Credit rating agencies and the European sovereign debt crisis

Credit Rating Agencies and the European Sovereign Debt Crisis

This book was typeset by the author using LaTeX

Cover designed by Vera Bossel
Background image designed by Freepik

ISBN 978 94 6159 560 7
published by Datawyse - Universitaire Pers Maastricht

Credit Rating Agencies and the European Sovereign Debt Crisis

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday, May 12th, 2016 at 10.00 hrs

by

Lennart Freitag

Supervisor:
prof. dr. Bertrand Candelon

Assessment Committee:
prof. Dr. Joan Muysken (Chairman)
prof. Dr. Eric-Paget Blanc (Université d'Evry)
prof. Dr. Christophe Hurlin (Université d'Orléans)
prof. Dr. Stefan Straetmans

Acknowledgements

The process of writing a thesis is never a one man show, although in the end there is only one author printed on the thesis. This is the moment to thank all of those that supported and encouraged me during the time I wrote my PhD. I want to thank first and foremost my supervisor Bertrand. You always were supportive of my ideas, gave me guidance, especially during the first year of my PhD, but also gave me enough freedom to pursue my own research agenda.

I also want to extend my thanks to the AE2 department of SBE at Maastricht University. Special thanks go to our former head of department Joan Muysken, Erik de Regt and Tom van Veen with whom I taught multiple courses and of course our great secretaries Fleur, Silvana and Marion. Also I would like to thank Silvia for her work at the department. All of you were helpful and supportive, especially during my last year when Bertrand was already in Paris.

My friends from the research master occupy a special place throughout the last four years (and hopefully in the future as well). Aline, with whom I have been to Israel and had long talks, both serious and funny, over the last four years. Flo, who always motivated us for various activities (rowing, going to Pizza Hut, etc.). Oana, who always listened to you, and had good advice for all kind of situations. Nadine, watching Champions League with you is amazing. Gabri, your cooking events were great, and so was going out with you. Christine, we had great talks about everything, from research issues to personal things. All of you made the experience of writing a PhD so much more fulfilling.

I of course also want to thank my fellow PhD candidates from AE2. My former office mates Ehsan, Ron and Cai Cai, with whom I shared a room for one year. You gave me a great start into my PhD, helped me orientate myself and were overall amazing. Dominik, Tom and Phillip, who were my office mates from my second year on. We supported each other throughout the last few years. Also, thanks go out to Thomas Götz and Lenard Lieb with whom I discussed research matters, especially Chapter 3 of this thesis.

The support my Family gave me during my PhD was invaluable. Thank you Mom and dad, for being such amazing parents, also thanks to my siblings Marle and Jonas for everything you did for me during my PhD.

Lastly, I want to thank my girlfriend Vera. We got to know each other at the beginning of my PhD, at the end of 2011. Ever since, you were always there for me. You supported me during both good and bad times, on a personal as well as on an academic level. I doubt this thesis would be in the form it is right now without you. You are an amazing person and I'm looking forward to the moment when we'll again live in the same city.

Lennart Freitag, Gießen, 11.4. 2016

Contents

Chapter 1

Introduction

"There are two superpowers in the world today in my opinion. There's the United States and there's Moody's Bond Rating Service."

Thomas Friedman, Journalist, Pulitzer price winner, and Author (1996)

Credit Rating Agencies (CRAs) are corporations that provide an opinion on the probability of default of firms, financial assets, and countries. Although the agencies stress that their opinion is not an investment advice,[1] this is treated as such throughout most of the world's financial markets.

Recently, the CRAs have come under severe criticism for their role in the financial crisis in 2008 as well as their role during the escalation of the European sovereign debt crisis. During the former, rating agencies gave high grades to structured financial vehicles such as Collateralized Debt Obligations (CDO), calling them essentially a safe investment. The problem with this was that while initially most of these CDOs were of good quality, they deteriorated severely in 2007. However, the CRAs chose to still give them high ratings, and a lot of investment decisions were based on these grades. These vehicles defaulted in big numbers during the 2008 financial meltdown,[2] leading to massive bank bailouts in a sizable number of countries throughout the whole world. The second wave of criticism was triggered by decisions made by the agencies from 2010 onwards with regard to country ratings in Europe. This came mainly from European politicians who raised complaints regarding sudden and violent downgrades. It went as far as policymakers circulating the idea of an European rating agency, see for example Bartels and Weder (2013).

To understand why there was such a discussion and outrage about the decisions of the CRAs, we need to understand their history and how they are embedded in the financial system. The three major agencies Moody's Investors Service (Moody's),

[1] See for example Moody's disclaimer
[2] see White (2009)

Standard and Poor's (S & P), and Fitch Ratings (Fitch) were found at the beginning of the 20th century and managed to survive and thrive until today.

After the big crash of 1929, and the subsequent tight regulation of capital markets, ratings by all three agencies were de facto made compulsory for banks to use when evaluating bonds. These regulations were further extended during the 1970s. During the 1990s, the first Basel accord came into being and required the use of capital to risk weighted asset ratios. That is, the more risky the asset held by a bank is, the more capital it needs to retain for this asset, in order to be prepared for a default. One possibility to evaluate the long run risk was to rely on the ratings of CRAs. Indeed, it is often easier for banks to simply use external investment grades than to develop own estimates of asset riskiness. This cemented the top three agencies as the dominant entities in the rating business. We refer to White (2010) for a good overview of the history of CRAs.

During the 1970s the CRAs also changed their business model. While initially CRAs charged investors for booklets containing ratings from them, the advent of the copy machine put this model in danger. Therefore, CRAs started to charge the entities they rated for the rating. This was possible because the ratings were seen as a quality signal; having none while auctioning bonds was uncommon and thus suspicious to investors.

However, this new approach to the rating market triggered criticism: the CRAs were accused of no longer rating for investors, but actually for their new clients, leading to inflated ratings. This was for example the case during the financial crisis in 2008, when a lot of rather sketchy assets received AAA status from the agencies, only to default later on. The question of rating inflation has been the subject of multiple studies, see for example Frenkel (2015), or Jiang et al. (2012).

The market for sovereign credit ratings can best be described as an oligopoly. The three big agencies - Moody's, S & P, and Fitch - control the vast majority of the market.[3] Nearly all of these ratings are unsolicited; that is, the countries receiving the ratings do not pay the agencies for them. This counters the allegation of inflated ratings for countries. Still, in recent years, the CRAs have come under criticism for their rating actions for countries that were caught in the currents of the European sovereign debt crisis.

Often voiced criticism, especially by politicians, was that the downgrades were very sudden, appeared without warning, and stood in no relation to reality. When the Euro was introduced we saw a convergence of bond yields all over its member states. The difference between a Greek and a German bond was negligible after only a few years. The ratings were simply following bond yields. What drove this process is still up for debate. On the one hand, there was the convergence hypothesis, which states that all members of the currency union would be moving towards a similar level of GDP and productivity, and this expectation, although false, was reflected in the yields. On the other hand, investors might have priced in the possibility of a bail out for any country in the Euro that experiences fiscal

[3] See IMF (2010).

stress. This significantly reduces individual country risk and thus would justify higher ratings. However, in 2010, when financial market participants as well as rating agencies realized that there is a chance that none of these scenarios will take place, the agencies were quick to downgrade countries while investors asked for an increased risk premium on bonds for fiscally stressed countries.

In this thesis, we want to take a closer look at the behavior of credit rating agencies during the last decade. Specifically, we want to analyze sovereign ratings in Europe. This research is of interest to policymakers, investors, and researchers alike. To policymakers, because it would give a scientific argument to their complaints about CRAs. To investors, because it helps assessing the reliability of ratings. Finally, to researchers, because the current literature in the field of sovereign ratings is rather sparse, especially when compared to research on corporate ratings. This thesis has six chapters, the first being an introduction and the last providing a conclusion. The other four chapters analyze different aspects of Credit Rating Agencies and the European Sovereign Debt Crisis. All of them are of empirical nature.

In chapter two, we explore whether or not rating agencies have assigned country ratings in a coherent manner. Specifically, we look at two phenomenons: procyclicality and path dependence. The former refers to whether the CRAs take the business cycle systematically into account, while controlling for macroeconomic fundamentals. The latter effect refers to the probability that a country is downgraded again, given its macroeconomic characteristics and the fact that it has already been downgraded. Procyclicality is a violation of rating through the cycle, that is assigning ratings on a longterm basis, rather than following short term market movements. Path dependence, also referred to as sluggishness, can produce a vicious cycle of repeated downgrades. We find no systematic evidence for procyclicality, but we find evidence for path dependence. Also, our regressions explain downgrades better than upgrades, which give rise to the suspicion that CRAs might have assigned upgrades in a careless manner prior to 2008.

The third chapter investigates whether ratings during the recent crisis for European countries reflect default probabilities. If we would perform this exercise with corporate ratings, it would be easy. We would have millions of rating changes for hundreds of thousands of firms available. One can calculate transition matrices from this and compare them to a subsample to see whether firms that are similar in size and other characteristics are treated similarly over time, or whether there are anomalies in the data. The data base for country ratings is much smaller and contains fewer defaults. Thus, we need to use another methodology. We use Credit Default Swap (CDS) data and correlate it with rating data to see whether there is a significant positive relationship. To reduce the noise in the CDS data we use a new variant of a Mixed Data Sampling (MIDAS) estimator. We find that in the majority of cases CDS data and rating data predict a similar default probability. However, there is a subgroup of countries in our data set located in eastern Europe where the correlation is not positive significant.

Chapter four concentrates only on the second part of our title. That is we concentrate on the sovereign debt crisis aspect and leave the agencies out of the picture for the time being. We examine the impact of austerity programs that have been implemented in Greece, Portugal, and Spain from 2010 onwards. To do so, the synthetic counterfactual estimator by Abadie and Gardeazabal (2003) is used. We construct a synthetic Greece, Portugal, and Spain and compare the evolution of GDP per capita between the actual data and the synthetic counterpart. We find that the programs have done extensive damage to Greece and Portugal, however we find no effect for Spain. This can be explained by the fact that the Spanish economy is in a process of change ever since the collapse of the construction sector and that this is the major driver of the current economic climate in Spain.

In the fifth chapter we analyze the relationship between the three big agencies. Specifically, we look for a possible leader-follower relationship in their rating history using frequency-domain analysis. To do so, we study the rating history of five countries in Europe during the crisis. While for Greece, Portugal, and Spain the CRAs act independently, there is a leader-follower relationship for the case of Ireland and Italy, suggesting that the agencies might be unsure about their judgment regarding these two countries.

The last chapter provides a conclusion to our research, where the results from the different chapters are put into perspective and overall policy implications are drawn.

Chapter 2

Procyclicality and Path Dependence of Sovereign Credit Ratings

2.1 Introduction

In this chapter we investigate whether Credit Rating Agencies (CRAs) assign ratings to European sovereigns in a procyclical manner as well as whether or not ratings are path dependent. Procyclicality means that ratings depend upon the business cycle even after controlling for macroeconomic factors. Path dependence is defined as ratings being dependent upon past ratings after controlling for macroeconomic fundamentals. Both are examples of distortions in ratings. So far, the literature on credit ratings for emerging countries is abundant, however the literature on developed countries is sparse and the topic of procyclical behavior or dynamic components in ratings for developed countries has not been investigated yet. Credit rating agencies (CRAs) and their actions are playing an increasingly significant role, not only in financial markets but also by affecting decisions of policymakers. Ever since the global financial crisis that started in 2007 and the recent sovereign debt crisis, the actions of CRAs have come again under scrutiny, not only by the public but also by academics, see for example Arezki et al. (2011) or Baum et al. (2013). Therefore, CRAs and their actions have become an important topic for discussion for both market participants as well as the general public.

The first characteristic of sovereign ratings analyzed in this paper is the procyclical

This chapter is based on Procyclicality and Path Dependence of Sovereign Credit Ratings: The Example of Europe, published in Economic Notes, special issue on the Economics of Rating Agencies. We would like to thank Bertrand Candelon, Elena Dumitrescu, Oana Floroiu, and Simone Vermeend. We would also like to thank all particpants of the 8th Conference on Risk, Banking and Financial Stability, September 2013, Bali, Indonesia, and the particpants of the XXII International Conference on Money, Banking and Finance, December 2013, Rome, Italy.

behavior. For the purpose of this paper, ratings are considered to move procyclically if they are positively correlated with economic or credit cycle fluctuations. In particular, procyclicality relates to the sensitivity of ratings towards these cycles and their conditionality upon macroeconomic fundamentals. This directly collides with rating through the cycle (see Altman and Rijken 2004), which implies that ratings are based upon issuer or issue specific characteristics and are independent from the economic cycle.

When CRAs assign a long term sovereign rating, they formulate their opinion on the creditworthiness of a sovereign and express this opinion by using an established and defined ranking system of rating categories. In doing so, CRAs do not necessarily need to calculate an absolute measure of default, but rather to assess the riskiness of default of an issue or an issuer through time. That is exactly why we may observe that default definitions vary among CRAs. In particular, Standard and Poor's Ratings Services (S & P) treats payment defaults on financial obligations, certain distressed exchanges and breaches of imputed payment promises as a default. Moody's Investors Service's (Moody's) definition of debt default constitutes various events, including missed or delayed disbursements, bankruptcy filings or legal receiverships, distressed exchanges as well as changes in the payment terms imposed by a sovereign. Fitch Ratings (Fitch) distinguishes between restricted default (when an uncured payment default occurs, which does not lead to a formal winding-up procedure or ceasing of business) and default (an uncured payment default which leads to a formal winding-up procedure or ceasing of business). Moreover, other smaller CRAs have their own definitions of default, which may include, but are not limited to the above elements from the definitions of the three larger CRAs.

One role of CRAs is to provide long term information to investors in order for them to optimize their risk-return trade off. This information, which is embedded in the rating, should have a long term nature and not a short or medium term scope. Therefore, long term ratings should be independent of the economic cycle and, by definition, are not supposed to vary in a procyclical manner. Looking at methodological documents and criteria articles across CRAs, we can easily identify that CRAs indeed claim that they rate through the cycle. A straightforward way to assess how ratings move vis-à-vis the economic cycle, is to look at widely used metrics, such as transition matrices and rating statistics and examine rating movements conditional upon the cycles. For example, we can review European sovereign ratings of the three larger CRAs, regarding upgrades and downgrades for the period of January 2006 to December 2008 and January 2009 to December 2011 (constituting a growth and a downturn period respectively). Through these statistics on sovereign ratings, we can see whether or not ratings tend to show higher occurrence of downgrades during an economic downturn or higher occurrence of upgrades during an economic growth.

Indeed looking at Table 1, for the period between 2006 to 2008, there were more upgrades than downgrades for all the three large CRAs. During the sovereign crisis period (2009-2011) downgrades occurred relatively more frequently. For the

Table 2.1: EU 27 Ratings 2006-2011

Agency	Moody's 06-08	S & P 06-08	Fitch 06-08
Upgrades	10	5	9
Downgrades	3	9	8
Agency	Moody's 09-11	S & P 09-11	Fitch 09-11
Upgrades	1	4	4
Downgrades	26	25	19

three CRAs the number of downgrades was more than three times greater than the number of upgrades.

However, even though tools like descriptive statistics and transition matrices can provide an initial view on whether or not ratings move procyclically across different periods, they fail to show that ratings are indeed assigned in a procyclical manner. In this study we investigate whether CRAs actually assign investment grades for sovereign solvency procyclically.

The second characteristic of sovereign rating that is analyzed in this paper is that of path dependence. This means that current ratings are depending on past ratings. This phenomenon is closely connected to procyclicality. Procyclicality basically asks the question whether or not ratings are better than expected during economic boom time and vice versa while controlling for macroeconomic characteristics. Therefore, the idea behind procyclicality is that analysts might get overly optimistic when good news from the economy are pouring in. Conversely, path dependence simply means that the current rating is dependent upon the previous rating. This would imply that rating levels are very sluggish, that is they only move very slowly compared to the information released about the economic conditions in the sovereign country that is rated. This would lead to over and under evaluations of countries depending on where they start on the rating scale.

To understand the exact implication of this research question we need to be aware of the actual role CRAs play in the current financial system. The Basel agreements have been the cornerstones of the current financial architecture since the beginning of the nineties. These agreements regulate the amount of capital banks have to retain for an investment of a given riskiness. One possibility to assess the risk of an individual asset is to rely on the judgment of CRAs. The big three agencies (S & P, Moody's and Fitch) work with an issuer-pays model. This means that the banks do not pay for using the ratings and it implies that from a business perspective it is more efficient for banks to use ratings provided by the CRAs instead of maintaining an own risk assessment division.[1] This makes the CRAs dominant entities in the risk assessment business. Furthermore, a change in the rating of an asset might require banks to increase or decrease the amount of capital held (depending on whether it is an upgrade or a downgrade). Also, pension funds are often bound by national law to CRA recommendations.

[1] Please note that there are also a number of banks who use the IRB approach and construct their own rating scheme.

Additionally the ratings have a strong signaling impact, which also moves capital in institutions that have no strict legal requirement for it, such as hedge funds. The fact that sovereign rating changes have indeed this impact is for example documented by Gande and Parsley (2010).

In this paper, we use (ordered) probit regressions, since the assumptions underlying this regression model match the nature of the data. We find that no agency is completely consistent in assigning ratings. There is evidence of procyclicality, sluggish rating assignments and country-group based discrimination even after controlling for typical macroeconomic variables. We also find a substantial difference in explanatory power for upgrades and downgrades. To the best of our knowledge there exists no other study which looks into the procyclicality issue from a sovereign debt perspective for developed countries. Furthermore, only a very limited number of studies were conducted using mostly developed country data, mainly because these countries were seldom downgraded prior to 2008.

The question of how the business cycle gets amplified by reckless lending in times of boom and by a credit crunch in times of recession is a hotly debated issue. Borio et al. (2001) gives an overview of the topic. So far there have been numerous proposals for actual reasons of this bank behavior. These include the idea that an actual fall in demand for loans is looking like a credit crunch as in Bernanke et al. (1991), who study the 1990-1991 recession in the United States. Also the idea of institutional memory leading to the deterioration of the skill of bank staff in Berger and Udell (2004) has been put forward, or that it is inherently anchored in the current financial architecture as concluded by Lowe (2002). The idea that this cycle could be supported by rating agencies has gotten little academic attention despite the strong criticism that the CRAs are facing since 2007. Ferri et al. (1999) investigate the behavior of rating agencies in the Asian financial crisis of 1997 to 1998, and conclude that procyclical behavior can be shown during this crisis. Furthermore Bar-Isaac and Shapiro (2012) show that in a dynamic model quality of ratings correlates negatively with the business cycle, which gives a theoretical foundation to the idea of procyclicality in credit ratings. Amato and Furfine (2004) conduct an analysis of procyclicality for corporate ratings. With this chapter we attempt to fill the literature gap on procyclicality for sovereign ratings and extend the literature on developed country sovereign debt ratings.

The rest of the chapter is structured as follows. The methodology will be discussed in section 2, section 3 is devoted to the empirical analysis of the research questions, and section 4 concludes.

2.2 Methodology

2.2.1 General Framework

We start by assuming that there exists a latent process of the form:

$$Y_{i,t}^{*} = X_{i,t}\beta' + E_{i,t} \tag{2.1}$$

where $Y_{i,t}^{*}$ is the dependent variable. $X_{i,t}$ is a matrix of regressors, and $E_{i,t}$ represents the errors which are standard normally i.i.d distributed and independent of the regressors $X_{i,t}$. The actual realization of the left hand side variable $Y_{i,t}^{*}$ is being governed by the rules:

$$Y_{i,t} = \begin{cases} 1 & \text{if } \; Y_{i,t}^{*} > 0 \\ 0 & \text{if } \; Y_{i,t}^{*} \leq 0 \; , \end{cases} \tag{2.2}$$

or

$$Y_{i,t} = \begin{cases} 0 & \text{if } \; Y_{i,t}^{*} \geq \alpha_1 \\ 1 & \text{if } \; \alpha_1 < Y_{i,t}^{*} \geq \alpha_2 \\ \vdots & \\ J & \text{if } \; Y_{i,t}^{*} > \alpha_J \; . \end{cases} \tag{2.3}$$

The α's denote the threshold parameters. The first rule gives rise to a binary regression model and the second rule to a categorical or ordered regression model. The choice of the particular regression method employed depends on the assumptions made on the distribution of $E_{i,t}$. Here we assume a standard normal distribution which indicates the use of a probit regression, whereas a standard logistic one would indicate the use of a logit regression.

We assume that the process described in equation (2.1) is a good approximation of the actual rating assignment process, since ratings constitute the informed opinion of an analyst who assembles as much (macroeconomic) data as possible on a given country and consequently forms an opinion on the risk of insolvency of this country. As a last step, the analyst transforms this judgment into a rating scale. Thus the first two steps of this process are captured by equation (2.1). $X_{i,t}$ represents the data assembled on the country at hand, whereas the $Y_{i,t}^{*}$ is the actual decision on the risk of insolvency of the country. The process of transforming this decision into a rating is represented by equation (2.3). Therefore, we need a statistical method that accounts for this framework. The obvious choice for this setting is a binary or ordered choice regression. This chapter makes use of both types of regressions. The binary choice regression is employed to model the rating changes of countries. When doing this, we must split the rating change sample into upgrades and downgrades, otherwise the regression would identify rating changes into different directions as rating changes in the same direction. Additionally, splitting the ratings gives the opportunity to investigate asymmetries between upgrades and downgrades. Another advantage of a binary choice regression is that specifying a dynamic model is much less troublesome for the binary case compared to the ordered case. A detailed explanation for this can be found in section 2.2.2. The disadvantage of using a binary regression is that it comes with the implicit assumption that rating changes are done indiscriminate of whether a country gets downgraded from AA to A, or from C to default. Also, we have to realize that a series with upgrades as well as downgrades might be problematic due to the fact that there is asymmetric movement in the regressors, compared to the dependent variable. For example, when analyzing upgrades there is a distinct time span when upgrades are relatively sparse, however, downgrades are relatively frequent. In this period, the regressor variables are moving in line with the downgrades, but the binary upgrade series does not exhibit any change. This might lead to a bias in the regression.

We use ordered probit when analyzing the actual ratings. Ratings are an example of an ordered series, as triple A clearly is better than BB+. One of the advantages of this approach is that, unlike the binary approach, it makes use of all information contained in the rating data, since now we have the level of the rating available. Also, an ordered probit is a non-linear model and thereby takes into account the nonlinearities of the rating scale. There are two factors contributing to it. First of all the marginal effects are differing depending on whether a country is more or less wealthy. Generally, better off countries have lower variation in their ratings. This feature is captured by the marginal effects. Second off all the estimation of the cut-off points, which serve as intercepts allow for differentiation by rating level. Since these points are conditional on which rating a country has, they are therefore also able to reflect the non-linearity of the rating scale.

2.2.2 Model and Dynamic Specification

In this section a dynamic specification for probit regressions is developed. Consider:

$$Y^*_{i,t} = X_{i,t-1}\beta' + E_{i,t} \ , \tag{2.4}$$

where $Y_{i,t}$ is a measure of the rating or rating change or a rating level, $X_{i,t-1}$ is a set of regressors, and $E_{i,t}$ are the errors. To specify a model with dynamics for a binary choice case we can simply set up a regression including a variable of autoregressive order one, that is we simply lag the rating change and use it as an explanatory variable. This allows to investigate whether the CRAs are taking past rating history into account, that is whether ratings are path-dependent or not. This is similar to the rating stickiness documented by Ferri et al. (1999), and it is interesting to see whether this extends to European countries as well. De Jong and Woutersen (2011) show the validity of Maximum Likelihood Estimations (MLE) for a dynamic setting.

This straightforward approach of lagging the dependent variable unfortunately does not work for the ordered probit regression. Generally, in order to qualify for an autoregression, the ordered variable should be interval scaled. This is clearly not the case for the rating series at hand. Also, when actually estimating, the dynamic ordered probit exhibits massive problems to converge. Therefore, we follow a different approach when specifying a dynamic model for an ordered probit, building on the work of Kauppi and Saikkonen (2008). They propose to include a lagged probability of the dependent variable as an additional regressor such that:

$$\pi_t = \Phi(X_{i,t}\beta') \ , \tag{2.5}$$

where $\Phi(.)$ is the cumulative normal distribution. In this case we follow the method employed by Candelon et al. (2014), estimating the lagged probability in a first step regression by using Ordinary Least Squares for the following equation:

$$\ddot{Y}_{i,t} = X_{i,t-1}\beta' + E_{i,t} \ , \tag{2.6}$$

where $\ddot{Y}_{i,t}$ is the demeaned version of the rating data. For this method to yield actual probabilities, we start by determining $\Phi(X_{i,t}\beta')$, which creates the probability $\pi_{i,t}$. As a second step we introduce $\pi_{i,t-1}$ into the ordered probit as an additional regressor, thereby estimating a dynamic model. Furthermore this approach will be used in some cases for the dynamic binary probit estimations, instead of relying on the simple autoregressive approach. This is due to the breakdown of the MLE when dealing with a regressor which predicts the independent variable perfectly, which is the case in some of the estimations. However, a word of caution needs to be added here since this method is only an imperfect proxy for the autoregressive process, because the regressions only partly explain the variation in the dependent variable and therefore the lagged probability given by these estimations can also only partly model the true dynamics.

We want to make a remark about the overall approach of using a simple pooled panel model. Generally introducing dynamics into a probit model is as shown in the preceding paragraphs not completely straightforward and has some pitfalls. Therefore, it was decided to go with the simplest model, in order to bypass most of the problems involved when estimating a more complex model.

2.3 Empirics

2.3.1 Data

This chapter uses quarterly data retrieved from Eurostat. The series start in the first quarter of 2001 and end in the last quarter of 2010. It contains data from 21 EU countries, namely Belgium, Czech Republic, Denmark, Germany, Estonia, Ireland, Spain, France, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, the Netherlands, Austria, Poland, Portugal, Slovenia, Finland, and the United Kingdom. This makes for a total of 924 observations, of which 903 are used in the estimations. The data set consists of the following variables: GDP per capita (in logarithm), GDP growth, inflation rate, government debt as percentage of GDP, net current account as percentage of GDP, and the primary balance as percentage of GDP. These variables were used in earlier studies such as Afonso et al. (2007) or Bennell et al. (2006) and found to be significant in explaining sovereign ratings. For the primary balance we have to interpolate for quarterly data, which was also the case for the current account for Belgium in the year of 2001. Using a Bry-Boschan algorithm as described in Bry and Boschan (1971) a business cycle indicator is created. This variable is set equal to one during a boom-time and equal to minus one in a recession. This design was chosen because it yields an intuitive regression coefficient. Furthermore we create a GIPS variable which indicates whether a country is part of the European troubled economies, that is it is equal to one in case of Spain, Ireland, Portugal, and Italy. Gärtner et al. (2011) find a significant relationship between ratings and such a variable. For the dependent variable, the respective ratings are Long Term: Issuer Rating (Foreign) from Moody's, Issuer Credit Rating Foreign Long Term by Standard and Poor's, and Long Term Issuer Default Rating by Fitch. These are all the equivalent of a long term default assessment. These ratings are transformed into two different scales. First, there is a binary rating-event series where a one corresponds to a rating change and a zero to no change in rating in the respective quarter. Separate series for upgrades and for downgrades are constructed. Second, we convert all rating data into a numerical scale according to Table 2.2 and 2.3.

Table 2.2: Rating Scale Part 1

Numerical	Moody's	S & P	Fitch
22	Aaa	AAA	AAA
21	Aa1	AA+	AA+
20	Aa2	AA	AA
19	Aa3	AA-	AA-
18	A1	A+	A+
17	A2	A	A
16	A3	A-	A-
15	Baa1	BBB+	BBB+
14	Baa2	BBB	BBB
13	Baa3	BBB-	BBB-
12	Ba1	BB+	BB+

Multiple rating events in one quarter are treated differently for each series. The purpose of the first series is to separate quarters with a rating event from quarters without one. Thus, multiple rating changes except for the first one are ignored. For the numerical scale, a two-notch downgrade is completely reflected in the data, the only problem present is that we cannot distinguish between a two notch downgrade at one date and two single rating changes within a given quarter. However, the latter is seldom the case in the investigated time span. Furthermore a pooled version for the ratings is constructed, by averaging all three series and rounding mathematically. The latter is also useful for detecting information contamination, that is a CRA is only looking at the actions of its competitors and is acting in line

Table 2.3: Rating Scale Part 2

Numerical	Moody's	S & P	Fitch
11	Ba2	BB	BB
10	Ba3	BB-	BB-
9	B1	B+	B+
8	B2	B	B
7	B3	B-	B-
6	Caa1	CCC+	-
5	Caa2	CCC	CCC
4	Caa3	CCC-	-
3	Ca	CC	CC
2	Ca	C	C
1	C	D	D

with their evaluations. If the pooled estimations differ vastly from the individual agency estimations, we might have a case of information contamination.

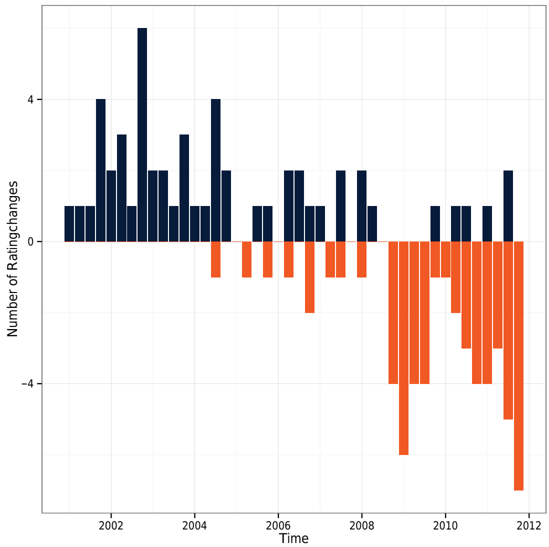

Figure 2.1: Pooled Rating Changes

In Figures 2.1 to 2.4 we can see the number of quarters with rating changes aggregated for all members in our panel. S & P is the most active rating agency, while Moody's is the least active in the timespan covered in this chapter. Also, we see a distinct increase in downgrades after 2008, as expected.

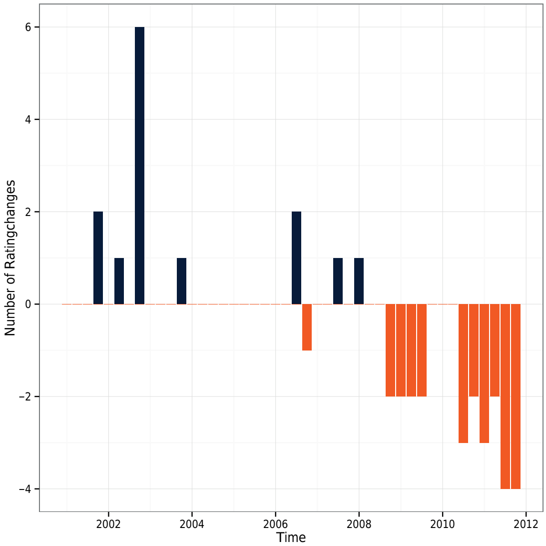

Figure 2.2: Moody's Rating Changes

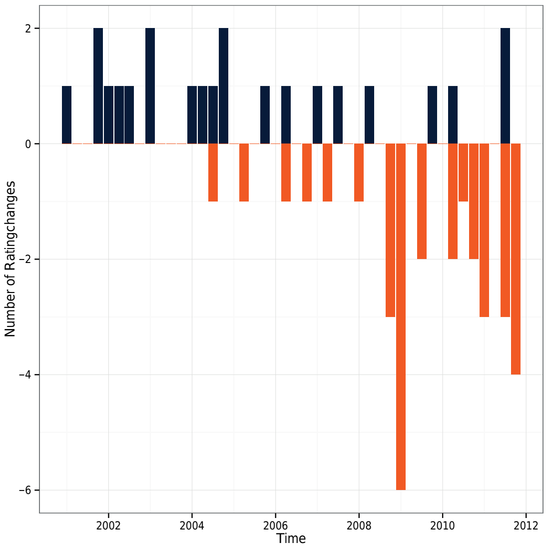

Figure 2.3: Standard & Poor's Rating Changes

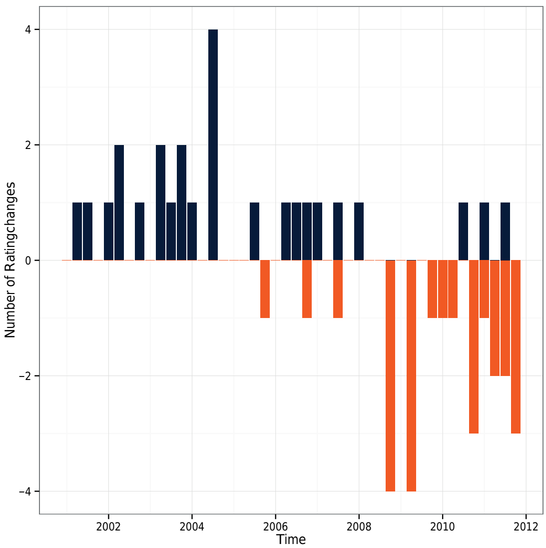

Figure 2.4: Fitch Rating Changes

2.3.2 Preliminary Tests

We start with some preliminary tests. First, a test for cross sectional dependence is conducted. Cross sectional dependence (CD) means that there is a correlation structure between different countries. This is often the case for country panels, since countries, especially in close geographic proximity, are often highly dependent on each other. Estimation-wise, this will lead to distorted standard errors as well as biases in dynamic models. For the CD-test every regression specification used in the paper is estimated and the corresponding test developed by Hsiao et al. (2012) is used.[2] It tests whether the average correlation between the generalized residuals (see Gourieroux et al. 1987) is different from zero, to investigate whether there is a systematic relationship in the cross sectional dimension. The results can be found in Table 2.4. We only test the static specification since the results should be similar for the dynamic case. Only in one case the null hypothesis of no cross sectional dependence is rejected. Therefore, it is concluded that no systematic cross sectional dependence is present in the dataset. A possible reason for not detecting cross sectional dependency could be the rather short time-dimension of our dataset which is merely a decade long.

Table 2.4: Cross Sectional Dependence Test

	Static Upgrades	Static Downgrades
Pooled	0.349	1.649
critical value at 5%	(4.748)	(2.947)
Moody's	4.800**	2.567
critical value at 5%	(2.966)	(4.646)
S & P	0.909	2.198
critical value at 5%	(5.365)	(4.874)
Fitch	0.567	1.660
critical value at 5%	(5.824)	(4.886)

This means we can go ahead with the proposed estimation strategy and use standard probit to tackle the question at hand, instead of trying to correct for cross sectional correlation. This also means that established robust standard errors can be used. Also, a test for heteroskedasticity is conducted. This is on the one hand crucial for discrete response regressions, on the other hand not exactly straightforward as in the linear Ordinary Least Squares case. The importance of this test lies in the fact that one assumption for identifying the parameters of a discrete response regression is a constant variance. Therefore, when having a case of heteroskedasticity not only are the estimates inconsistent but actually the point estimates make no sense (Wooldridge, 2001). However, the actual test is not simple because there are no clearly defined residuals for this regression-type, which we could inspect. Instead we reestimate each regression using a heteroskedasticity-

[2] Please note that instead of using individual regressions, a pooled version is employed due to the invariancy of the dependent variable for some elements of the cross section.

robust probit method introduced by Harvey (1976) and conduct a Wald test for significance of the coefficient modeling the variance. Using this approach implies that variables that might model the variance need to be selected. There is no reason to suspect heteroskedastic errors in the ordered series, however for the binary series, the business cycle variable could have an influence on the variance, since upgrades and downgrades are separated. For upgrades there is always more activity during boom times and for downgrades there is always more activity during a recession. The results of the test can be found in Table 2.5; the null hypothesis of the Wald test is no heteroskedasticity.

Table 2.5: Heteroskedasticity Test for Binary Probit

	Upgrades	Downgrades
Pooled	0.085	0.770
Moody's	0.379	0.014
S & P	0.000	0.835
Fitch	0.000	0.572

As it can be seen, there are three cases of heteroscedasticity present. In those cases the heteroskedastic probit regression introduced by Harvey (1976) is used.

As a last step, we want to make sure that no multicollinearity is diluting the results. To do so, a correlation table is produced which can be found in the appendix. Table 2.14 shows that only in two cases is the (absolute) correlation between two variables above 0.4. Therefore, multicollinearity should not be a problem in the estimations.

2.3.3 Rating levels

The first regression estimated is:

$$\begin{aligned} rating_{i,t} = \beta_0 &+ \beta_1 sebt_{i,t-1} + \beta_2 inflation_{i,t-1} \\ &+ \beta_3 growth_{i,t-1} + \beta_4 primarybalance_{i,t-1} \\ &+ \beta_5 currentaccount_{i,t-1} + \beta_6 businesscycle_{i,t-1} \\ &+ \beta_7 GIPS * businesscycle_{i,t-1} + \beta_8 GIPS_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{2.7}$$

This is done for every rating agency separately as well as a fourth probit, pooling the ratings by the average (rounded) rating assigned to the countries on the left hand side. This regression approach is essentially analyzing the rating level of each country. We could therefore expect that only stock variables are significant. However, there is a considerable amount of literature for example Afonso (2003) or Gärtner et al. (2011) who find stock as well as flow variables to be significant in their regressions. Therefore, both types of variables are used as regressors.

Table 2.6: Rating Level Estimations using Ordered Probit

	Pooled	Moody's	S & P	Fitch
GDP	3.607***	2.172***	4.596***	3.826***
	(0.436)	(0.466)	(0.606)	(0.500)
Government Debt	−0.029***	−0.018***	−0.038***	−0.029***
	(0.005)	(0.006)	(0.006)	(0.006)
Inflation	-0.014	-0.008	-0.029	-0.024
	(0.017)	(0.015)	(0.018)	(0.018)
Growth	0.021	0.017	0.025	0.035
	(0.026)	(0.024)	(0.039)	(0.027)
Primary Balance	0.118**	0.080	0.140***	0.107**
	(0.047)	(0.057)	(0.037)	(0.045)
Current Account	1.275	0.583	3.063***	1.241
	(0.791)	(1.250)	(1.109)	(0.944)
Business Cycle	−0.243***	−0.213***	-0.098	−0.189**
	(0.075)	(0.077)	(0.116)	(0.082)
Interaction	0.676*	0.462	0.592*	0.674
	(0.356)	(0.318)	(0.333)	(0.442)
GIPS	0.071	0.331	-0.254	0.003
	(0.578)	(0.575)	(0.596)	(0.588)
R-squared	0.395	0.270	0.474	0.410

Standard errors in all estimations are clustered robust.

Looking at the GDP variable in Table 2.6, we can see that for all three agencies as well as the pooled rating the coefficients are significant and the sign is positive. This means that a higher output per capita increases the likelihood of a higher rating which makes economic sense. Turning to government debt, all coefficient signs are negative and in all four regressions the impact of government debt on ratings is statistically significant. This means that a higher indebtedness of the public sector increases the probability of a lower rating, ceteris paribus. These findings are largely in line with Afonso et al. (2009), who studies ratings for a country panel involving developing nations as well as developed ones. In row three and four, inflation and the growth rate of GDP are both statistically not distinguishable from zero. This is contrary to the before mentioned paper which does find these variables to be statistically correlated to the level of rating. However, it does make sense from an intuitive perspective that the level of rating should only be connected to the level of the economic indicators. Some mixed signals are received when looking at the primary balance and the current account. Two individual CRA regressions as well as the pooled regression exhibit a significant coefficient for the primary balance whose direction also makes sense from an economic perspective, a higher primary balance is associated with the probability of a higher assigned rating. The current account is only significant in the Standard and Poor's regression, which hints towards the fact that overall, rating agencies do not see the trade balance as an important factor for determining the solvency of a country within the EU. This makes sense in so far that EU countries normally can issue bonds in their own currency. Comparing this to developing countries who issue a significant amount of debt often in US Dollar, the latter need to have trade surpluses in order to service the non-home-currency denominated debt. Thus, we see here a fundamental distinction between developing and developed countries. Lastly, there are in three cases negative significant coefficients for the business cycle variable as well as two out of four positive significant coefficients for the interaction between the business cycle and the GIPS countries. The former indicates that CRAs are rating countercyclical, meaning that in boom times the probability of getting a higher ratings assigned are a bit lower than they should be, and in bad times the rating level is likely a bit higher. The positive sign of the interaction variable means that this effect is the opposite for the GIPS country group, hinting either towards discrimination or a fundamental difference between those countries and the rest of the sample. However, these results do not hold when taking dynamic behavior into account. The (pseudo) R-squares are on the lower end of what is common in the literature, but given our smaller dataset, this is not unexpected. Note that Moody's is a bit of an outlier with and R-square of only around 0.27.[3] In general, a discussion of the economic effects of the marginal effects of the probit estimations are not in order. This is because marginal effects can only be evaluated when taking the dataset into account. Typically we see marginal effects evaluated at the average observation in a dataset. However, it does not seem appropriate to talk about the average European country.

In Table 2.7, we estimate:

$$\begin{aligned} rating_{i,t} = \beta_0 &+ \beta_1 debt_{i,t-1} + \beta_2 inflation_{i,t-1} + \beta_3 growth_{i,t-1} \\ &+ \beta_4 primarybalance_{i,t-1} + \beta_5 currentaccount_{i,t-1} \\ &+ \beta_6 businesscycle_{i,t-1} + \beta_7 GIPS * businesscycle_{i,t-1} \\ &+ \beta_8 GIPS_{i,t-1} + \beta_9 estimatedprobability_{i,t-1} + \epsilon_{i,t}. \end{aligned} \tag{2.8}$$

The main difference to the regression estimated in Table 2.6 is the addition of a dynamic term which takes the form of an estimated probability (see equation (2.6). This is similar to an autoregressive estimation, but tweaked to work well in an ordered probit setup. If we compare the results of the dynamic estimations with the static ones from Table 2.6, we can see that GDP as well as government debt continue to play a significant role in explaining rating levels, and the magnitudes and signs are also similar. Also, the flow variables, that is inflation and GDP growth are not statistically significant, except for inflation in the regression on S & P ratings. In this case, a higher inflation rate is increasing the probability to have a lower rating, which seems reasonable, since higher inflation is often connected to an increasingly unstable economic situation. The primary balance

[3]Often regression models are evaluated by their out of sample performance, however with probit or logit regression this is not possible, since the coefficients can only be translated into probabilities affecting the dependent variable.

Table 2.7: Dynamic Rating Level Estimations using Ordered Probit

	Pooled	Moody's	S & P	Fitch
GDP	2.499***	1.354**	3.431***	2.809***
	(0.613)	(0.653)	(0.578)	(0.512)
Government Debt	−0.030***	−0.018***	−0.039***	−0.030***
	(0.005)	(0.006)	(0.007)	(0.007)
Inflation	-0.013	-0.007	−0.028*	-0.024
	(0.015)	(0.014)	(0.016)	(0.017)
Growth	0.006	0.011	0.006	0.020
	(0.027)	(0.023)	(0.037)	(0.026)
Primary Balance	0.067	0.034	0.092**	0.065
	(0.056)	(0.075)	(0.046)	(0.054)
Current Account	0.640	0.150	2.357**	0.663
	(0.847)	(1.175)	(1.106)	(0.862)
Business Cycle	-0.087	-0.096	0.052	-0.045
	(0.091)	(0.122)	(0.087)	(0.089)
Interaction	0.464	0.328	0.392	0.478
	(0.421)	(0.352)	(0.383)	(0.519)
GIPS	0.144	0.227	-0.054	0.058
	(0.634)	(0.622)	(0.652)	(0.657)
Dynamics	2.140***	1.656	2.205***	1.934***
	(0.739)	(1.441)	(0.815)	(0.663)
R-squared	0.407	0.276	0.486	0.421

Standard errors in all estimations are clustered robust.

is now only significant for the Standard and Poor's regression, so is the current account balance. Again both coefficient signs make economic sense, a higher primary balance is connected to a higher rating level. The same logic applies to the current account. When now looking at the business cycle, the GIPS indicator and the interaction between these two, none of them are significant anymore. Instead these effects are picked up by the dynamic variable. Indeed, when removing the business cycle and the interaction term from the equation, the R-squared barely drops. The interpretation of this variable is that given that one has a high rating, it is also more likely to stay in that rating. This relationship is also true for a low rating, thereby leaving countries with a lower rating, than they would actually deserve, given the economic fundamentals. The R-squares are slightly larger compared to the static regression. This is an example of path dependence, that is rating agencies are taking the rating history into account instead of just assigning ratings based on macroeconomic factors.

2.3.4 Rating Changes

This section investigates rating changes. So far this has not been done often in the literature. The main advantage is that a dynamic model can easily be specified without relying on the estimated probability procedure used for estimating rating levels. As a first step the following regression equations are estimated:

$$\begin{aligned} ratingchange_{i,t} = {} & \beta_0 + \beta_1 debt_{i,t-1} + \beta_2 inflation_{i,t-1} + \beta_3 growth_{i,t-1} \\ & + \beta_4 primarybalance_{i,t-1} + \beta_5 currentaccount_{i,t-1} \\ & + \beta_6 businesscycle_{i,t-1} + \beta_7 interaction_{i,t-1} + \beta_8 GIPS_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{2.9}$$

Table 2.8: Downgrades Estimations using Probit

	Pooled	Moody's	S & P	Fitch
Government debt	0.006	0.016***	0.005	0.005
	(0.006)	(0.006)	(0.005)	(0.006)
Inflation	0.052	0.072*	0.058***	0.060**
	(0.034)	(0.041)	(0.017)	(0.028)
Growth	−0.090***	−0.150***	−0.063*	−0.124***
	(0.027)	(0.043)	(0.032)	(0.034)
Primary Balance	−0.183***	−0.111***	−0.097***	−0.108***
	(0.042)	(0.032)	(0.022)	(0.026)
Current Account	−6.453***	−7.554**	−5.353***	−4.855***
	(2.230)	(3.070)	(1.748)	(1.727)
Business Cycle	-0.067	-1.387	−0.210**	0.276
	(0.129)	(0.910)	(0.096)	(0.342)
Interaction	0.563**	0.700*	0.309	0.344
	(0.219)	(0.395)	(0.191)	(0.362)
GIPS	-0.053	-0.496	0.236	-0.013
	(0.365)	(0.682)	(0.254)	(0.453)
Constant	−2.890***	−4.644***	−2.895***	−3.143***
	(0.365)	(0.829)	(0.259)	(0.497)
R-squared	0.413	-	0.308	0.346

Standard errors in all estimations are clustered robust, a heteroskedastic probit regression was used for Moody's.

When looking at Table 2.8 we see that as hypothesized, for rating changes the main significant variables are changes or growth rates. Starting with government debt (a stock variable), only Moody's seem to take it into account, whereas in all other cases it is insignificant. The coefficient has the correct sign, that is a higher level of government debt is associated with an increased probability of getting a downgrade. Turning to inflation, we can see that it is significant in three out of four regressions. Also, the coefficient points into the right direction, meaning that a higher level of price changes is correlated with an increased probability of a downgrade. For the variables growth and primary balance, regressors are significant in all equations. Again, the signs of the variables make economic sense, an increased primary balance and increased GDP growth is connected to a decreased risk of getting downgraded. The same pattern holds for the current account variable, a higher current account is associated with a decreased downgrade risk. It should be noted that the magnitude of the current account is normally smaller than the other variables set in relation to GDP. Therefore, the coefficient also needs to be greater in magnitude in order to have a similar impact upon rating changes. This explains the relatively big current account coefficients. The S & P regression has a significant business cycle variable, whose sign points towards procyclicality. That is, in an economic downturn is the chance of a downgrade increased, although the regression already takes the reduced economic activity via the growth variable into account. Contrary to that, in two regressions the interaction variable between the GIPS countries and the business cycle is positive significant, which means that the GIPS countries have a decreased chance of downgrades in bad times and an increased chance of downgrades in boom times. The GIPS variable itself is insignificant in every regression, which is different from Gärtner et al. (2011), however, Gärtner only looks at levels of ratings. The R-squares are on a similar level as the ones explaining the rating level. Note that there is no R-squared for the Moody's regression since there is none defined for the heteroskedastic probit.

The estimations for upgrades are in Table 2.9. In this case government debt is significant for the pooled regression as well as the Standard and Poor's regression. Also the economic growth variable is significant in three out of four cases. In all significant cases point the coefficients into the right direction, that is a higher government debt is associated with a decreased chance for an upgrade and a higher growth-rate means a higher chance of getting upgraded. Contrary to the downgrade case, inflation is insignificant in two of the four estimations. The primary balance is insignificant for all regressions. For the current account, we can see that it is significant in two cases, however the sign of the coefficient points into the wrong direction, meaning that an increase in the current account decreases the chances of an upgrade, which does not make much sense from an economic perspective. The business cycle regressor is relevant in the pooled estimations as well as in the Standard and Poor's regression. In the pooled case it gives rise to a procyclical interpretation, that is a decreased chance of getting upgrade in economic boom times and an increased chance of upgrades in a recession, whereas for the S & P regression it is the opposite. Finally, the interaction variable as well as the GIPS variable are statistically significant in two cases. The interaction variable indicates that GIPS-countries upgrades have a more procyclical component in the Moody's regression, compared to the rest of Europe, and a more countercyclical component in the Fitch regression. The negative coefficient for the GIPS variable means that an upgrade is less likely for such a country. Does this mean that the CRAs are discriminating against the GIPS countries? We propose to look at the results with caution, since the explanatory power is rather low, and the results might be due to the very distinct downgrade environment in the last two years.

Table 2.9: Upgrades Estimations using Probit

	Pooled	Moody's	S & P	Fitch
Government Debt	-0.009^{*}	0.001	-0.032^{***}	-0.006
	(0.005)	(0.004)	(0.008)	(0.011)
Inflation	-0.032	-0.004	-0.047^{***}	-0.011^{*}
	(0.021)	(0.022)	(0.023)	(0.059)
Growth	0.052^{**}	0.080^{***}	0.025	0.160^{***}
	(0.024)	(0.022)	(0.019)	(0.050)
Primary Balance	0.005	-0.003	0.016	0.020
	(0.027)	(0.041)	(0.077)	(0.074)
Current Account	-2.039^{*}	-1.724	-4.457^{***}	-3.592
	(1.043)	(1.292)	(2.256)	(2.400)
Business Cycle	0.240*	-0.130	-1.334***	-
	(0.138)	(0.134)	(0.405)	-
Interaction	-	1.736***	0.338	-0.682***
	-	(0.188)	(0.280)	(0.358)
GIPS	-0.214	-1.810***	-0.657***	-0.016
	(0.226)	(0.206)	(0.237)	(0.380)
Constant	-1.513***	-2.343***	-1.996***	-4.313***
	(0.255)	(0.255)	(0.419)	(0.572)
R-squared	0.101	0.072	-	-

Standard errors in all estimations are clustered robust, Fitch lacks the business cycle variable due to perfect failure prediction, the pooled estimations lack the interaction term due to non-concavity of the log-likelihood, the heteroskedastic probit was used for S & P as well as Fitch.

Therefore, we should take a look and split the sample just after the beginning of the financial crisis, to get a better idea from where these results are coming. This is done in the robustness section.

For the dynamic regression the following equation is estimated:

$$\begin{aligned} ratingchange_{i,t} = \beta_0 &+ \beta_1 debt_{i,t-1} + \beta_2 inflation_{i,t-1} + \beta_3 growth_{i,t-1} \\ &+ \beta_4 primarybalance_{i,t-1} + \beta_5 currentaccount_{i,t-1} \\ &+ \beta_6 businesscycle_{i,t-1} + \beta_7 GIPS * businesscycle_{i,t-1} \\ &+ \beta_8 GIPS_{i,t-1} + \beta_9 ratingchange_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{2.10}$$

The results can be found in Table 2.10. Comparing the dynamic estimations to the static estimations the control variables remain at large similar. Again, most of the flow variables are significant, while a majority of the stock variables have no influence. Also, the magnitude of the regressors are close to their respective counterparts from the static regression. The newly introduced feature for these estimates is the dynamic component in the model. Indeed, we can observe that in three out of four regressions, the dynamic component is significant, and it should

Table 2.10: Downgrades Dynamic Estimations using Probit

	Pooled	Moody's	S & P	Fitch
Government Debt	0.006	0.015**	0.005	0.004
	(0.005)	(0.006)	(0.004)	(0.005)
Inflation	0.045	0.079**	0.050***	0.028
	(0.030)	(0.037)	(0.017)	(0.025)
Growth	-0.075***	-0.138***	-0.054*	-0.085***
	(0.023)	(0.046)	(0.028)	(0.029)
Primary Balance	-0.150***	-0.101***	-0.066*	0.023
	(0.034)	(0.032)	(0.037)	(0.034)
Current Account	-5.841***	-7.390**	-5.309***	-4.694***
	(1.978)	(3.025)	(1.679)	(1.573)
Business Cycle	-0.015	-1.568	-0.210**	0.264
	(0.152)	(1.021)	(0.094)	(0.318)
Interaction	0.422**	0.638	0.404**	0.813**
	(0.213)	(0.409)	(0.185)	(0.414)
GIPS	-0.040	-0.347	0.026	-1.062*
	(0.314)	(0.673)	(0.292)	(0.596)
Dynamics	0.854***	0.668*	9.391	41.750***
	(0.240)	(0.352)	(7.417)	(13.317)
Constant	-2.866***	-4.810***	-7.648**	-24.329***
	(0.297)	(0.832)	(3.689)	(6.841)
R-squared	0.435	-	0.308	0.346

Standard errors in all estimations are clustered robust, a heteroskedastic probit regression was used for Moody's, the lagged probability approach was used for Fitch as well as S & P, due to perfect correlation with the dependent variable

further be noted that in the one regression where it is not significant, the lagged probability approach needed to be used, which might be a less than perfect proxy for an autoregressive estimation. The interpretation of the dynamic component in this case is the following: given that a country was downgraded, there is an increased chance of getting downgraded in the next quarter, even if we account for economic factors. However, the latter should be the only variables driving the downgrade. This is an example of path dependency of ratings. That means in this case past rating changes have a significant impact on current rating changes.

Table 2.11 introduces dynamics into the upgrade estimations.[4] Turning to the variables, except for economic growth there are no systematically significant economic variables that explain upgrades. Otherwise all comments on individual variables from the non-dynamic probit regression directly apply to the dynamic version as well. The newly introduced dynamic coefficient is contrary to the case of downgrade-regressions only in the pooled case significant. This adds to the

[4]The heteroskedastic probit was not used for Standard and Poor's estimation, because in this case the maximum likelihood algorithm did not converge.

Table 2.11: Upgrades Dynamic Estimations using Probit

	Pooled	Moody's	S & P	Fitch
Government Debt	-0.007*	0.001	-0.015	0.000
	(0.004)	(0.004)	(0.009)	(0.014)
Inflation	-0.026	-0.003	-0.023	-0.091
	(0.019)	(0.021)	(0.022)	(0.062)
Growth	0.046**	0.079***	0.020	0.088
	(0.023)	(0.021)	(0.035)	(0.089)
Primary Balance	0.002	-0.003	0.012	0.017
	(0.029)	(0.041)	(0.030)	(0.082)
Current Account	-1.794*	-1.682	-2.032**	-2.441
	(0.974)	(1.320)	(0.862)	(2.330)
Business Cycle	0.211*	-0.129	0.266*	-
	(0.128)	(0.133)	(0.148)	-
Interaction	1.599***	1.733***	1.507***	-0.838*
	(0.156)	(0.189)	(0.208)	(0.462)
GIPS	-1.745***	-1.803***	-1.638***	0.439
	(0.181)	(0.206)	(0.184)	(0.523)
Dynamics	0.782***	0.275	-2.791	54.351
	(0.172)	(0.353)	(35.420)	(63.771)
Constant	-1.633***	-2.354***	-0.250	-32.540
	(0.222)	(0.254)	(18.393)	(33.007)
R-squared	0.134	0.074	0.103	-

Standard errors in all estimations are clustered robust, Fitch lacks the business cycle variable due to perfect failure prediction, the heteroskedastic probit has been used for Fitch

evidence gathered from the non-dynamic estimations that in this framework explaining upgrades is extremely difficult. There are two possible explanations for that. First, the sample period is not suited to explaining upgrades. In this case the regressors are exhibiting a strong downward tendency in the last three years of the sample, which is the financial crisis, but are not accompanied by any movement in the dependent variable, possibly diluting the precision of the estimations. However, if this is the case, an ordered probit model, investigating rating changes as a variable taking values of -1 (for a downgrade), 0 (for no change), and 1 (for an upgrade) should do a very good job of explaining the overall variation, which is addressed in the robustness section. Furthermore, removing the financial crisis period from the estimation should significantly improve the performance of the probit. This is also done in the following section.

2.3.5 Robustness Tests

In this section robustness tests are discussed. We investigate logit against probit specifications, institutional factors, as well as subsamples. We start by testing whether a probit or logit specification is correct. The consequence of choosing the wrong specification results in inconsistent estimates. Unfortunately, to the best of our knowledge, there exists no test for this in the literature. Therefore, we have estimated all (non-ordered) regressions with probit specifications as well as with logit specifications. When examining the standard errors, we can see no systematic difference between the estimates.[5] Therefore, the probit approach seems to be valid.

Looking at the explained variation of our data, it might be the case that our statistics are lacking some fundamental variables to explain sovereign rating levels. Often mentioned possibilities are institutional variables such as effectiveness of the government or the ease of doing business. In fact, since we use long term ratings in this analysis, it makes sense to use these factors, since they should significantly determine the long run prosperity of a nation. To check this, the Worldwide Governance Indicators, supplied by the World Bank are used. Unfortunately this database has only been established in 1996 and furthermore the sampling frequency is on a yearly basis (and that only since 2002, beforehand being biannually). However, this is one of the most reliable database on this topic, and therefore, it is used for this robustness test. The ratings are converted into yearly ratings, by simply taking the rating of the fourth quarter. We estimate an ordered probit with time series dimension of nine (from 2002 until 2010) and cross sectional dimension of 21 countries with the following variables as regressors: Voice and Accountability, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. These regressions are estimated separately for every CRA. It turns out that all independent variables are insignificant at the 10 % level.[6] Therefore we conclude that institutional factors do not play a significant role when evaluating developed sovereign debtors.

Next, we go back to the question about the validity of the upgrade results with respect to the financial crisis time frame. An ordered probit with the dependent variable defined as taking values of -1 (for a downgrade), 0 (for no change), and 1 (for an upgrade) is estimated. The results can be found in the appendix in Table 2.12. Overall the performance of this model is not impressive, with very few significant variables and an R-square that is closer to the upgrade probit ones, than it is to the downgrade estimations. Therefore, the estimations obtained beforehand are not fundamentally biased by the fact that the binary rating change series is asymmetric, but instead upgrades are indeed less well explained in our estimations than downgrades are.

As a next step, the financial crisis is excluded from the sample, and the simple static regression is re-estimated. However, it is not a priori clear where to make the cut off. One possibility is to take the collapse of Lehman Brothers as the starting point of the financial crisis. Another possibility is to take the date when Greece's fiscal problems were first made public, in spring 2010. The latter would make more sense, since we analyze sovereign data exclusively. To settle the argument,

[5]The logit estimates are available upon request.

[6]The estimates are available upon request.

a breakpoint-test due to Andrews (1993) is conducted. It detects breaks in three out of four series (pooled, Moody's, and S & P) in the last quarter of 2008, which is coinciding with the Lehman Brothers bankruptcy. Therefore, the (static) probit-upgrades regression is re-estimated excluding the years 2009 and 2010. The results can be found in the appendix in Table 2.13. It turns out that only using the subsample is not improving the estimations, which means that the explanation of the financial crisis diluting the sample needs to be discarded. Therefore, it seems that the CRAs were assigning upgrades to countries without thoroughly looking at the fundamentals of these countries.

2.4 Conclusion

In this chapter we investigated the behavior of CRAs assessing sovereign solvency of European Nations. Probit regressions are used to investigate upgrades and downgrades. While controlling for typical macroeconomic variables (GDP growth, sovereign debt level and change, etc.) frequently used in the literature we come to the following results. There are only a few cases when rating agencies take the business cycle into account, however, there is no clear pattern identifiable. What we find is that there is a dynamic element in the rating change, usually leading to an increased probability of rating changes in the future, given that there is a rating change happening now. This might lead to a vicious circle, where financially troubled economies are further downgraded, leading to a higher cost of borrowing which in turn leads to more financial distress. Also employing ordered probit regressions to analyze actual sovereign debt ratings, we find that while controlling for typical macroeconomic factors, CRAs take the rating history into account as well. Furthermore we find that there is a distinct asymmetry between upgrades and downgrades. While downgrades can be very well explained by macroeconomic characteristics, this is not the case for upgrades. A possible explanation for this is that CRAs were ignoring country specific risks after the introduction of the Euro, which implies that these upgrades were handed out too freely. This is also in line with the convergence of bond yields between Euro countries during the early 2000s. All these findings are robust to a number of specification tests. Due to the extensive impact these ratings have on capital allocation as well as sovereign budgets, the current overreliance on ratings needs to be addressed. The EU regulation 1060/2009 is a good first step into that direction. Other countries should follow this example on a national regulatory basis, but also international cooperation is needed to address the problem.

2.5 Appendix

Table 2.12: Ordered Probit estimation Rating Changes

	rating change
GDP	-0.465***
	(0.124)
Government Debt	-0.004*
	(0.003)
Inflation	-0.048
	(0.030)
Growth	0.060***
	(0.017)
Primary Balance	0.109***
	(0.029)
Current Account	1.857
	(1.702)
Business Cycle	0.168
	(0.120)
Interaction	-0.252
	(0.189)
GIPS	0.028
	(0.233)
R-squared	0.182

Standard errors in all estimations are clustered robust.

Table 2.13: Upgrades Estimations using Binary Probit subsample until Q4 2008

	Pooled	Moody's	S & P	Fitch
Government Debt	0.006**	0.001	-0.011***	0.000
	(0.004)	(0.006)	(0.006)	(0.005)
Inflation	-0.027	-0.007	-0.013	-0.043
	(0.019)	(0.024)	(0.026)	(0.027)
Growth	0.059*	0.065	0.032	0.076
	(0.032)	(0.045)	(0.043)	(0.040)
Primary Balance	-0.010	-0.019	0.027	-0.016
	(0.031)	(0.044)	(0.044)	(0.039)
Current Account	-1.955*	-1.771	-1.941	-1.445
	(1.096)	(1.456)	(1.468)	(1.406)
Business Cycle	0.122	-0.144	0.089	1.813
	(0.177)	(0.208)	(0.251)	(24.651)
GIPS	-0.217	-0.146	-0.121	-0.319
	(0.224)	(0.315)	(0.318)	(0.305)
Constant	-1.541***	-2.217***	-1.768***	-3.797
	(0.290)	(0.408)	(0.404)	(24.653)
R-squared	0.073	0.051	0.080	0.075

Standard errors in all estimations are retrieved from the Hessian.

Table 2.14: Correlation Table of Independent Variables

	GDP	Government Debt	Inflation	Growth	Primary Balance	Current Account	Business Cycle
GDP	1						
Government Debt	0.283	1					
Inflation	-0.233	-0.204	1				
Growth	-0.215	-0.288	0.328	1			
Primary Balance	0.291	0.041	0.124	0.339	1		
Current Account	0.124	0.065	-0.066	-0.028	0.132	1	
Business Cycle	0.101	0.045	0.356	0.550	0.460	0.038	1

Based on 903 observations.

Chapter 3

Default probabilities, CDS Premiums, and Downgrades

3.1 Introduction

This chapter investigates whether Credit Rating Agencies (CRAs) are correctly assessing the default probability of a country. Since there are very few instances of a sovereign nation defaulting, we cannot simply search for patterns in ratings given to countries prior to defaulting. Therefore, this chapter takes a different route and extracts implied default probabilities from CDS data and regresses them on downgrades. A sovereign CDS is an insurance on the default of a government issued bond, while a sovereign rating gives the likelihood of the default of a country. Thus, the premium of a CDS essentially measures the same as a rating. The main reason for only analyzing downgrades is that in the time-frame for which there are sufficient sovereign CDS premiums available (i.e. from 2006 onwards), we see nearly exclusively downgrades for European countries.

The initial seller of a CDS is a bank or insurance company. The buyer pays a fee to the seller, and in return receives the nominal value of the underlying asset, in case the asset defaults. This fee may be paid in installments (which is called a spread) or at once (which is called a premium). The fee is calculated using no-arbitrage arguments. The literature on the determinants of sovereign CDS prices is quickly growing. There are two main strands of the literature. The first one claims that most of the variation of sovereign CDS can be explained by global factors such as the state of US financial markets and its economy. Pan and Singleton

We would like to thank Thomas Götz, Lenard Lieb, Alain Hecq, and Bertrand Candelon for feedback and discussion while writing this chapter. We would also like to thank the particpants of the Spring Meeting of Young Economists, April 2014, Vienna, Austria, especially my discussant Willi Mutschler, as well as the participants of the IAAE 2014, June 2014, London, United Kingdom.

(2008) find systemic risk in the credit-event arrival for three countries using a one-factor lognormal model. Ang and Longstaff (2013) use a multifactor affine model to disentangle systemic and local effects on CDS spreads. They come to the conclusion that macroeconomic variables are less important than the systemic risk created by global financial markets. The second strand of the literature claims that local risk factors are mainly driving CDS spreads. Alter and Schüler (2012) look at interdependencies between sovereign CDS and financial institution CDS using multi-equation regressions and conclude that the interaction between actors in the local financial industry is driving sovereign CDS spreads in the post financial crisis world. Several papers merge the two strands of literature. Using panel regression techniques, Arghyrou and Kontonikas (2012) find that it was only after 2007 that EMU CDS spreads are driven by macroeconomic fundamentals, while before 2007 EMU fundamentals were not significant. Remolona et al. (2008) decompose CDS spreads into expected losses and market risk premiums and find that the former is explained by macroeconomic variables, while the latter is explained by global risk factors.

Ratings are created in a completely different fashion. A rating agency is a corporation which is specialized in assessing the probability of default of entities (including but not limited to firms and states) or financial assets (such as collateralized debt obligations). This assessment is published via a rating which is similar to grades. Therefore, the probability is not explicitly communicated. The market for ratings has been dominated for decades by three big rating agencies, namely Moody's, Fitch, and Standard and Poor's (S & P).[1] Recently, these agencies have come under severe criticism for their role in the 2007/2008 financial crisis as well as their behavior in the European debt crisis.

Research about the determinants of sovereign credit ratings started in the nineties. Cantor and Packer (1996) are among the first to investigate them and established a set of macroeconomic variables that explain sovereign ratings. Afonso (2003) is refining this line of research. In his paper the author is proposing a list of variables that are significantly correlated with ratings as well as looking at different transformation of ratings. Mora (2006) extends this research by analyzing the east-Asian financial crisis and comes to the conclusion that CRAs are rather conservative in times of crisis. Contrary to this conclusion, Ferri et al. (1999), who also analyze the east-Asian crisis, find that CRAs behave in a procyclical fashion during times of financial distress.

So far there are only a few papers investigating the connection between sovereign CDS and sovereign ratings. Arezki et al. (2011) look at the impact of sovereign rating changes on several financial markets, including the CDS market. Ismailescu and Kazemi (2010) conduct an event study on the reaction of sovereign CDS on sovereign rating changes for emerging economies. Afonso et al. (2012) also conduct an event study using European CDS, bond yield, as well as rating data. They conclude that the effects between ratings and the other two series is significant, but also bi-directional. From corporate rating research, we know that CDS tend to be

[1] See White (2009)

able to forecast rating changes in advance, see for example Hull et al. (2004).

When analyzing ratings, the estimation method usually employed is a binary or ordered choice regression with ratings as a dependent variable and variables explaining defaults as regressors. Typical examples for this are Gande and Parsley (2005), Mora (2006) or Freitag (2014). In order to tackle our research question, a similar methodology is needed. The difference between the papers mentioned and my approach is that they use macroeconomic variables that are published on a quarterly basis, while we make use of financial data which comes at a daily frequency. One could argue that ratings are also published at a daily frequency, since at every given trading day, a rating can be published, but this ignores the work that a CRA invests into a rating. Indeed, the typical rating process takes at least around a month for the big three agencies, as documented in Fuchs and Gehring (2013). This means that an estimation method is needed that is able to correlate data which is sampled at different frequencies. There are several options to approach this issue. The first possibility is to simply average over the high frequency data. This is a convenient solution, however it causes an enormous loss of information contained in the high frequency series. Another possibility is a distributed lag model in which all current and past values (up to a certain point) are used as regressors to explain the low frequency variable. The problem with this approach is that it quickly leads to overfitting. The most recent innovation in the field of mixed frequency data was the introduction of MIDAS by Ghysels et al. (2004). The main innovation of MIDAS is that it lets the data determine its lag-length and weighting. This is done by using a polynomial with a small amount of parameters to weight the high frequency data. The values of these parameters are determined by optimizing them together with a slope coefficient using a minimum distance-type estimation. Therefore, MIDAS methods still use a substantial amount of information contained in high frequency data, while at the same time being a relatively parsimonious model to estimate.

Early examples of MIDAS application is Ghysels et al. (2005). It shows that the theoretical sound, but empirically rather elusive relationship between risk and return can be found in a time series setting by using MIDAS estimations. More recent applications of MIDAS are focused on forecasting of GDP or other economic indicators by using financial assets as regressors in addition to the more traditional explanatory variables. Typical examples are Clements and Galvão (2008) or Schumacher and Breitung (2008). Also, a wide array of different MIDAS derivatives were developed. Guérin and Marcellino (2013) construct a Markov-Switching MIDAS and apply it to US GDP-growth as well as US industrial production. A Factor MIDAS was developed by Marcellino and Schumacher (2010) who introduce three different MIDAS estimation methods and join them with three methods for factor analysis. Their methods are then applied to German GDP. Foroni et al. (2012) introduce an unrestricted MIDAS. This is a method of estimating a MIDAS with OLS, which works well if the difference in sampling frequency is relatively small. A smooth transition MIDAS is formulated by Galvão (2013) and applied to out-of-sample US and UK output growth. Foroni and Marcellino (2013) give a good overview of the current development in Mixed Data sampling. However, up to our knowledge, no binary choice MIDAS estimator has been developed, which would be the estimator of choice for this research. The contribution to the existing literature is twofold. First, we develop a MIDAS estimator for binary data. This is similar to papers such as the above mentioned Guérin and Marcellino (2013). Second, we try to asses the appropriateness of sovereign ratings by correlating them with filtered CDS data. This approach is connecting the two literature strands of event study analysis, such as Ismailescu and Kazemi (2010), and categorical regression models, for example Mora (2006).

The rest of the chapter is structured as follows. In section 2 CDS and ratings are discussed, in section 3 a short introduction to MIDAS-estimator is given, followed by the probit-MIDAS estimator. Section 4 contains simulations for the estimator, and in section 5 we estimate the relationship between sovereign CDS and sovereign rating changes. Section 6 concludes.

3.2 The Economics of CDS and Credit Ratings

In this section, the economics of CDS and the interaction with credit ratings are explained. CDS are a derivative that is insuring the buyer against the default of the asset underlying the CDS. The CDS works in the following way: The seller issues a contract promising to pay whoever is holding the contract up on default of the underlying asset, the nominal value of that asset. The buyer of a CDS is paying the seller a price which might be paid in one installment, in which case it is called the premium, or in staggered installments which are referred to as spread. The premium or spread can be calculated using:

$$\sum_{i=1}^{N} e^{-rt} Q(t_i) p = \int_0^{t_N} e^{-rt} (100 - M_q) q(t) dt, \tag{3.1}$$

where r is the risk free rate, $Q(t)$ is a risk neutral survival probability, M_t is the market value, and p is the CDS premium. The left-hand-side of this equation is the spread of the CDS, the right-hand-side is the insurance part which guarantees the reimbursement of any losses incurred after the underlying asset goes bankrupt. The only component that needs to be determined is $Q(t)$. This is done by matching the underlying asset with a risk-free counterpart of the asset. For example, the counterpart of an Italian government bond, would most likely be a German government bond with otherwise identical characteristics. The difference in price between these two bonds can be used to calculate the default risk that the market attaches to the non-risk-free bond. For further details, see Hull and White (2000).

In contrast to CDS which are a recently introduced derivative, Credit Rating Agencies have existed since the beginning of the 20th century. Starting in 1909, with the founding of Moody's, CRAs became important actors in financial markets after the great depression in 1929. During that time the ratings of CRAs became binding for a large number of market participants due to regulation by the Security Exchange Commission (SEC). These regulations also helped establishing the oligopoly that is currently present in the markets structure by making merely the ratings of Moody's, Fitch, Standard, and Poor's (back then still two separate entities) binding. In the seventies, the CRAs switched from a user-pays to an issuer pays system. This change came mainly because of the advent of copy machines which threatened the original business model of CRAs of selling booklets containing their ratings. From 1980 onwards, the CRAs became more involved into the business of sovereign ratings, mainly because developing countries were trying to access international capital market and needed a seal of approval for their bonds. In the last decade the rating agencies played a major role in the subprime debacle and the financial crisis by assigning top ratings to collateralized debt obligations (CDOs) consisting mainly of US mortgages. A lot of these mortgages were defaulting during the financial crisis despite the AAA rating given to the CDOs in which they were packaged.[2]

CDS as well as credit ratings are measuring very similar things. While a rating is an estimate about the default probability of an entity, given by a CRA, in case of a CDS this default probability is reflected by its premium. However, the way ratings and CDS premiums are created is very different. A rating is constructed by a specialist at a CRA who takes into account macroeconomic factors as well as other country specific factors. CDS premiums are constructed using equation 3.1 and therefore the main driver are bond prices. This means that the premium of a CDS is driven by financial markets. This difference in construction might also dilute the correlation between these two variables. What follows is a discussion of possible reasons.

The first reason that needs to be mentioned is liquidity in the financial markets. Indeed, the lower the trading volume of a bond is, the more likely it is to have price distortion. The CDS premium is a nonlinear transformation of this price. Therefore, any reduced liquidity in the bond market directly affects the CDS premium. This would imply that CDS premium are increasing disproportionate in times of liquidity crunches. In terms of correlation this would imply that there is either a negative correlation or no correlation between CDS premium and ratings. However, this is not what we see in the majority of the results. The next issue that needs to be discussed is a possible decoupling of individual country characteristics from bond prices. Oliveira et al. (2012) provide some evidence that this is the case before 2007. Aßmann and Boysen-Hogrefe (2012) shows that coefficients for macroeconomic factors in the Eurozone are time varying, with a noticeable increase after the crisis. However, the data used in this chapter has its earliest point in 2006, and so this study uses mainly crisis or post-crisis data. Therefore, this issue should have no major impact in the estimations.

[2] For an overview of the history of Credit Rating Agencies see White (2010)

3.3 Methodology

In this section the Probit-MIDAS estimator is described. It is an extension of the existing MIDAS estimator first introduced by Ghysels et al. (2004). The idea behind the MIDAS estimator is to use high frequency data, such as asset returns, to explain low frequency variables such as GDP or inflation rates. This is done by imposing a parametric weighting scheme on the high frequency variable(s). The main advantage of using MIDAS is that the weighting can extract information relevant for the estimation from extremely noisy data.

3.3.1 The MIDAS estimator

We start by describing the MIDAS estimator. A simple MIDAS data generating process (DGP) can be written as:

$$y_t = \beta x_t^{(m)}(\theta) + u_t, \tag{3.2}$$

where:

$$x_t^{(m)}(\theta) = \sum_{j=1}^{q} w_j(\theta) L^{j/m} x_{t-j/m}^{(m)}, \tag{3.3}$$

with m being the sampling frequency of the high frequency variable, L is the lag-operator, and $w_j(\theta)$ is a polynomial function for weighting the regressor data. In order to estimate this Andreou et al. (2010) show that only two assumptions are needed:

1) u_t is white noise.

2) $0 \leq w_j(\theta) \leq 1$ and $\sum_{j=1}^{q} w_j(\theta) = 1$.

This allows the identification of the slope coefficient. A widely used functional form for the weighting polynomial is the Almon lag polynomial:

$$w_j^a(\theta_1, \theta_2) = \frac{a(\frac{j}{j^m}, \theta_1, \theta_2)}{\sum_{j=1}^{m} a(\frac{j}{j^m}, \theta_1, \theta_2)}, \tag{3.4}$$

with

$$a(\frac{j}{j^m}, \theta_1, \theta_2) = exp(\theta_1 j + \theta_2 j^2). \tag{3.5}$$

The estimations are executed using Nonlinear Least Squares (NLS). Andreou et al. (2010) show that the estimations are asymptotically efficient and in most cases superior to other approaches which are able to deal with data sampled at different frequencies such as averaging the high frequency data, or using distributed lag models. However, the main advantage of MIDAS compared to these other models is that on the one hand it is data-driven, that is the weighting scheme is estimated and therefore relatively little assumption on the exact weighting of the data is needed. On the other hand it is also a rather parsimonious approach which only needs to estimate one or two additional parameters. Ghysels et al. (2007) note that in principle non-linear MIDAS estimator are possible. Several examples are mentioned in the introduction of this chapter such as the Markov-Switching MIDAS estimator by Guérin and Marcellino (2013). In that framework, our estimator is another nonlinear MIDAS estimator.

3.3.2 The probit estimator

Binary choice methods in general, and the probit estimator in specific has long been a working horse of empirical microeconomics, although it has also seen some application in macroeconomics, for example Candelon et al. (2014). Probit estimations assume a latent variable DGP such that:

$$y_t^* = x_t\beta + \epsilon_t, \tag{3.6}$$

and

$$y_t = \begin{cases} 1 & \text{if } \; y_t^* > 0 \\ 0 & \text{if } \; y_t^* \leq 0 \, , \end{cases} \tag{3.7}$$

In order to conduct estimations on this kind of data, a maximum likelihood approach is often used. The log-likelihood function(LLF) is:

$$l_t(\beta) = y_t \;\, ln[\Phi(x_t\beta)] + (1 - y_t) \;\, ln[1 - \Phi(x_t\beta)], \tag{3.8}$$

where $\Phi(.)$ is the cumulative standard normal distribution function. Apart from the standard maximum likelihood assumptions, the following additional assumptions are needed:

1) The error ϵ_t needs to be standard normal.

2) The data needs to follow a latent process from equation 3.7.

These can be found in Maddala (1986).

3.3.3 The probit-MIDAS estimator

Consider now a combination of the MIDAS DGP and the probit DGP in the form of

$$y_t^* = x_t(\theta)\beta + e_t, \tag{3.9}$$

where e_t has a normal distribution with mean zero and variance one. However, y_t is not actually observable but instead also governed by equation 3.7. In order to conduct estimations on this kind of data, one has to combine the probit and the MIDAS approach. The most straightforward way to do so is to embed the optimization routine for the MIDAS polynomial into the log likelihood function of a probit regression which yields the following log-likelihood function:

$$l_t(\beta, \theta) = y_t \;\, log[\Phi(x_t^{(m)}(\theta)\beta)] + (1 - y_t) \;\, log[1 - \Phi(x_t^{(m)}(\theta)\beta)]. \tag{3.10}$$

This estimation method needs to fulfill all assumptions for both approaches. This means the following assumptions need to be fulfilled:

1) The error e_t needs to be standard normal.

This is the standard error assumption for probit regressions. Note that it is placing additional restriction on the MIDAS assumption number one. Instead of requiring only finite variance, it is now restricted to be equal to one. Also, the distribution of the errors is now specified.

2) $0 \leq w_j(\theta) \leq 1$ and $\sum_{j=1}^{q} w_j(\theta) = 1$.

This is the same assumption as in the standard MIDAS model and the polynomials are always constructed in such a way that this assumption is fulfilled.

3) The data needs to follow a latent process described in equation 3.7.

Additionally the standard MLE assumptions are needed.

3.3.4 Testing the $\beta = 0$ Hypothesis

MIDAS regressions are often used for forecasting of various economic series, such as in Clements and Galvão (2008) or Götz et al. (2014). But an economist might also be interested in the significance of the slope coefficient(s) of the high frequency variable. However, MIDAS is problematic when we want to test the significance of said coefficient. This is due to the joint estimation of slope coefficients and weighting parameters and the consequent non-identification of the weighting parameters under the null hypothesis. To understand the problem, let us assume that the slope coefficient of a MIDAS regression is actually equal to zero. This means that the weighting parameters are no longer uniquely identified. Looking at the alternative hypothesis it is obvious that the weighting parameters can only take the value determined by our estimations, and are therefore uniquely identified. This means we have a lopsided scenario in which not all parameters are identified under one hypothesis but all of them are identified under the other hypothesis. Clearly, this is not a desirable scenario to conduct inference. Therefore, a recipe for the probit-MIDAS, which is bypassing this problem, is presented in this section. This involves a two-step procedure as well as a bootstrap. This approach is proposed by Ghysels et al. (2007). However, to the knowledge of the author this chapter is one of the first to apply the bootstrap and provide simulation results for it.

Davies (1987) recognizes the problem of non-identified nuisance parameters under one type of hypothesis. He proposes to estimate over a grid of the nuisance parameters and take the supremum instead of conducting standard inference. His paper provides asymptotic behavior for a simple test under these conditions. Hansen (1996) takes this approach and generalizes it. He notes that the distributions of these tests depends upon a large number of parameters which makes deriving asymptotic behavior tedious as well as difficult to generalize. Instead he proposes a bootstrap approach. For the case of a probit-MIDAS this procedure can be implemented as follows:[3]

1) Estimate a probit-MIDAS regression using equation 3.10.

2) Construct a grid of c different θ combinations.

3) Use this grid to weight the regressor data, producing c different regressor series, and estimate c standard probit regressions.

4) Take the supremum over the t-statistics of the coefficient of these regressions as explained in Davies (1987).

Now we have a t-statistic. As a next step the critical value for this t-statistic needs to be computed. This is done by bootstrapping this procedure under the null hypothesis. What follows is the bootstrap loop:

5) Start by drawing bootstrap samples with replacement from the residuals.

6) Construct the dependent variable by adding up residuals and intercept.

7) Repeat step 3) and 4) from the first part with the bootstrap sample.

8) Repeat step 5) to 7) k times, where k is the amount of bootstrap samples.

9)Take the $95th$ percentile of the t-statistics to produce a critical value against which the t-statistic from step 4) can be evaluated to test the significance of the slope coefficient.

We think that one remark with respect to the actual bootstrapping is in order. A potential problem with the bootstrap is that there are no straightforwardly defined residuals in a probit regression. The standard solution in the literature to this issue is to use generalized residuals due to Gourieroux et al. (1987), see for example Hsiao et al. (2012). However, one assumption when estimating a probit is that the errors are distributed as a standard normal random variable. Generalized residuals seem to consistently have a significant lower variance than one in this setup, and therefore seem to be unsuited for constructing the bootstrap sample. Possible other candidates are Pearson residuals, response residuals, and deviance residuals (see Hinkley et al. 1991). In section 3.4.2, a small-scale simulation study is done concerning the behavior of the different residual options available. In general, it is advised to inspect the residuals before starting the bootstrap.

3.4 Simulations

3.4.1 Finite Behavior of the Estimator

We start by investigating how the probit-MIDAS estimator behaves in finite sample given different weighting schemes and different sampling frequencies. Then we

[3]This example uses only one regressor series

investigate the behavior of the bootstrap. Starting with the first part, two different polynomials will be investigated. Apart from the Almon polynomial detailed in equation 3.5, we also include the Beta polynomial first introduced in Ghysels et al. (2007) which is:

$$w_j^b(\theta_1, \theta_2) = \frac{b(\frac{j}{j^m}, \theta_1, \theta_2)}{\sum_{j=1}^{m} b(\frac{j}{j^m}, \theta_1, \theta_2)}, \tag{3.11}$$

with

$$b\left(\frac{j}{j^m}, \theta_1, \theta_2\right) = \frac{(\frac{j}{j^m})^{\theta_1 - 1}(1 - \frac{j}{j^m})^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \tag{3.12}$$

where Γ represents the gamma-function. It is as flexible as the Almon polynomial and offers a wide variety of shapes. Both parameters need to be positive. Also both parameters are constrained from above (by 15), to allow for better convergence. The Almon polynomial parameters are constrained from above and below by -1 and 1, because otherwise the function becomes non-smooth. When comparing the two weighting schemes, the following Data Generating Process (DGP) is used:

$$y_t = \beta_1 x_t^{(m)}(\theta) + \beta_2 z_t + e_t. \tag{3.13}$$

Whereas $x_t^{(m)}(\theta)$ is a high frequency regressor, z_t is sampled at the speed as the dependent variable, and e_t is $N(0, 1)$. The reason for designing the DGP in such a way is that we can compare the estimations of the high frequency estimator with that of a low frequency regressor by comparing the performance of the first regressor to that of the second regressor. The simulations for every case investigates three different weighting schemes. These are downward sloping weights, upward sloping weights and a hump shape. The weighting in the DGP is always generated by the same polynomial that is used for the estimation. That is when using the Almon polynomial as a weighting polynomial in a probit-MIDAS estimation, the high frequency variable in the DGP is also generated by the Almon polynomial.

Typically MIDAS estimator are working with weekly data ($m = 5$), monthly data ($m = 22$), or quarterly data ($m = 66$). In this chapter we only look at the first two, that is monthly frequencies, and weekly frequencies. The reason for this lies in the characteristics of the rating variable. In most cases, when considering a MIDAS approach, it is a priori clear at which sampling frequency the low frequency variable is sampled at. For example, quarterly GDP estimates have a predefined date each quarter on which they are published. Ratings do not have a pre-specified date when they are announced. Thus, it is better to work with weekly or monthly frequency. The intuition behind this is that if one uses quarterly data, the ratings that are published in consecutive quarters could be up to 6 month apart. This seems to be an unacceptable difference between estimation and reality.

The first case to investigate is the monthly sampling frequency case, that is, $m = 22$. The DGP for the simulation is given by equation 3.13 with $\beta_1 = 0.5$ and $\beta_2 = 0.3$.The Almon polynomial parameters are $(0.035, -0.085)$ for generating a downward sloping weighting, $(-0.02, 0.005)$ for an upward sloping weighting, and $(0.005 - 0.0005)$ for having a hump shaped weighting scheme. For the Beta polynomial we use $(0.6, 2.1)$ to get a downward sloping weighting scheme, $(1.1, 1)$ for generating an upward slope, and $(4, 2.1)$ to have a hump shaped pattern. The parameters are chosen in such a way that they generate the appropriate shape in the given sampling frequency.[4]

All simulations are done with 2000 replications, the sample sizes are 50, 100, 250, 500, 750, and 1000. The tables are based on $\frac{\beta}{\beta^t}$ with β^t being the true parameter underlying the latent process. This should, due to the identification of the probit coefficient which is $\frac{\beta}{\sigma_e}$, and the fact that the errors are normally distributed with $\sigma = 1$, converge to 1. The deviations from this are denoted in percentages. The θ-parameter are not reported.

We start with the Almon polynomial. All three cases are displayed in Table 3.1. As we can see the three cases are quite different from each other. The estimator performs reasonably well in a downward sloping weighting scheme, however the sample size to achieve full convergence seems quite high. When looking at the other two cases, we can see that the estimator does not perform well at all. In both cases there are still significant biases in the slope coefficients, even at a sample size of 1000. Fortunately these two cases are much rarer in economic applications than the downward sloping one.

The second polynomial that is investigated is the Beta polynomial. Starting again with the downward sloping case, one can see that the estimator is approaching the true value rather quickly. With a sample size of 250 the coefficient approaches very closely the true value and fluctuates around it when increasing the sample size. These results resemble the results of the Almon polynomials when it has a sample size of 1000 to estimate on. In line 3 and 4 of Table 3.2 an upward sloping weighting scheme is investigated. The estimations are performing similar to the downward sloping weighting scheme.

[4] Additionally there were some conflicts with the procedure supplying the initial values, and it was deemed safer to use parameter that are not exactly the same as those covered by this procedure.

Table 3.1: Almon polynomial simulations monthly frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping	β_1^{HF}	-7.3 %	-15.6 %	-19.0 %	-6.2 %	-3.1 %	-1.6 %
	β_2	12.2%	5.4 %	3.3 %	1.5 %	0.5 %	0.3 %
upward sloping	β_1^{HF}	-22.2 %	- 33.4 %	-39.0 %	-26.5 %	-18.9 %	14.1 %
	β_2	14.8 %	4.4 %	2.1 %	1.9 %	1.0 %	- 0.0 %
hump-shaped	β_1^{HF}	-49.5 %	- 67.9 %	-77.0 %	- 72.4 %	-65.8 %	-62.4 %
	β_2	14.4 %	6.4 %	2.4 %	0.9 %	0.1 %	0.2 %

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.2: Beta polynomial simulations monthly frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping weighting	β_1^{HF}	11.2 %	0.4 %	2.2 %	2.2 %	4.4 %	0.7 %
	β_2	17.5 %	7.1 %	2.4 %	1.1 %	1.1 %	0.7 %
upward sloping weighting	β_1^{HF}	10.4 %	6.2 %	1.9 %	-3.1 %	-1.6 %	-1.8 %
	β_2	17.1 %	8.2 %	3.3 %	1.1 %	1.4 %	0.9 %
hump-shaped weighting	β_1^{HF}	11.2 %	-1.0 %	0.3 %	0.3 %	0.3 %	0.4 %
	β_2	12.9 %	6.9 %	2.5 %	0.8 %	0.7 %	0.4 %

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.3: Almon polynomial simulations weekly frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping weighting	β_1^{HF}	38.5%	16.6%	7.8%	4.8%	0.0%	0.0%
	β_2	13.3%	5.2%	2.5%	0.9%	3.8%	1.6%
upward sloping weighting	β_1^{HF}	-4.0%	-5.5%	0.9%	5.1%	4.8%	4.2%
	β_2	10.8%	4.9%	2.3%	1.0%	0.8%	0.5%
hump-shaped weighting	β_1^{HF}	19.6%	9.6%	4.5%	2.1%	2.7%	1.5%
	β_2	12.6%	5.9%	1.9%	0.8%	0.1%	0.5%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.4: Beta polynomial simulations weekly frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping weighting	β_1^{HF}	-15.1%	-21.6%	-24.3%	-23.2%	-20.1%	-18.7%
	β_2	14.3%	6.9%	1.8%	0.7%	0.5%	0.7%
upward sloping weighting	β_1^{HF}	-15.1%	-28.0%	-27.5%	-24.5%	-24.3%	-23.2%
	β_2	10.5%	2.0%	2.1%	0.7%	0.3%	0.1%
hump-shaped weighting	β_1^{HF}	8.8%	4.6%	1.1%	1.4%	1.3%	0.5%
	β_2	14.9%	6.6%	3.2%	0.1%	0.9%	0.7%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

This is clearly superior to the downward sloping scheme using the Almon polynomial. As a last case a hump-shaped polynomial is investigated. Observing a very similar pattern as in the downward sloping case, the estimator achieves convergence to the true value relatively quickly and with a sample size of 250 the coefficient is already very close to the true parameter. However, there is also a disadvantage to this performance. The Beta polynomial has a relatively high number of cases (around 10 %) where the algorithm exits without finding a proper value for the slope coefficients and parameters.

As a next step a weekly sampling frequency is investigated, that is $m = 5$. The DGP is the same as in the monthly simulations. For the Almon polynomial we use $(-0.5, 0.085)$ to create a downward sloping weighting scheme, $(0.5, 0.085)$ to build an upward sloping weighting scheme, and $(0.5, -0.085)$ to have an inverted-U shaped weighting scheme. For the Beta polynomial $(0.9, 1)$ is used for a downward sloping weighting, $(1.1, 1)$ is creating an upward sloping scheme, and $(4.1, 2.1)$ is for an inverted-U shaped weighting scheme. Otherwise the setup is identical to the previous simulation case, 2000 replications are done and the sample sizes are 50, 100, 250, 500, 750, and 1000. The results for the Almon polynomial are in Table 3.3

The results in this set of simulations looks good. The slope coefficient is converging for the downward sloping case, as the number of observation increases to no detectable bias at all. The distortion in the estimation when assessing the hump shaped pattern is relatively small. There is still some bias in the upward sloping weighting scheme, however, when comparing these results to the results obtained when using a monthly frequency, we can see a significant improvement over those results.

The second set of simulations investigates the Beta polynomial. Also in this case do we see a significant difference compared to the previous set of simulations. However, we now have a worsening of the performance of the estimator. For the first two weighting schemes, there is a heavy bias, even when the sample size is at 1000. Only in the hump shaped case does one see an acceptable result for the estimator. Furthermore, compared to the monthly frequency case, the number of simulations where the algorithm exits the simulation without finding acceptable results is vastly increased, and in some cases up to 50 %.

Therefore, we come to the following conclusions. For the monthly sampling frequency, the best weighting scheme seems to be produced by the Beta polynomial. It needs significantly fewer observations in converging to the true parameter. Furthermore it does a good job in converging when dealing with a hump-shaped pattern. The Almon polynomial does not converge to its true parameter in this case. Contrary to the results for monthly frequency, for the weekly sampling frequency, the Almon polynomial seems to be doing the best job. Convergence to the true parameter is achieved with a modest sample size, and it can model a variety of shapes.

The obvious question is why does the Almon polynomial do such a bad job in the monthly setup, compared to the Beta polynomial and vice versa for the weekly setup. Starting with the monthly estimations, it should be noted that for the upward sloping weighting and the hump-shaped weighting, the Almon polynomial tends to prefer corner solutions. That is, at least one of the weighting parameter is equal to one. This in turn tends to lower the value of the regression coefficient, which explains the large negative deviations from the true coefficient value in Table 3.1. The biggest difference between the two polynomials is a significantly increased forced exit rate when using the Beta polynomial in the monthly sampling frequency. A forced exit means that the optimizations algorithm is unable to converge in the given amount of iterations. With the Almon polynomial, this rate was around 0.1%, however using the beta polynomial this went up to around 10 %. This means it is more obvious for the algorithm when the estimates are clearly off and it counteracts this by eliminating the faulty estimates. Therefore, the Beta polynomial does not suffer from the corner-solution case that can be found with the Almon polynomial. The forced exit rate is most likely also the reason for the difference in performance in the weekly sampling case. The Beta polynomial already has slight problems converging with significantly more observations, so it seems logical that for the weekly case the forced exit cases are increased. Indeed, the forced exit rate for the Beta polynomial shoots up to 50 %, which causes the drop in point-estimate performance.

One interesting fact worth mentioning is that the estimator seems to do reasonable well in small sample sizes, performs relatively bad in medium sized sample sizes, and better again in large sample sizes. However we do not have a decent explanation as to why this is the case.

So far we have seen that the estimator has overall good finite sample properties, but the estimator has the tendency to select corner solutions. This is obviously a problem if it happens in an empirical application. What follows are approaches to circumvent this problem. If we look again at the estimations, the downward-sloping weighting scheme seems to perform best. Therefore, it would be good if we could impose such a weighting on the estimator, in case of non-convergence. There is such a polynomial in use in the MIDAS-literature. Ghysels et al. (2009) introduce the hyperbolic polynomial:

$$w_j^h(\theta_1) = \frac{h(\frac{j}{j^m}, \theta_1)}{\sum_{j=1}^{m} b(\frac{j}{j^m}, \theta_1)}, \tag{3.14}$$

with $h(\frac{j}{j^m}, \theta_1)$ being:

$$h(\frac{j}{j^m}, \theta_1) = \frac{\Gamma(j + \theta_1)}{\Gamma(j+1)\Gamma(\theta)}. \tag{3.15}$$

Only one parameters needs to be estimated. This parameter is constrained such that $0 < \theta < 0.5$ in order to guarantee stationarity (see Tanaka 1999). Since there are no different weighting pattern to be investigated, we simply look at three different parameter values. For the monthly frequency these are: (0.11), (0.251), and (0.41). The results are in Table 3.5.

Table 3.5: Hyperbolic Polynomial Simulations Monthly Frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
$\theta = 0.11$	β_1^{HF}	36.4 %	25.6 %	18.1 %	15.8 %	10.4 %	9.5 %
	β_2	7.5 %	3.5 %	1.8 %	1.0 %	0.6 %	0.6 %
$\theta = 0.251$	β_1^{HF}	19.6 %	14.6 %	9.8 %	8.7 %	6.2 %	5.6 %
	β_2	10 %	4.3 %	1.4 %	0.8 %	0.2 %	0.9 %
$\theta = 0.41$	β_1^{HF}	4.8 %	10 %	1.2 %	2.0 %	-0.7 %	-1.8 %
	β_2	1.28 %	3.6 %	1.5 %	0.5 %	0.6 %	0.0 %

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.6: Hyperbolic Polynomial simulations weekly frequency

Weighting-scheme	Coefficients	50	100	250	500	750	1000
$\theta = 0.11$	β_1^{HF}	19.6%	10.5%	5.4%	3.4%	2.5%	2.4%
	β_2	7.0%	4.4%	1.3%	0.3%	0.0%	0.0%
$\theta = 0.3$	β_1^{HF}	11.9%	4.8%	1.4%	-1.5%	-1.4%	-1.9%
	β_2	10.6%	4.5%	1.1%	0.6%	1.1%	0.4%
$\theta = 0.49$	β_1^{HF}	15.5%	7.6%	2.3%	1.0%	1.5%	0.3%
	β_2	11.5%	4.7%	2.2%	0.9%	0.4%	0.0%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

As it can be seen the results for a rather low parameter coefficient are distorted even with 1000 observations. However, the higher the parameter-value gets, the better the estimates are. When inspecting closer the θ parameter, it turns out that also the hyperbolic polynomial tends to corner solution. However, having this polynomial at ones disposal gives another possible approach to the data.

The results for the weekly frequency are in Table 3.6. They are very similar to the results for the Almon polynomial for this frequency[5]. The estimator improves over the monthly frequency and behaves nice in finite sample.

As a next step, we explore alternative optimization algorithms. All estimations so far were conducted using a combination of the BHHH algorithm and the Newton-Raphson algorithm, implemented in GAUSS 12, using the Constrained Maximum Likelihood package. Normally these methods are sufficient for solving a Maximum Likelihood problem. However, there are also possibilities to solve more troublesome optimization problems. What follows are simulations using an alternative method. Corana et al. (1987) introduced Simulated Annealing into statistics. It is a gradient free optimization approach particularly good at escaping local minima.[6] What follows is a short description of how the algorithm works and simulation results. Simulated Annealing works in the following way:

1) Draw from a uniform distribution with endpoints -1 and 1, and scale it by a step size for all parameters to be estimated, to produce a step.

2) Accept or reject the steps by a probabilistic rule which is governed by a global variable called temperature (which is a positive number). The lower the temperature, the higher the probability that the optimal step gets accepted.

3) Do this N_s times.

4) Adjust the step size in such a way that about 50 % of the current moves would be accepted as moves next time.

5) Decrease temperature, check for exit conditions, otherwise go back to 1).

Simulate Annealing is closely related to the Metropolis-Hasting Algorithm due to the accept-reject rule in step 2), which is a special case of Metropolis-Hasting. Next, simulation results are presented. The setup is the same as before. That is, all simulations are done with 2000 replications, the Almon as well as the Beta polynomial are investigated with parameters $(0.035, -0.085)$ for generating a downward sloping weighting, $(-0.02, 0.005)$ for an upward sloping weighting, and $(0.005 - 0.0005)$ for a hump shaped weighting scheme. For the Beta polynomial $(0.6, 2.1)$ is used to get a downward sloping weighting scheme, $(1.1, 1)$ for generating an upward slope, and $(4, 2.1)$ for a hump shaped pattern.

As we can see, the results in Table 3.7 are very similar to the results obtained by standard algorithms earlier. The downward sloping weighting scheme behaves

[5] Note the slight change in θ to accommodate the change in frequency

[6] Also a Covariance-Evolutionary Algorithm was investigated as an alternative, however the results were similar to the standard algorithms

well, whereas the other two schemes pose a challenge to the estimator. However, the amount of corner solutions in the θ parameters are lower compared to the standard optimization methods.

In contrast to the earlier results on the Beta polynomial, in this Table we see a clear deterioration in the quality of the estimates. This happens because the rejection rate is now close to zero, compared to the case of standard optimization algorithms, where it was close to 10 %. This indeed confirms the assumption that the rejection rate is driving the preciseness of the estimates for the Beta polynomial.

In Tables 3.9 and 3.10, the results for weekly-frequency simulations can be found. Overall the results are remarkably similar to the ones using standard optimization tools. However, again fewer corner solutions are observed, which puts the Simulated Annealing method ahead. Next, the hyperbolic polynomial is investigated.

When looking at the results for the hyperbolic polynomial, we can see marginal improvement over the Quasi-Newton methods used earlier. However, it seems generally advisable to use Simulated Annealing, instead of standard optimization routines when estimating a probit-MIDAS regression, because the occurrence of corner solutions is reduced.

Table 3.7: Almon polynomial simulations monthly frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping	β_1^{HF}	28.5%	23.5%	16.0%	9.4%	6.1%	6.1%
	β_2	14.6%	6.7%	2.3%	0.6%	1.1%	0.3%
upward sloping	β_1^{HF}	-26.5%	-19.6%	-18.9%	-16.2%	-13.2%	-13.2%
	β_2	13.7 %	6.5%	1.9%	0.6%	0.9%	0.3%
hump-shaped	β_1^{HF}	-17.6%	-24.2%	-18.3%	-17.7%	-19.8%	-14.9%
	β_2	17.2%	5.9%	0.8%	1.5%	1.1%	0.2%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.8: Beta polynomial simulations monthly frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping	β_1^{HF}	2.3%	2.3%	-3.2%	-7.4%	-5.8%	-6.5%
	β_2	16.6%	7.2%	2.5%	0.1%	0.8%	0.8%
upward sloping	β_1^{HF}	-9.9%	-22.9%	-25.2%	-24.7%	-22.1%	-21.3%
	β_2	15.2 %	3.4%	2.1%	1.1%	0.7%	-0.1%
hump-shaped	β_1^{HF}	-38.4%	-36.5%	-35.1%	-28.1%	-27.7%	-13.4%
	β_2	15.2%	5.7%	2.2%	0.1%	0.7%	0.5%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.9: Almon polynomial simulations weekly frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping	β_1^{HF}	20.9%	7.3%	4.3%	0.6%	1.3%	0.8%
	β_2	14.5%	3.9%	2.5%	1.9%	0.7%	0.7%
upward sloping	β_1^{HF}	33.8%	17.4%	9.5%	5.0%	4.7%	3.8%
	β_2	10.1 %	5.1%	0.6%	1.4%	0.8%	0.8%
hump-shaped	β_1^{HF}	17.6%	11.8%	4.0%	1.7%	1.4%	1.9%
	β_2	16.5%	6.0%	2.0%	1.4%	0.6%	0.9%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.10: Beta polynomial simulations weekly frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
downward sloping	β_1^{HF}	-11.0%	-21.3%	-22.9%	-20.2%	-19.9%	-19.6%
	β_2	11.9%	4.7%	2.2%	-0.2%	0.5%	0.2%
upward sloping	β_1^{HF}	-14.7%	-26.7%	-25.3%	-23.6%	-23.4%	-22.6%
	β_2	14.3%	4.6%	2.4%	0.2%	0.4%	0.6%
hump-shaped	β_1^{HF}	-3.8%	-4.8%	-1.4%	0.2%	1.7%	0.3%
	β_2	14.2%	7.2%	1.4%	0.9%	1.0%	0.1%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.11: Hyperbolic Polynomial Simulations Monthly Frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
$\theta = 0.11$	β_1^{HF}	36.3 %	25.9 %	15.3 %	13.0 %	10.4 %	7.1 %
	β_2	9.3 %	4.5 %	1.6 %	1.4 %	1.6 %	0.0 %
$\theta = 0.251$	β_1^{HF}	24.1 %	13.6 %	6.7 %	8.5 %	3.8 %	1.2 %
	β_2	8.2 %	6.5 %	1.9 %	0.6 %	0.8 %	0.8 %
$\theta = 0.41$	β_1^{HF}	20.7 %	7.5 %	-0.1 %	-0.1 %	-2.5 %	-1.4 %
	β_2	8.0 %	4.5 %	2.6 %	-0.1 %	1.0 %	0.5 %

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

Table 3.12: Hyperbolic Polynomial simulations weekly frequency with Simulated Annealing

Weighting-scheme	Coefficients	50	100	250	500	750	1000
$\theta = 0.11$	β_1^{HF}	18.4%	12.0%	6.7%	3.9%	3.7%	2.8%
	β_2	7.6%	6.4%	1.4%	1.2%	-0.1%	-0.1%
$\theta = 0.3$	β_1^{HF}	16.5%	6.3%	1.9%	1.8%	0.0%	0.5%
	β_2	8.2%	4.6%	1.7%	0.7%	0.6%	0.4%
$\theta = 0.49$	β_1^{HF}	9.1.5%	2.9%	-1.3%	-0.7%	-2.5%	-1.0%
	β_2	12.4%	5.2%	1.5%	1.4%	0.1%	0.5%

Simulations are done with 2,000 replications. The Table shows deviations from the true coefficient in percentages. HF denotes the high frequency regressor.

3.4.2 Finite Behavior of the Bootstrap

In this section the finite behavior of the bootstrap approach is investigated. Two issues are addressed. First of all, we investigate the bootstrap itself. Second of all we look at the residuals in the bootstrap procedure. The reason for this is that residuals in a probit are not as straightforwardly defined as they are in standard Ordinary Least Squares (OLS) regression. However, due to the nonlinear nature of the probit regression, there are several possibilities to define the residuals. The most popular definition is due to Gourieroux et al. (1987), who define the generalized residuals as:

$$u_t = \frac{\phi(\beta x_t^{(m)}(\theta))}{\Phi(\beta x_t^{(m)}(\theta))(1-\Phi(\beta x_t^{(m)}(\theta)))}(y - \Phi(\beta x_t^{(m)}(\theta))). \tag{3.16}$$

Also, there are three other residual versions for a probit. These were introduced in Hinkley et al. (1991). These are the response residuals defined as:

$$u_t = y_t - \Phi(\beta x_t^{(m)}(\theta)). \tag{3.17}$$

The Pearson residuals are defined as:

$$u_t = \frac{(y_t - \Phi(\beta x_t^{(m)}(\theta)))}{\sqrt{(\Phi(x_t^{(m)}(\theta))(1-\Phi(x_t^{(m)}(\theta))}}, \tag{3.18}$$

and the deviance residuals:

$$u_t = \begin{cases} \sqrt{(-2log\Phi(x_t^{(m)}(\theta)))} & \text{if } \; y_t = 0 \\ \sqrt{(-2log(1-\Phi(x_t^{(m)}(\theta))))} & \text{if } \; y_t = 1 \; . \end{cases} \tag{3.19}$$

In the probit case the econometrician has to work under the assumption that the errors are distributed as a standard normal variable. All the point estimates are based on this assumption. Therefore, we would also want to draw from a residual distribution which is close to the assumed distribution. Since there is an abundant choice of residual specification for this regression, we have to take a closer look at the properties of the different residuals. We are mainly interested in the variance of the different residual specifications. For this, three sets of simulations are conducted, one set each for one of the three main weighting schemes. The downward sloping weighting scheme is conducted with an Almon polynomial, the upward sloping polynomial and hump shaped one are estimated using the beta polynomial. We simply recycle the values for the weighting parameters from the simulations in section 3.4.1. The frequency is monthly, and there are 1000 observations and 250 replications. The results for the standard deviations of the different residuals can be found in Table 3.13.

As it can be seen, there are clear differences between the various residual types. Since we prefers the residuals to have a variance of one, it should be clear that

Table 3.13: Residual Standard Deviation

	Pearson	Generalized	Response	Deviance
downward	1.001	0.659	0.366	0.133
upward	1.001	0.662	0.366	0.053
hump	1.001	0.662	0.366	0.059

Simulations are done with 250 replications and sample sizes 1,000, $m = 22$, and the Beta polynomial is used

the Pearson residuals are preferred over the other types.[7] However, it should be stressed again that these simulations in no way allow general conclusion about the optimal choice of the residual-specification for the bootstrap. Therefore, before actually conducting the bootstrap, the variance of the extracted residuals should be investigated.

As a next step the size and power of the bootstrap was investigated. For this investigation, the bootstrap was simulated 250 times with 1000 observations in each replication in a monthly frequency setting. The Almon polynomial was used. To make sure that the estimator is converging correctly, a downward sloping scheme was used to serve as parameters, thereby setting θ to $(0.035, -0.085)$. The first run analyzes the size of the test. Thus, the DGP is generated by white noise. In this setup it turns out that the test is moderately oversized, since for a 10 % confidence level, the null hypothesis is rejected 34 times (out of a total of 250). The second run investigates the power of the bootstrap. The data is determined by the following DGP:

$$y_t = \beta_1 x_t^{(m)}(\theta) + e_t, \tag{3.20}$$

with $\beta_1 = 0.5$. Again, 250 bootstraps were simulated. In this case the results are not looking as good as in the size test. It turns out that in 250 replications, we see a Type II error in 130 of the cases, given an $\alpha = 0.1$ the probability of a Type II error is around 0.52. This seems to be a rather low power and generally means that one should be suspicious when not being able to reject the null hypothesis. Indeed, when comparing the setup to a standard probit with the same slope coefficient, the probability of a Type II error is close to 0. Therefore, the test is definitely not performing well on the power side. For further robustness we also consider a parametric version of the bootstrap, where the errors are standard normally distributed. This does improve the power, but only marginally to 0.65.[8] The results improve significantly (only 11 % false positives) if we use a weekly frequency, instead of a monthly one. This does reflect the

A last remark in this section shall be devoted to the grid. So far no systematic investigation has been done to determine properties and optimal composition of the grid. Here, a brute force approach is used. Both the Almon polynomial as well

[7]Exploratory simulations show that the very same problem is present in standard probit regressions

[8]The same problem does not seem to persist in standard MIDAS estimations

as the hyper polynomial are bounded from above and below. Thus it is possible to simply create a monotonically increasing series between the two bounds, and in case that more than one weighting parameter is estimated, to compute the Cartesian product between them to construct a grid. For the beta polynomial, which is not bounded from above, we simply have to set a stop at the increasing series at some point. The experience so far with the construction of the grid reveals that a grid with a number of combinations in the low double digit (10-20) is too small, but a grid in the low triple digit (100-150) is giving acceptable results. A denser grid (400) does not yield superior results.

3.5 Estimations

3.5.1 Empirical Strategy

In this section we analyze whether ratings and CDS data agree on the default probability of a sovereign nation. This is done by first regressing lagged differenced CDS premiums on downgrades using the probit MIDAS estimator. The CDS data is lagged to avoid obvious endogeneity problems inherent in using data that gets published at the same time. However, there is a less obvious endogeneity problem at work here as well. Theoretically the analysts from a CRA could look at the evolution of CDS series and rate a country accordingly. This is prevented with the MIDAS estimator that extracts the default probability from noisy data that the analyst is unable to observe when only looking at the CDS data. Using lagged CDS series gives us the possibility to test the agreement on default probabilities using past data. However, it is also possible that CDS series do not expect a rating. To test for this, we will re-estimate each regression with lead instead of lagged CDS series. This gives an indication whether the implied default probability increases, decreases, or stays constant vis-a-vis the rating change. The first case indicates that the downgrade delivers new information to financial markets, the other two mean that there is a disagreement between CRAs and CDS data, about the default probability: CRAs believe it to be higher than financial markets do. This is done for all cases under investigation.

3.5.2 Data

Three European datasets are analyzed. First, we investigate 10-year CDS data. The availability of the series differs drastically amongst countries, therefore to maximize the amount of observations, two panels were created. One which is named the western European panel, containing Austria, Belgium, Cyprus, Finland, France, Germany, Greece, Ireland, Italy, Portugal, and Spain, which spans June 2009 until December 2012. A second panel named the eastern European panel has Bulgaria, Czech republic, Estonia, Hungary, Lithuania, Poland, Romania, Slovakia, and Slovenia in it, and covers April 2006 until September 2009. Thirdly, we have a dataset of 5-year maturity CDS, consisting of Austria, Belgium, Cyprus, Czech Republic, Denmark, Estonia, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Poland, Portugal, Slovenia, and Spain. The panel starts in June 2009 and ends in June 2013. All data has been taken from Datastream. It should be noted that the CDS premiums are not trading data produced by financial markets, but the premiums that buyers of CDS pay to acquire a newly issued CDS which are calculated as explained in section 3.2. The frequency of ratings throughout the time period for each panel is depicted in Figures 3.1, 3.2, and 3.3.

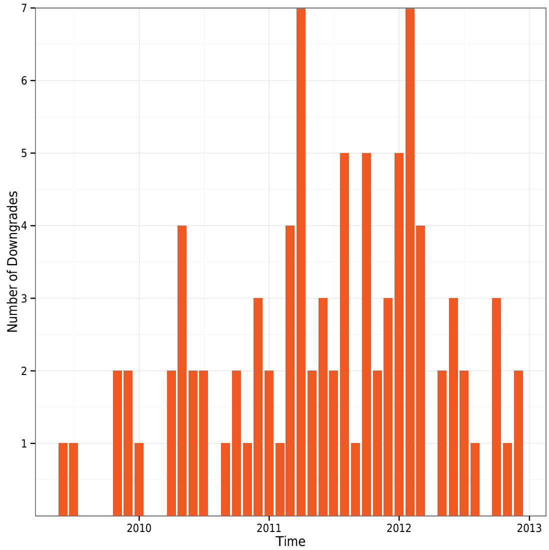

Figure 3.1: West Panel Downgrade Frequency

As we can see there is a substantial number of downgrades in all three panels. Unsurprisingly they are a bit clustered around time between 2011 and 2012 which was a point in time when several countries were asking for financial assistance from the EU, the ECB, or the IMF. The East panel has the lowest number of months with a downgrade, whereas the 5-year-maturity panel has the highest. This is unsurprising, since the former has the lowest number of countries amongst the three datasets, and the latter has the highest number of countries.

3.5.3 Empirics

We start with a probit-MIDAS regression with downgrades of the country as the dependent variable, sampled at monthly frequency, and the lagged differenced[9] sovereign CDS of the respective nation as an independent variable, such that we

[9]Since downgrades are essentially differenced ratings, the CDS data also needs to be differenced

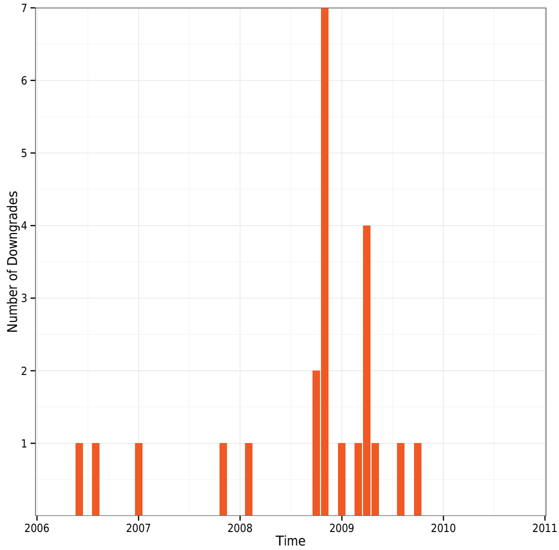

Figure 3.2: East Panel Downgrade Frequency

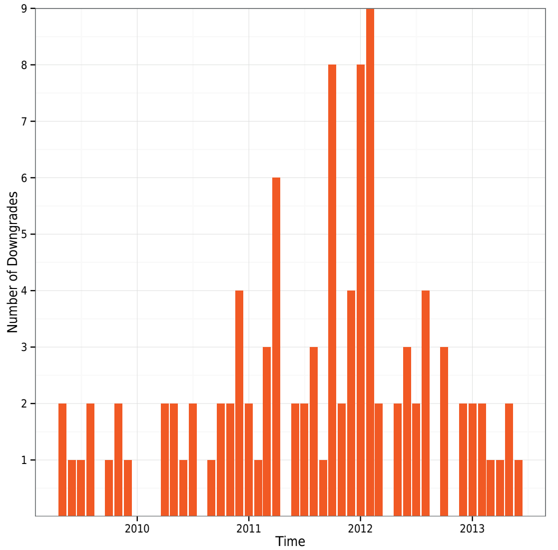

Figure 3.3: Five Year Maturity Panel Downgrade Frequency

have:

$$downgrade_{i,t} = \beta_0 + \beta_1 \Delta CDSpremium^{(m)}_{i,t-1} + u_{i,t}. \tag{3.21}$$

If β is significant this shows that CDS have the same opinion on default probabilities as CRAs.

A panel pooled approach is employed. This has the advantage that we have more information at hand to estimate the relationship and the disadvantage that it is assumed that the CDS markets for all countries are reacting in the same way to rating changes. The analysis starts with the monthly CDS data. Every rating movement that happened from the 20th of each month is instead assigned to the following month. The significance of the slope coefficient is tested with the bootstrap described in section 3.4.2. However, instead of sampling with replacement over all residuals, sampling takes only place over the cross section, following Hansen (1999). For the estimations where $m = 22$, the hyper polynomial is used since the Almon polynomial tends to select corner solutions in a lot of the estimations conducted. The optimization method is Simulated Annealing, due to its slightly better performance. The results can be found in Table 3.14.

Table 3.14: Monthly Estimations

Moody's	West Panel	East Panel	5-year maturity
intercept	-1.521	-2.274	-1.746
CDS	0.006**	-0.076**	0.049**
θ	0.474	0.497	0.003
S & P			
intercept	-1.523	-2.198	-1.745
CDS	0.007**	0.061	0.060**
θ	0.436	0.002	0.498
Fitch			
intercept	-1.586	-2.256	-1.729
CDS	0.007**	0.060	-0.024**
θ	0.461	0.001	0.420

In the first column the analysis for the western panel can be found. For all three CRAs we do see positive significant results as expected by theory. Therefore, CDS spreads are increasing prior to a downgrade. The second column contains the eastern European panel. Here, we see a different pattern. Only for one agency is the coefficient significant. In the other two regressions the coefficients are statistically not different from zero. This implies two possibilities. First, financial markets are actually surprised by the downgrade and are reacting afterwards to it, or second, financial markets do not agree with CRAs on the default probability of a country and therefore ignore their decisions. To test this, another set of probit-MIDAS regression is estimated using lead instead of lagged CDS data. If the first explanation is true we should see a positive significant coefficient, if the second one is true we should see a non-significant or negative coefficient. For the 5-year maturity data, we see that in two cases there is the expected positive significant case, and in one case the coefficient is negative. The regressions will also be re-estimated using lead CDS data. To get a feeling for the magnitude of the coefficient, we have calculated the marginal effect. For example, for Moody's, the CDS data predicts an increase for the probability of a Greek downgrade of roughly 56% for August 2011. Conversely, it predicts an increase of less than 0.01% for July 2009, which is the first observation for which we can do this calculation.

Next, the lead estimations are conducted. The results can be found in Table 3.15. Note that all regressions are re-estimated. The reason for this is that even in cases with a positive significant coefficient, there might still be some reaction after the downgrade, that is the downgrade is expected with a certain probability, but financial markets are not completely sure about the actions of the CRAs.

Table 3.15: Monthly Lead Estimation

Moody's	West Panel	East Panel	5-year maturity
intercept	-1.518	-2.242	-1.685
CDS	0.007**	-0.070**	-0.013
θ	0.020	0.484	0.035
S & P			
intercept	-1.504	-2.161	-1.698
CDS	0.008	-0.093***	-0.020*
θ	0.004	0.493	0.490
Fitch			
intercept	-1.569	-2.131	-1.710
CDS	0.002	-0.059	0.016
θ	0.399	0.484	0.023

We can see that in nearly all cases there are no reactions after the downgrade, indicated by non-significant point estimates. Only the estimations for Moody's with the west panel indicate a further premium increase for CDS. Therefore, it can be concluded that on a monthly level, the decisions of CRAs are expected by markets, for the cases where we see a positive significant coefficient for the lagged estimations. It should be additionally noted that the cases where we had significant negative coefficients in the lagged estimations, do not exhibit positive significant coefficients for the lead data case. Thus, there seems a difference in opinion on the default probability between CRAs and financial markets.

As a next step all regressions are re-estimated with rating changes sampled on a weekly frequency, therefore $m = 5$. The reason for doing this is that a month does not have exactly 22 trading days as assumed by the framework for the monthly estimations. Instead in this sample the average month has 21.75 trading days. This means that as t increases, the lagged independent variable is slowly moving towards the same t as the dependent variable, which might lead to an endogeneity problem. Fortunately in our data-set, every week has exactly five days which circumvents this problem. Additionally, we can now look at the impacts on CDS premiums closer to the downgrade, which is especially interesting in the lead-estimations. Therefore, all regressions are again estimated with weekly frequency samples for the ratings. The western European panel now spans from the 15th of May 2009 until December 21st 2012, the eastern Panel covers 17th of March 2006 to 24th of September 2010, and the five-year maturity CDS panel goes from June 2009 until June 14th 2013. The results are in Table 3.16.

Table 3.16: Weekly Estimations

Moody's	West Panel	East Panel	5-year Maturity
intercept	-2.116	3.095	-2.353
CDS	0.002**	-0.059**	-0.012
θ_1	-0.042	0.965	0.568
θ_2	-0.015	-0.154	-0.167
S & P			
intercept	-2.179	-2.726	-2.367
CDS	0.003**	0.030**	0.018*
θ_1	0.966	0.823	-0.31
θ_2	-0.197	-0.176	-0.050
Fitch			
intercept	-2.246	-2.785	-2.498
CDS	0.002	-0.021**	-0.005
θ_1	0.193	0.919	0.531
θ_2	-0.001	-0.573	0.008

The weekly estimations for the west panel paint a similar picture as the monthly estimations. In two cases we see positive significant coefficients, in one case it is insignificant. Thus, in one case a positive significant coefficient was changed to a non-significant coefficient. Therefore, the interpretation made for the monthly estimations stands.

For the east panel, we see that the coefficient for Fitch are switched from a negative to a positive one. Therefore, the decisions of this agency are not expected for the monthly horizon, however when we look at the weekly data, markets are expecting the downgrades simply within a shorter time frame. For the other two cases, there are significant negative coefficients which indicate that there is a disagreement between CRAs and CDS implied default probabilities.

The five-year maturity data shows in all three cases agreement with the monthly estimations. As a next step, all regressions are again re-estimated using lead CDS data. The results can be found in Table 3.17.

Table 3.17: Weekly Lead Estimations

Moody's	West Panel	East Panel	5-year Maturity
intercept	-2.130	-2.830	-2.351
CDS	0.000	-0.026**	-0.014**
θ_1	0.366	0.896	0.918
θ_2	0.049	-0.017	-0.732
S & P			
intercept	-2.153	-2.740	-2.350
CDS	0.000	-0.030**	-0.010*
θ_1	-0.285	0.947	-0.303
θ_2	-0.49	-0.441	-0.882
Fitch			
intercept	-2.251	-2.898	-2.528
CDS	-0.003*	0.058	-0.006**
θ_1	0.396	0.977	-0.555
θ_2	-0.042	-0.227	-0.854

For the west panel lead estimation, we see that they are more or less in line with the monthly estimations. In one case we do see a decrease of CDS premiums, where we saw no significant movement for the monthly estimations. In another case there was a positive significant coefficient in the monthly estimations, where the corresponding weekly coefficient is insignificant. For the east panel we see exactly the same coefficients as in the monthly lead regressions. The third column has three negative significant coefficients, whereas beforehand we saw that in only one case for the monthly estimations.

It should be noted that even with these differences in coefficients, generally the lead estimations for both frequencies have rather small coefficients even if they are significant. Also, they overall have the same tendency. Markets tend to calm down after a downgrade. Therefore, the lead estimations do not give rise to the theory that markets are surprised by downgrades and are reacting afterwards. Instead we have two cases. There is the evaluation of eastern European countries, where CRA decisions and implied default probability by CDS are clearly diverging. In the two other cases, ratings and CDS premiums are moving into the same direction and therefore financial markets agree with CRAs about default probabilities of sovereigns. A possible explanation for the divergence of the eastern panel is that the rating business is a reputation-based one. The three big rating agencies have a very long history of rating countries in the western world. However, rating former communist countries has only started in the mid-nineties. Therefore, financial markets could mistrust the agencies assessment, because the CRAs have not yet demonstrated reliably that they are able to rate these countries.

As a robustness test, we are doing the same regressions but pooling all the agencies into one dependent variable.[10] For the monthly series we see that in no case is the

[10] The results are available upon request.

coefficient significant. Thus, the markets seem to look at individual CRA decisions rather than combining them together into one information set. As a further test we look at the rating watch status of countries. Having a rating on watch means that it is due for a re-evaluation. Furthermore the CRAs mostly attach a tendency to this watch as either negative or positive. Therefore, it could be the case that CDS premiums do not react to downgrades, but to the announcement of a rating coming under scrutiny. Therefore all regressions are re-estimated using negative rating watch as a dependent variable. For the monthly case there are only three cases of significant coefficients. However, two of them are from the east panel which indicates that for these countries the rating-watch announcements are more interesting.[11] The weekly regressions have two significant regression coefficient for the 5-year maturity data, but otherwise everything else is statistically not significant. This further supports the hypothesis that rating watch status is not important for the CDS market.

The interesting question is, whether one even needs CRAs, when it seems that CDS premiums can yield similar information about default probability already before the rating is released. However, for this claim one should keep in mind that CDS premiums are rather volatile, and that the employed MIDAS scheme is reducing the noise in the data significantly before correlating this data with downgrades. Therefore, it is unlikely to extract these information by simply looking at CDS premiums, and the CRAs are thus still giving valuable information to the market.

3.6 Conclusion

In this chapter we investigate whether rating agencies and CDS data agree on the default probability of a country. To do so a probit-MIDAS estimator is developed to account for the characteristics of CDS time series data, which is available at rather high frequency on the one hand, and the event-type nature of rating changes on the other hand. While probit regressions are a long time workhorse of modern econometrics, MIDAS is a recent type of estimator introduced by Ghysels et al. (2004). The idea behind a MIDAS estimator is to weigh the data using a parsimonious weighting scheme in which the parameters of the weighting functions are minimized jointly with the regression coefficient. The estimator is investigated using Monte Carlo simulations. Also, multiple weighting functions are explored. It turns out that the probit-MIDAS estimator performs differently for different weighting schemes and sampling speeds, but if the optimal weighting scheme is employed for a given sampling frequency, it behaves well in finite sample. Also, we want to test the significance of the slope-coefficient. This chapter is one of the first to implement and investigate a bootstrap approach proposed by Ghysels et al. (2007). The test has reasonable size but is lacking power.

[11] Fitch was excluded due to no published negative rating watch, the results are available upon request.

When investigating the relationship between sovereign CDS and sovereign downgrades, this chapter uses various European countries with data stretching from 2006 until 2013. 10-year-maturity contracts as well as 5-year-maturity contracts are investigated. It turns out that for western European countries, CDS premiums and ratings do exhibit a positive significant relationship in nearly all of the investigated cases. However, for eastern European countries this is not the case. A possible explanation might be that rating agencies are a reputation based business, and these countries are simply not long enough rated, such that investors trust the judgment of the agency. We also conduct multiple robustness checks, such as changing sampling frequency or exploring credit rating announcements to confirm this. Additionally, it should be noted that CDS premiums are expecting downgrades in advance in a substantial amount of cases. However, this does not mean that raw CDS premiums can give a direct indication of the default probability of the underlying asset, because the MIDAS estimator removes most of the noise from the data and this allows the extraction of default indication from CDS data. Therefore, rating agencies are still providing a valuable service to financial markets.

Chapter 4

Growth Consequences of Austerity Programs in Europe

4.1 Introduction

In this chapter we want to quantify the damage that fiscal consolidation programs did in Greece, Portugal and Spain. To do this, we use a synthetic control method developed by Abadie and Gardeazabal (2003). We find that while the austerity treatment has done extensive damage in Greece and Portugal, it is difficult to establish such an impact on the Spanish economy.

How and whether government actions can influence GDP has been an important topic ever since the Great Depression. Until the 1960s, the Keynesian view of active government intervention was dominant. From then on, neoclassical economics was favoring Ricardian Equivalence as an argument that active fiscal policy cannot stimulate the economy. Recent research paints a more diversified picture. Giavazzi and Pagano (1990) investigated Ireland and Denmark during the 1980s and claim that in some cases governments can influence GDP, and in other cases it cannot. The outcome of this question depends on the macroeconomic environment. Blanchard and Perotti (2002) investigate the US economy in a VAR and use institutional information to identify fiscal shocks. They conclude that on the one hand government can influence GDP, on the other hand the multipliers are in general slightly less than one. An event study approach is used in Romer and Romer (2010), who identify exogenous fiscal policies via speeches and congressional reports, and plot economic output following these speeches. The authors come to the conclusion that tax increases are significantly contractionary. Barro and Redlick (2011) use military buildups in the US to identify exogenous fiscal shocks and concludes that multipliers of these events are significantly below one.

We would like to thank Leonard Wolk, William DuPont, and Bertrand Candelon for feedback and discussion while writing this chapter.

Alesina and Ardagna (2010) find that fiscal consolidation via tax increases are more likely to induce a recession compared to the slashing of spending. The IMF report by Leigh et al. (2010) concludes that a fiscal consolidation is more damaging when several countries are participating in it or when it is happening close to the zero-lower bound of interest rates.

The austerity programs in Europe were a direct response to the sovereign debt crisis that escalated at the beginning of 2010. After the collapse of Lehman Brothers in late 2008, the world economy was plunged into a recession of a magnitude not seen since 1929. What followed were fiscal stimulus programs in nearly every country, as well as rescuing of systemically important banks, especially in Europe. Spain and Ireland came under exceptional pressure due to a property bubble collapsing at the same time. In Europe debt level significantly increased throughout the different member states of the EU. This lead investors to doubt that the weaker members of the union would be able to pay back their debts, leading to an increase in spreads between European bonds, after a long convergence of yields. In response to this increase in yields, the European Central Bank (ECB), the Eurozone countries, the International Monetary Fund (IMF), and the European Union (EU) provided emergency loans that were conditional on austerity programs and economic reforms. So far Greece, Portugal, Ireland, Latvia, and Spain have requested these loans with varying loan providers and different conditions attached to it. The exact unfolding of events are discussed in section 4.3.

To quantify the impact of austerity programs in Europe we optimally would have two identical countries available of which one is subject to an austerity program. Then, in order to quantify the impact of the programs introduced, we would need to compare the results from a country undergoing the treatment to the same country not being treated. Now unfortunately (or fortunately), economists cannot conduct randomized experiments on a whole economy. Instead, we will use an approach called synthetic control method or synthetic counterfactual developed by Abadie and Gardeazabal (2003) to construct an untreated version of the countries under investigation. This artificial country can serve as comparison in the assessment of the impact of the program. This is done by assembling a panel of untreated countries which are called the donor pool and find the combination of countries from that pool that best reflects the time series of the outcome variable (in this case GDP per capita). Also the estimator optimizes over a set of covariates that explain the outcome variable. This is done up until the point when the intervention occurs.

Consider the following example for applying the synthetic counterfactual method. Suppose Greece is subject to an austerity program starting in 2010. Therefore, we want to construct a synthetic counterfactual that allows us to asses the impact of this program on Greek GDP. As an outcome variable, we chose GDP per capita, since it gives us a standardized unit of output. We also need to select covariates which are possible explanatory variables to our outcome variable. This needs to be done in order to prevent the estimator from overfitting only on the outcome variable. As a last step we need to select a donor pool of countries from which we construct the synthetic Greece. The main restriction when selecting the pool is that their GDP needs to be uncorrelated with the shock induced by the austerity program. Thus, in this case excluding the EU countries from the OECD countries will yield a donor pool. The synthetic Greece is then constructed by assigning weights between zero and one (which overall add up to one) to the countries in the donor pool. These weights are determined by minimizing the distance between the actual Greek GDP per capita and the weighted combination of GDP per capita from the donor pool, as well as minimizing the distance between (the averages of) the covariates of Greece and the weighted combinations. This estimation is done using observations up until the austerity program actually is implemented. The comparison between the outcome variable of Greece and its synthetic counterfactual after the intervention gives us a measure of the impact of the program for Greece.

When investigating such a question we need to be careful about how factors other than the treatment have affected GDP at the time the treatment is administered. That is, we need to devote some time to an exogeneity discussion. As we can see when looking at earlier work on fiscal consolidation and economic stimulation via government purchases, it is important to find events that are exogenous to the economic situation. The question here is slightly different. What is relevant is, whether the imposition of austerity programs is actually exogenous to output or not. We argue that this is the case. First, the official mission of the troika (the three institutions that provide loans and guarantees to the treated countries) was to make sure that the government will be able to pay back these loans, by modernizing administration, privatizing state owned businesses and increasing flexibility on the labor market. This program would have been imposed regardless of the state of the economy. Additionally, the length of the downturn speaks against a normal recession. The last time such a long contraction of output occurred in developed countries was during the Great Depression. In that case the recession was aided by contracting monetary and fiscal policy. Also, the general narrative is that the aftermath of the 2008 financial crisis was a wake up call to investors. This indicates that the actual debt crisis was not triggered by a recession, but rather by investors being afraid that the respective countries could not pay back its loans. In line with this is the fact that European bond yields were converging until the crisis with basically no risk premium left. Lastly, if we look at the evolution of the individual countries, the experience of the 2009 recession are not markedly different to other countries that did not need to request financial assistance. Greece experienced a recession comparable to its European neighbors in 2009, yet asked only 113 days later for a bailout. Portugal actually experienced economic growth in 2010, yet in the second quarter of 2011 it requested financial aid. The Spanish case is a bit less clear cut. After flat growth in 2010 and a slight contraction in 2011, the bailout was requested in late 2012. However, during the analysis, we will see that Spain is indeed distinct from the other two cases. In general we argue that the request for financial support was not coinciding with (another) deep recession. This means the treatment imposition is exogenous to the outcome variable. As we can already see, it is more difficult to replicate such a discussion for voluntary imposition of austerity as for example in the case of the Netherlands in recent years. Therefore, no comparison with this group of countries can be made.

The rest of the chapter is structured as follows: Section two provides a more in-depth overview of the synthetic control method, section three presents the data, and section four the empirical results. Section five gives discussion and implications, and section six concludes.

4.2 Methodology

In this section we introduce the synthetic control method. We start by giving a quick overview of the relevant literature. The method was pioneered by Abadie and Gardeazabal (2003) who investigate the impact of the terror of Euskadi Ta Askatasuna (ETA) in the Basques province in Spain and finds significant impact compared to the counterfactual that was not under attack by terrorists. Furthermore, Abadie et al. (2010) evaluate California's smoking reduction program. They find a significant decrease in cigarette sales per capita. When investigating economic liberalization, Billmeier and Nannicini (2013) find in a majority of cases a positive impact on economic growth. Moretti et al. (2013) analyzes whether accession to the EU stimulates economic growth. For most of the countries they can confirm this, except for Greece which is the only country with a lower output compared to its synthetic control. Abadie et al. (2014) look at the impact of the German reunification and come to the conclusion that there are significant losses in terms of GDP per capita. Another popular application of the synthetic control method is the evaluation of the impact of natural catastrophe. See for example DuPont and Noy (2012), who investigate the impact of the Kobe earthquake in 1995.

In the rest of this section we will give a more detailed introduction to the estimator. Let us start by considering

$$Y_{i,t}^{u} = \alpha + \beta_j X_{i,t,j} + e_{i,t}, \tag{4.1}$$

where the error term is zero mean, and

$$Y_{1,t}^{t} = Y_{i,t}^{u} + \theta_1 D_{1,t}. \tag{4.2}$$

The equations describes a set of outcomes of a variable ($Y_{i,t}^{u}$), which is determined by an intercept (α), a set of J covariates ($\beta_j X_{i,t,j}$), and an error term. Furthermore, there is a country 1 in the dataset ($Y_{1,t}^{t}$) that is receiving a treatment (for example an austerity program) from some period k onwards which is indicated by $\theta_1 D_{i,t}$, where $D_{1,t}$ is a vector that is one if $t > k$, and zero otherwise. In total we have N countries in our dataset. This setup naturally gives rise to use the difference-in-difference (DD) estimator. However, this estimator requires a certain set of assumptions. First of all, it requires to have countries that are treated and countries that are not treated available. Second of all, it is required that all countries in the group follow a common trend, as can easily be seen from equations 4.1 and 4.2. In macroeconomic research this is often difficult to find. The methodology used in this chapter does not require this assumption.

Let us start by assuming that we have one country (for simplicity assume again it is the first one) which is subject to an austerity program, just as described in equation (4.2). Also, we observe a group of other countries that might or might not be following the common trend assumption, which implies that their covariate matrix $X_{i,t,j}$ is subject to variation. Assuming that the latter group, which is called the donor pool, is big enough, we can construct a synthetic control for individual 1 before the treatment as:

$$Y_{1,t} = \sum_{i=2}^{N} w_i Y_{i,t}, \quad \forall t < k, \tag{4.3}$$

and

$$\sum_{i=2}^{N} w_i = 1, \quad w_i \geq 0. \tag{4.4}$$

Similarly, we can construct the set of covariates as:

$$X_{1,t,j} = \sum_{i=2}^{N} w_i X_{i,t,j}, \quad \forall t < k. \tag{4.5}$$

However, since a priori, we cannot be sure which covariates are important to the outcome variable, the estimator needs a method to select the appropriate variables. This is done by constructing a matrix:

$$Z_1 = (Y_{1,t}, X_{1,t,j}) \quad \forall\, j, \forall\, t < k, \tag{4.6}$$

in which we collect all variables from the treated unit prior to the intervention. We can do the same for the units that are not subject to a treatment

$$Z_0 = (Z_2, Z_3 ... Z_N), \tag{4.7}$$

which are identically constructed as Z_1. Finally, let us collect the individual weights in a matrix

$$W = (w_2, w_3 ... w_N). \tag{4.8}$$

Then, we can use:

$$min(Z_1 - Z_0 W)' V (Z_1 - Z_0 W) \tag{4.9}$$

subject to

$$\sum_{i=2}^{N} w_i = 1, \;\; w_i > 0, \;\; Y_{1,t} = \sum_{i=2}^{N} w_i Y_{i,t}, \;\; \forall t < k. \tag{4.10}$$

The V matrix is a symmetric positive semi-definite matrix that gives the relative importance of the different predictor variables. Therefore, V and W are jointly determined to solve equation (4.9). Note that the third constraint does not require that all data points in the outcome variable are perfectly matched, and instead requires a zero-mean error for the individual observations.

The synthetic counterfactual is constructed using data only up until the treatment effect takes place. Therefore, we can evaluate the effect of a treatment on a country by comparing the evolution of GDP per capita for the actual country and its synthetic counterfactual after the treatment takes place. We require two assumptions to be met. First of all, the treatment has no effect prior to its implementation. Second of all, the donor pool needs to be independent of the treatment effect. The latter does mean that when looking at the austerity programs in Europe, the donor pool should not contain any EU countries.

The size of the donor pool merits some discussion. The estimator is converging with t, since the more time periods we have prior to the treatment, the more precise we can construct a counterfactual. N, the number of countries in the donor pool should not be viewed as observations in the traditional sense, that is the estimator does not converge as n goes to infinity. Rather it should be interpreted as a parameter for model complexity, such as adding additional independent variables to a regression model. That is, the more countries we add to the donor pool, the more options the estimator has to fit the data before the treatment. This inevitably leads to overfitting. Therefore, the donor pool should not be made as large as possible, but should rather be populated by countries that are similar to the treated country.

Conducting inference in such an estimator is not straightforward. We do not have standard errors at our disposal as it would be the case with regression analysis. Instead we will resort to a permutation test that was proposed by Abadie et al. (2010). The basic idea behind this test is to assign the treatment to each member of the original donor pool one by one, while the actually treated unit is put into the donor pool and not assigned the treatment. For example, we want to investigate the significance of the austerity program for Greece. We start by estimating a synthetic counterfactual for Greece and calculate the difference between the Greek GDP per capita series and the synthetic counterfactual. Next, we put Greece into the donor pool and instead select a member of the original donor pool. Then we apply exactly the same procedure to this country as was done with Greece. This is done for all countries in the donor pool. Afterwards, we plot the results for all countries. We expect that if there is a treatment effect for Greece that its output path, after the imposition of the austerity treatment, is the most negative. If this is not the case we can conclude that the impact seen in the data after the treatment is only due to randomness, and thus is not connected to austerity measures. This test has several disadvantages. First of all, we will not be able to replicate all members of the donor pool. Mathematically speaking these countries are not in the convex hull of the donor pool. For example, imagine we want to replicate the GDP per capita of the USA with this method, which is the highest in the panel. Therefore, there exist no solution for the estimator, due to the fact that the weights are constrained between zero and one. Second of all, the usefulness of the test is restricted by the size of the donor pool. A rather large donor pool can give a good indication about the significance of the results obtained. However, the smaller the donor pool, the less sure can we be about the test.

To partly counter the shortcomings of the permutation test, we now introduce a second inferential test. Instead of testing along the cross section, we will now test in the time dimension. This is called a placebo-in-time test and was introduced by Abadie et al. (2014). However, in their paper, the authors only tested one time period against the actual treatment. We will is formalize this test such that it gives more explanatory power. Therefore, we will test over all possible time periods, and compare the evolution of the outcome over the same amount of time periods as the original treatment to said treatment effect. The placebo-in-time will be administered to all time periods except for the following:

1) The first 10% of the data

2) $2 * (t - k)$ time periods cut off at the end of the panel

The first restriction is needed to have a working estimator for the first few time periods to which the placebo is applied. The second one guarantees that there is no contamination, that is the GDP per capita evolution after the last placebo-in-time administration is not overlapping with the actual treatment. Since GDP per capita generally changes over time, the results will be detrended for economic growth before comparing them to the actual outcome. For the comparison we use the root mean squared error (RMSE).

4.3 Data

In this section we discuss the selection of variables and countries. For the synthetic counterfactual method to work, we need two types of countries, a treated country, and a group of countries that form the donor pool from which the synthetic control is constructed. It is advisable that a variable is coming exclusively from one source, since the estimator matches the data from the donor pool to the treated country, and thus the measurement and the handling (de-seasonalize, smoothing), should be exactly the same. This cannot be guaranteed when using data from different sources. This of course puts some constraints on the selection of suitable countries and variables.

For the treated countries, we choose the most prominent examples in Europe that were subject to an austerity program, which are Greece, Portugal and Spain. Greece received a rescue package from the Eurozone countries, the ECB, and the IMF in May 2010. Portugal received aid by the IMF, the European Financial Stability Facility (ESFS), and the European Financial Stabilization Mechanism in April 2011. Finally, Spain requested financial assistance from the Eurozone countries at the end of June 2012, to conduct a major restructuring of its domestic banking sector. There are some more countries that were subject to an intervention by the IMF and EU institutions, such as Ireland, Latvia or Cyprus. The first two were excluded because of the rather unique characteristics of the two economies, especially the rapid growth at the end of the nineties and beginning of the century. The estimator was unable to replicate these features with the donor pool. Cyprus was excluded because the data source for the output series did not contain Cyprus.

As mentioned, the donor pool has the requirement that it is uncorrelated with the treatment effect. This excludes every Euro-zone country from the pool, and to be on the safe side we also exclude European countries. Thus, we assemble a dataset of developed non-European countries, as well as some developing countries. All data was taken from the OECD. The donor pool consists of: Australia, Canada, Chile, Israel, Japan, Korea, Mexico, New Zealand, Turkey, USA, Argentina, Brazil, India, Indonesia, Russia, and South Africa.

For the actual data selection we follow a mixed frequency approach. That means, the (dependent) GDP variable is available on a quarterly basis, while the predictor variables are available on a yearly one. This poses no problem because the estimator is trying to match the outcome variable, while simultaneously matching the mean of the predictor variables between the country and its artificially constructed counterpart. Therefore, it does not matter whether the predictors are available on a quarterly or yearly basis. Additionally, using the dependent variable on a quarterly basis allows us to make rather accurate observations about the evolution of the post-treatment period. This is especially important since the treatment has only been administered three to five years ago. Also, it allows for a better determination of the start of the treatment, compared to using yearly frequency. As an outcome variable we use real GDP per capita as an annualized version, that is the output is calculated as if it were yearly data. The real GDP data, measured in constant 2005 US Dollars and adjusted for PPP, was taken from the OECD database. The data stretches from the first quarter of 1997 to the second quarter of 2014. We use population data from the Worldbank to construct GDP per capita from it. The population series is only available at a yearly frequency, thus there is some fuzziness in the second to fourth quarters of the year, however, since it is present in all series this should not be a problem. Also, there are not yet population data available for 2014, and consequently we use the average of the last five years of population growth, to construct an estimate for 2014.

The selection of predictor variables for this estimator should be happening on the basis of existing literature, since there is no straightforward mechanism to select variables in this set up. We follow Barro (1996) in the selection of variables, and use the share of manufacturing to value added in the economy, the Balance of Payment as percentage of GDP, gross capital formation as percentage of GDP, and percentage of tertiary educated people in the labor force as predictors. The share of manufacturing is representing the structure of the economies, since a service based economy might behave completely different compared to an industrial-based economy in the face of a shock. The Balance of Payment represents the openness of the economy, while the gross capital formation variable measures investment activity. Finally, the percentage of tertiary educated persons gives an indication of the education level in the country. This also gives a good indicator of the structure of the economy. Moretti et al. (2013) use a similar set of predictor variables in their synthetic counterfactual study which analyzes at the impact of EU membership on output. Also, Abadie and Gardeazabal (2003) uses comparable co-variates in their study of the impact of ETA terrorism on economic growth.

4.4 Results

In this section we present the results of the synthetic counterfactual for Greece, Portugal, and Spain. All three were subject to fiscal consolidation programs in the recent years. We assign the treatment of the different countries as the quarter in which the respective government officially requested financial aid. For Greece this is the second quarter of 2010, for Portugal it is the second quarter of 2011, and for Spain the treatment took place in 2012, third quarter.[1] We start with the analysis of Greece. The black vertical line in Figure 4.1 indicates the point in

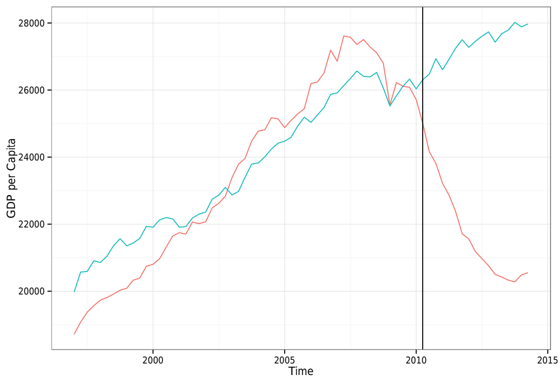

Figure 4.1: GDP per capita evolution of Greece and its synthetic counterfactual from 1997Q1 to 2014Q2.

time when the government asked for financial support. The red series is the Greek GDP per capita variable. The blue line is the artificial GDP per capita figure that is constructed from a combination of multiple countries that make up the donor pool. The effects of the imposed austerity program can be clearly seen in the graph. We see a very close tracking of the economy prior to the intervention by the estimator. From the treatment onwards, we see a sharp drop in the GDP per capita for Greece. There is already a small contraction one quarter prior to the treatment that cannot be mimicked by the synthetic counterfactual. This shows that the Greek economy had some problems coming out of the recession, however, it is so small that its impact should be negligible. The effect of the program have led to a huge drop in output for the Greece economy. Also, once the GDP per

[1]Technically the intervention took already place in June 2012, but since the decision was only made on the 21st we decided to move the treatment to the third quarter

capita stopped contracting, there is basically no growth in the Greece, while the synthetic Greek economy is growing throughout the crisis. If we want to quantify this, we need to calculate the difference between the synthetic Greece and the actual development of the country, that is the area between the two series after the treatment. This adds up to around 22,000 US Dollars that on average every Greek citizen has lost since the second quarter of 2010 in unrealized output. Multiplying this with the numbers for the Greek population, this results in unrealized GDP of 244 billion US Dollars since the introduction of the programs by the troika.

Table 4.1: Weights for Constructing Synthetic Greece

Country	Weights
Australia	0.696
Canada	0.000
Chile	0.000
Israel	0.000
Japan	0.000
Korea	0.000
Mexico	0.000
New Zealand	0.000
Turkey	0.000
United States	0.000
Argentina	0.000
Brazil	0.000
India	0.304
Indonesia	0.000
Russia	0.000
South Africa	0.000

Table 4.2: Comparison Treated and Actual Predictors Greece

	Treated	Synthetic
Manufacturing	9.558	12.588
Trade	-10.027	-1.679
Investment	23.911	27.681
Education	22.071	23.799

Table 4.3: Predictor Weights Greece

Variable	Weight
Manufacturing	0.338
Trade	0.129
Investment	0.000
Education	0.533

As we can see the synthetic counterfactual method does quite a good job in replicating other features of the Greek economy as well. The counterfactual is composed of 70% Australia, and 30 % India. With regards to the covariates, there are some interesting observations. Trade is notably different between Greece and its counterfactual. Also, the estimator gives zero weights to the investment variable, as seen in table 4.3. However, given the close tracking of the GDP variable for the original Greece in the pre-treatment period this seems to be no issue.

The second country that we look at is Portugal. The evolution of Portugal's GDP per capita and its synthetic counterfactual is depicted in Figure 4.2

The vertical line depicts the request for financial aid, the red line is the synthetic Portuguese economy, while the blue line is the GDP per capita in Portugal. Compared to the Greek counterfactual, we can see that Portugal is more difficult to match by the estimator. The output of the actual Portugal is since 2004 always slightly below its synthetic counterpart. However, it should be noted that both series are following the same general pattern remarkably well. Also, the financial

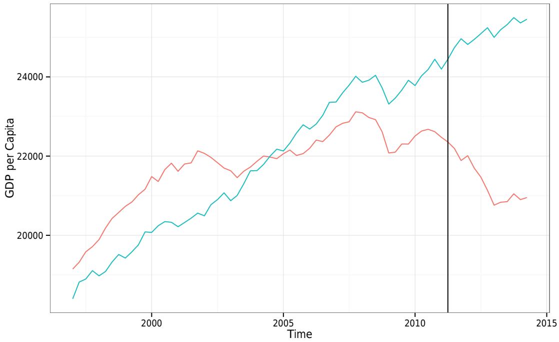

Figure 4.2: GDP per capita evolution of Portugal and its synthetic counterfactual from 1997Q1 to 2014Q2.

crisis in 2008 can be very well observed in both series. Two to three quarters before the treatment, we can see the beginning of a divergence between the Portuguese output and the synthetic counterfactual. This is similar as in the Greek case. Again, the effect seems too small to influence the post-treatment period significantly. After the treatment, the difference between the two series becomes much more pronounced, and it is easy to see that the austerity program pushed down income in Portugal significantly. In the last few quarters one sees a stabilizing effect of GDP. The damage done to the Portuguese economy can be calculated in a similar fashion as for the Greek case. The average Portuguese citizen has lost close to 8,600 US Dollars of unrealized GDP since the second quarter of 2011 until the second quarter of 2014, and the cumulative foregone income is around 91 billion US Dollar.

The synthetic counterfactual is mainly constructed from US and Indian data, with each being weighted around a half. Trade and eduction are different in the counterfactual, compared to the original Portugal, while the other variables are fairly well replicated. The estimator needs to trade off tracking the GDP variable closely, for decreased matching of the predictor variables. Overall, we can very well see the impact that austerity programs had in these two countries, with a very clear picture emerging from the Greek case, and a slightly more fuzzy picture from Portugal. Nevertheless, these two examples show the extent of the damage done by such programs.

The last country that is analyzed is Spain. The results can be found in Figure 4.3. Here the picture is much less clear cut. This is because Spain suffered especially harsh consequences from the 2009 recession. Therefore, the Spanish economy and its synthetic counterfactual also start to diverge in 2009. This is probably the effect of the Spanish property bubble bursting, which had two main effects. First of all, Spanish banks were severely hit. Second of all, the Spanish construction sector which had been a major driver of economic growth since around 2000 suffered. The problem is that while several countries had a major house price bubble prior to the

Table 4.4: Weights for Constructing a Synthetic Portugal

Country	Weights
Australia	0.000
Canada	0.000
Chile	0.000
Israel	0.000
Japan	0.000
Korea	0.000
Mexico	0.000
New Zealand	0.000
Turkey	0.000
United States	0.480
Argentina	0.000
Brazil	0.000
India	0.520
Indonesia	0.000
Russia	0.000
South Africa	0.000

Table 4.5: Comparison Treated and Actual Predictors for Portugal

	Treated	Synthetic
Manufacturing	15.377	14.664
Trade	-8.379	-3.241
Investment	24.570	26.474
Education	12.660	20.622

Table 4.6: Predictor Weights Portugal

Variable	Weight
Manufacturing	0.006
Trade	0.730
Investment	0.038
Education	0.226

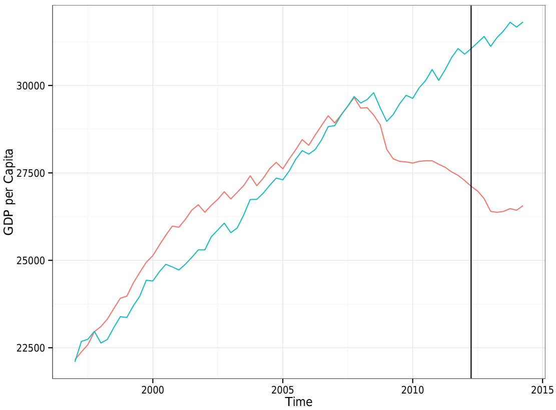

Figure 4.3: GDP per capita evolution of Spain and its synthetic counterfactual from 1997Q1 to 2014Q2.

financial crisis, the only country that had a similarly important construction sector that collapsed was Ireland which cannot be part of the donor pool due to being not only European, but also requesting aid from the EFSF as well. Therefore, the method has difficulties replicating the GDP evolution of Spain. If we look at the time after the treatment, we see a sustained drop for several quarters and afterwards virtually no increase in output per capita. However, this drop could also be a continuation from the damage done to the construction sector in the financial crisis, and thus we should not calculate the loss in potential GDP for Spain as we did it for the other two countries. Instead we can argue that the damage of the austerity programs is actually rather limited in Spain. The economy is now starting to tackle the transformation from construction to a more diversified economy. This of course leads to turbulences in the economy, and is possibly a major driver of the current state the Spanish economy is in.

Table 4.7: Weights for Constructing a Synthetic Spain

Country	Weights
Australia	0.000
Canada	0.002
Chile	0.001
Israel	0.003
Japan	0.002
Korea	0.168
Mexico	0.001
New Zealand	0.001
Turkey	0.000
United States	0.511
Argentina	0.001
Brazil	0.000
India	0.308
Indonesia	0.000
Russia	0.001
South Africa	0.000

Table 4.8: Comparison Treated and Actual Predictors for Spain

	Treated	Synthetic
Manufacturing	16.756	16.754
Trade	-2.428	-2.353
Investment	26.299	26.134
Education	29.694	25.112

Table 4.9: Predictor Weights Spain

Variable	Weight
Manufacturing	0.765
Trade	0.154
Investment	0.073
Education	0.008

Looking at the composition of the synthetic Spain, we can see that it is more diversified than the counterfactuals for Greece or Portugal. South Korea, India, and the United States make up the vast majority of the artificial Spanish economy. Education is not well matched by the estimator (but given low weights in table 4.9), but all other variables are rather nicely mirrored by the synthetic counterfactual.

As a next step we want to know whether the effects that we see are actually a significant treatment effect, or whether these might be due to random variation. To do so, we use the permutation test by Abadie et al. (2010) that was described in the methodology section. The basic idea of this test is to assign the treatment to each unit in the donor pool, while putting the treated unit into the donor pool. This allows to distinguish whether the postulated treatment effect we see is actually induced by the treatment, or whether it is due to random variation in the data.

We will only present the test for Greece and Portugal.

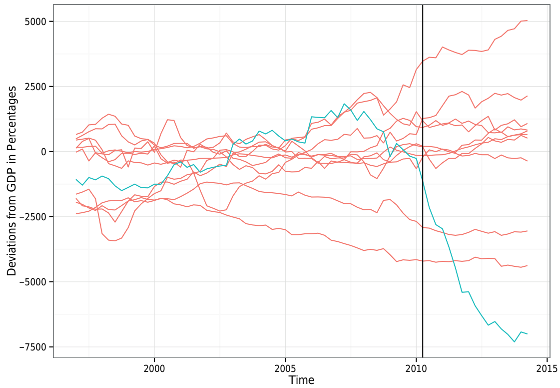

Figure 4.4: Permutationtest for Greece

In figure 4.4, the blue path is the difference between the actual output and the synthetic counterfactual for Greece, while the red paths are the same difference for each donor pool country having the treatment assigned.[2] All series are centered around zero with small deviation until the treatment takes effect, as expected. After the austerity program takes place, it is clear that there is a significant impact on the Greece economy, depicted as the blue line. This impact is stronger than for any other country in the donor pool. Thus, there is a high probability that the impact on the Greek economy after the treatment is not caused by randomness. Next, we discuss the test for Portugal. The results for this test can be found in figure 4.5.

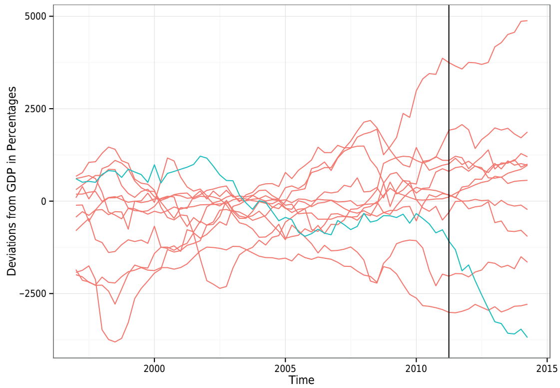

Figure 4.5: Permutationtest for Portugal

For Portugal, we observe a similar pattern as for Greece. Prior to the intervention

[2]A substantial number of donor pool countries needed to be removed because they were not in the convex hull of the donor pool, therefore, the estimator was unable to match the GDP per capita properly with a synthetic counterfactual

the difference to the synthetic control fluctuates around zero. Afterwards we can see a significant drop in the series, which is more pronounced compared to the same procedure involving the donor pool countries. As in the previous case, we needed to drop a significant amount of countries. As a next step we will repeat the test with a slightly different dependent variable. Instead of using deviations in GDP per capita in absolute terms, we will use deviations in percentages. This gives a better method of comparison and allows us additionally to include some countries we needed to drop from the test, since some deviations might be large in absolute terms, but small in percentages. The results for Greece can be found in Figure 4.6.

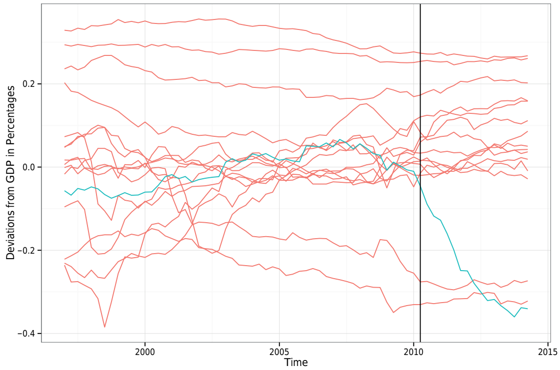

Figure 4.6: Permutationtest in Percentages for Greece

Basically, the test confirms what we already saw with the test in absolute deviations. The impact is clearly significant and assigning the treatment to any other country does not produce the same deep recession. Next, we take a look at Portugal. The results are in Figure 4.7.

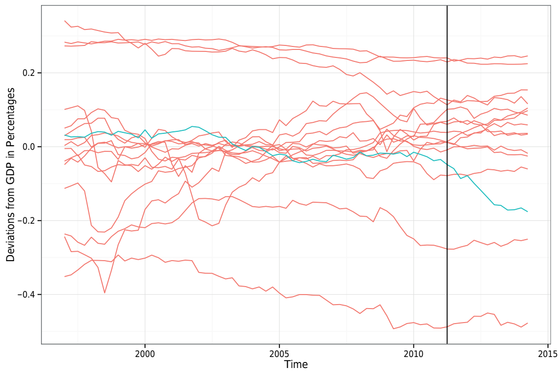

Figure 4.7: Permutationtest in Percentages for Portugal

Again, the results are similar to the previous test for Portugal. In this case it is not as clear cut as for Greece, as there are two countries that experience a deeper recession, but Portugal still belongs to the top three countries by decrease in GDP relative to the counterfactual. However, we need to keep in mind that even with the test based on percentage deviations, we are still constrained by the number of cross sectional units, that is fifteen[3]. Therefore, the explanatory power of this test is also limited. To complement this, we will now conduct a placebo-in-time test as already discussed in the methodology section. Due to the fact that the number of time series observation is relatively large, we hope to gain more insight into the significance of the treatment effect. The intuition behind this test is that the treatment effect is assigned to every time period prior to the treatment and observe whether the resulting difference between the synthetic counterfactual and the GDP per capita series is bigger than the difference after the actual treatment.[4] The results are detrended for economic growth.

Table 4.10: Placebo in Time Test, RMSE

	Portugal	Greece
Observations	39	31
Treatment	3171.723	6303.632
Maximum in test	1674.810	2670.722

For Portugal one time period needed to be dropped

As we can see for both countries, the gap caused by the actual treatment is bigger than the biggest gap caused by the placebo treatments. This gives further evidence for the hypothesis that the austerity measures indeed decreased economic growth in the affected countries. Also, with 31 and 39 observations respectively, the explanatory power of this inferential test is notably bigger compared to the permutation test.

As robustness check we excluded for every estimation the most important country from the donor pool. For Greece and Spain, there are hardly any differences. The fit for Portugal decreases slightly, but the overall movement of the estimator is similar to the one including India in the donor pool. Thus, it seems that the estimations are rather robust to changes to the donor pool. We also included more covariates into the estimation. Namely money (M1) growth and public sector spending as percentage of GDP. For Spain and Greece there is no visible difference. For Portugal the fit becomes slightly worse. This confirms further that we selected a proper set of covariates for constructing the synthetic counterfactual.[5]

[3]For this test we still needed to exclude India

[4]The last time period of the test is chosen in such a way that there is no overlap between the placebo-in-time results and the actual treatment effect

[5]All robustness check results are available upon request

4.5 Discussion and Policy Implications

In this section we will further discuss the results obtained in the preceding section. First of all, it should be made clear what the synthetic counterfactual is actually representing in this chapter. Since it is built from countries outside the Euro-zone, it has a floating exchange rate, compared to the semi-fixed exchange rate that the Euro-area countries have. Also, since we did not optimize on debt levels, the synthetic counterfactual experiences no fiscal stress. This would be the equivalent of a complete bailout of the debts for the treated countries without strings attached. There were several possible ways for this to happen. The ECB could have monetarized the debt by buying the respective government bonds. This option was discussed especially at the beginning of the crisis. Another option would have been the collectivizing of debts, which was discussed under the names of Eurobonds. Finally a fiscal bailout (as with the ESFS) with no strings attached would have been another possibility. We want to make clear that all of these options are politically difficult (monetarizing debt, Eurobonds) to unfeasible (full bailout with no strings attached). However, an agenda of debt support and reforms without austerity would have led to a better outcome than we see today, and would probably be close to the synthetic counterfactual.

Furthermore, the results obtained for Greece and Portugal should not be viewed as an upper bound of the potential damage done to economies via austerity. This is, because some of the enforced reforms such as making the labor market more flexible should have positive effects on GDP. Thus, the impact of only slashing spending might potentially be even more devastating.

Another thing to note is the case of Spain that stands apart from Greece and Portugal. While for the first two we can see a rather obvious impact of the treatment, for Spain the whole situation is much less obvious. Indeed when looking at the actual output evolution of Spain, the development after the treatment looks like a linear interpolation of the trend prior to the request for assistance. Thus, it is very difficult to establish how much damage austerity has done to the Spanish economy, if any at all. An alternative explanation is that the Spanish economy is still in a stage of transition after the collapse of the construction sector and thus we see the contracted GDP.

Also, when looking at Portugal, it should be noted that the synthetic counterfactual is tracking Portugal above its actual GDP per capita in the last few years before the treatment. This is because the estimator is tracking the outcome variable over the period such that it hits the target on average, which in this case implies that over a sustained period of time the estimator is slightly underestimating the outcome series and later on slightly overestimate the variable. Thus, in this case the damage to the economy might be slightly overstated.

Overall, this analysis begs the question whether the damage inflicted upon Greece and Portugal can be justified. Especially when looking at the debt levels of Greece, which are not very far away from the 260 billion dollar that was lost in potential output.

4.6 Conclusion

In this chapter we evaluated the economic impact of austerity programs for Greece, Portugal, and Spain. We do so by constructing a synthetic control method, to obtain an estimate of a the respective country that was not subject to the treatment. It turns out that there are significant differences between the countries. While Portugal and Greece exhibit clear signs of a massive contraction of output due to the fiscal consolidation program, it is difficult to attribute an effect from the program to the Spanish economy.

The estimator used in this chapter was introduced by Abadie and Gardeazabal (2003). It uses a pool of countries (called the donor pool) to construct a synthetic version of a treated unit. This is done by assigning weights from zero to one which add up to one to the donor pool members such that the distance between the outcome variable (which is affecting by the treatment) and the weighted combination of the outcome variable of the members of the donor pool is minimized. Simultaneously, the estimator matches a number of characteristics that explain the outcome variable. The optimization is carried out until the point in time where the treatment takes place. Afterwards, the evolution of the outcome variable of the synthetic counterfactual is compared to the actual outcome variable, to give an indication of the effect of the treatment.

We investigate the fiscal consolidation programs for Greece, Portugal and Spain, imposed due to the sovereign European debt crisis in this chapter. Due to the fact that the donor pool should be independent of the treated unit, it mainly consists of non-European OECD countries as well as some developing countries. The outcome variable is GDP per capita measured on a quarterly basis, while the covariates are variables connected to economic development and the structure of the economy such as education, investment rate, and value added by the manufacturing sector. The results show for Greece and Portugal a good matching of the synthetic control method with the actual data and a stark divergence of GDP per capita between treated unit and synthetic unit after the austerity programs take place. These also turn out to be statistically significant. For the Spanish economy, it is notable that the synthetic counterfactual is unable to track the evolution of GDP per capita from 2009 onwards, that is from the beginning of the crisis. It seems that the Spanish economy is still suffering from the massive turmoils of 2008/2009, when next to the financial sector, also the booming construction sector, which was a long-time driver of the Spanish economy, was severely hit. Thus, it is difficult to establish whether the fiscal consolidation did any significant damage to Spain.

The results for Greece and Portugal allow to calculate the loss in potential output, which is 90 billion (2005) US Dollar for Portugal, and 260 billion (2005) US Dollars for Greece since the request for aid. Especially when looking at the latter figure, we see that the accumulated economic damage is very close to the current (start of 2015) debt level of the country, and it should be re-evaluated whether this policy is a reasonable path to follow.

Chapter 5

Leader-Follower Relationships for Sovereign Credit Ratings

5.1 Introduction

In this paper we investigate whether there exists a leader follower-relationship in the market for sovereign credit ratings during the recent sovereign debt crisis in Europe. The market for ratings is best characterized as an oligopoly, with three agencies controlling 95% of the market. Therefore, any interdependency between agencies would further decrease the variation in credit ratings that are vital for correcting mistakes by individual agencies. Due to the importance of credit ratings for investment decisions and economic policy, this research question is of high importance and interest to policymakers, agents acting in financial markets, as well as researchers in the field of economics and finance. We explore the possibility of a leader-follower relationship for the last decade in Europe, and find mixed results. While the most prominent crisis countries are evaluated independently by the Credit Rating Agencies (CRAs), we find interdependence when analyzing rating decisions from Italy and Ireland. A possible reason for this is that both of these countries are not clear cut cases, compared to Greece, Portugal and Spain, and the agencies are thus unsure of their decisions.

Credit ratings are an opinion on the probability of default of an asset, published by a corporation known as Credit Rating Agency. These ratings are used in financial markets to determine the riskiness of investments. The Basel approach allows banks to use them for determining appropriate risk weights such that the capital ratio stays within the allowed limits. Therefore, the agencies have the ability to move huge amounts of capital between different assets by changing ratings, see for example Gande and Parsley (2010). Additionally, a lot of other financial institution, such as pension funds, are encouraged or forced by government regulation to use these ratings as risk indicators, which further amplifies their decisions.

The market for Credit Rating Agencies has always been quite concentrated. Currently it is dominated by three big agencies, namely Moody's Investor Service (Moody's), Standard and Poor's (S & P), and Fitch Ratings (Fitch). All of them were founded at the beginning of the 20th century, and all of them managed to survive and thrive until today through a combination of reliable ratings, innovative business decisions, and government regulation. Recently, the CRAs have come under severe criticism. On the one hand for their role in the 2008 financial crisis, when they rated financial products as very safe, which went into default only months later. On the other hand, it was for their role in the European sovereign debt crisis that escalated in 2010. The agencies were criticized for downgrading countries in a sudden manner, mainly by European politicians, that led to plans in the EU to establish an own rating agency. White (2010) gives an excellent summary of the history of the CRAs, with focus on the past ten years. The rating system of all three agencies is rather similar and actually comparable. Tables 5.1 and 5.2 give an overview.

Table 5.1: Rating Scale Part 1

Moody's	S & P	Fitch
Aaa	AAA	AAA
Aa1	AA+	AA+
Aa2	AA	AA
Aa3	AA-	AA-
A1	A+	A+
A2	A	A
A3	A-	A-
Baa1	BBB+	BBB+
Baa2	BBB	BBB
Baa3	BBB-	BBB-
Ba1	BB+	BB+

Table 5.2: Rating Scale Part 2

Moody's	S & P	Fitch
Ba2	BB	BB
Ba3	BB-	BB-
B1	B+	B+
B2	B	B
B3	B-	B-
Caa1	CCC+	-
Caa2	CCC	CCC
Caa3	CCC-	-
Ca	CC	CC
Ca	C	C
C	D	D

Optimally, we would like to have many agencies that provide ratings on a country. The average rating would then give a rather accurate description of the probability of default of said country. The fewer rating agencies are active in this market, the less accurate the picture will be. Unfortunately we only have three main agencies that are active in the market. This already significantly reduces the possibility that if one agency makes a wrong judgment call, the other two are able to counterbalance it, such that the average of the ratings is correct. Next, averaging the ratings only works if the agencies are actually independent of each other. If they are not, the error of one agency might induce the other two to follow with similar decisions and fundamentally distort the ratings. Rating changes by all three agencies are often announced in a limited time interval. One could argue that this is already evidence that there must be some interdependence. However, we should keep in mind that the ratings are made up of public information which all three agencies receive simultaneously. Therefore, it makes sense that all decisions are also made closely to each other. This of course means that we first need to take this into account. There is a body of literature that explains ratings with fundamental (macroeconomic) factors. The first to do so are Cantor and Packer (1996). They find that a substantial amount of variation in ratings can be explained by relatively few macroeconomic factors. Afonso (2003) extends this analysis and confirms the results from Cantor and Packer. Ferri et al. (1999) use such a model to predict ratings in the east-Asian financial crisis and compares them to actual rating development. They find that agencies were procyclical in their assignment of grades. Contrary, Mora (2006) comes with a slightly different methodology to the opposite result.

In this paper we will use downgrades to evaluate the leader-follower relationship. An important reason for this is that we want to investigate the behavior of the agencies in the current crisis in Europe, in which downgrades were handed out almost exclusively. We will use a binary measure for downgrades, that is when a downgrade is issued, the variable becomes a one, while it is zero otherwise. Due to the low number of upgrades happening, treating upgrades and no rating changes symmetric should not distort the analysis. This makes our approach closest to Freitag (2014), who also uses binary rating changes.

So far the literature on leader-follower relationship in the rating market is rather limited. One paper that does such a test is Gande and Parsley (2010). However, they simply look at the release dates of rating changes and do a relatively straightforward test to detect such a relationship. Our paper is probably closest to Alsakka and Gwilym (2010), who are conducting a granger-style causality test with ordered probit to tackle the question. They find that Moody's is generally leading for upgrades, while there is no clear pattern for downgrades. Both studies do not take into account macroeconomic fundamentals. Also, it is difficult to make any claims about long-run relationships from these approaches. In this paper, we will provide a long-run analysis of the leader-follower relationship for five European countries over the last decade. We use a frequency domain test due to Breitung and Candelon (2006) to do so. The test separates the economic factors that influence the rating decision of the respective agencies from any leader-follpwer pattern and systematically tests for this behavior in the long run. Thus, it allows for an additional vector of control compared to a simple granger-causality test. Also, since we will sample at a quarterly frequency, our test is looking at a true long-run relationship between the agencies, instead of giving weights to a difference of a few days between the rating decisions of different agencies, as it would be the case in Alsakka and Gwilym (2010).

The rest of the paper is organized as follows. The next section will introduce the methodology, section three details the empirical strategy, section four presents the data, section five is devoted to results, and section six concludes.

5.2 Methodology

Let us consider the binary variable ($M_{i,j,t}$), representing the occurrence of a downgrade published by the three CRAs ($i =$ (Fitch, S & P, Moody's)) for the country j at time t, $t = 1, ..., T$. To evaluate the existence of potential collusions between the CRAs, it is necessary to control in a first instance for the common information (F_t) simultaneously available at time t and use to fix the ratings. Without such a preliminary step, the causality test would be biased because of informational endogeneity issues. Practically, we estimate a panel logit model where F_t is composed by several macroeconomic variables as GDP growth, government debt as percentage of GDP, inflation rate, primary balance. It follows thus the following equation:

$$M^*_{it} = b' F_t + e_{it}, \tag{5.1}$$

where $M_{it} = 1$[1] if $M^*_{it} > 0$, 0 otherwise, and e_t is a 3-dimensional residual vector.

Once the commonly available information is retrieved, it is possible to test for collusion from the residuals e_{it}. This is done via the causality test in the frequency domain proposed by Breitung and Candelon (2006) (BC hereafter). This test presents the advantage of being widely accepted and to allow the separation between causality in the short and the long run. Such an ability is crucial for our test. It would be completely logical that ratings issued by the three CRAs move in the same direction in the long-run. On the contrary, finding causality in the short run would be the signal of collusion between the CRAs. To present BC methodology, let us consider $v_t = [e_{it}\prime = [x_t, y_t, z_t]$ to be a three-dimensional vector of time series observed obtained in the previous stage corresponding to the part of the rating downgrade which is not explained by the common information (F_t). It is assumed that z_t has a finite order vector autoregressive (VAR) representation of the form:

$$\Theta(L)v_t = \varepsilon_t \ , \tag{5.2}$$

where $\Theta(L) = I - \Theta_1 L - \cdots - \Theta_p L^p$ is a 3×3 lag polynomial with $L^k v_t = z_{t-k}$. We assume that the error vector ε_t is white noise, with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Sigma$, where Σ is positive definite. For ease of exposition, we do not include any deterministic terms in (5.1) although in empirical applications the model typically includes a constant.

Here, y_t is Granger causal for x_t if the forecast variance of x_{t+1} conditional on $\mathcal{X}_t = \{x_t, x_{t-1}, \ldots\}$ is larger than forecast variance of x_{t+1} conditional on $\mathcal{X}_t \cup \mathcal{Y}_t$, where $\mathcal{Y}_t = \{y_t, y_{t-1}, \ldots\}$. In other words $\mathcal{Y}_t$ contains information to predict the one-step ahead value of x_t.

BC propose a causality test from y_t to x_t in the frequency domain in such a trivariate system. Ir requires an intermediate step in which the VAR is conditionalized

[1] Subscript j has been dropped by ease of notation.

with respect to z_t. In other words, y_t and x_t are projected on z_t and the three dimension system is reduced to two using expectations ($\tilde{x}_t$ and $\tilde{y}_t$) instead of the raw variables.

Let G be the lower triangular matrix of the Cholesky decomposition $G'G = \Sigma^{-1}$ such that $E(\eta_t \eta_t') = I$ and $\eta_t = G\varepsilon_t$. If system (1) is assumed to be stationary, the MA representation of the system is

$$\begin{aligned} v_t &= \Phi(L)\varepsilon_t = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \\ &= \Psi(L)\eta_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}, \end{aligned} \tag{5.3}$$

where $\Phi(L) = \Theta(L)^{-1}$ and $\Psi(L) = \Phi(L)G^{-1}$.

The measure of causality suggested by Geweke (1982) and Hosoya (1991) is the following:

$$M_{y\to x}(\omega) = \log\left[1 + \frac{|\Psi_{12}(e^{-i\omega})|^2}{|\Psi_{11}(e^{-i\omega})|^2}\right]. \tag{5.4}$$

Several methods have been proposed to test for the nullity of $|\Psi_{12}(e^{-i\omega})| = 0$, corresponding to the case where y does not cause x at frequency ω.

Breitung and Candelon (2006) propose the simplest approach to test for the null hypothesis of non-causality (i.e. $M_{y\to x}(\omega) = 0$) based on the necessary condition $|\Psi_{12}(e^{-i\omega})| = 0$, using $\Psi(L) = \Theta(L)^{-1}G^{-1}$ and

$$\Psi_{12}(L) = -\frac{g^{22}\Theta_{12}(L)}{|\Theta(L)|},$$

where g^{22} is the lower diagonal element of G^{-1} and $|\Theta(L)|$ is the determinant of $\Theta(L)$. It follows that y does not cause x at frequency ω if [2]

$$|\Theta_{12}(e^{-i\omega})| = \left|\sum_{k=1}^{p} \theta_{12,k}\cos(k\omega) - \sum_{k=1}^{p} \theta_{12,k}\sin(k\omega)\, i\right| = 0 .$$

Their empirical procedure consists of testing for these *linear* restrictions. To simplify the notation, we let $\alpha_j = \theta_{11,j}$ and $\beta_j = \theta_{12,j}$, so that the VAR equation for x_t is written as

$$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \varepsilon_{1t} . \tag{5.5}$$

The hypothesis $M_{y\to x}(\omega) = 0$ is equivalent to the linear restriction

$$H_0: \qquad R(\omega)\beta = 0 , \tag{5.6}$$

[2]Note that g^{22} is positive due to the assumption that Σ is positive definite.

where $\beta = [\beta_1, \ldots, \beta_p]'$ and

$$R(\omega) = \left[\begin{array}{cccc} \cos(\omega) & \cos(2\,\omega) & \cdots & \cos(p\,\omega) \\ \sin(\omega) & \sin(2\,\omega) & \cdots & \sin(p\,\omega) \end{array} \right] .$$

This restriction tests that (5.5) is an ordinary F statistic and is asymptotically distributed as $F(2, T-2p)$ for $\omega \in (0, \pi)$. Such a method can be extended to higher dimensional systems or to cointegrated VARs (see Breitung and Candelon, 2006). The comparison with the causality test in time domain is not straightforward.

5.3 Empirical Strategy

This section will describe the empirical strategy that we will employ in this paper. As we saw in the preceding sections, we want to setup a VAR in order to apply the frequency domain test. However, the dependent variable is of binary nature. Therefore, we cannot simply estimate a standard VAR using OLS. The binary extension of these multi-equation regressions is a multivariate probit that could be estimated using Simulated Maximum Likelihood Estimation. For the case at hand we would estimate a multivariate probit for every single country, and apply the frequency domain test to it. However, this will be difficult with the data at hand. First of all, we only have very few downgrades available for each country. Second of all, when introducing lagged dependent variables, there is a high chance that they can perfectly forecast the dependent variable at the quarterly frequency we have used to sample the data. This leads to a breakdown of the Maximum Likelihood Estimator.[3] We could be changing the frequency of downgrades to monthly, but since we want to control also for macroeconomic characteristics which are typically sampled at quarterly frequency, this is unattractive. Furthermore, this does not get rid of the problem of having too few downgrades per country on which to conduct the estimation.

Therefore, we will use a two-step approach. In a first step, we control for macroeconomic characteristics. This is done by estimating three logit regressions (one for each CRA) with downgrades as dependent variable and GDP growth, government debt, the primary balance, and inflation as independent variables. For this regression, the cross section will not only be made up of the countries under investigation but a broad panel of European countries. Specifically, these are: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, the Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, and the UK. By including a broad cross section, we gain stable results for the first step regressions.[4] Additionally, this is good practice when trying to estimate the relationship between macroeconomic factors and ratings, see for example

[3]Indeed, nearly all multivariate probits that were estimated failed to converge

[4]We also tried time series estimations, but these regularly came up with most of the regressors insignificant

Bennell et al. (2006), or Afonso et al. (2007). In a second step, we extract the residuals for the countries to be investigated. In order to reduce the dimension of the problem from three to two, we next estimate a regression of:

$$y_{1t} = \beta_1 y_{2,t} + \beta_2 y_{3,t} + \sum_{j=1}^{p} \beta_j y_{1,t-j} + e_t, \tag{5.7}$$

where the y's represent the residuals from the first stage. This gives us the ability to investigate the action of one CRA on another, given the action of the third one (number 1 in this case). As a last step, we estimate a VAR on data containing two of the residual series from the first stage as well as the residuals from equation (5.7) as exogenous variable. This will be used to conduct the frequency domain test on. This procedure is also described in Breitung and Candelon (2006).

5.4 Data

In this section we describe the dataset that we use. For the first step estimation the cross section consists of: Belgium, Czech Republic, Denmark, Germany, Estonia, Ireland, Spain, France, Italy, Cyprus, Latvia, Lithuania, Luxembourg, Hungary, The Netherlands, Austria, Poland, Portugal, Slovenia, Finland, and the United Kingdom. The variables are government debt as percentage of GDP, growth rate of GDP, the primary balance as percentage of GDP, and the inflation rate. The data spans the first quarter of 2006 up until the second quarter of 2013. All data is retrieved from Eurostat.

In this paper we will use what we call a pseudo-vintage approach. A Vintage is the very first iteration of a data that gets published. Therefore, it represents the information that is available at time t. In order to understand market reactions or decisions of CRAs, it would be optimal to always use vintages, since these are the information that the CRAs are actually basing their decisions on. Using Vintages is generally not done in the literature. There are several reasons for this. First of all, data availability is a problem. Very few statistical agencies actually supply vintages to the general public.[5] Second of all, a researcher might argue: "why use an outdated version of a dataset, when there is an update available". The problem that we face in this paper is that the European national statistical agencies needed to do a significant overhaul of their data after the sovereign debt crisis hit. Especially in the crisis countries, a lot of GDP development prior to the crisis has now been identified as being produced by bubbles, which should not be reflected in GDP figures. However, the CRAs made judgments according to the numbers that were issued first, that is the ones that were biased. Thus, if we use the current numbers provided by Eurostat, we would be getting a highly diluted

[5] A notable exception is the St. Louis Fed with its vast vintage database

result. [6]So, to solve this we take a Eurostat dataset prior to the main upgrade from 2011/2012 onwards, and add data from 2012 until 2013 from the current database. Since the data prior to 2012 is not the very first iteration, and we combine it with other updated data, we call this a pseudo vintage approach.

The ratings have been taken from the agencies. They cover the same span as the macroeconomic variables. We describe in the tables on the next page the downgrades for all countries under investigation in the respective time period.

[6]Indeed when estimating the first step regression described in section 5.3 with the current Eurostat database, the R-square is tenfold smaller and most variables are insignificant.

Table 5.3: Downgrades for Portugal

Agency	Rating Change	Date
Moody's	Aa2 to A1	July 13th 2010
	A1 to A3	March 15th 2011
	A3 to Baa1	April 5th 2011
	Baa1 to Ba2	July 5th 2011
	Ba2 to Ba3	February 13th 2012
S & P	AA- to A+	January 21st 2009
	A+ to A-	April 27th 2010
	A- to BBB	March 24th 2011
	BBB to BBB-	March 29th 2011
	BBB- to BB	January 13th 2012
Fitch	AA to AA-	March 24th 2010
	AA- to A+	December 23rd 2010
	A+ to A-	March 24th 2011
	A- to BBB-	April 1st 2011
	BBB- to BB+	November 12th 2012

Table 5.4: Downgrades for Ireland

Agency	Rating Change	Date
Moody's	Aaa to Aa1	July 2nd 2009
	Aa1 to Aa2	July 19th 2010
	Aa2 to Baa1	December 17th 2010
	Baa1 to Baa3	April 15th 2011
	Baa3 to Ba1	July 12th 2011
S & P	AAA to AA+	March 30th 2009
	AA+ to AA	June 8th 2009
	AA to AA-	August 24th 2010
	AA- to A	November 23rd 2010
	A to A-	February 2nd 2011
	A- to BBB+	April 1st 2011
Fitch	AAA to AA+	April 8th 2009
	AA+ to AA-	November 4th 2009
	AA- to A+	October 6th 2010
	A+ to BBB+	December 9th 2010

Table 5.5: Downgrades for Greece

Agency	Rating Change	Date
Moody's	A1 to A2	December 22nd 2009
	A2 to A3	April 22nd 2010
	A3 to Ba1	June 14th 2010
	Ba1 to B1	March 7th 2011
	B1 to Caa1	June 1st 2011
	Caa1 to Ca	July 25th 2011
	Ca ro C	March 2nd 2012
S & P	A to A-	January 14th 2009
	A to BBB+	December 16th 2009
	BBB+ to BB+	April 27th 2010
	BB+ to BB-	March 29th 2011
	BB- to B	May 9th 2011
	B to CCC	June 13th 2011
	CCC to CC	July 27th 2011
	CC to SD	February 27th 2012
	CCC to SD	December 5th 2012
Fitch	A to A-	October 22nd 2009
	A- to BBB+	December 8th 2009
	BBB+ to BBB-	April 9th 2010
	BBB- to BB+	January 14th 2011
	BB+ to B+	May 20th 2011
	B+ to CCC	July 13th 2011
	CCC to C	February 22nd 2012
	C to RD	March 9th 2012
	B- to CCC	May 17th 2012

Table 5.6: Downgrades for Spain

Agency	Rating Change	Date
Moody's	Aaa to Aa1	September 30th 2010
	Aa1 to Aa2	March 10th 2011
	Aa2 to A1	October 18th 2011
	A1 to A3	February 13th 2012
	A3 to Baa3	June 13th 2012
S & P	AAA to AA+	January 19th 2009
	AA+ to AA	April 28th 2010
	AA to AA-	October 13th 2011
	AA- to A	January 13th 2012
	A to BBB+	April 26th 2012
	BBB+ to BBB-	October 10th 2012
Fitch	AAA to AA+	May 28th 2010
	AA+ to AA-	October 7th 2011
	AA- to A	January 27th 2012
	A to BBB	June 7th 2012

Table 5.7: Downgrades for Italy

Agency	Rating Change	Date
Moody's	Aa2 to A2	October 4th 2011
	A2 to A3	February 13th 2012
	A3 to Baa2	July 13th 2012
S & P	AA- to A+	October 19th 2006
	A+ to A	September 19th 2011
	A to BBB+	January 13th 2012
Fitch	AA to AA-	October 19th 2006
	AA- to A+	October 7th 2011
	A+ to A-	January 27th 2012
	A- to BBB+	March 8th 2013

In all these tables we see that the decisions of the different agencies are remarkably similar. Also, time wise the downgrades are rather close to each other. However, by only looking at the tables we are of course unable to make a formal argument for any kind of leader-follower relationship. Therefore, a more rigorous approach is presented in the next section.

For the analysis we transform the ratings into a binary variable. We only take into account downgrades, in every quarter when there is a downgrade, the variable is equal to one, and zero otherwise. One might argue that this approach is problematic because multiple downgrades are not taken into account. However, in most cases there is only one downgrade per quarter, and when this is not the case, often all agencies are dishing out multiple downgrades, see for example Greece, or Portugal.

5.5 Results

In this section we present the results of our testing procedure. We start with the first step, that is with the logit regression of macroeconomic factors on ratings. The results can be found in table 5.8.

Table 5.8: Downgrade Estimations using Logit Regression

	Moody's	S & P	Fitch
Growth	-0.141^{***}	-0.134^{***}	-0.207^{***}
Debt	0.020^{***}	0.018^{***}	0.016^{***}
Inflation	0.019	-0.062	0.012
Primary Balance	-0.079^{***}	-0.053^{**}	-0.058^{**}

The results are in line with what we would expect. The negative significant coefficient for economic growth indicates that a higher growth rate decreases the chance of a downgrade. The same logic applies for the Primary Balance. Having more money available for debt servicing decreases the chance of getting downgraded. Higher government debt has a positive significant effect on the probability of receiving a downgrade. Inflation is not significant in all three regression. This makes sense in so far that European countries have not experienced high inflation since the 80s, and thus CRAs might have decided to not treat it as an important indicator for this set of countries. These results are very much in line with Freitag (2014), who presents similar results. The interested reader might note that we do not have any variable that is covering trade relations. There are two reasons for this. First of all, Freitag (2014) who uses a similar panel than we do reports that in quite a few cases the trade variable is not significant. Second of all, we have a lack of data. The current account data for which we have prior to 2012 the pseudo vintage data available is reported in millions of currency units, and the conversion to a percentage-of-GDP variable was calculated by hand. Unfortunately we do not have GDP figures for all the countries after 2012 available due to a continuing updating process by Eurostat and the national statistical agencies. Therefore, we have decided to drop this variable.

For the second step, we have extracted the (Pearson) residuals from the regression. In order to reduce the dimension of the causality question from three to two, we first estimate a regression to abstract from one CRA, as described in equation(5.7). Lastly, we estimate a VAR. The order of this VAR is three for every case under investigation, which is the minimum amount of lags that we can use due to the construction of the test.[7] We use this to apply the frequency domain test. We will only present the results of the test. We test for causality at different frequencies in the data. This ranges from $pi/2$ which is four quarter to $\pi/8$ which is approximately four years. The critical value for rejecting the null hypothesis of no predictability is 5.99.

Table 5.9: Frequency Domain Test Results - Greece

	Moody's → Fitch	Moody's → S & P	Fitch → S & P	Fitch → Moody's	S & P → Moody's	S & P → Fitch
$\pi/8$	1.839	0.985	4.795	1.627	2.308	4.835
$\pi/4$	1.753	0.937	4.308	1.480	2.156	4.302
$3\pi/8$	1.564	0.922	3.497	1.174	1.881	3.434
$\pi/2$	1.221	1.210	3.447	0.712	1.746	3.393

The results for Greece are displayed in table 5.9. As we can see there is no causality between the different agencies at any frequency. This is indicated by the fact that the case of downgrading Greece seems to be rather clear cut for every individual Credit Rating Agency. Therefore, we can conclude that in this case the agencies are not observing each others action.

Table 5.10: Frequency Domain Test Results - Ireland

	Moody's → Fitch	Moody's → S & P	Fitch → S & P	Fitch → Moody's	S & P → Moody's	S & P → Fitch
$\pi/8$	6.082**	9.856**	3.151	4.185	13.814**	8.804**
$\pi/4$	6.244**	9.228**	2.775	3.999	14.612**	8.587**
$3\pi/8$	6.811**	8.535**	2.432	3.590	16.275**	8.084**
$\pi/2$	7.378**	11.585**	3.903	3.087	18.116**	7.466**

The situation changes quite a bit when we look at Ireland in Table 5.10. Here we see a significant amount of interaction between the CRAs. Fitch is a follower for this country, being influenced by the other two agencies. Conversely, Standard and Poor's and Moody's are reacting to each others' actions in this case. Overall this indicates that the CRAs are unsure about their decisions. We argue that this is because Ireland used to be the economic role model for Europe, especially for the poorer countries. Being itself a poor countries for centuries, it was on an

[7] The laglength was determined by the AIC.

economic growth path ever since the eighties, and was dubbed the Celtic Tiger during the nineties. The Irish did everything according to economic textbook. Few regulations, low taxes, and business friendly environment. After the crisis, the CRAs need to answer the question whether the Irish economy has fundamentally changed from the state before the crisis. This is obviously difficult to evaluate and thus, from this perspective it is understandable that the analysts at the CRAs were hesitant to downgrade the country without the agreements of their colleagues. This can lead to such a leader-follower pattern that we observe here.

Table 5.11: Frequency Domain Test Results - Spain

	Moody's → Fitch	Moody's → S & P	Fitch → S & P	Fitch → Moody's	S & P → Moody's	S & P → Fitch
$\pi/8$	0.612	1.283	1.054	3.299	1.711	0.156
$\pi/4$	0.597	1.264	1.088	2.893	1.564	0.156
$3\pi/8$	0.536	1.170	1.027	2.469	1.537	0.143
$\pi/2$	0.400	0.943	0.680	3.138	2.157	0.099

Next, we take a look at the result for Spain which are displayed in Table 5.11. As in the case with Greece, the results are rather clear cut. There is no leader-follower relationship. Our results suggest that the CRAs are evaluating the probability of default independently of each other.

Table 5.12: Frequency Domain Test Results - Italy

	Moody's → Fitch	Moody's → S & P	Fitch → S & P	Fitch → Moody's	S & P → Moody's	S & P → Fitch
$\pi/8$	23.469**	11.579**	50.514**	25.795**	27.992**	51.304**
$\pi/4$	24.239**	13.212**	50.638**	26.004**	28.924**	51.230**
$3\pi/8$	24.922**	16.953**	51.024**	26.983**	31.252**	50.914**
$\pi/2$	21.820**	18.562**	52.867**	29.272**	36.493**	49.799**

The next country under investigation is Italy. Here we see, similar to Ireland, a very strong interdependency between the agencies. In fact in this case all agencies are watching each other and their respective action. Therefore, none of the rating decisions for Italy are completely independent of the actions of the CRA's peers. This can be explained by the fact that first of all Italy is one of the biggest economies in the Euro-area. Any premature decisions by rating agencies might have huge political implications and economic repercussions which not only impact Italy itself, but the EU as a whole. Naturally, no agency is willing to be the first to trigger such an event with a downgrade. Second of all, Italy has historically been a country with a high debt level for quite some time, and has always managed to keep the level stable, mainly by having a primary surplus most of the years. This distinguishes Italy from most of the other countries that we analyze. Also, Italy is the only country that we analyze in this paper that has not requested financial aid from any of the various EU and international institutions that have provided support to Greece, Spain, Ireland, and Portugal. All of this understandably makes CRAs very hesitant when making downgrading decisions. For some other countries the case is simply much more clear cut and thus the agencies have more confidence in their decisions in those cases.

Table 5.13: Frequency Domain Test Results - Portugal

	Moody's → Fitch	Moody's → S & P	Fitch → S & P	Fitch → Moody's	S & P → Moody's	S & P → Fitch
$\pi/8$	3.389	1.450	1.602	3.551	2.293	0.771
$\pi/4$	2.945	1.571	1.205	3.823	2.185	0.674
$3\pi/8$	2.320	2.065	1.038	4.100	1.923	0.469
$\pi/2$	2.284	2.906	2.079	3.536	1.551	0.227

The last set of tests that we perform is analyzing Portugal. Here, we see what we also saw in the Spanish and Greek case. All three agencies are acting independently of each other, which indicates that the agencies must be relatively sure of their ratings.

To sum up, we have tested for a leader-follower relationship in the market for sovereign credit ratings. We investigated five European countries from 2006 to mid 2013. These are Greece, Spain, Ireland, Italy, and Portugal. Our interpretation is that generally the CRAs make independent decisions for the more clear cut cases. These are Spain, Greece, and Portugal. All of these countries requested financial aid from European and international institutions. Also, all of these countries have economies that are struggling to stay competitive for quite some time (Portugal, Greece), or the main pillar of economic growth has suffered heavily (Spain and its construction sector).

For the less obvious cases the agencies seem to watch each other's actions closely. In our analysis this is the case for Ireland and Italy. We argue that the CRAs are simply not sure about how to rate these countries for a variety of reasons. Ireland was the European economic miracle of the nineties, often dubbed the Celtic Tiger. Its government did everything right according to standard economic theory. Low corporate taxes did increase economic growth to levels of developing countries. Additionally, it could convince several multi-national companies to transfer their headquarters to Dublin. The question when rating Ireland post-crisis is of course, whether the economy has changed fundamentally, such that one should expect that the solvency of the government is in danger, or whether it has remained identical in structure compared to before the crisis. There are good arguments for both positions. For Italy, the main question is whether there is a fundamental change from the situation it was in since the early nineties. Since then, the economy had anemic growth and the debt level has been relatively high. However, Italy nearly always managed to run a primary surplus to service its debts. Also, compared to other European countries, Italy's output contraction in 2009 was relatively small. This all might explain why the agencies are more careful when assigning investment grades to Italy.

5.6 Conclusion

In this paper we analyzed leader-follower relationships in the market for sovereign ratings. The market for these ratings is already an oligopoly, and any sort of interdependence would lead to further concentration. Also, such a leader-follower relationship would indicate that the CRAs are not doing their job properly. Instead of evaluating countries only with respect to macroeconomic characteristics, they look also at the decisions of their peers. To conduct this analysis, we use a frequency-domain test developed by Breitung and Candelon (2006). This test allows us to decompose the rating decisions (after taking into account macroeconomic fundamentals) into its cycles and thereby gain the possibility to test for a leader-follower relationship. We tackle the problem with a two-step approach, since we need to estimate a VAR for this test. However, this was not directly possible with the characteristics of the dataset.

We analyze the decisions of the three leading ratings agencies, Moody's S & P, and Fitch, when rating five European countries, namely Greece, Portugal, Spain, Ireland, and Italy, over the last decade. We find that in the cases of Greece, Portugal, and Spain, there is no interdependency, but for Ireland and Italy, there is heavy interdependence at all frequencies tested. We argue that these two groups exist for the following reason. The first group is rather straightforward to evaluate. All three countries have struggled for quite some time with their competitiveness (Greece, Portugal) or recently had their main economic growth driver crumbling (the Spanish construction sector). Therefore, the evidence for the CRAs seems to be so overwhelming that they are confident in their judgments and do not need to look at the decisions of their competitors. Italy and Ireland form the second group. In their case, the agencies are not sure of their verdict and thus eye the competitors, whether they come to similar judgments. The reason is that both countries are not as straight forward to evaluate as the first group. Italy has had a debt to GDP ratio of over 100 % for more than two decades, and was always able to service its debts. Ireland's high level of debt comes from rescuing several banks which in turn were in trouble mainly because of a property bubble. It is difficult to asses whether the Irish economy is fundamentally damaged, while keeping in mind that not long ago Ireland had one of the highest growth rates in western Europe.

Overall these results offer some mixed signals. While it is laudable that the CRAs acted independently when evaluating Greece, Portugal, and Spain, it is of course problematic that in two out of the five cases investigated this is not the case. Keeping in mind the importance that ratings have in the current financial architecture, the relevant authorities should have a closer look at the rating creation process, especially for countries where the decision is difficult to make for the agency.

Chapter 6

Conclusion

This dissertation examines the decisions of Credit Rating Agencies during the European sovereign debt crisis, as well as the implementation of austerity programs in response to the crisis. We have done four studies, each looking at a different aspect of the subject.

In chapter two we examine whether ratings are assigned in a procyclical and path dependent pattern. Procyclicality means that ratings are influenced by the business cycle, although macroeconomic fundamentals are controlled for. Path dependence implies that conditional on country characteristics, your current rating is also dependent on your rating history. Both violate the through-the-cycle rating assignment, that CRAs claim to do. We do not find clear cut evidence for procyclicality, however we find path dependence for nearly all agencies. Also, there is a discrepancy between upgrades and downgrades in our sample, indicating that upgrades were given out too freely in Europe prior to the financial crisis.

The next chapter investigates whether ratings for European countries are actually reflecting default probabilities. Due to the fact that the amount of sovereign ratings is rather constrained (compared to corporate ratings), we use CDS data for our analysis. We use a new version of the MIDAS estimator to separate the noise from the signal. We find that in most cases the ratings are actually in agreement with financial markets on the default probability, however, we have a subset of eastern European countries where CRAs and CDS data disagree on the implied default probability.

Chapter four looks at the second component of our title. We analyze the economic impact austerity programs implemented in the last few years in Greece, Portugal, and Spain by constructing a synthetic counterfactual that serves as a comparison. We find that the programs did extensive damage to Greece and Portugal, however it is difficult to identify such an impact for Spain. A possible reason might be the turmoil caused by the transformation of the Spanish economy when the construction sector collapsed.

In the last chapter we look at the distribution of sovereign ratings over time. Specifically, we want to investigate whether there is a leader-follower relationship for multiple European countries. We use a frequency domain test to tackle the question and find that there are two categories among the countries investigated. On the one hand, there are countries such as Greece, where the CRA verdict seem to be relatively straightforward, and there is no interdependence. On the other hand we find cases such as Italy, where the CRAs seem to be less sure of their decisions and thus eye the decisions of other agencies.

Of course all these studies have limitations. An important one was already discussed in chapter five. We use updated data instead of vintage data (that is the first release of the data) for our studies. The problem is that CRAs make decision based on the first iterations of data, and thus updated data might reflect a completely different situation. Utilizing vintages is mainly hindered by a comprehensive vintage data base for Europe. However, it could be possible to assemble a vintage data base from press releases of Eurostat, and reestimate the regressions for chapter two and five. For chapter three we could go deeper into different models for estimating binary dependent variables. A logit-MIDAS regression or a MIDAS version of a linear probability model (with a modified bootstrap) might be options worth exploring.

Overall this thesis paints a mixed picture of the actions of Credit Rating Agencies. While we see problems with the rating assignment process in terms of rating sluggishness (chapter two), we also have to acknowledge the fact that CRAs seem to be able to evaluate default probabilities properly (chapter three). We see that upgrades might have been handed out too freely prior to 2008 (chapter two), but also that ratings for the most obvious fiscal sinners were assigned independently of the other agencies in the market (chapter five). A similar line of conclusions can be drawn from chapter four. While we see huge economic losses from austerity programs in Greece and Portugal, the medicine seemed to have worked in Spain.

The main takeaway of these conclusions is that CRAs are only imperfectly measuring default probabilities. They are prone to multiple fallacies that we have documented in this thesis. Therefore, it seems appropriate to call for more regulation in the market of CRAs. The stimulation of competition in the market might be one possible idea which was also discussed recently. Another one might be a further modification of the Basel accord.

Bibliography

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490).

Abadie, A., Diamond, A., and Hainmueller, J. (2014). Comparative politics and the synthetic control method. *American Journal of Political Science*.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, pages 113–132.

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

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

Afonso, A., Gomes, P., and Rother, P. (2007). What'hides' behind sovereign debt ratings? *ECB Working Paper No. 711.*

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

Alesina, A. and Ardagna, S. (2010). Large changes in fiscal policy: taxes versus spending. In *Tax Policy and the Economy, Volume 24*, pages 35–68. The University of Chicago Press.

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

Alter, A. and Schüler, Y. S. (2012). Credit spread interdependencies of european states and banks during the financial crisis. *Journal of Banking & Finance*, 36(12):3444–3468.

Altman, E. and Rijken, H. (2004). How rating agencies achieve rating stability. *Journal of Banking & Finance*, 28(11):2679–2714.

Amato, J. and Furfine, C. (2004). Are credit ratings procyclical? *Journal of Banking & Finance*, 28(11):2641–2677.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856.

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

Arezki, R., Candelon, B., and Sy, A. (2011). Sovereign rating news and financial markets spillovers: Evidence from the european debt crisis. *IMF working papers*, pages 1–27.

Arghyrou, M. G. and Kontonikas, A. (2012). The emu sovereign-debt crisis: Fundamentals, expectations and contagion. *Journal of International Financial Markets, Institutions and Money*, 22(4):658–677.

Aßmann, C. and Boysen-Hogrefe, J. (2012). Determinants of government bond spreads in the euro area: in good times as in bad. *Empirica*, 39(3):341–356.

Bar-Isaac, H. and Shapiro, J. (2012). Ratings quality over the business cycle. *Journal of Financial Economics*.

Barro, R. J. (1996). Determinants of economic growth: a cross-country empirical study. Technical report, National Bureau of Economic Research.

Barro, R. J. and Redlick, C. J. (2011). Macroeconomic effects from government purchases and taxes. *The Quarterly Journal of Economics*, 126:51–102.

Bartels, B. and Weder, B. (2013). A rating agency for europe a good idea?

Baum, C. F., Krpava, M., Schafer, D., and Stephan, A. (2013). Credit rating agency announcements and the eurozone sovereign debt crises. *Discussion Papers, DIW Berlin*.

Bennell, J., Crabbe, D., Thomas, S., and Gwilym, O. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications*, 30(3):415–425.

Berger, A. and Udell, G. (2004). The institutional memory hypothesis and the procyclicality of bank lending behavior. *Journal of Financial Intermediation*, 13(4):458–495.

Bernanke, B., Lown, C., and Friedman, B. (1991). The credit crunch. *Brookings papers on economic activity*, 1991(2):205–247.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.

Blanchard, O. J. and Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*, 117(4):1329–1368.

Borio, C., Furfine, C., and Lowe, P. (2001). Procyclicality of the financial system and financial stability: issues and policy options background paper of the bis. *BIS papers*, 1:1–57.

Breitung, J. and Candelon, B. (2006). Testing for short-and long-run causality: A frequency-domain approach. *Journal of Econometrics*, 132(2):363–378.

Bry, G. and Boschan, C. (1971). Front matter to" cyclical analysis of time series: Selected procedures and computer programs".

Candelon, B., Dumitrescu, E., and Hurlin, C. (2014). Currency crises early warning systems: Why they should be dynamic. *International Journal of Forecasting*, 30(4):1016–1029.

Cantor, R. and Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic Policy Review*, 2(2):37–53.

Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the united states. *Journal of Business & Economic Statistics*, 26(4):546–554.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Transactions on Mathematical Software (TOMS)*, 13(3):262–280.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43.

De Jong, R. M. and Woutersen, T. (2011). Dynamic time series binary choice. *Econometric Theory*, 27(04):673–702.

DuPont, W. and Noy, I. (2012). What happened to kobe? a reassessment of the impact of the 1995 earthquake in japan. *SEF Working paper*.

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

Foroni, C. and Marcellino, M. (2013). A survey of econometric methods for mixed-frequency data. *Working Paper*.

Foroni, C., Marcellino, M., and Schumacher, C. (2012). U-midas: Midas regressions with unrestricted lag polynomials. *CEPR Discussion Paper No. DP8828*.

Freitag, L. (2014). Procyclicality and path dependence of sovereign credit ratings the example of europe. *GSBE research memorandum*.

Frenkel, S. (2015). Repeated interaction and rating inflation: A model of double reputation. *American Economic Journal: Microeconomics*, 7(1):250–280.

Fuchs, A. and Gehring, K. (2013). The home bias in sovereign ratings. *Working Paper*.

Galvão, A. B. (2013). Changes in predictive ability with mixed frequency data. *International Journal of Forecasting*, 29(3):395–410.

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

Gande, A. and Parsley, D. (2010). Sovereign credit ratings, transparency and international portfolio flows. *MPRA Paper*.

Gärtner, M., Griesbach, B., and Jung, F. (2011). Pigs or lambs? the european sovereign debt crisis and the role of rating agencies. *International Advances in Economic Research*, pages 1–12.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, 77(378):304–313.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. *Working Paper*.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3):509–548.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Ghysels, E., Valkanov, R., and Serrano, A. (2009). Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches. In *EFA 2009 Bergen Meetings Paper*.

Giavazzi, F. and Pagano, M. (1990). Can severe fiscal contractions be expansionary? tales of two small european countries. In *NBER Macroeconomics Annual 1990, Volume 5*, pages 75–122. MIT Press.

Götz, T. B., Hecq, A., and Urbain, J.-P. (2014). Forecasting mixed-frequency time series with ecm-midas models. *Journal of Forecasting*, 33(3):198–213.

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987). Generalised residuals. *Journal of Econometrics*, 34(1-2):5–32.

Guérin, P. and Marcellino, M. (2013). Markov-switching midas models. *Journal of Business & Economic Statistics*, 31(1):45–56.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the Econometric Society*, pages 413–430.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of econometrics*, 93(2):345–368.

Harvey, A. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society*, pages 461–465.

Hinkley, D. V., Reid, N., and Snell, E. (1991). *Statistical theory and modelling: in honour of Sir David Cox, FRS.* Chapman and Hall.

Hosoya, Y. (1991). The decomposition and measurement of the interdependency between second-order stationary processes. *Probability theory and related fields*, 88(4):429–444.

Hsiao, C., Pesaran, M., and Pick, A. (2012). Diagnostic tests of cross-section independence for limited dependent variable panel data models*. *Oxford Bulletin of Economics and Statistics.*

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

Hull, J. C. and White, A. (2000). Valuing credit default swaps 1: No counterparty default risk. *NYU Working Paper No. FIN-00-021.*

IMF (2010). Global financial stability report, october 2010.

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

Jiang, J. X., Stanford, M. H., and Xie, Y. (2012). Does it matter who pays for bond ratings? historical evidence. *Journal of Financial Economics*, 105(3):607–621.

Kauppi, H. and Saikkonen, P. (2008). Predicting us recessions with dynamic binary response models. *The Review of Economics and Statistics*, 90(4):777–791.

Leigh, D., Devries, P., Freedman, C., Guajardo, J., Laxton, D., and Pescatori, A. (2010). Will it hurt? macroeconomic effects of fiscal consolidation. *World Economic Outlook.*

Lowe, P. (2002). Credit risk measurement and procyclicality. *BIS working paper.*

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics.* Number 3. Cambridge university press.

Marcellino, M. and Schumacher, C. (2010). Factor midas for nowcasting and forecasting with ragged-edge data: A model comparison for german gdp. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

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

Moretti, L., Campos, N., and Coricelli, F. (2013). Economic growth and european integration: A counterfactual analysis. In *AFSE Meeting 2013.*

Oliveira, L., Curto, J. D., and Nunes, J. P. (2012). The determinants of sovereign credit spread changes in the euro-zone. *Journal of International Financial Markets, Institutions and Money*, 22(2):278–304.

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

Remolona, E. M., Scatigna, M., and Wu, E. (2008). The dynamic pricing of sovereign risk in emerging markets: Fundamentals and risk aversion. *The Journal of Fixed Income*, 17(4):57–71.

Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, 100:763–801.

Schumacher, C. and Breitung, J. (2008). Real-time forecasting of german gdp based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3):386–398.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric theory*, 15(04):549–582.

White, L. (2009). The credit-rating agencies and the subprime debacle. *Critical Review*, 21(2):389–399.

White, L. (2010). Markets: The credit rating agencies. *The Journal of Economic Perspectives*, 24(2):211–226.

Wooldridge, J. (2001). *Econometric analysis of cross section and panel data.* MIT press.

Valorization

In this part of the thesis I will discuss the valorization of the thesis. A valorization is the process by which scientific knowledge can shape and influence the world and have practical application. In this thesis I discuss a subject that is of immediate concern not only to the academic community but also to policy makers and even to the public at large. Credit Rating Agencies were heavily discussed during the recent crisis and their behavior has sparked controversy. This was not only the case during the financial crisis, when previously rated AAA financial products defaulted and dragged banks and subsequently the world economy into the biggest recession since 1929, but also during the more recent sovereign debt crisis that can be regarded as a continuation of the financial crisis.

During the sovereign debt crisis, criticism of CRAs was again increasing. This came mainly from European politicians who accused the agencies of not rating European countries fairly and went as far as the accusation that the (US) credit rating agencies were serving a political purpose. Thus, the discussion of establishing a European rating agency as a result of the sovereign debt crisis that up until this date did not come to fruition. As it can be deduced from this narrative, the role of CRAs is an important topic that not only matters to academia, but also to political actors, policy makers in the economic sense and the society as a whole. To understand the impact of this thesis, we also need to review the rules and regulations that have enshrined CRAs in their current position within the financial system. Ratings serve to banks (and depending on the country also to other financial institutions) as a guideline to how much capital they need to retain for a given asset, dictated by the Basel rules. Thus, any change in ratings will have an impact on the price and interest rate of the rated financial asset due to changing supply and demand for said asset. This means the ratings have a direct impact on the deficit and debt evolution of a sovereign nation. This fact should give a strong indication why studying rating agencies is generally of value. Next, I will outline the valorization of the individual chapters.

In chapter 2 I discuss the issue of path dependence and procyclicality. Both should not be present in sovereign ratings. The latter because any rating that is procyclical will give too high ratings during boom time and too low ratings during a recession. The latter because it alleviates ratings of already highly rated countries and conversely makes it more difficult for countries that have low ratings to achieve higher ratings. This has a direct impact upon the fiscal position of a country. This chapter gives policy makers an indication that there are flaws in the rating process. It should give an incentive to review this process. Also, given the flaws in the process, it should impact the future discussion of capital requirements and the connected usage of ratings for its determination.

Chapter 3 discusses the fundamental question whether sovereign ratings actually reflect default probabilities. This is done by introducing a new estimator which can be potentially of use in other applications such as testing effectiveness of central bank communication. The fundamental question is of high importance to the discussion on rating agencies and its role in the sovereign debt crisis. While the CRAs were accused of not rating correctly, this chapter shows that there is indeed a connection between default probability and rating. This again gives arguments for the discussion by policymakers on minimum capital requirements in the Basel rules. It should be noted that changing these rules has profound impact on banking behavior and thus on the overall economy.

Chapter 4 addresses a different aspect of the sovereign debt crisis. I investigate the economic impact of austerity programs for three different European countries. The results of this chapter should have profound implications on the discussion of how to approach the sovereign debt crisis in the Euro-area. Specifically its analysis gives arguments in the debate to use austerity programs when saving a country using one of the rescue mechanisms established by the ECB, the European Commission and the IMF. Given the vast impact that these programs had in some countries, as shown by the analysis, a change of direction in this matter will have big economic implications.

In Chapter 5 I analyze the behavior of rating agencies between each other. Specifically, I investigate leader-follower pattern. Similarly to chapter 2 and chapter 3, the results of this chapter can shape the debate around the Basel accord and the question whether ratings should be used for minimal capital requirements. As already stated, any change to these rules could have an enormous impact on the financial system and thus on the world economy.

In conclusion, this thesis elaborates on two separate topics that are of high importance. First of all, the question whether ratings should be given the weight that they have in the current financial architecture, where I mainly supply policymakers with argument, due to the very technical nature of the topic. Second, the question whether austerity measures are actually effective, which is a question that gets debated not only amongst policymakers, but also in political circles and in the society as a whole.

CV

Lennart Freitag was born on November, 28th 1984 in Oldenburg, Germany where he also grew up. After attaining his Abitur in 2004 from the Herbartgymnasium, he did his civil service at the Rehazentrum Oldenburg.

In 2005 he enrolled at Maastricht University School of Business and Economics, attaining his Bachelor in International Economic Studies. During his studies, he spent a semester abroad at Paris-Sorbonne University. He attained his Master Degree in Economic and Financial Research with specialization Econometrics also from Maastricht University taking additional courses at Université Catholique Louvain-La-Neuve as well as at multiple other Dutch universities.

He continued his academic career by enrolling into the PhD Program at Maastricht University. He presented his research in Rome, Vienna and London amongst others and was awarded the best paper award at the Spring Meeting of Young Economists in 2014. Lennart is currently working as a Senior Consultant for the Data Analytics Team of Deloitte's Corporate Finance Division in Frankfurt.