0,13	0,41	-0,25	-0,17	0,28	0,32	-0,42	1,00	-0,23	-0,11	0,01	0,14	0,28	-0,16	0,34
MKT	-0,12	-0,12	0,10	-0,02	-0,46	0,21	0,07	0,11	-0,65	0,24	-0,23	1,00	0,31	0,09	0,01	-0,25	0,01	-0,25
SMB	0,02	-0,01	0,05	-0,05	0,10	-0,05	-0,15	0,34	0,04	-0,17	-0,11	0,31	1,00	0,05	0,27	0,17	-0,08	0,04
HML	0,11	0,11	0,04	-0,04	-0,08	0,06	0,11	0,03	-0,05	-0,10	0,01	0,09	0,05	1,00	-0,15	0,32	-0,24	-0,15
GPZ Politics	-0,20	-0,17	0,00	-0,03	-0,06	0,06	0,01	-0,22	-0,12	0,26	0,14	0,01	0,27	-0,15	1,00	-0,03	0,15	-0,01
GPZ Policy	0,14	0,12	-0,10	0,03	0,37	-0,22	-0,16	0,15	0,29	-0,33	0,28	-0,25	0,17	0,32	-0,03	1,00	-0,12	0,26
Sentiment	-0,39	-0,36	-0,24	0,07	-0,13	-0,22	-0,56	0,17	-0,06	0,02	-0,16	0,01	-0,08	-0,24	0,15	-0,12	1,00	0,47
CDS	-0,07	-0,02	-0,40	-0,14	0,50	-0,61	-0,83	0,65	0,54	-0,55	0,34	-0,25	0,04	-0,15	-0,01	0,26	0,47	1,00

Figure B.2: Correlation matrix

Appendix C

SFFS Algorithm

Figure C.1: Pseudo code for SFFS variable selection technique. Line 7 describes the conditional exclusion (floating) step

	SFFS Algorithm for variable selection
	Input: $Y=\{y_1, y_2, \ldots, y_p\}$, where p is the number of variables included, and J is a scoring function
	Output: $X_k=\{x_j \mid j=1,2,\ldots,k;\ x_j \in Y\}$, where $k=(0,1,2,\ldots,p)$
1	$X_0=\emptyset$, $k=0$
2	**while** (k not equal to number of desired features) **do**:
3	$x^+ = \arg\max J(x_k+x)$, where $x \in Y-X_k$
4	$X_{k+1} = X_k+x^+$
5	$k = k+1$
6	$x^- = \arg\max J(x_k-x)$, where $x \in X_k$
7	**if** $J(x_k-x)>J(x_k-x)$ **then**
8	$X_k-1 = X_k-x^-$
9	$k = k-1$
10	**return** $X_k=\{x_j \mid j=1,2,\ldots,k;\ x_j \in Y\}$, the optimal set of variables to include

Appendix D

Principal Component Analysis (PCA)

The familiar algebraic form of PCA was presented by Hotelling (1933), however Pearson (1901) had earlier given a geometric derivation. PCA is a statistical method that transform a set of variables (possibly correlated) into a set of linearly independent variables (principal components). The main idea of PCA is to perform a linear transformation of the data in the high-dimensional space into a space of fewer dimensions (dimensionality reduction), while preserving as much of the variability in the data as possible.

More intuitively, PCA attempts to draw straight explanatory lines through the data (much like linear regression). Each straight line represents a principal component. There will be as many principal components as there are dimensions in the data (given that there is no perfect multicollinearity in the set of variables). PCA's role is then to prioritize the principal components. The first principal component is a straight line in a N-dimensional space that explains the most variance (i.e. reduces the error). The second principal component must then cut through the data perpendicular to the first line, fitting the errors produced by the first. The third component would fit the errors from the first and second principal components, and so on.

All principal components will have an associated eigenvector and eigenvalue. The eigenvector represent the direction of the principal component in the feature space, while the eigenvalue represent how the data spreads along this direction.

We let X be a matrix of shape $T \times K$. This matrix represents 7 column vectors K (the seven news indices) and 29 time periods, T. The mean for every column and the covariance of the whole data set was then computed. From these, the eigenvectors and corresponding eigenvalues can be computed. The eigenvectors are sorted by decreasing order of eigenvalues to form a $K \times K$ dimensional matrix W. The full principal component analysis can then be written as in Equation D.1 below.

$$T = XW \tag{D.1}$$

PCA solves the problem of multicollinearity by creating linearly independent principal components from the set of explanatory variables. In addition to handling multicollinearity, PCA provides the opportunity to reduce the dimensionality of the data, which may reduce the risk of overfitting. Dimensionality reduction through PCA is done by selecting a subset of the principal components, and perform a transformation using only these components.

An issue related to this is how to select a good number of principal components. Jolliffe (2002) suggests to base the decision on the percentage of total variance explained by the set of selected principal components. Typical threshold values are 80 % or 90 %, and so if the three first variables account for 90 % of the variation, it should suffice with three principal components. Another solution proposed by Cattell (1966) is to plot the value of the eigenvalues of the principal components (they will decrease for each component since higher components have less explanatory power). The idea is to select k principal components, where the slopes of lines joining the plotted points are 'steep' to the left of k, and 'not steep' to the right (i.e. an elbow is formed at k).

An issue with PCA is the difficulty in interpreting the representation of the principal components. In our case, however, having made sure that all indices measure increasing risk, the expected relationship with CDS spreads should be positive for all principal components as well. Each principal components is in fact a linear equation of seven specific news sentiment, and so each principal component can be interpreted as a broader measure of news sentiment risk.

On the other hand, a few principal components capture almost all variability in the data, and has no multicollinearity and reduces the number of dimensions (parameters to estimate). Therefore, including principal components in models could be a good way to optimize the trade-off between goodness of fit and simplicity. This was also seen in Section 7.4 where we included one principal component for news sentiment.

NTNU

Norwegian University of
Science and Technology