Chasing the Usual Suspect:
Sovereign Credit Ratings and Borrowing Costs before and during the Crisis

By

Attila Gáspár

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of

Master of Arts in Economics

Supervisor: Professor Péter Benczúr

Budapest, Hungary

2012

ABSTRACT

I study the widespread suspicion that credit rating agencies contribute to financial crises through incorrectly assigning sovereign credit ratings, and thus exacerbating financing difficulties. First I investigate whether an autonomous effect of rating changes on sovereign spreads can be identified, having controlled for changes in the macroeconomic environment. I find an upper bound for such an effect, but it is very small in magnitude, and indirect evidence suggests that its true value could be negligible, which is in contrast with the body of literature that blames credit rating agencies for contributing to the financial turmoil. Then I model sovereign ratings themselves in order to contrast their actual behaviour with model predictions. The estimations suggest neither a distinct pattern of co-movement between model generated ratings and actual ones, nor a big systematic deviation between the two after the onset of the current crisis. On the average, ratings seem to (very mildly) lag behind macroeconomic data, but this does not hold for the period directly preceding the crisis. These findings suggest that the role of rating agencies in exacerbating the crisis (at least regarding sovereign debt financing) is probably exaggerated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

When the calendar turned to 2012, the world saw the start of the beginning of fifth year of what is now called the Great Recession, the global economic meltdown that is widely believed to have started by the financial sector of the United States, but in these days, at least for us, Europeans, it has become a debt crisis as well. As public debt is the common foe, the role of those institutions that are meant to rate this debt by riskiness has become the topic of a lively debate.

Credit rating agencies are supposed to tell how good an investment the debt of a given economic agent is, whether it is a firm, a local government or a sovereign state. This activity is supposed to be increasing market efficiency, as it reduces the costs of acquiring information for investors. If markets do care about these judgements, they will affect the costs at which these agents can borrow. On the other hand, getting a worse rating, if markets react accordingly, could increase the cost of financing on the scale of billions of dollars for a sovereign debtor. No wonder these institutions, the privately owned credit rating agencies came into the focus of a heated debate that is taking place at both an academic and a political level. What if, one might ask, rating agencies fail to assess correctly the riskiness of a sovereign? What if they pursue their own goals instead of what they claim to do? Is it desirable that three privately owned, profit-maximizing firms dominate the market for credit ratings? Is it a good thing that regulators increasingly “outsource” their mandates by building rules for market players upon the credit quality assessments of Moody’s, Fitch and Standard & Poor’s? Political upheaval spurred people to come up with ideas like the one of the common European credit rating agency, but the problem itself also paved the way for academic research.

The reason for this is that the questions about the effect and the performance of credit rating agencies cannot be answered on a theoretical basis; *a priori* beliefs should be confronted with the body of data available. The thesis contributes to this field by addressing two particular issues. The first is the effect rating changes have on sovereign borrowing costs. The second issue is modelling the ratings themselves and what we can infer from these models. How did the agencies perform during the crisis? Did they fail to predict? Did they overreact?

Sovereign credit ratings first came to be studied in the mid-90s with the works of Cantor and Packer (1995, 1996). The intensity of research on the topic has a very distinct time pattern, regularly gaining momentum after periods of major economic and financial turmoil, when the reasons behind the given crisis are being analyzed. Cantor and Packer wrote their papers after the Mexican crisis; the next big wave came after the Asian collapse at the end of the 90s, while interest in the topic has been rising again in the light of the 2008 global downturn. The original 1996 Cantor-Packer paper looked at a cross section of sovereign borrowing costs which they tried to explain with a handful of macroeconomic variables, and by including changes in credit ratings in the equation they tried to estimate the effect they make on yields. They also estimated an equation for explaining which factors affect sovereign ratings. These two approaches became two strands of literature afterwards, which I am going to briefly look at now.

One way to look at credit ratings is to conduct event studies, which in this context means measuring how a variable that is related to sovereign risk behaves in a given time window around a change in rating (a review of the methodology can be found in Binder, 1998). The dependent variable is in most cases the difference between the yield of a riskless asset (usually a US bond) and the secondary market yield of a reference debt instrument of the given sovereign. Rating changes as explanatory variables are chosen either from the same country as the spreads (Larraín et al. 1997; Reisen and Maltzan 1998, 2002; Sy 2002; Hill and Faff, 2010), different countries to study spillover effects (Kaminsky and Schmukler 1999; Gande and Parsley 2005), or both (Kaminsky and Schmukler 2002). In some cases primary market yields were chosen to be modelled; one paper, for example models the ratings and the market entry decision jointly a la Heckman (Eichengreen and Mody, 1998). Sometimes stock price indices are the explained variables (Kaminsky and Schmukler, 1999; Hooper et al. 2008). The newer literature often applies CDS spreads instead of (or together with) interest rate spreads, as the CDS market is more active, and is supposed to react more sensitively to changes in investors' expectations. Ismailescu and Kazemi (2010) use percentage changes in the CDS price as dependent variables; Afonso, Furceri and Gomes (2012) use both CDS and bond spreads as dependent variables in their fixed effects regressions; Arezki, Candelon and Sy (2011) estimate VAR models for the CDS and stock market indices. There are more "exotic" variables to model: Kraussl creates an "index of speculative pressure", a weighted average of one countries' exchange rate, short term interest rate and its stock market index (Kraussl, 2005). A few papers also concern the changes in volatility of spreads (or stock returns) after credit events, modelling it jointly with spreads in a GARCH structure (Andritzky et al, 2007), or in separate equations (Hooper et al, 2008).

The other question researchers frequently look at following the tracks of the Cantor and Packer paper is that of the possible determinants of sovereign credit ratings, by trying to fit adequate statistical models for them. The difference is that *qualitatively* all the factors are known, as agencies make public what issues they care about (e.g. political stability, soundness of the financial system etc.), but they do not disclose their exact methodologies, so researches need to reverse engineer the underlying model.

Researchers implement various methodologies to model ratings. Ferri, Liu and Stiglitz estimate a random effects panel model with the Moody's rating transformed to a numerical scale as dependent variable (Ferri et al, 2003), while Mora does similarly with a fixed effects specification (Mora, 2006). Afonso compares three cross-sectional regressions of ratings on a set of explanatory variables (Afonso, 2003), with the relationship between the variables is modelled as linear, as logistic and as exponential. [1]

A more complicated way to model ratings became standard after the paper of Hu, Kiesel and Perraudin (2002). They estimated an ordered response model to analyze the determinants of credit ratings on a cross-section of data. These models generalize the standard probit or logit techniques to be able to handle ordinal scales of more than just two categories (see, for example Greene, 2011). Another instance for the use of these types of models is that of Bissoondoyal-Bheenick, who tried to estimate a separate model for every year in the interval between 1995 and 1999 to track the possible changes in the estimated parameters over time (Bissoondoyal-Bheenick, 2005)[2]. A better solution to the same problem was given by those who estimated the model in panel structure (Mora 2006, Afonso et al. 2009, 2010).

The thesis proceeds as follows. The second section describes the sources and the basic patterns of the data and the sample used, while also addresses the main issues that have to be dealt with regarding them. The third section contains the estimation strategies and results of the calculations. The fourth section concludes.

[1] The result is that different specifications do not improve the model significantly compared to the baseline linear model. However, he only looks at highly rated countries in the first place, so similar results for similar countries are not really surprising even for different specifications.

[2] Data turned out to be too few to estimate the model on a yearly basis and get significant coefficients. This (lack of) result mainly remained when the author re-coded the ratings into only 9 wide categories (AAA, AA, B etc.)

2 DESCRIPTION OF THE DATA

2.1 DATA SOURCES

For the purpose of the research I relied on publicly available data. I obtained the main macroeconomic time series from the International Financial Statistics database of the International Monetary Fund, with some exceptions: the data on government revenue and spending are from IMF's World Economic Outlook, while the external debt statistics come from the body of data collected at the World Bank.[3]

2.2 ISSUES WITH CREDIT RATINGS

The three main credit rating agencies that provide ratings for the majority of the sovereigns and whose opinions frequently make it to the front page of the news are Fitch Ratings, Moody's and Standard & Poor's, of which the latter two have a 40%-40% market share each (Hill, 2002), while Fitch has 15%. For sovereign debtors, agencies publish ratings for debt denominated in foreign and in domestic currency, and for long and short term debt as well. They also provide rating outlooks to see if sovereign debtors are likely to be downgraded or upgraded in the future by current tendencies, and credit watch warnings, that respond to sudden events that might change underlying tendencies. The researches (including mine) normally focus on the ratings on long term foreign currency debt, because agencies are more permissive in terms of local currency denominated debt (as countries seem to be more likely to default on foreign liabilities)[4], and the rating scale for local debt is much less refined, involving only 4-7 categories. Another convenient factor in the choice is that in terms of the

[3] IFS data obtained through IMF eLibrary: http://elibrary-data.imf.org/
WEO data is obtained through the WEO online database:
http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/download.aspx
The debt data comes from the Quarterly External Debt Statistics database at World Bank:
http://data.worldbank.org/data-catalog/quarterly-external-debt-statistics-ssds

[4] Exchange rate risk is also out of concern of foreign investors when foreign currency denominated debt is in question.

long term foreign debt, all three apply a very similar scale to rate sovereigns, with only slight differences in notation, and even less in theory (see Table 1 for the S&P scale).

Table 1: The rating categories for sovereign debt

Investment grade	Non-investment grade
AAA	BB+
AA+	BB
AA	BB-
AA-	B+
A+	B
A	B-
A-	CCC+
BBB+	CCC
BBB	CCC-
BBB-	CC/C/D

Note: Fitch applies less categories for the extremely speculative range (CCC+ and less), while Moody's uses different letters. Source: (Standard & Poor's, 2012).

Even if the use of this kind of data is agreed upon, there are some questions that still need to be addressed. The first is how to use them correctly. Credit ratings are measured on an ordinal scale. The only meaningful comparison one can make given the rating scale of CRAs is ordering them by default risk. An AA- is a riskier bond than an AA, but we cannot say anything about the difference in their level of riskiness (that would mean an interval scale), nor their ratio (meaning a ratio scale). It cannot be established a priori that an AA- is exactly that much riskier than an AA as a BB+ is riskier than a BBB-, and it is also counterintuitive, given that the latter difference is considered the difference between "investment grade" and "non investment grade" debt.

One has to model this feature of the data explicitly (by modelling ratings with ordered choice models, that are designed to involve an ordinal variable for outcome), or make simplifying assumptions, of which the most prevalent is that of converting the AAA to CCC scale into a 16 to 1 scale (or, for example, 21 to 1, depending on how differently the worst debtors are handled), and work with the numerical value (I use both approaches). This is done frequently with modelling effects of ratings, but also with modelling ratings themselves. Although the method is very appealing because of its simplicity, it pretends that ratings are measured on a ratio scale instead of ordinal, and also assumes that a credit rating notch represents a fixed credit risk increase. So riskiness is supposed to be a linear function of the ratings, defined over the range from "completely safe" to "in default".

This assumption is on one hand quite implausible, but it makes the life of the researcher a lot easier, so many studies make it nevertheless, which is, of course, understandable (see Larraín et al. 97, Kamin and von Kleist 1999, Kaminsky and Schmukler 2002, Gande and Parsley 2005, Sy 2002 **etc.**). There are steps to alleviate the problem by manually breaking the linearity. Kraussl uses a modified linear scale where the distance between the lowest investment grade rating and the highest non investment grade rating is 3 notches instead of one (Kraussl, 2005). Afonso tries to fit exponential and logistic curves to the ratings (Afonso, 2003). However, empirical results from my research in Section 3 suggest that this is not such a big issue, as scales that endogeneously emerge from ordered response models are of a rather linear fashion.

The second important issue is to decide which data source to use, as even if one only deals with the Big Three, there are up to three potentially different credit ratings for any given country. Luckily, as I have mentioned above, the rating scales are rather similar in nature, so comparability of them means no additional concern. Even then, there are different strategies to follow, and each of them implies making some additional assumptions about the data.[5]

The most obvious method is to pick one of the three, and not to be concerned with the other two. Quite a few points can be made against this strategy. First, it basically cuts the potential sample size of rating events into third, endangering the statistical significance of

[5] This problem spurred another direction of research, namely the study of interactions between ratings decisions, which I will now by-pass (e.g. Hill et al., 2010).

estimations. Second, it neglects the effects of agreements and disagreements between agencies, which cannot be ruled out *a priori*. If ratings mean useful information for market participants, it is arguable that when the three main agencies put a sovereign into three different categories the markets will charge a higher price for debt financing given this uncertainty of knowledge, compared to another sovereign whose "average rating" is similar, but CRA-s can come to accord about it. An effect of the opposite sign is also plausible. It might be the case that some investors make decisions contingent on the joint move of credit rating agencies. For example, it is common knowledge that some institutional investors have to sell an asset when, for example, two out of three big rating agencies cast it out of the investment grade category. This feature of the data is completely missed by the first estimation strategy. Third, the choice of rating agency is necessarily ad hoc, even if it is based on some statistical evidence from the past.

A second path to follow is to estimate different models for the three agencies. If it is the rating itself that is being modelled, this creates an opportunity to compare agencies by the emphases they put on one factor against the other. It might be the case that in the judgement of Agency 1 the current account balance is much more important than it is for Agency 2 (and thus has a bigger coefficient in the estimated regression), while Agency 2 is more keen on the terms of trade of a given country. On the other hand, studying bond yields equations might reveal that the market cares more about (i.e. reacts more harshly to) events by one agency than the other. Of course the researcher who opts for this method still faces the second problem mentioned above.

Throughout this thesis I am using data from Standard and Poor's and Fitch Ratings, for two reasons. The practical reason is that the possibility to extend the research to cover Moody's is seriously impeded by the availability of their data. The more theoretical reason is that among the Big Three, S&P and Moody's decisions are much more similar, and the smaller Fitch is mostly invoked as a “third opinion” if they disagree (Hill, 2004). So by looking at S&P and Fitch, I almost cover the whole spectrum.

If one does not estimate different models, pooling all the rating events seems to be a better idea for studying their effects on sovereign yield spreads. In this case, sample size is as big as it gets, while agreements and disagreements between agencies can be described by dummies, interaction terms, etc. Also, not discriminating between data sources *a priori* naturally boosts the credibility of any result. This method, however, cannot be used to model ratings as outcomes. As ratings decisions are made separately at the separate agency, there is no sense in estimating a common model for the whole sample of rating events.

Perhaps the least common method is to construct a variable of “average ratings”, that involves first converting the ordinal ratings into numbers, then taking their averages across agencies and then using the resulting data as their “common opinion” (e.g. Cantor and Packer, 1996; Afonso et al, 2012, Mora 2006). This method tries to avoid losing information in the sense that all ratings data will be used in the calculations, but will not use information on CRA-s agreeing or disagreeing over the credit quality of a sovereign. In fact disagreements can be arbitrarily large as long as they average out each other, so if a country is rated 10 at all three agencies, the market is supposed to react just the same as if was rated 6,10,14, even if this means that one agency ranks the country as superior quality, while the other considers it near to default.

As I already mentioned above, actual upgrades and downgrades constitute only a part of all rating events. All three agencies have standard procedures for reviewing their ratings, and in the meantime they are signalling what their future steps will be, should the course of events remain on the current path for the country. If credit ratings indeed drive market yields, it is plausible that credit watches and rating outlooks also do, so they should also be modelled somehow. There are four ways to do this. One possible strategy is to drop these signals altogether, which is of course a considerable loss of information, but is very simple and easy to support with the argument that at the end of the day, it is downgrades that make it to the front page of the newspaper, not outlook changes. Outlooks can also be used to “refine” the original scale, so a positive outlook would, for example mean an additional 0.3 notch upgrade (as in the paper of Cantor and Packer). Of course this decision is quite arbitrary. The third solution is to use a set of dummies that are accounting for these events. This makes it possible to model the effect of a rating change conditional on being anticipated or not, calling an event “anticipated” if it is in line with previous outlooks (as in Kraussl, 2005)**.** This third approach is accepted in this research. Some papers model outlooks as wholly separate events, but that misses any possible relationship between rating events and outlook changes (Arezki et al, 2011; Kaminsky and Schmukler, 1999).

2.3 THE SAMPLE

It is a crucial issue how to choose the set of countries to look at, and the existing literature shows huge differences in terms of sorting procedures. Most papers focus on 1) a mixed sample of emerging and developed nations; 2) emerging markets only; or 3) countries selected on some specific criteria. The first choice is justified by the intention to look at as many countries as possible. The rationale for the second is that most developed countries are rated as high quality debtors, so most variation in credit quality will be found in emerging markets. But some authors have other concerns than those. There are studies that use a geographic criterion (only Asia: Kaminsky-Schmukler 1999, only the European Union: Afonso et al, 2012), others leave out low quality debtors intentionally (Afonso 2003).

I decided to choose a sample that is as wide as possible and covers a very wide range of countries by credit ranking, and managed to obtain a sample of 40-50 countries, depending on model specification (the "core" of the sample is presented in Table 2). Of course, I was heavily constrained by data availability, so selection will be an issue, because apparently the quantity and quality of the data a country publishes are positively correlated with its level of development, which in turn is correlated with credit ranking, so better rated countries are overrepresented. However, this selection issue is alleviated by the fact that the least developed countries are not really affected by the activities of credit rating agencies, because they are not very actively present on the international sovereign debt market. Also, the calculations carried out in Section 3 suggest that the fact that better debtors are relatively better represented in the sample does not bias the calculations.

Table 2: countries in the sample

Country group 1		**Country group 2**	**Country group 6**
United States	Greece	Brazil	Morocco
United Kingdom	Iceland	Chile	**Country group 9**
Austria	Malta	Colombia	Bulgaria
Belgium	Portugal	Mexico	Russian Federation
Denmark	Spain	Peru	Ukraine
France	Turkey	Uruguay	Czech Republic
Germany	Australia	**Country group 4**	Slovak Republic
Italy	New Zealand	Cyprus	Estonia
Luxembourg	South Africa	Israel	Latvia
Netherlands		**Country group 5**	Hungary
Norway		Hong Kong	Lithuania
Sweden		Korea, Republic of	Croatia
Switzerland		Malaysia	Macedonia, FYR
Canada	.	Philippines	Poland
Japan	.	Singapore	Romania
Finland	.	Thailand	Slovenia

Note: the list applies for the countries appearing in a typical regression (fixed effects for Fitch ratings). Country groups are formed by the first digit of the three digit country codes of the IMF.

Figures 1 and 2 show how ratings and income are distributed over the sample, which includes most industrially developed countries, many post-socialist countries, a fair amount of Latin-American countries, and also countries from the Middle-East, Africa and South-East Asia. Figure 3 shows that rating events do not concentrate on one particular part of the rating scale (e.g. it is not worse rated countries that get re-rated all the time).

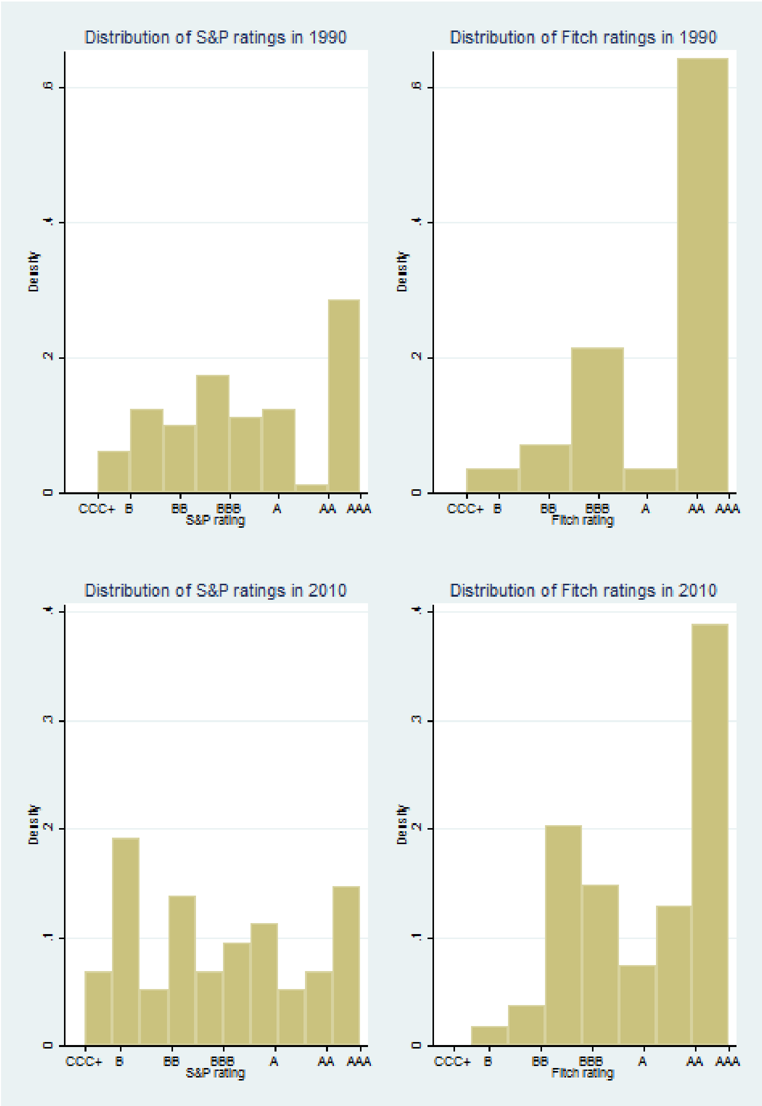

Figure 1: Distribution of ratings (source: own sample of data)

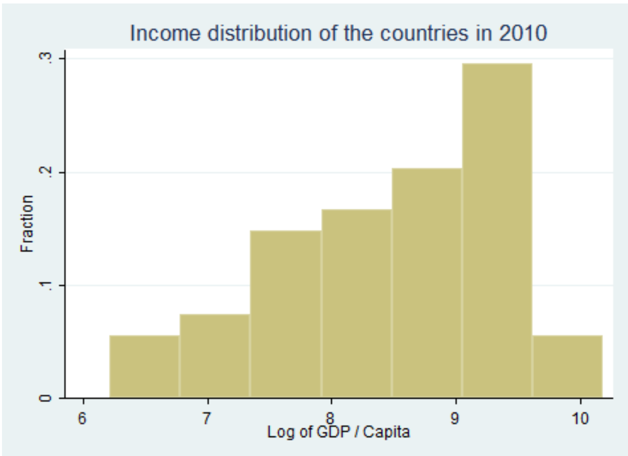

Figure 2: Income distribution of countries in 2010 (source: own sample of data)

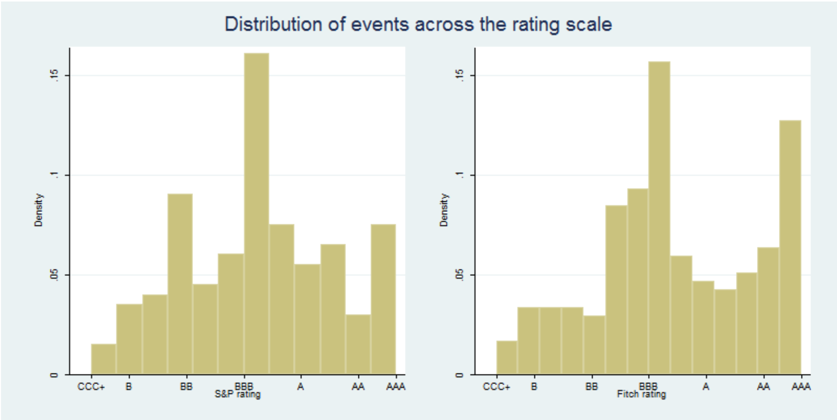

Figure 3: Distribution of events across the ratings scale (source: own sample)

I use quarterly data from the early nineties[6] to the 4th quarter to 2011. Papers that study how ratings are assigned usually go with a yearly database, while those that study the effects of rating events on interest rate spreads prefer daily ones. Both approaches have their merits. A yearly database for this goal is convenient, because some variables that CRA-s care about are indeed coming on a yearly basis; in this case, seasonality is not an issue either, and data from more countries is available.

Most event study papers look at the effect of ratings on a database of daily frequency in a time window ranging from 10 to 30 days, and the effect they find within this time window is on the scale of tenth of percentage points in the case of yield spreads. Daily data for event studies is good because the more frequent the data is, the more observations one has, and both variables of interest (spreads and ratings) are indeed available on a daily basis. The drawback is that this time window is too short to give a guess on how real a pressure they create on financing the budgets of governments. Also, authors cannot really include macroeconomic fundamentals as explanatory variables (day are not available on a daily

[6] There are slight differences in the exact starting date of the time series

basis), so one cannot really tell apart the effect of changing macroeconomic circumstances and the genuine effect of changes in credit ratings, should there be any.[7]

Quarterly data seem to be a sensible middle ground in the sense that it is possible to obtain a sample of reasonable size, timing of the events can be traced much more accurately than with yearly data, while it is also available to control for most of the macroeconomic variables of interest. Also, it is reasonable to assume that a quarterly average of spreads is much more relevant from the point of view of financing the budget of a given state than the changes that occur in a 10 to 30 days time window that is customary for event studies.

[7] One notable exception is the paper of Andritzky et al. (2007). Their explanatory variables are rating changes and macroeconomic announcements. They find that the latter tend to decrease, the former tend to increase market volatility, and the market puts less emphasis on ratings in crises than usual.

3 ESTIMATION OF THE MODELS

3.1 THE EFFECTS OF RATING CHANGES ON SOVEREIGN INTEREST RATE SPREADS

3.1.1. THE DATA GENERATING PROCESS

To assess the effect credit agencies can make on the market for sovereign debt, one has to look at the whole process in which it is determined why countries face different costs when they are borrowing resources from the market. The reason, of course, is the different magnitude of perceived risk. If sovereigns always repaid their debts, the need for credit rating and risk assessment would not have emerged, and each sovereign debtor would face a rather similar interest rate at which it can borrow. But they do default, and when they do, the debt contract embodied in the bond is not really coercible, and the possibility to at least partly regain the money in question is much more limited than in the case of a firm (countries' assets are very rarely liquidated through a legal procedure). So when a risk-neutral representative investor makes a decision about the price of his funds, he or she faces a lottery:

$$Z = (1 - p) \times r \times 1 + p \times 0, \tag{1}$$

where *Z* is the amount of future money the investor wants in return for his present day one dollar lent to the sovereign; *p* is the probability that the debt is never serviced (meaning that 0 default payoff for simplicity, though there is regularly a positive recovery rate); and *r* is the interest rate that is requested from the sovereign. Should one express *r* from the equation, it will be visible that *r* is increasing in default risk *p*, and also in *Z*. So the factors determining *r* are those that determine the possibility that a sovereign will not repay its debt (which are in turn determined by idiosyncratic characteristics of the country), and those that determine the overall "appetite" of investors for returns (these can be driven by time preferences, or by the global business cycle; I will only assume that they are different than the previous factors). In this case the cross-sectional differences in borrowing costs emerge from cross-sectional differences between perceived default risks.

But what drives the risk of defaults? One could imagine basically two general reasons for a government not to service its debt: lack of capability or lack of willingness to do so. If the debt levels are already very high, or the government is running big deficits, the risk of not being able to repay it will be higher, while the willingness factor comes in when the government has to make a decision between, for example, laying off a great number of state employees to cut costs, or defaulting on foreign debts. So in the end, the factors affecting sovereign default risk are boiled down to the macroeconomic fundamentals with other social and institutional characteristics of a country. Moreover, a rational investor would also take into account the information he or she has about how these circumstances are expected to change in the future.

Credit rating agencies come into the picture as they compress this plethora of possibly useful information into a very simple credit rating that has a meaningful interpretation and it is comparable across countries and through time. One could think of many potential stories how these ratings are useful for the investor: those who make market decisions find gathering information costly, so they prefer to use a publicly available information set as it is processed and evaluated by experts on the field, like credit rating agencies; it is also possible that information from credit rating agencies is deemed more reliable than the gossip and advice circulating on the market, as they are specialized on the field and can make judgments about the validity of such pieces of information. It might be the case that market participants care about credit ratings because they have to – they are obliged by external regulations (e.g. the capital adequacy requirements of banks; pension funds regulations)[8] or their own investment

[8] The historical trend of "outsourcing" regulatory authority to credit rating agencies by explicitly including credit ratings of the big agencies into financial laws and regulations is explored by Kruck (2011).

policies (e.g. a decision not to hold non investment grade assets). In any case, there are many possible explanations for credit rating agencies to have an important responsibility in the determination of sovereign borrowing costs. And where there is responsibility, there is always a possibility of failure to live up to it, and to cause trouble by doing so.

The most prominent argument of the literature that is blaming credit rating agencies for financial turmoil is that credit rating agencies make financial crises worse by working in a "procyclical" fashion by being unaware of increasing risks in boom periods and then downgrading sovereigns too late and too excessively after the bust arrives (stated, for example, in Ferri et al. 2003 for the case of the Asian crisis of the late 90s).[9] This is, however, is not one statement but two, an explicitly stated and an implicitly assumed one, namely that the market "cares" about ratings decisions and they are eventually transformed into risk premia. To evaluate this underlying statement, let's assume the following underlying data generating process for the determination of sovereign risk spreads):

$$spread_{it} = \alpha + \beta * Y_{it} + f(E_t[Y_{it+1}], E_t[Y_{it+2}] \dots) + \delta * rating_{it}, \qquad (2)$$

hence, the spread over the riskless asset is determined by the vector of observable macrovariables, an unobservable function f of the current period expectations of the future values of the same vector, and also a function of the sovereign ratings. These are both in the equation to account for default risk, based on the argument provided above. The question is whether there is a nonzero δ; this will determine whether markets care about the decisions made by rating agencies, or their co-movement is due to the fact that ratings change along with the changes in current and future values of Y. This feature of ratings, that they are also a (different) function of Y and its expectations is captured by the following econometric specification:

[9] The critics now project their arguments back in time to explain earlier financial crisis scenarios as well, for example the interwar foreign debt crisis (Flandreau et al., 2011)

$$rating_{it} = g(Y_{it}, \dots, \overline{E}_t[Y_{it+1}], \overline{E}_t[Y_{it+2}] \dots), \quad (3)$$

where the bars refer to the fact that these are the expectations of a different agent.[10] Most of the literature simply looks at the "effect of rating changes on spreads" by estimating a regression of the spreads on the ratings (or dummy variables representing changes in ratings), and interpreting the coefficient as the effect in question. So they estimate something like this:

$$spread_{it} = \hat{\alpha}_i + \hat{\tau}_t + \hat{\delta} * rating_{it} + u_{it} \quad (4)$$

However, in this case the error term will contain everything that is left out of the true model (macrovariables and its expectations), so it will obviously be correlated with the dependent variable, resulting in an omitted variable bias in $\hat{\delta}$. The bias can be reduced by including Y_{it}, the publicly available and widely observable information at time *t* into the regression, but this will only partly eliminate the bias, as the unobserved expectations will still be correlated with ratings. However, even if the estimate for $\hat{\delta}$ is biased, it will be informative how it changes in terms of size and significance. In the "true model" we can be quite sure that *f* is not the zero mapping (e.g. market participants do have their own expectations on the future), but we cannot be sure on theoretical grounds that δ also is. In the case of an extended, but still biased regression:

$$spread_{it} = \hat{\alpha}_i + \hat{\tau}_t + \hat{\beta} * Y_{it} + \hat{\delta} * rating_{it} + u_{it} \quad (5)$$

Now if $\hat{\delta}$ is small enough, one can argue that it can be divided into a part that is sure to be different from zero (due to expectations) and the part that is not so sure (the autonomous effect of ratings, the "true" δ), so the latter is probably minuscule or better yet it is not there at all. If the coefficient on ratings is rendered insignificant, it can be argued that the observable

[10] This theoretical relationship is further elaborated in the next sub-section where the determination of sovereign ratings is investigated.

variables in use proxy well enough for the expectations themselves, and establish that ratings do not move markets.

3.1.2 *Estimation of the basic models*

Now I present the estimation of the models above. To obtain spreads I take the reference treasury bill and government bond yields from each quarter as reported in the International Financial Statistics database, and subtract from it the reference US treasury bill or government bond yield for the same quarter. The vector of control variables in Y are the same that I use in the next subsection for explaining credit ratings, assuming that the market and the credit rating agencies look at the same factors in assessing country risk (the description of the variables is available there). I choose to estimate a fixed effects model with robust standard errors, because cross sectional fixed effects account for idiosyncratic differences between countries that are driving yields but are stable over time, while time fixed effects account for global trends in "risk appetite" that are driven by the same factors for all country.

The baseline model (that does not include macroeconomic controls, and is presented in Table 3) shows a highly significant *ceteris paribus* increase (decrease) of 0.6-0.7% percentage points in government bond yields after a downgrade (upgrade) episode. The effect for treasury bills is more pronounced for Fitch (0.8%), but an almost insignificant 0.6% for S&P. Including controls (can be found in Table 4) dampens the magnitude of effects (0.5-0.7%) for bond yields, while the coefficients in the treasury bill spread equations are essentially cut in half. For the purpose of comparison, in this case I also ran an "empty" regression, where only the control variables were present, not the credit ratings. I found that the inclusion of ratings increased the goodness of fit of regressions only slightly (8-12% gain in R^2 for government bonds, 1-2% for treasury bills).

Table 3: Regression of spreads on ratings, without country specific macroeconomic controls

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
rating_{it}	-0.764***	-0.802**	-0.647***	-0.634*
	(0.125)	(0.369)	(0.139)	(0.355)
Constant	16.08***	17.78***	13.43***	15.87**
	(2.242)	(6.163)	(2.531)	(6.122)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Macro controls	No	No	No	No
Observations	2,109	1,635	2,129	1,770
R-squared	0.567	0.306	0.491	0.304
Number of countries	44	39	42	39

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1;

Table 4: Regression of spreads on ratings, with country specific macroeconomic controls

	Fitch		S&P		Empty model	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	gov. bond.	tr. bill	gov. bond.	tr. bill	gov. bond.	tr. bill
$\text{rating}_{\text{it}}$	-0.727***	-0.490***	-0.570***	-0.341*		
	(0.137)	(0.172)	(0.175)	(0.195)		
Constant	15.61**	11.89	18.19**	9.143	15.99*	12.90
	(7.433)	(10.54)	(7.996)	(11.06)	(8.642)	(10.73)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Macro controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,109	1,635	2,129	1,770	2,295	1,811
R-squared	0.625	0.678	0.582	0.677	0.509	0.660
Number of countries	44	39	42	39	44	40

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; coefficients on control variables can be found in Appendix 1 (Table A.1)

I also estimate the regressions with a richer lag structure, looking at the effect of ratings 1, 2 and 3 quarters ago:

$$spread_{it} = \hat{\alpha}_i + \hat{\tau}_t + \hat{\beta} * Y_{it} + \hat{\delta}_1 * rating_{it} + \hat{\delta}_2 * rating_{it-1} + \hat{\delta}_3 * rating_{it-2} + \hat{\delta}_4 * rating_{it-3} + u_{it} \quad (6)$$

The results in Table 5 for government bond yields reveal an interesting pattern: the sign for the effect of ratings is negative in the first two periods, but changes its sign for the next two quarters. That is, the contemporary effect of one notch downgrade by Fitch is a 0.9% increase now, a 0.3% increase next quarter, followed by a 0.2% and a 0.5% decrease in the third and firth quarter. The pattern is similar for S&P. This means that the effect of rating changes is at least partly reverted, so the average effect for a whole year is smaller, around 0.5% percentage points. In case of the treasury bills, the signs of the coefficients are similar, but the significance levels of the results are much weaker, and only the contemporary values are significant.

Table 5: Regression of spreads on ratings with richer lag structure

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
rating_{it}	-0.885***	-0.756**	-0.758***	-0.530*
	(0.0885)	(0.328)	(0.121)	(0.309)
rating_{it-1}	-0.317***	-0.0643	-0.271**	-0.160
	(0.117)	(0.185)	(0.102)	(0.108)
rating_{it-2}	0.189**	0.239	0.115	0.170
	(0.0931)	(0.278)	(0.111)	(0.206)
rating_{it-3}	0.456***	0.210	0.479***	0.220
	(0.0986)	(0.178)	(0.127)	(0.207)
Constant	13.89**	14.04	16.76**	10.42
	(6.785)	(10.42)	(7.089)	(11.20)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Macro controls	Yes	Yes	Yes	Yes
Observations	2,056	1,595	2,126	1,767
R-squared	0.658	0.681	0.601	0.680
Number of countries	44	39	42	39

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; coefficients on control variables can be found in Appendix 1 (Table A.2)

3.1.3 Event dummy regressions

I also model the effect of ratings in a more complex way, by looking at a set of dummies that describe rating events more comprehensively. Now there is a dummy for upgrades, and another for downgrades, that take the value of 1 in the quarter when such an event occurs, and 0 otherwise. This should capture any asymmetry between reactions to downgrades and upgrades. The dummy *catchup* takes the value of 1 in the quarter when by the effect of an event at one agency, the risk judgement of S&P and Fitch on a given country becomes identical. The variable *decouple* in turn takes the value of 1 if an event results in that the previously identical ratings become different for the two agencies. The rationale behind including these dummies is that markets might see disagreement between two rating agencies as a source of uncertainty, and thus they might penalize disagreement by higher yields. These two variables are also interacted with the change in the level of ratings (0 if there is no change, 1 for a one notch upgrade, –1 for a one notch downgrade), allowing for different downgrade/upgrade effects in these situations. Another dummy (*inv*) tells whether the country is in the investment grade category at the given credit rating agency. It might be the case that some investors have such investment policies that they have to get rid of an asset if it dips below the investment grade barrier. As selling these assets might be time consuming in order to avoid losses resulting from a fire sale, I also include a one period lag for the same variable. The last dummy (*double_inv)* tells whether the country is investment level debtor at both agencies or not, also included for concerns of investors' policies that are pegged to the investment grade status of the asset.

Table 6: Event dummy regressions

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
upgrade_t	-0.409*	-0.519*	-0.267	-0.115
	(0.238)	(0.277)	(0.165)	(0.363)
downgrade_t	1.610***	1.200	1.269***	0.765
	(0.454)	(0.737)	(0.406)	(0.597)
$\text{rating}_t * \text{catchup}_t$	0.0960	-0.482	0.345	-0.0692
	(0.282)	(0.525)	(0.358)	(0.492)
$\text{rating}_t * \text{decouple}_t$	0.536	0.469	0.202	0.0797
	(0.389)	(0.487)	(0.322)	(0.273)
$\text{rating}_t * \text{inv}_t$	-0.876	-1.214	-1.679	-2.817***
	(0.640)	(0.779)	(1.359)	(1.033)
$\text{rating}_{t-1} * \text{inv}_{t-1}$	0.218	-1.750***	0.000800	-0.315
	(0.568)	(0.508)	(1.015)	(1.499)
catchup_t	0.149	0.495	0.183	0.275
	(0.162)	(0.384)	(0.176)	(0.358)
decouple_t	-0.0595	-0.125	-0.0654	-0.181
	(0.251)	(0.401)	(0.153)	(0.418)
double_inv_t	-1.998**	-0.497	-1.263	0.550
	(0.786)	(0.695)	(1.117)	(0.857)
double_inv_{t-1}	-0.0490	1.581*	0.167	0.503
	(0.622)	(0.919)	(0.918)	(0.733)
Constant	18.38**	13.60	19.96**	11.36
	(7.873)	(10.71)	(7.628)	(11.23)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Macro controls	Yes	Yes	Yes	Yes
Observations	2,091	1,622	2,128	1,769
R-squared	0.583	0.680	0.575	0.682
Number of countries	44	39	42	39

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; coefficients on control variables can be found in Appendix 1 (Table A.3)

The results in Table 6 reveal considerable asymmetry in the effects of downgrades and upgrades, the prior being 4 times as big in magnitude (1.6-1.2%), and more significant in the case of government bonds, while upgrades are hardly significant, which is peculiar, since they constitute the bigger part of the events in the sample (see later in Table 7). Catching up and decoupling of ratings by different agencies seem not to effect yields. The investment grade dummy has a strongly significant and big effect (1.8%-2.8%) for treasury bills, while there is mixed evidence for the importance of the dummy that indicates being graded for investment at both agencies (2% effect in the Fitch equation for government bonds, but no significant results elsewhere).

The evidence so far does not provide enough evidence to reject the hypothesis that ratings drive spreads to some extent. On the other hand, the values that I have inferred for effects are to be treated as upper bounds in absolute value, because the omitted variable bias could not have been completely eliminated, and for an upper bound, the estimated effects are not at all huge. The bias is sure to be there because if there is any knowledge that is relevant in terms of country risk in the intersection of the information sets of credit rating agencies and the market that is not captured by current values of macroeconomic fundamentals, it will be attributed to the rating agency by the model, if it acts upon it and changes the rating.

3.1.4 ANTICIPATED AND UNANTICIPATED RATING EVENTS

However, there remains another way to obtain indirect evidence on these phenomena. Assume now that the market players work as such that they "use" ratings information through two channels: they have a policy for asking more money from worse graded debtors (through δ), but they also use any other information from the rating agencies to formulate their own expectations on the future values of the observables (these information directly enter f in the true model). The trick is that agencies not only give ratings, they give hints about how they would change ratings given the current conditions persist, these are credit watch procedures on the short run, and ratings outlooks on a longer run. Given that I am dealing with quarterly data, I consider only the latter. Hill et al. (2010) argue that outlook procedures are strong predictors of rating changes relative to other public data, so we can use them as proxy for current state of expectations before the rating change comes out.

Suppose that the outlook is positive, and an upgrade comes. If the model is correct, the market maintains its current expectations, and adjusts the spread according to the "policy-implied" value (δ). On the other hand, if the outlook is positive and stable, and a downgrade occurs, the market in this case does not just make the policy-implied move, but also adjusts its expectations accordingly, so the change in spreads will be bigger in absolute value, as it will be the sum of the change in $\delta * ratings$ and the change in *f*. So an "anticipated" step will be associated with a slighter change that is close to the true value of δ (that is here dissected into four different variables), while an "unanticipated" step will be followed by a relatively big change in spreads. I estimate the following model:

$$spread_{it} =$$
$$\alpha_i + \tau_t + \beta * downgrade\_anticipated_{ti} + \gamma * downgrade\_unanticipated_{ti} + \zeta *$$
$$upgrade\_anticpated_{ti} + \eta * upgrade\_unanticipated_{ti} + \theta * Y_{it} + u_{it}, \qquad (7)$$

where downgrade (upgrade) stands for a dummy that takes the value of 1 if the sovereign is downgraded (upgraded) in the given period and 0 otherwise, while anticipated is 1 if the credit outlook in the previous period was negative (positive), and it is zero if it was positive or neutral (negative or neutral). The variables in the model are the full set of interactions between these two variables.

I also estimate a model where "semi-unanticipated" moves are treated as distinct events. I call an event a "semi-unanticipated" downgrade (upgrade), when the rating event follows a neutral, or stable credit outlook. Of course, sample size is a big issue here (as it is seen in Table 7), especially for unanticipated downgrades, which are quite rare. But they are also supposed to be very negative outlier episodes, which arguably move spreads upwards, so if there is any "true" effect, it should be smaller in magnitude and less significant than what we see in the results.

For the purpose of the calculations anticipation is defined as outlook status in the previous month, that is, if credit outlook in June is neutral, and in July an upgrade takes place, it will be considered as an unanticipated upgrade in July in the first model, and a "semi-unanticipated" one in the second.

Table 7: Events in the given categories

N	Fitch upgrade	Fitch downgrade	S&P upgrade	S&P downgrade
Anticipated	49	40	71	42
Unanticipated	0	1	2	1
Neutral (semi-unanticipated)	39	5	33	2

Note: the rows 2 and 3 are taken as the same event in the 2nd equation.

The results (presented in Table 8) do not support the surprise-hypothesis; in fact, anticipated moves are the only ones to have significant coefficients in any setting, and even the insignificant numbers do not show any common pattern that would hint that surprise was an issue. In fact, unanticipated (or neutral) upgrades, which are the second most numerous events, seem to increase spreads (though this is not significant either), while the strongly significant and spread-increasing anticipated downgrades are paired with unanticipated ones of insignificant and close to zero coefficients for government bonds, and ones lacking a common pattern for treasury bills. This evidence is, of course, indirect and rather weak (in many cases, events categorized as unanticipated in the second model were so few that calculations for the coefficients could not be provided). However, the argument from the previous paragraph should be considered here. Even if the events when moves are somewhat unanticipated are quite rare, if CRAs did drive market prices for debt, these episodes, one could argue, would be quite "shocking" events, resulting in outliers in the data that would produce big and significant coefficients. This feature is completely missing from the data.

As a side note, these results are also inconsistent with the approach in the literature (as proposed by Cantor and Packer) to model outlooks as partial rating changes (e.g. a positive outlook is modelled as +0.5 rating), because in this case an anticipated change is a single-notch rating change, while an unanticipated is a change from 1.5 to 2 notches, so it should have a proportionally bigger coefficient as well.

Table 8: Regression on anticipations

	2 categories				3 categories			
	Fitch		S&P		Fitch		S&P	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	gov. bond.	tr. bill	gov. bond.	tr. bill	gov. bond.	tr. bill	gov. bond.	tr. bill
ant. upgrade	-0.559***	-0.575*	-0.305**	0.0677	0.0678	-0.575*	-0.305**	-0.559***
	-0.188	-0.316	-0.143	-0.303	-0.304	-0.316	-0.143	-0.188
semi-ant. upgrade					-0.289	0.167	0.375	0.0938
					-0.655	-0.346	-0.321	-0.129
ua. upgrade	0.0937	0.167	0.375	-0.288	-0.271	0	0	0
	-0.129	-0.346	-0.321	-0.638	-1.037	0	0	0
ant. dgrade	1.598***	0.978*	1.596***	1.202*	1.202*	0.978*	1.596***	1.598***
	-0.526	-0.523	-0.419	-0.631	-0.631	-0.523	-0.419	-0.526
semi-ant. dgrade					-0.026	4.671	0.0249	0.18
					-0.515	-2.888	-0.198	-0.754
ua dgrade	0.0249	-0.026	0.15	4.671	0	0.0302	0	0
	-0.198	-0.515	-0.632	-2.888	0	-0.271	0	0
Constant	16.93**	10.29	17.45*	9.536	16.93**	17.45*	10.29	9.536
	-8.374	-11.1	-8.831	-10.99	-8.374	-8.833	-11.11	-10.99
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macro controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,278	1,772	2,278	1,772	2,278	2,278	1,772	1,772
R-squared	0.526	0.671	0.526	0.677	0.526	0.526	0.671	0.677
Number of countries	44	39	44	39	44	44	39	39

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; coefficients on control variables can be found in Appendix 1 (Tables A.4-A.5)

3.2 MODELLING SOVEREIGN CREDIT RATINGS

Up to this point I have studied the possible relationship between sovereign ratings and spreads on government bonds and treasury bills; the following section I am investigating into the process through which these ratings are assigned. The study of rating agency behaviour sheds light to the extent ratings are in line with actual available economic data; based on such modelling, it can be assessed whether rating events observed a different pattern in a specific time given their general behaviour (in times of crises), or if their actions diverged from an otherwise favourable course of actions and as such they can be judged as erroneous.

In order to decide over these dilemmas, there are some preliminary questions that need to be answered. How can one imagine the underlying process of the determination of ratings? How can this be modeled econometrically, what is the correct specification? What are the empirical caveats of the given specifications? What explanatory variables should one consider?

Once I have the model and the right hand side variables, I can look at how the models fit; relate model implied ratings to actual ones; use the models to evaluate the behaviour of sovereign ratings as the crisis hit, and perform tests on the quality of agency performance.

3.2.1 THE DATA GENERATING PROCESS

The meaning of a sovereign credit rating is most clearly defined by one particular issuer: "*All references to sovereign ratings in this article pertain to a sovereign's ability and willingness to service financial obligations to nonofficial, in other words commercial, creditors*." (Standard & Poor's, 2012) To put it the other way around, credit ratings rank debtors in terms of the reverse likelihood of default. The information set they use is a set of publicly available data and expectations about how the data will change over time. The outcome variable takes a discrete value measured on an ordinal scale between the lowest rank of "CCC" for worst debtor (or "D" for "already in default"), and "AAA" (most secure debtor). These features suggest the following structure of the data generating process:

$$G_t = g(Y_{it}, \dots, \overline{E}_t[Y_{it+1}], \overline{E}_t[Y_{it+2}] \dots)$$

$$rating_{it} = \begin{cases} best\ debtor & if\ G_t > cutoff_{first} \\ & \dots \\ worst\ debtor & if\ G_t < cutoff_{last} \end{cases}. \tag{8}$$

That is, agencies translate the vector of observable variables and the expectations formed about their future values (bars indicate that they are not necessarily the same expectations that the market had in the previous section) into some evaluation variable *G*. This continuous latent variable is then compared to the rating scale, which has *n–1* cut-off values for *n* different possible rating values. The agency re-calculates *G* in every period, and if it finds out that the debtor dipped below or emerged above a neighbouring cut-off level, makes the according move and downgrades or upgrades the given debt issuer.

3.2.2 Selecting the proper econometric specification

To proceed, one has to choose the model structure to estimate. The simplest way is to assume away the nontrivial relationship between the underlying *G* variable and the actual ratings and estimate an equation with a number on a left hand side that corresponds to the rating (with the value of 1 given to the worst possible assignable value), and various right hand side variables. Cantor and Packer used the basic OLS structure, while in the paper of Ferri a random effects panel model is chosen (Cantor and Packer, 1996; Ferri et al. 2003). The consistency of the OLS estimates is, of course, doubtful in such a setting, as for the estimates to be unbiased it is required that no explanatory variable is left out of the regression that is correlated with the ones in the regression. While RE gives a more accurate guess for the coefficients, its assumptions are not less restrictive than those of the OLS, so if one doubts the consistency of the OLS results, RE is not any surer (see, for example, in Wooldridge 2001). Thus, for linear models I estimate a fixed effects model (as suggested by Mora, 2006 for the same use), which makes life easier, as cross-sectional fixed effects account for all the idiosyncratic factors that are not varying over time (at least not within the given time window), and there seem to be many such factors explaining sovereign credit ratings (institutional factors, for example).

There are two potential caveats with the linear panel models. First, if there is significant nonlinearity in the rating scale, the model will not capture it in any way. Second, as the rating scale has an absolute minimum and an absolute maximum category, the sample is censored from both up and below. It can cause a biased estimate for the coefficients of the explaining variables if one country cannot have a better rating than AAA even if there is a difference in credit quality between itself and the next best AAA country that would account for a full notch upgrade otherwise (the coefficients will be downward biased in absolute value for everyone).

The other more complicated way is to model explicitly the above structure of the data generating process by using an ordered probit model (Greene, 2011). Now to every possible ordinal outcome of the dependent variable, an interval of the latent, numerical *G* variable is assigned. The value of the latent variable for every data point is in turn determined as a function of the explanatory variables. The parameters of the function and the limiting points of the intervals are then estimated jointly. With this, I let rating scale to be non-linear: if, for example losing the investment grade rating is harder than losing an AAA rating, it will show up in the threshold number estimates for rating categories. Also, these models account for the censored nature of the data (i.e. one cannot be better a debtor than AAA and worse than C, so even if one deserved a better category, it will receive only an AAA, introducing a bias into the estimate).

The problem is that extending the ordered probit model to panel data is nontrivial, so I decided to follow two strategies. First I “imitate” a fixed effects setting by introducing dummies for cross sectional and time units. However, as this model already estimates much more parameters from the same set of data then does a simple OLS-based model, I could not use a dummy for every cross-sectional observation and time period. Instead I introduced one for every three-digit country group defined by the IMF. As this categorization has remarkable within-group homogeneity (group 1 are the most industrialized states, group 2 is Latin-America, group 4 is the Middle-East and North Africa, group 5 is South Asia, group 9 are the post-socialist states, etc), I only made one alteration, swapping the places of Turkey and South Africa (originally in group 1), with Israel and Cyprus (originally in group 4), as this change increased the internal homogeneity of the groups in terms of level of development (per capita income, for example). On the time dimension I used yearly dummies instead of quarterly dummies to reduce the number of estimated coefficients. While in the full model the convergence of the algorithm was compromised, the reduced one turned out to be estimable. As another approach, I used the random effects ordered probit panel estimator (as in Greene, 2011) which was designed to handle panel data in the first place.

3.2.3 EXPLANATORY VARIABLES

The variables that Cantor and Packer included in their paper were: per capita income, GDP growth, inflation, fiscal balance, external balance, external debt and two dummies (economic development and default history). They argue that a bigger per capita income means a bigger tax base for the government, so the debts are more likely to be repaid. Growing GDP alleviates the burden of existing debt, while inflation is a symptom of structural problems within fiscal policy. Fiscal deficits indicate that governments are unable to collect enough taxes or they are not willing to do it, while current account deficits mean that countries rely on external financing – both a negative factors in determining sovereign risks. A higher external debt burden obviously goes with a higher risk of default. Industrial countries are supposed to be more reliable debtors, while a default event in the recent history is probably associated with less confidence and a bigger risk that it will happen again.

The literature mostly agreed with this set of variables, adding only a few other factors, terms of trade, for example (Hilscher and Nosbusch, 2010), or a dummy for being a member in the Euro-zone (Afonso et al., 2010). Also concerns were raised, because some macroeconomic variables can be endogeneously determined within these models, thus the estimates on their coefficients should be suspicious (Gonzalez-Rozada and Yeyati, 2008). For example the lack of indebtedness can be a sign of borrowing constraints, not necessarily sound finances, while current account deficits can proxy for booming investment. This is why the Cantor-Packer paper and its successors could not establish a robust pattern between external and fiscal balances and the credit rating, even though it is very intuitive that a big and persistent fiscal or external deficit should be associated with an increased default risk.

Since the first papers covered these issues, credit rating agencies have became somewhat more transparent in the sense that now much more details are available about the rating assignment process (see, for example, Standard & Poor's 2012). My choice of variables mostly echoes that of the Cantor-Packer paper and subsequent literature, with some modifications implied by this more accurate recent and direct information on the rating process, with some modifications implied by the selection of the models. The variables I use throughout the essay are the following:

- *Per capita nominal GDP in USD*, natural logs, seasonally adjusted. A smaller subset of the countries reports already adjusted time series, for the remaining countries I performed the adjustment by generating the 4 quarters moving average time series of each country.
- *Yearly real GDP growth, %*
- *Reserves in dollars to GDP* (proxying for financial buffers for the economic policy)
- *Inflation*, measured as the yearly % change in the consumer price index
- *Unemployment rate*, which is not directly linked to the ability to pay outstanding debts, but might be to the willingness to do so. It is arguable that politicians would rather default on foreign debt to avoid, for example, streamlining state bureaucracy in a high unemployment setting, as it would further exacerbate existing tensions.

- *Net external liabilities/CAR*: This is an index defined in the S&P manual to judge the ability to service the debts, more accurately than the external debt measure solely. The numerator is defined as the total external debt of the country plus the stock of direct and portfolio investment from abroad, minus the value of external assets, while the denominator is the sum of the current account receipts.
- *FDI to GDP*: The ratio of foreign direct investment into the country and the gross domestic product, both in dollars; this measure proxies the long run attractiveness of the country.
- *Government income and expenditure to GDP, and their squared terms*. Even though fiscal balances are not useful directly to model sovereign ratings, through separating the revenue and the expenditure side it is possible to gain meaningful regressors for the equations. Squared terms are there, because if there is an "optimal" level for incomes and expenditures they might catch it. By optimal level I mean that too small a government revenue can be understood as an ineffective state that cannot collect enough taxes to finance itself, while too high a revenue level can be understood as a signal of the presence of excessive bureaucracy. If this is true, the expenditure to ratings profile should be increasing on the first section of the scale, and decreasing on the second.
- *Gross government debt to GDP*

Except for the last two bullet points, all the data is quarterly. Income and expenditure levels, debts are unavailable on a quarterly level for most countries, however, the S&P manual that describes the rating processes explicitly establishes that the agency looks at yesteryear's yearly statistic for each country, so I felt authorised to plug yearly data into the quarterly regression and disregard the methodological concerns this might raise under different circumstances. Throughout all regressions, the explanatory variables are taken lagged by one period (4 periods for yearly data) so that the threat simultaneity is reduced, and only that information appears in the regression that was in fact publicly available in the time when the ratings were determined.

There are variables that were present in previous literature and I omit here, and also (much more) that are written down in the S&P guidelines but are not presented here. There are three reasons why I omitted particular variables.

First, I omitted some variables because of the lack of data. Terms of trade, the literature suggests, is, for example, a good explaining variable of sovereign risks for emerging countries, and so is its volatility (Hilscher and Nosbusch, 2010). However, since there is so little data in the IFS database on the unit value of imports and exports, adding this explanatory variable would have essentially cut the sample size in half, and mostly industrialized countries remained afterwards, for which this factor was of a lesser concern.

Second, some variables for financial soundness were left out because they were very much correlated with the variables I included, so they would not have added much explanatory power to the regression, or would have even worsened it. S&P reports a sizeable list of variables of which I selected only NEL/CAR, because it embodies a lot of information in a single number.

Third, other variables were left out because the lack of time variation and high enough frequency. S&P explicitly lists measures like the Corruption Perception Index of Amnesty International, or the World Bank's quality of government indices, but these are only published on a yearly basis. However, by mostly relying on fixed effects estimations, cross sectional fixed effects are supposed to suck up these information, because they do not vary much over the course of one or two decade. They are much more informative over the cross sectional ranking of countries, but this will be translated into different constant parameters in our regressions. The same applies to all dummy variables that were present in the Cantor-Packer paper: in a fixed effects regression the "industrial country" dummy or the "default history" dummy cannot be identified separately. "Industrial countries" (as defined by Cantor-Packer) will have a *ceteris paribus* better rating, while previously defaulted countries will have a *ceteris paribus* worse rating than others throughout the whole time window of the sample.

3.2.4 ESTIMATION OF THE MODELS – FIXED EFFECTS

The estimations are presented in Tables 9-10. The fixed effects model (Table 9) provides a quite good fit, the significant coefficients are having the expected signs. A one point *ceteris paribus* increase in the log GDP is followed by assigning a 1.5-2 times better rating, which seems a lot, but it is understandable if we bear in mind that it would mean a *ceteris paribus e*-time increase in the per capita GDP. Real growth is also a factor, though with less importance. Other things equal, about 15-20% of inflation and an unemployment rate of 9-10% are associated with a one notch worse rating each; both coefficients are strongly significant. Reserves do not have a significant effect, which is curious at first glance, because other things equal, a bigger reserve is associated with a stronger financial stability. However, a bigger reserve can mean, for example, that the country asked for (and received) external financial help, which does not happen when finances are sound, so its weak explaining power is not that much of a surprise after all.

Two variables have coefficients of an unexpected sign: the net external liabilities to current account receipts ratio and the foreign direct investment to GDP ratio (of which the second is only marginally significant. The reasons are similar: in the sample the ability to run big current account deficits is probably connected to being a developed economy, while the amount of FDI compared to GDP is probably bigger for the poorer, worse debtor countries and so these coefficients do not reflect the "true" relationships between ratings and the given variables. The estimates for effects of government revenue and expenditure indeed support the hypothesis for a nonlinear relationship between them and the ratings, and the coefficients imply that 35-40% of GDP is considered as optimal by CRAs (at least they seem to give the biggest rating for such countries) for both revenue and expenditure. Also, the optimal levels for revenue are higher, indirectly implying that the ratings penalize fiscal deficits. The government debt level is also a significant explaining variable, though it is somewhat surprising how small its coefficient is, as it says that out of two otherwise identical countries, the one with 100% debt to GDP ratio is only one notch worse rated than the other with merely half of its own stock.

Table 9: Fixed effects model

VARIABLES	(1) Fitch	(2) S&P
$\log(\mathrm{GDP})_{t-1}$	1.934***	1.617***
	(0.447)	(0.514)
reserves to GDP_{t-1}	-8.867e+08	-1.455e+08
	(6.325e+08)	(4.149e+08)
real growth of GDP_{t-1}	0.0723***	0.0438**
	(0.0150)	(0.0165)
$\mathrm{inflation}_{t-1}$	-0.0670***	-0.0544***
	(0.0119)	(0.0115)
unemployment rate_{t-1}	-0.102***	-0.132***
	(0.0281)	(0.0306)
$\mathrm{NEL/CAR}_{t-1}$	0.328***	0.350***
	(0.0811)	(0.0903)
$\mathrm{FDI/GDP}_{t-1}$	-0.0651	-0.0662*
	(0.0443)	(0.0342)
gov. $\mathrm{revenue/GDP}_{t-4}$	0.397***	0.446***
	(0.0620)	(0.103)
$(\mathrm{gov.\ revenue/GDP}_{t-4})^2$	-0.00494***	-0.00569***
	(0.000815)	(0.00134)
gov. $\mathrm{expenditure/GDP}_{t-4}$	0.160**	0.0660
	(0.0738)	(0.0816)
$(\mathrm{gov.\ expenditure/GDP}_{t-4})^2$	-0.00218***	-0.00102
	(0.000788)	(0.000896)
gov. $\mathrm{grossdebt}_{t-4}$	-0.0206***	-0.0199***
	(0.00740)	(0.00596)
Constant	-6.387	-3.235
	(4.092)	(4.918)
Country fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
Observations	2,664	2,723
R-squared	0.604	0.539
Number of countries	56	55

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

At this point I also carried out two robustness checks, which account for overall selection in the sample, and also for the potential bias that could come from neglecting the censored nature of the data in the fixed effects regression. In the first case, the fixed effects model is estimated for all countries except the ones in country group 1 (most industrialized countries, with best data quality). This is a very severe restriction, and if it does not change the coefficients very much, it is arguable that sample selection is not an issue in general, and because most AAA rated countries are in this group, there is a strong hint that censoring also does not cause a bias. In the second case, I estimate the model dropping out all countries with an AAA rating, without regard to country groups. If coefficients remain the same, AAA countries are not that better than non-AAA countries to cause a bias through censoring. The lower censoring is probably not an issue, as there are not many countries in the sample accumulated at the lower end of the distribution. The results (Appendix 1, Table A.6) show that there is basically no difference between the baseline FE model and the second robustness check equation, and very little with the first one; some coefficients lose from their significance as a big part of the sample is left out of the estimation, so the overall precision decreases, but the quantitative pattern of coefficients does not change. So it can be established not just that censoring is not an issue with the FE model, but also the whole body of calculations is spared of serious sample selection that would impair the coefficient estimates.

3.2.5 Estimation of the models – ordered response models

The coefficients in the ordered probit model with region and year dummies (Table 10, columns 1-2) are remarkably similar; the notable exceptions are that the net external liabilities became insignificant, while the reserves did exactly the opposite. The effect of government debt became smaller, and the neat quadratic pattern of revenue and expenditure is not that pronounced anymore. Same goes for the random effects ordered probit model (Table 10, column 3-4), that is having otherwise very similar coefficients to the previous models.

Table 10: Ordered probit regressions

	Country & year dummies		Random effects	
	1	2	3	4
VARIABLES	Fitch	S&P	Fitch	S&P
$\log(\mathrm{GDP})_{t-1}$	1.755***	1.739***	2.110***	1.615***
	-0.0688	-0.0647	-0.0583	-0.0476
reserves to GDP_{t-1}	-1.202e+09***	-1.085e+09***	3.750e+08***	9.615e+08***
	-1.49E+08	-1.38E+08	-1.38E+08	-1.41E+08
real growth of GDP_{t-1}	0.0414***	0.0058	0.0688***	0.0451***
	-0.00889	-0.00901	-0.00644	-0.0054
$\mathrm{inflation}_{t-1}$	-0.126***	-0.116***	-0.0641***	-0.0643***
	-0.0103	-0.0102	-0.00766	-0.00686
unemployment rate_{t-1}	-0.00265	0.00066	-0.0767***	-0.181***
	-0.00464	-0.00448	-0.00538	-0.00593
$\mathrm{NEL/CAR}_{t-1}$	-0.079	-0.0885	0.527***	0.450***
	-0.136	-0.122	-0.0607	-0.0578
$\mathrm{FDI/GDP}_{t-1}$	0.289	0.527*	-0.508***	-0.486***
	-0.271	-0.291	-0.0999	-0.0982
gov. $\mathrm{revenue/GDP}_{t-4}$	0.0833***	0.109***	-0.0560*	0.292***
	-0.03	-0.0319	-0.0328	-0.033
(gov. revenue/ $\mathrm{GDP}_{t-4})^2$	-0.000341	-0.000725*	0.00225***	-0.00254***
	-0.00039	-0.000387	-0.000421	-0.000412
gov. expenditure/ GDP_{t-4}	-0.109***	-0.134***	0.127***	-0.0472*
	-0.0278	-0.0273	-0.0275	-0.0277
(gov. expenditure/ $\mathrm{GDP}_{t-4})^2$	0.000485	0.000791**	-0.00300***	0.000339
	-0.000361	-0.000331	-0.000345	-0.000333
gov. $\mathrm{grossdebt}_{t-4}$	-0.00992***	-0.01000***	-0.0275***	-0.0266***
	-0.000809	-0.00081	-0.00108	-0.00101
Country dummy	Yes	Yes	No	No
Year dummy	Yes	Yes	No	No
Observations	2,664	2,723	2664	2,723

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; the estimated threshold levels can be found in Appendix 1 (Table A.7)

There is, however, one very important feature of both probit models: the estimated threshold levels for rating categories do not exhibit nonlinearity. In Figures 4-5 I report the threshold levels in a way that their first values are normalized to 1 to achieve comparability (original values can be found in Table 8 of Appendix 1). The lack of nonlinearity is very important, because the possibility that it would show up was one of the most important reasons to use ordered choice models in the first place. Since there is no evidence for this phenomenon, and because the robustness checks for the fixed effects models showed that not accounting for censoring does not impair the estimates, fixed effects models seem a much better choice for modelling ratings, as their other underlying assumptions are the least stringent of all.

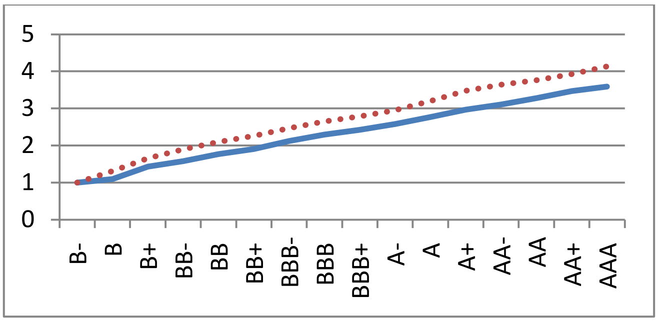

Figure 4: Rating thresholds for ordered probit model with country group and year dummies (Source: own calculations)

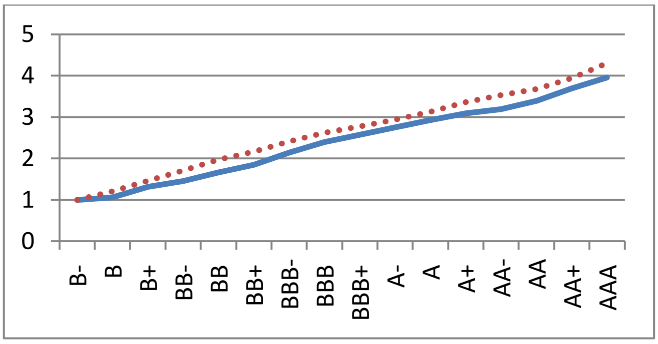

Figure 5: Rating thresholds for random effects ordered probit models.. (Source: own calculations)

Note: continuous line: Fitch; dotted line: S&P

3.2.6 Common time pattern of sovereign ratings

One way to look at how credit rating agencies reacted to the crisis is to check whether credit ratings exhibited a common time pattern or not. In Figures 6-9 I present the time dummy estimates (where applicable) with the approximate 95% confidence band (by the plus/minus twice the standard error rule of thumb).

At first glance, one could tell that the ratings are steadily declining over time, which could be interpreted as, for example, that the otherwise identical economy in 2010 would receive a worse rating in 2010 than in 1995. Of course, this is not the case, as the time average also proxies for the common movement created by the variables that are unobserved by the model, and all global factors that affects all countries in a similar way.

There are two things worth noticing at this point. First, the remarkable downward trend in ratings started *before* the crisis hit with a full blast in late 2008 (robust for all four cases). This means that that on the average credit rating agencies *did* to some extent predict the crisis, as on the average, factors unobserved by the models triggered a common downward adjustment. Second, for the more reliable fixed effects model, the 95% confidence band for most of the time includes the zero level, so this time pattern of ratings is not very significant (though the late 2007- early 2008 dip indeed is significant, so the previous finding remains valid). The confidence band, in the meantime, grows wider, indicating that sovereign ratings in the sample are diverging from each other to some extent.

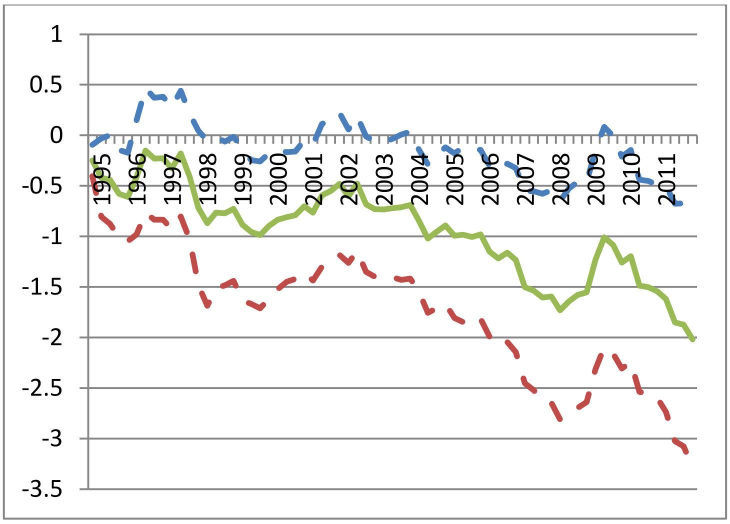

Figure 6: fixed effects: time effects for Fitch ratings (Source: own calculations)

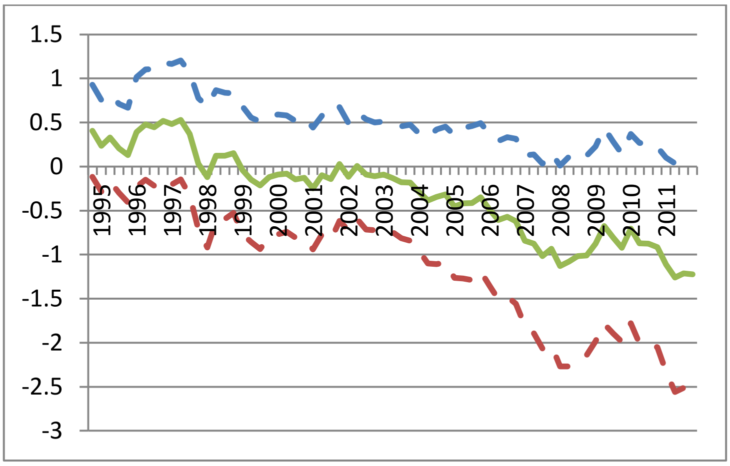

Figure 7; fixed effects: time effects for S&P ratings (Source: own calculations)

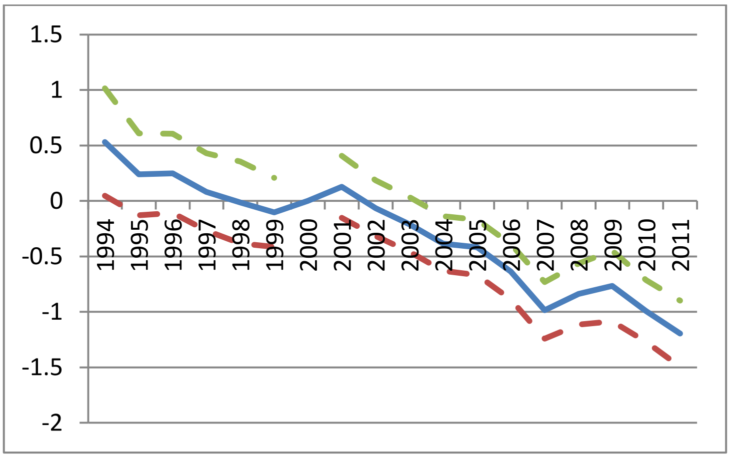

Figure 8: ordered probit: time effects for Fitch ratings (Source: own calculations)

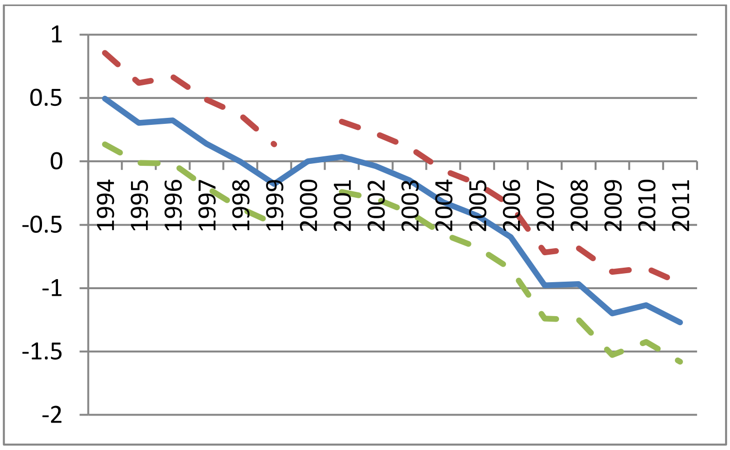

Figure 9: ordered probit: time effects for S&P ratings (Source: own calculations)

3.2.7 Agency behaviour in the pre-crisis and crisis years – a visual evidence

Ferri, Liu and Stiglitz made judgements about the procyclical fashion of ratings mainly by eyeballing, looking at the patterns of the model implied and the actual ratings around the Asian crisis (Ferri et al, 2003). Although the empirical rigour of this practice is questionable, it can be useful for gaining some first hint about the behaviour of rating agencies, so now I present some selected examples. I only look at Fitch ratings at this point (because modelling that resulted in a better fit) and also just the fixed effects models. The reason is that the visual of the ordered response models is impaired by the fact that the outcome variables can only take integer values, so if the latent variable is floating around a threshold value, the outcome variable will show a "thorny" pattern (fluctuating between the two integer values), and provides very little insight. The figures, because of the excessive use of space, are located in Appendix 2.

The most important lesson of this preliminary look at model generated ratings and true ones is that there is no distinct pattern of deviations between the two. There are clearly some countries for which by mere inspection the blamer argument seem to hold, that is, they upgraded too much and too early, and then downgraded excessively in the crisis, however, these are exactly the same countries, namely Greece and Portugal (see Appendix 2, Figure A.3), for which it is much more plausible that the expectations not reflected by current numbers but reflected by ratings are reasonably dim.

There are also countries for which ratings seem to predict correctly the pattern of the model generated ratings, so Fitch's predictions on macroeconomic fundamentals turned out to be valid (Appendix 2, Figure A.5). For other countries, rating changes lag behind changes in the data (Appendix 2, Figure A.4), while they are mostly moving together in other cases (Appendix 2, Figure A.6). So eyeballing is not conclusive.

3.2.8 *Agency behaviour in the pre-crisis and crisis years – predictions for the crisis*

To get a more empirically rigorous opinion on the behaviour of credit rating agencies in the crisis, I make two more estimations. First I re-estimate the fixed effects models for both set of rating data (but without time fixed effects), this time for the pre-crisis years only (so the sample ends at the second quarter of 2008), then use these coefficients to predict ratings for the crisis years. Then I relate the model predicted ratings to actual ones, and see if the two diverge from each other. If the difference in pattern seems to be systematic, that could mean that the agencies behave differently in the crisis and the coefficients from previous models no longer describe the behaviour of the dependent variable properly. An "overshooting" of downgrades in the crisis, for example, would provoke such a systematic divergence.

This calculation can be carried out by the fixed effects model only, because throwing out the last four years of the sample rendered it too small to do the ordered probit models as well (outputs can be found in Appendix 1, Table A.8). To avoid abusing with space, I put the related figures for the first estimation in Appendix 3. I look at the country groups where the subsample is big enough for the averaging to make sense; those are the most developed economies (group 1), Latin-America (group 2), the "adjusted" Middle-East (group 4), and the ex-Soviet bloc (group 9).

On the big average, for the first year of the crisis, ratings and fundamentals remain in line, but then, as expectations get worse, country ratings decouple from their model generated counterparts, and ratings become worse than the current data would suggest. However, the departure then is no wider than a single notch, even at its widest part, so the data does not suggest a big systemic failure or a general overhaul of the credit rating methodologies. While this pattern holds for industrial countries, for Latin-America it is exactly the opposite: after correcting for the overrating in the first period, ratings and model predictions move remarkably close to each other. The Middle-Eastern countries repeat the main pattern (somewhat more markedly), and the data for transition nations are inconsistent: while there is a case for a mild underrating of sovereigns at Fitch, Standard and Poor's seem to overrate them to some extent.

So the evidence is inconclusive – while there is some case for underrating sovereigns in the crisis, the departures on the average are smaller than a rating notch, and some subgroups even experience "overrating". According to my data, such phenomena like excessive downgrading hardly could be imagined as a driving force of the Great Recession, and there is no evidence of other systematically different rating behaviour for the crisis years.

3.2.9 Are credit ratings sticky?

The second estimation is suggested by the literature that suggests that credit rating agencies are mistakenly blamed for exacerbating crises as they are not leading markets but lagging behind and updated in a "sticky" fashion; they reflect otherwise widespread opinions on the market and do not contribute to formulating them (Mora, 2006; Gonzalez-Rozada and Yeyati, 2008; Gaillard, 2009**).** There is a method suggested by Mora (2006) to check whether ratings show a certain "sticky" behaviour, that is, are updated with a significant lag to what fundamentals would suggest (I already presented some sporadic visual evidence on that, now I determine if this is true on the average or not). The idea is to regress the change of ratings on lagged errors (model generated rating in last period less the true rating of last period). If the coefficient on the latter is positive and significant, that means that the fact that there is difference between true and predicted ratings helps to predict further rating changes, or, to put it differently, that the rating agencies move with a considerable lag compared to macroeconomic data.

The results for this inertia regression can be found in Table 11. The coefficient on the lagged error term significantly differs from zero 3 out of 4 times at the 95% confidence level, once at the 1% confidence level. The results are robust for re-adding cross sectional fixed effects, time fixed effects and the main explaining variables lagged by another period (the first robustness check is reported in Appendix 1, Table A.10).

Table 11: Inertia regressions

VARIABLES	(1) $\Delta \mathrm{Fitch}_t$	(3) $\Delta S\&P_t$
error_{t-1}	0.0533**	0.0248**
	(0.0217)	(0.0101)
Constant	0.0126**	0.0173***
	(0.00623)	(0.00602)
Observations	2,605	2,683
R-squared	0.012	0.006
Number of countries		

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1; a robustness check can be found in Appendix 1 (Table A.9)

The estimated coefficient is bigger than that calculated in Mora (2006), but it is of the same order of magnitude (0.02-0.03 there, 0.02-0.05 here), and somewhat less significant. It nevertheless indicates the presence of "sticky" rating adjustments, so on the average, credit rating agencies seem to react later than expected to changes in macroeconomic indicators. However, the size of the coefficient means that it would take an error of 10-15 rating notches to account for a 50% chance of adjusting the credit rating, so the stickiness is not a truly big concern.

4 CONCLUSION

My thesis went after the question whether credit rating agencies are deservedly claimed to have contributed to exacerbating the current crisis through misjudging sovereign risk (by "sleeping through" the onset of the crisis, or on the contrary, overreacting). To decide, I first looked at the question whether credit rating dynamics drive government bond spreads. The evidence is mixed, and it is impaired by the omitted variable problem, which cannot be eliminated completely. Although the possibility of ratings autonomously (over the common knowledge about macroeconomic health of the country that it reflects) affecting spreads cannot be ruled out completely, the estimated effects, which can be interpreted as upper bounds for "true effects" are quite moderate. Also, the indirect evidence obtained through proxying the anticipated nature of the rating change with rating outlook data suggest the co-movement between ratings and spreads should not be interpreted as causality.

Then I modelled ratings with a variety of model specifications, using the macroeconomic variables suggested by previous literature and agency handouts on the rating procedure. It seems that the more sophisticated ordered choice models do not do better at modelling ratings than simple fixed effect models, as they endogenously produce some of the features that one has to assume in the latter specifications (namely, that the rating scale is roughly linear in the explaining variables), and the possibly biggest drawback of the fixed effects model (missing the censored nature of the data) turns out not to be a empirically relevant. The magnitudes and significance levels of the right hand side variables are in line with those suggested in the literature. Both kind of regressions suggest a downward time trend in the average rating of the countries in the sample, and both show that this trend has started in late 2007 – early 2008 or even before, so in the big picture we cannot see that credit rating agencies would have spectacularly failed to account for the changing global circumstances.

Out of sample predictions for the crisis with coefficients retrieved from pre-crisis regressions show no clear pattern of systematic departures between predicted and actual ratings; some group of countries are overrated, others are underrated according to the model, but even the biggest average errors are within a 1 notch band. Empirical evidence suggests that sovereign credit ratings are inert to some extent and on the average respond in a slightly “sticky” manner, but not ostentatiously.

All in all, the role attributed to credit rating agencies seems to be exaggerated, both in terms of determining sovereign borrowing costs, and as a potential factor in exacerbating financial crises. I believe this has relevance for policymakers who tend to treat credit rating agencies as scapegoats for their fiscal worries. On the other hand, the same policymakers can break the *status quo* by putting more and more emphasis on credit ratings in newly drafted regulatory frameworks for various institutions, thus forcing market participants to bear credit ratings in mind (a tendency quite contradicting to the open communications of politicians). If markets now really do not care that much about credit ratings, it might be better to keep it that way.

REFERENCES

Afonso, António, Davide Furceri, and Pedro Gomes. "Sovereign Credit Ratings and Financial Markets Linkages: Application to European Data." *Journal of International Money and Finance* 31, no. 3 (April 2012): 606–638.

Afonso, Antonio, Pedro Gomes, and Philipp Rother. "Ordered Response Models for Sovereign Debt Ratings." *Applied Economics Letters* 16, no. 8 (2009): 769–773.

Afonso, António, Pedro Gomes, and Philipp Rother. "Short- and Long-run Determinants of Sovereign Debt Credit Ratings." *International Journal of Finance & Economics* 16, no. 1 (January 19, 2010): 1–15.

Afonso, Antonio, and Christophe Rault. "Long-Run Determinants of Sovereign Yields." *SSRN eLibrary* (August 17, 2010). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1660368.

Afonso, Antonio. "Understanding the Determinants of Sovereign Debt Ratings: Evidence for the Two Leading Agencies." *Journal of Economics and Finance* 27, no. 1 (2003): 56–74.

Andritzky, Jochen R., Geoffrey J. Bannister, and Natalia T. Tamirisa. "The Impact of Macroeconomic Announcements on Emerging Market Bonds." *Emerging Markets Review* 8, no. 1 (March 2007): 20–37.

Arezki, Rabah, Bertrand Candelon, and Amadou N. R. Sy. "Sovereign Rating News and Financial Markets Spillovers : Evidence from the European Debt Crisis Rabah Arezki Sovereign Rating News and Financial Markets Spillovers : Evidence from the European Debt Crisis Abstract." *October* 11/68 (2011): 1–18.

Binder, John. "The Event Study Methodology Since 1969." *Review of Quantitative Finance and Accounting* 11, no. 2 (1998): 111–137.

Bissoondoyal-Bheenick, Emawtee. "An Analysis of the Determinants of Sovereign Ratings." *Global Finance Journal* 15, no. 3 (February 2005): 251–280.

Cantor, Richard, and Frank Packer. “Determinants and Impacts of Sovereign Credit Ratings.” *FRBNY Economic Policy Review* (October 1996): 37–54.

———. “Sovereign Credit Ratinfs.” *Current Issues in Economics and Finance* 1, no. 3 (1995): 1–6.

Eichengreen, Barry, and Ashoka Mody. “What Explains Changing Spreads on Emerging-Market Debt: Fundamentals or Market Sentiment?” *National Bureau of Economic Research Working Paper Series* No. 6408 (1998). http://www.nber.org/papers/w6408.

Ferri, G., L.-G. Liu, and J. E. Stiglitz. “The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis.” *Economic Notes* 28, no. 3 (December 2, 2003): 335–355.

Ferri, Giovanni, Li-Gang Liu, and Giovanni Majnoni. “The Role of Rating Agency Assessments in Less Developed Countries: Impact of the Proposed Basel Guidelines.” *Journal of Banking & Finance* 25, no. 1 (2001): 115–148.

Flandreau, Marc, Norbert Gaillard, and Frank Packer. “To Err Is Human: US Rating Agencies and the Interwar Foreign Government Debt Crisis.” *European Review of Economic History* 15, no. 3 (December 1, 2011): 495–538.

Gaillard, Norbert. “Fitch, Moody’s and S&P’s Sovereign Ratings and EMBI Global Spreads: Lessons from 1993-2007.” *International Research Journal of Finance and Economics*, no. 26 (2009).

Gande, Amar, and David C. Parsley. “News Spillovers in the Sovereign Debt Market.” *Journal of Financial Economics* 75, no. 3 (March 2005): 691–734.

González-Rozada, Martín, and Eduardo Levy Yeyati. “Global Factors and Emerging Market Spreads.” *The Economic Journal* 118, no. 533 (November 1, 2008): 1917–1936.

Greene, William H. *Econometric Analysis*. 7th ed. Prentice Hall, 2011.

Hill, Claire A., "Rating Agencies Behaving Badly: The Case of Enron", *Conn. L. Rev.* 35, (2002): 1145

Hill, Claire, "Regulating the Rating Agencies", *WASHINGTON UNIVERSITY LAW QUARTERLY*, 82(43), (2004): p44

Hill, Paula, Robert Brooks, and Robert Faff. “Variations in Sovereign Credit Quality Assessments Across Rating Agencies.” *Journal of Banking & Finance* 34, no. 6 (June 2010): 1327–1343.

Hill, Paula, and Robert W. Faff. “Do Credit Watch Procedures Affect the Information Content of Sovereign Credit Rating Changes?” *SSRN eLibrary* (September 14, 2007). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=968273.

Hilscher, Jens, and Yves Nosbusch. “Determinants of Sovereign Risk: Macroeconomic Fundamentals and the Pricing of Sovereign Debt*.” *Review of Finance* 14, no. 2 (April 1, 2010): 235 –262.

Hooper, Vince, Timothy Hume, and Suk-Joong Kim. “Sovereign Rating changes—Do They Provide New Information for Stock Markets?” *Economic Systems* 32, no. 2 (June 2008): 142–166.

How We Rate Sovereigns. Standard and Poor’s, 2012. http://www.standardandpoors.com/spf/ratings/How_We_Rate_Sovereigns_3_13_12.pdf.

Hu, Yen-Ting, Rudiger Kiesel, and William Perraudin. “The Estimation of Transition Matrices for Sovereign Credit Ratings.” *Journal of Banking & Finance* 26, no. 7 (July 2002): 1383–1406.

Ismailescu, Iuliana, and Hossein Kazemi. “The Reaction of Emerging Market Credit Default Swap Spreads to Sovereign Credit Rating Changes.” *Journal of Banking & Finance* 34, no. 12 (December 2010): 2861–2873.

Kamin, Steven B., and Karsten Von Kleist. “The Evolution and Determinants of Emerging Markets Credit Spreads in the 1990s.” *SSRN eLibrary* (May 1999). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=850104.

Kaminsky, Graciela L., and Sergio L. Schmukler. “What Triggers Market Jitters?: A Chronicle of the Asian Crisis.” *Journal of International Money and Finance* 18, no. 4 (August 1999): 537–560.

Kaminsky, Graciela, and Sergio L. Schmukler. “Emerging Market Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns?” *The World Bank Economic Review* 16, no. 2 (August 1, 2002): 171–195.

Kräussl, Roman. “Do Credit Rating Agencies Add to the Dynamics of Emerging Market Crises?” *Journal of Financial Stability* 1, no. 3 (April 2005): 355–385.

Kruck, Andreas. *Private Ratings, Public Regulations - Credit Rating Agencies and Global Financial Governance*. 1st ed. Transformations of the State. Palgrave Macmillan, 2011.

Larraín, H. Reisen, and J. Von Maltzan. “Emerging Market Risk and Sovereign Credit Ratings.” *European Financial Management* 124, no. 124 (1997).

Mora, Nada. “Sovereign Credit Ratings: Guilty Beyond Reasonable Doubt?” *Journal of Banking & Finance* 30, no. 7 (July 2006): 2041–2062.

Reisen, Helmut, and Julia Von Maltzan. “Boom and Bust and Sovereign Ratings.” *International Finance* 2, no. 2 (December 16, 2002): 273–293.

Reisen, Helmut, and Julia von Maltzan. “Sovereign Credit Ratings, Emerging Market Risk and Financial Market Volatility.” *Intereconomics* 33, no. 2 (1998): 73–82.

Sy, Amadou N.R. “Emerging Market Bond Spreads and Sovereign Credit Ratings: Reconciling Market Views with Economic Fundamentals.” *Emerging Markets Review* 3, no. 4 (December 1, 2002): 380–408.

Wooldridge, Jeffrey M.: Econometric Analysis of Cross Section and Panel Data, Cambridge, MIT Press, 2011

APPENDICES

APPENDIX 1 – EXTENDED ESTIMATION OUTPUTS AND ADDITIONAL TABLES

Table A.1: baseline model with controls

$$spread_{it} = \hat{\alpha}_i + \hat{\tau}_t + \hat{\beta} * Y_{it} + \hat{\delta} * rating_{it} + u_{it}$$

	Fitch		S&P		Empty model	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	gov. bond.	tr. bill	gov. bond.	tr. bill	gov. bond.	tr. bill
$\text{rating}_{\text{it}}$	-0.727***	-0.490***	-0.570***	-0.341*		
	(0.137)	(0.172)	(0.175)	(0.195)		
$\log{(\text{GDP})_{t-1}}$	0.827	-0.477	0.381	-0.169	-0.217	-0.495
	(0.832)	(0.993)	(0.816)	(1.047)	(0.884)	(1.042)
$\text{reserves to GDP}_{t-1}$	-2.422e+08	-6.288e+08	4.178e+08	-3.178e+08	6.663e+08	-1.427e+08
	(4.656e+08)	(4.929e+08)	(4.578e+08)	(3.454e+08)	(6.780e+08)	(3.553e+08)
$\text{real growth of GDP}_{t-1}$	-0.0710**	-0.0214	-0.0981***	-0.0297	-0.125***	-0.0404
	(0.0273)	(0.0522)	(0.0293)	(0.0507)	(0.0380)	(0.0502)
inflation_{t-1}	0.159***	0.674***	0.199***	0.707***	0.226***	0.693***
	(0.0513)	(0.0905)	(0.0550)	(0.0870)	(0.0514)	(0.0881)
$\text{unemployment rate}_{t-1}$	0.0281	0.0349	0.00622	0.0344	0.103***	0.0707
	(0.0193)	(0.0583)	(0.0389)	(0.0560)	(0.0310)	(0.0612)
NEL/CAR_{t-1}	0.184	0.901***	0.0875	0.823***	-0.0816	0.651***
	(0.132)	(0.184)	(0.145)	(0.183)	(0.153)	(0.156)
FDI/GDP_{t-1}	-0.0686	-0.958	-0.0532	-0.731	-0.00734	-0.546
	(0.0630)	(0.660)	(0.0575)	(0.709)	(0.0545)	(0.616)
$\text{gov. revenue/GDP}_{t-4}$	-0.195	-0.334	-0.205	-0.152	-0.403*	-0.495*
	(0.208)	(0.307)	(0.212)	(0.301)	(0.207)	(0.292)
$(\text{gov. revenue/GDP}_{t-4})^2$	0.00142	0.00441	0.00171	0.00126	0.00417*	0.00581
	(0.00247)	(0.00506)	(0.00257)	(0.00492)	(0.00243)	(0.00466)
$\text{gov. expenditure/GDP}_{t-4}$	-0.0725	0.279*	-0.215	0.0847	-0.243*	0.0423
	(0.102)	(0.155)	(0.137)	(0.186)	(0.131)	(0.165)
$(\text{gov. expenditure/GDP}_{t-4})^2$	0.000460	-0.00270	0.00248	-0.000314	0.00276*	-0.000250
	(0.00106)	(0.00182)	(0.00156)	(0.00206)	(0.00138)	(0.00178)
$\text{gov. grossdebt}_{t-4}$	0.00351	-0.00755	-0.00249	-0.00536	0.0106	0.00670
	(0.00873)	(0.0209)	(0.0106)	(0.0209)	(0.00841)	(0.0162)
Constant	15.61**	11.89	18.19**	9.143	15.99*	12.90
	(7.433)	(10.54)	(7.996)	(11.06)	(8.642)	(10.73)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Macro controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,109	1,635	2,129	1,770	2,295	1,811
R-squared	0.625	0.678	0.582	0.677	0.509	0.660
Number of countries	44	39	42	39	44	40

Note: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1;

Table A.2: **Rich lag structure**

$$spread_{it} = \hat{\alpha}_i + \hat{\tau}_t + \hat{\beta} * Y_{it} + \hat{\delta}_1 * rating_{it} + \hat{\delta}_2 * rating_{it-1} + \hat{\delta}_3 * rating_{it-2} + \hat{\delta}_4 * rating_{it-3} + u_{it}$$

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
rating_{it}	-0.885***	-0.756**	-0.758***	-0.530*
	(0.0885)	(0.328)	(0.121)	(0.309)
rating_{it-1}	-0.317***	-0.0643	-0.271**	-0.160
	(0.117)	(0.185)	(0.102)	(0.108)
rating_{it-2}	0.189**	0.239	0.115	0.170
	(0.0931)	(0.278)	(0.111)	(0.206)
rating_{it-3}	0.456***	0.210	0.479***	0.220
	(0.0986)	(0.178)	(0.127)	(0.207)
$\log(\text{GDP})_{t-1}$	0.532	-0.818	0.149	-0.368
	(0.768)	(1.025)	(0.731)	(1.096)
reserves to GDP_{t-1}	-1.125e+08	-4.414e+08	4.262e+08	-2.835e+08
	(3.934e+08)	(3.985e+08)	(4.565e+08)	(3.168e+08)
real growth of GDP_{t-1}	-0.0719**	-0.0156	-0.0955***	-0.0224
	(0.0269)	(0.0517)	(0.0293)	(0.0494)
inflation_{t-1}	0.126**	0.676***	0.175***	0.710***
	(0.0490)	(0.0950)	(0.0537)	(0.0891)
unemployment rate_{t-1}	0.0437**	0.0478	0.0238	0.0375
	(0.0200)	(0.0591)	(0.0407)	(0.0563)
NEL/CAR_{t-1}	0.160	0.901***	0.0885	0.869***
	(0.121)	(0.192)	(0.136)	(0.185)
FDI/GDP_{t-1}	-0.0566	-0.874	-0.0559	-0.879
	(0.0549)	(0.733)	(0.0526)	(0.676)
$\text{gov. revenue/GDP}_{t-4}$	-0.166	-0.415	-0.190	-0.160
	(0.192)	(0.301)	(0.207)	(0.309)
$(\text{gov. revenue/GDP}_{t-4})^2$	0.00119	0.00559	0.00168	0.00139
	(0.00222)	(0.00493)	(0.00245)	(0.00501)
$\text{gov. expenditure/GDP}_{t-4}$	-0.0486	0.263*	-0.226	0.0678
	(0.0875)	(0.154)	(0.135)	(0.183)
$(\text{gov. expenditure/GDP}_{t-4})^2$	0.000138	-0.00278	0.00245	-0.000229
	(0.000901)	(0.00192)	(0.00152)	(0.00203)
$\text{gov. grossdebt}_{t-4}$	0.00638	-0.00568	0.00127	-0.00362
	(0.00801)	(0.0213)	(0.00958)	(0.0204)
Constant	13.89**	14.04	16.76**	10.42
	(6.785)	(10.42)	(7.089)	(11.20)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	2,056	1,595	2,126	1,767
R-squared	0.658	0.681	0.601	0.680
Number of countries	44	39	42	39

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.3a: Dummy regressions

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
$\mathrm{upgrade}_t$	-0.409*	-0.519*	-0.267	-0.115
	(0.238)	(0.277)	(0.165)	(0.363)
$\mathrm{downgrade}_t$	1.610***	1.200	1.269***	0.765
	(0.454)	(0.737)	(0.406)	(0.597)
$\mathrm{rating}_t * \mathrm{catchup}_t$	0.0960	-0.482	0.345	-0.0692
	(0.282)	(0.525)	(0.358)	(0.492)
$\mathrm{rating}_t * \mathrm{decouple}_t$	0.536	0.469	0.202	0.0797
	(0.389)	(0.487)	(0.322)	(0.273)
$\mathrm{rating}_t * \mathrm{inv}_t$	-0.876	-1.214	-1.679	-2.817***
	(0.640)	(0.779)	(1.359)	(1.033)
$\mathrm{rating}_{t-1} * \mathrm{inv}_{t-1}$	0.218	-1.750***	0.000800	-0.315
	(0.568)	(0.508)	(1.015)	(1.499)
$\mathrm{catchup}_t$	0.149	0.495	0.183	0.275
	(0.162)	(0.384)	(0.176)	(0.358)
$\mathrm{decouple}_t$	-0.0595	-0.125	-0.0654	-0.181
	(0.251)	(0.401)	(0.153)	(0.418)
$\mathrm{double\_inv}_t$	-1.998**	-0.497	-1.263	0.550
	(0.786)	(0.695)	(1.117)	(0.857)
$\mathrm{double\_inv}_{t-1}$	-0.0490	1.581*	0.167	0.503
	(0.622)	(0.919)	(0.918)	(0.733)
$\log(\mathrm{GDP})_{t-1}$	-0.390	-1.038	-0.519	-0.563
	(0.874)	(0.996)	(0.854)	(1.088)
reserves to GDP_{t-1}	2.948e+08	-2.431e+08	5.754e+08	5.317e+07
	(4.756e+08)	(3.773e+08)	(5.919e+08)	(2.909e+08)
real growth of GDP_{t-1}	-0.0923***	-0.0229	-0.101***	-0.0235
	(0.0280)	(0.0480)	(0.0312)	(0.0460)
$\mathrm{inflation}_{t-1}$	0.159***	0.685***	0.193***	0.709***
	(0.0466)	(0.0936)	(0.0496)	(0.0948)
unemployment rate_{t-1}	0.0875***	0.0700	0.0699**	0.0571
	(0.0283)	(0.0639)	(0.0325)	(0.0626)
$\mathrm{NEL/CAR}_{t-1}$	0.0221	0.702***	-0.0596	0.624***
	(0.111)	(0.171)	(0.117)	(0.155)
$\mathrm{FDI/GDP}_{t-1}$	-0.0381	-0.743	-0.0172	-0.665
	(0.0503)	(0.593)	(0.0462)	(0.692)
$\mathrm{gov.\,revenue/GDP}_{t-4}$	-0.401**	-0.516*	-0.293	-0.313
	(0.184)	(0.294)	(0.181)	(0.294)
$(\mathrm{gov.\,revenue/GDP}_{t-4})^2$	0.00429*	0.00722	0.00316	0.00358
	(0.00219)	(0.00498)	(0.00216)	(0.00472)
$\mathrm{gov.\,expenditure/GDP}_{t-4}$	-0.135	0.199	-0.281**	0.0639
	(0.102)	(0.149)	(0.124)	(0.173)
$(\mathrm{gov.\,expenditure/GDP}_{t-4})^2$	0.00120	-0.00196	0.00306**	-0.000344
	(0.00106)	(0.00181)	(0.00132)	(0.00188)
$\mathrm{gov.\,grossdebt}_{t-4}$	0.0137	0.00239	0.00792	0.00301
	(0.00864)	(0.0196)	(0.00874)	(0.0189)

Note: Continued on the next page. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.3b: Dummy regressions

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond	tr. bill
Constant	18.38**	13.60	19.96**	11.36
	(7.873)	(10.71)	(7.628)	(11.23)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	2,091	1,622	2,128	1,769
R-squared	0.583	0.680	0.575	0.682
Number of countries	44	39	42	39

Note: Continued from the previous page. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Two categories for anticipation (anticipated and unanticipated)

$$spread_{it} = \alpha_i + \tau_t + \beta * downgrade\_anticipated_{ti} + \gamma * downgrade\_unanticipated_{ti} + \zeta * upgrade\_anticpated_{ti} + \eta * upgrade\_unanticipated_{ti} + \theta * Y_{it} + u_{it}$$

	Fitch		S&P	
	(1)	(2)	(3)	(4)
VARIABLES	gov. bond.	tr. bill	gov. bond.	tr. bill
ant. upgrade	-0.559***	-0.575*	-0.305**	0.0677
	-0.188	-0.316	-0.143	-0.303
ua. upgrade	0.0937	0.167	0.375	-0.288
	-0.129	-0.346	-0.321	-0.638
ant. dgrade	1.598***	0.978*	1.596***	1.202*
	-0.526	-0.523	-0.419	-0.631
ua dgrade	0.0249	-0.026	0.15	4.671
	-0.198	-0.515	-0.632	-2.888
$\log(\mathrm{GDP})_{t-1}$	-0.361	-0.594	-0.409	-0.544
	(0.872)	(1.100)	(0.903)	(1.095)
reserves to GDP_{t-1}	6.634e+08	-1.799e+08	6.236e+08	-2.019e+08
	(6.934e+08)	(2.732e+08)	(6.784e+08)	(2.768e+08)
real growth of GDP_{t-1}	-0.120***	-0.0510	-0.119***	-0.0518
	(0.0365)	(0.0510)	(0.0363)	(0.0514)
$\mathrm{inflation}_{t-1}$	0.218***	0.721***	0.214***	0.716***
	(0.0526)	(0.0956)	(0.0506)	(0.0950)
unemployment rate_{t-1}	0.0980***	0.0776	0.0931***	0.0743
	(0.0313)	(0.0622)	(0.0321)	(0.0630)
$\mathrm{NEL/CAR}_{t-1}$	-0.0812	0.655***	-0.0852	0.662***
	(0.150)	(0.164)	(0.147)	(0.162)
$\mathrm{FDI/GDP}_{t-1}$	-0.0153	-0.701	-0.0121	-0.696
	(0.0517)	(0.681)	(0.0512)	(0.655)
$\mathrm{gov.\,revenue/GDP}_{t-4}$	-0.381*	-0.363	-0.393*	-0.341
	(0.203)	(0.288)	(0.212)	(0.291)
$(\mathrm{gov.\,revenue/GDP}_{t-4})^2$	0.00381	0.00429	0.00396	0.00402
	(0.00236)	(0.00464)	(0.00247)	(0.00471)
$\mathrm{gov.\,expenditure/GDP}_{t-4}$	-0.243*	0.0643	-0.235*	0.0640
	(0.127)	(0.170)	(0.130)	(0.171)
$(\mathrm{gov.\,expenditure/GDP}_{t-4})^2$	0.00274**	-0.000299	0.00267*	-0.000345
	(0.00134)	(0.00182)	(0.00137)	(0.00186)
$\mathrm{gov.\,grossdebt}_{t-4}$	0.0105	0.00126	0.00994	0.00250
	(0.00793)	(0.0184)	(0.00822)	(0.0186)
Constant	16.93**	10.29	17.45*	9.536
	(8.374)	(11.10)	(8.831)	(10.99)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	2,278	1,772	2,278	1,772
R-squared	0.526	0.671	0.526	0.677
Number of countries	44	39	44	39

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Three categories (anticipated, unanticipated, neutral)

	Fitch		S&P	
	(5)	(6)	(7)	(8)
VARIABLES	gov. bond.	tr. bill	gov. bond.	tr. bill
ant. upgrade	0.0678	-0.575*	-0.305**	-0.559***
	-0.304	-0.316	-0.143	-0.188
semi-ant. upgrade	-0.289	0.167	0.375	0.0938
	-0.655	-0.346	-0.321	-0.129
ua. upgrade	-0.271	0	0	0
	-1.037	0	0	0
ant. dgrade	1.202*	0.978*	1.596***	1.598***
	-0.631	-0.523	-0.419	-0.526
semi-ant. dgrade	-0.026	4.671	0.0249	0.18
	-0.515	-2.888	-0.198	-0.754
ua dgrade	0	0.0302	0	0
	0	-0.271	0	0
$\mathrm{FDI/GDP}_{t-1}$	-0.0153	-0.0121	-0.701	-0.696
	(0.0517)	(0.0513)	(0.682)	(0.655)
$\mathrm{gov.\,revenue/GDP}_{t-4}$	-0.381*	-0.393*	-0.363	-0.341
	(0.203)	(0.212)	(0.288)	(0.291)
$(\mathrm{gov.\,revenue/GDP}_{t-4})^2$	0.00381	0.00396	0.00429	0.00402
	(0.00236)	(0.00247)	(0.00465)	(0.00471)
$\mathrm{gov.\,expenditure/GDP}_{t-4}$	-0.243*	-0.235*	0.0643	0.0640
	(0.127)	(0.130)	(0.170)	(0.171)
$(\mathrm{gov.\,expenditure/GDP}_{t-4})^2$	0.00274**	0.00267*	-0.000299	-0.000345
	(0.00134)	(0.00137)	(0.00182)	(0.00186)
$\mathrm{gov.\,grossdebt}_{t-4}$	0.0105	0.00994	0.00125	0.00250
	(0.00793)	(0.00823)	(0.0186)	(0.0186)
Constant	16.93**	17.45*	10.29	9.536
	(8.374)	(8.833)	(11.11)	(10.99)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	2,278	2,278	1,772	1,772
R-squared	0.526	0.526	0.671	0.677
Number of countries	44	44	39	39

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Selection equations

	Not including group 1		Not including AAA rated	
	(1)	(2)	(6)	(7)
VARIABLES	Fitch	S&P	Fitch	S&P
$\log(\mathrm{GDP})_{t-1}$	2.340***	1.830**	2.251***	2.015***
	(0.704)	(0.831)	(0.506)	(0.598)
reserves to GDP_{t-1}	2.181e+09	3.631e+09	-7.323e+08	-1.340e+08
	(1.559e+09)	(2.190e+09)	(5.179e+08)	(3.134e+08)
real growth of GDP_{t-1}	0.0448***	0.0252**	0.0719***	0.0460**
	(0.0131)	(0.00977)	(0.0147)	(0.0174)
$\mathrm{inflation}_{t-1}$	-0.0375***	-0.0390***	-0.0582***	-0.0539***
	(0.0131)	(0.0103)	(0.0110)	(0.0100)
unemployment rate_{t-1}	-0.0587*	-0.0903**	-0.109***	-0.116***
	(0.0304)	(0.0391)	(0.0305)	(0.0342)
$\mathrm{NEL/CAR}_{t-1}$	0.499**	0.345	0.397***	0.371**
	(0.224)	(0.277)	(0.118)	(0.139)
$\mathrm{FDI/GDP}_{t-1}$	-0.323	0.547	-0.670**	-0.362
	(0.411)	(0.722)	(0.274)	(0.282)
$\mathrm{gov.\,revenue/GDP}_{t-4}$	0.389***	0.526***	0.437***	0.560***
	(0.117)	(0.159)	(0.0983)	(0.116)
$(\mathrm{gov.\,revenue/GDP}_{t-4})^2$	-0.00450**	-0.00676***	-0.00526***	-0.00740***
	(0.00182)	(0.00244)	(0.00143)	(0.00162)
$\mathrm{gov.\,expenditure/GDP}_{t-4}$	0.355**	0.295**	0.246**	0.101
	(0.141)	(0.142)	(0.0960)	(0.108)
$(\mathrm{gov.\,expenditure/GDP}_{t-4})^2$	-0.00594***	-0.00491**	-0.00356***	-0.00157
	(0.00186)	(0.00183)	(0.00118)	(0.00126)
$\mathrm{gov.\,grossdebt}_{t-4}$	-0.0294**	-0.0327**	-0.0252**	-0.0307***
	(0.0117)	(0.0142)	(0.00967)	(0.0112)
Constant	-17.38**	-16.16	-15.60***	-13.42**
	(7.349)	(9.897)	(5.136)	(5.801)
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	1,203	1,255	1,866	1,987
R-squared	0.713	0.625	0.658	0.564
Number of countries	30	31	47	47

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Threshold levels for ordered models

		Dummies		Random effects	
Threshold number	Corresponding rating	Fitch	S&P	Fitch	S&P
1	B-	3.751	3.196	5.78	5.1
2	B	4.108	4.186	6.142	6.147
3	B+	5.376	5.289	7.593	7.46
4	BB-	5.922	6.052	8.381	8.707
5	BB	6.618	6.698	9.618	10.04
6	BB+	7.16	7.227	10.66	11.05
7	BBB-	7.954	7.869	12.38	12.3
8	BBB	8.608	8.457	13.84	13.37
9	BBB+	9.099	8.891	14.89	14.16
10	A-	9.659	9.415	15.88	14.96
11	A	10.38	10.22	16.94	15.97
12	A+	11.14	11.12	17.86	17.14
13	AA-	11.64	11.63	18.45	18.01
14	AA	12.29	12.01	19.59	18.77
15	AA+	13	12.54	21.34	20.07
16	AAA	13.45	13.21	22.86	21.94

Table A.8: Truncated sample FE regression (data used until 2008Q3), no time fixed effects

VARIABLES	(1) Fitch	(2) S&P
$\log(\mathrm{GDP})_{t-1}$	1.262***	0.879***
	(0.212)	(0.207)
reserves to GDP_{t-1}	-3.985e+08	8.319e+07
	(5.628e+08)	(4.749e+08)
real growth of GDP_{t-1}	0.0354**	0.0286
	(0.0149)	(0.0180)
$\mathrm{inflation}_{t-1}$	-0.0750***	-0.0678***
	(0.0153)	(0.0165)
unemployment rate_{t-1}	-0.0819***	-0.0976***
	(0.0251)	(0.0295)
$\mathrm{NEL/CAR}_{t-1}$	-0.125***	0.0160
	(0.0409)	(0.0358)
$\mathrm{FDI/GDP}_{t-1}$	0.0353	-0.00923
	(0.0269)	(0.0264)
$\mathrm{gov.\,revenue/GDP}_{t-4}$	0.210***	0.285***
	(0.0666)	(0.0765)
$(\mathrm{gov.\,revenue/GDP}_{t-4})^2$	-0.00276***	-0.00368***
	(0.000916)	(0.000927)
$\mathrm{gov.\,expenditure/GDP}_{t-4}$	0.133**	0.00532
	(0.0649)	(0.0830)
$(\mathrm{gov.\,expenditure/GDP}_{t-4})^2$	-0.00150**	7.48e-05
	(0.000695)	(0.000958)
$\mathrm{gov.\,grossdebt}_{t-4}$	-0.0228***	-0.0276***
	(0.00463)	(0.00394)
Constant	-2.747	6.587***
	(2.308)	(2.295)
Country fixed effects	Yes	Yes
Time fixed effects	No	No
Observations	1,937	2,003
R-squared	0.610	0.530
Number of countries	55	55

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Inertia regressions

VARIABLES	(1) ΔFitch	(2) ΔFitch	(3) $\Delta S\&P$	(4) $\Delta S\&P$
Error_{t-1}	0.0533**	0.0501*	0.0248**	0.0462**
	(0.0217)	(0.0300)	(0.0101)	(0.0201)
Constant	0.0126**	0.0126***	0.0173***	0.0213***
	(0.00623)	(1.36e-05)	(0.00602)	(0.00376)
Country fixed effects	No	Yes	No	Yes
Time fixed effects	No	Yes	No	Yes
Observations	2,605	2,605	2,683	2,683
R-squared	0.012	0.011	0.006	0.013
Number of countries		56		55

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

APPENDIX 2 – VISUAL EVIDENCE FOR THE BEHAVIOUR OF THE RATING AGENCIES

Figure A.1: Too much optimism and then too much pessimism: Cyprus and Thailand

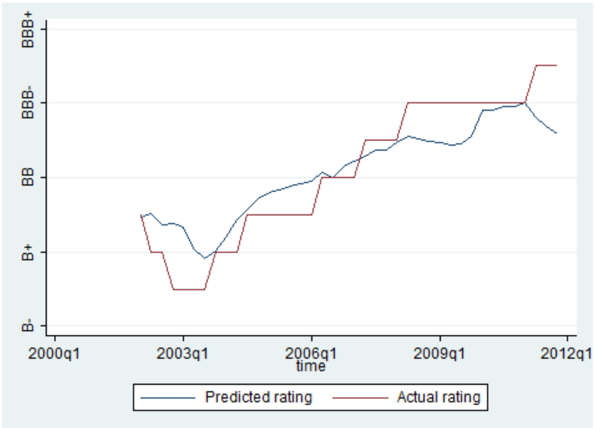

Figure A.2: Too much early confidence, than diverging ratings: Brazil

Figure A.3: Too harsh reaction: Greece and Portugal

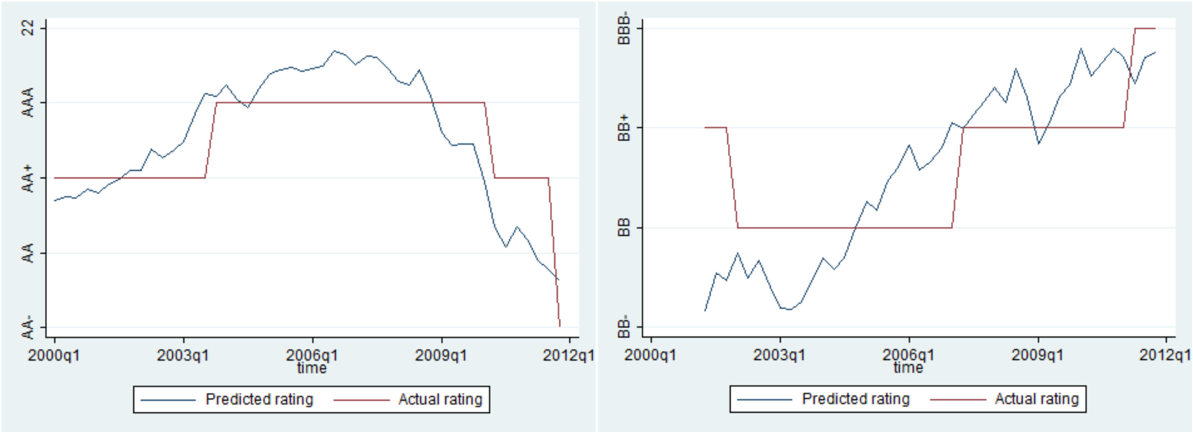

Figure A.4: Late reaction to changing fundamentals: Spain and Colombia

Figure A.5: Ratings correctly predicting fundamentals: Latvia and Hungary

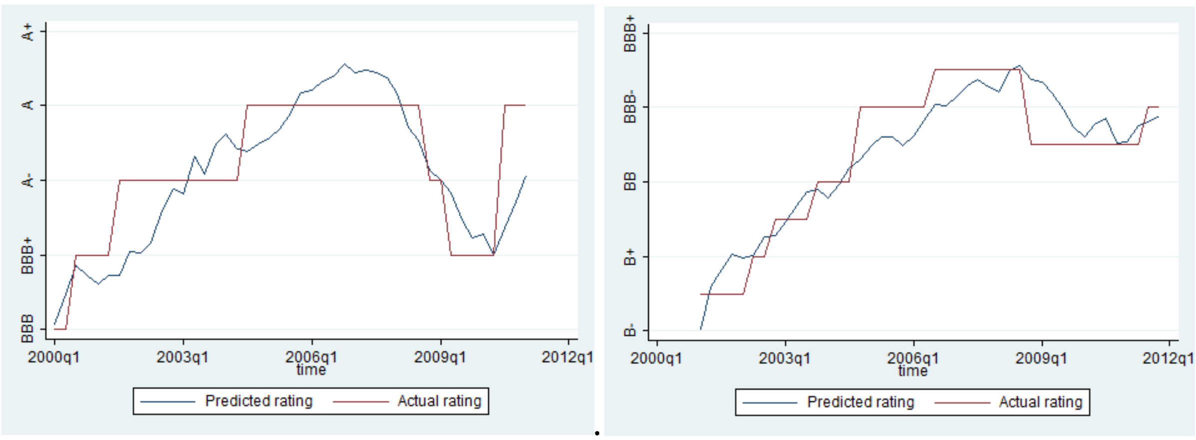

Figure A.6: Ratings in line with predictions: Estonia and Romania

:

APPENDIX 3 – OUT OF SAMPLE PREDICTIONS FOR THE CRISIS

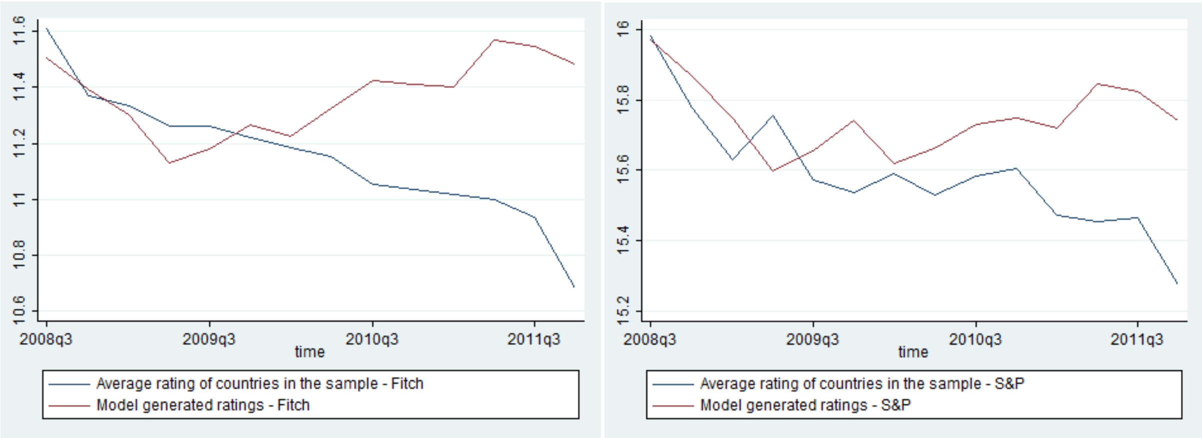

Figure A.7: Average ratings

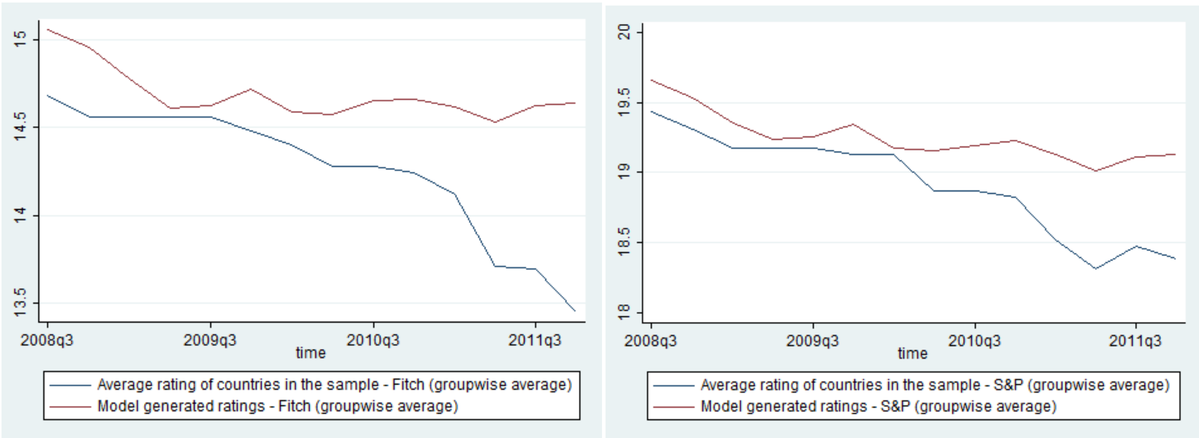

Figure A.8: Country group 1: industrial countries

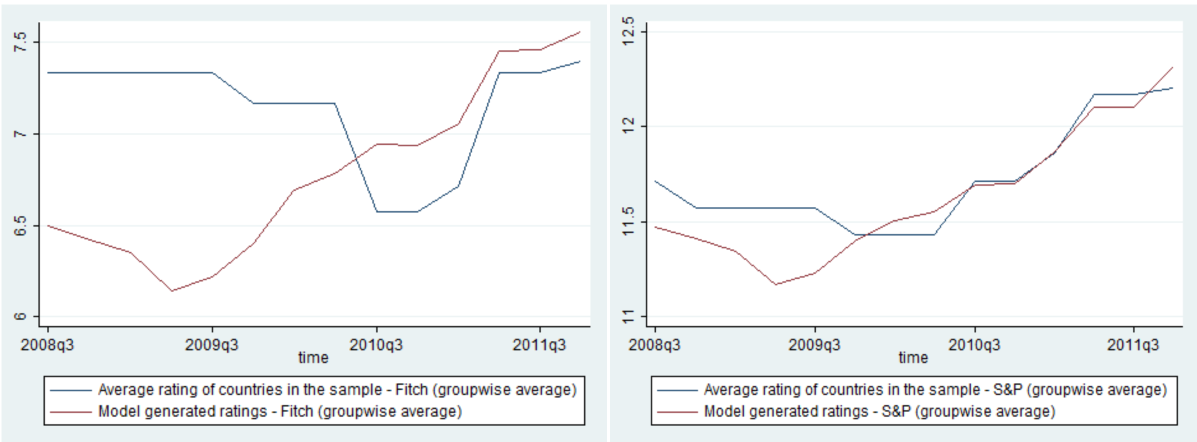

Figure A.9: Country group 2: Latin-America

Note: I preserved the numerical values for the Y axes on these graph to see the differences in terms of rating notches. Here the 16 corresponds to "AAA" for Fitch ratings and 21 to "AAA" for S&P ratings due to technical reasons (numbering starts at the worst debtor; S&P differentiates between more categories there).

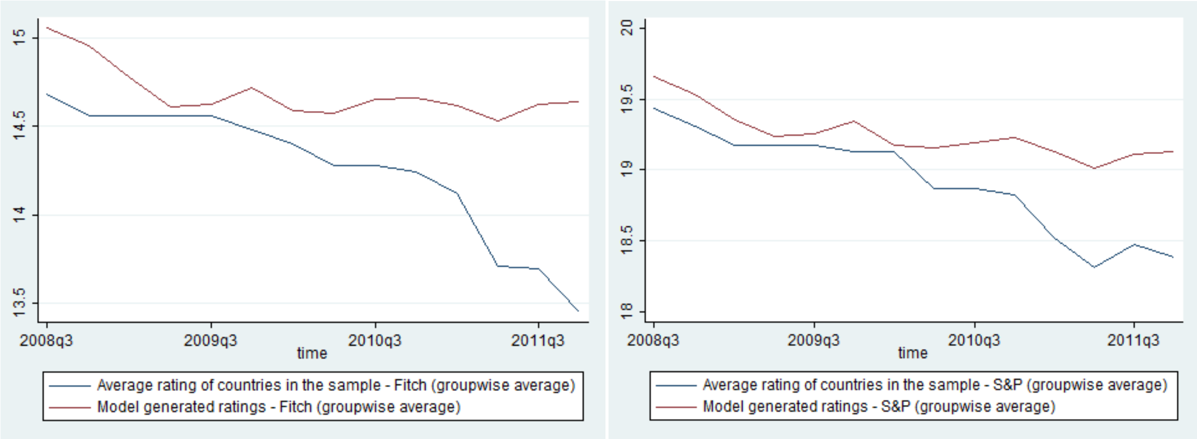

Figure A.10: Country group 4: Middle-East

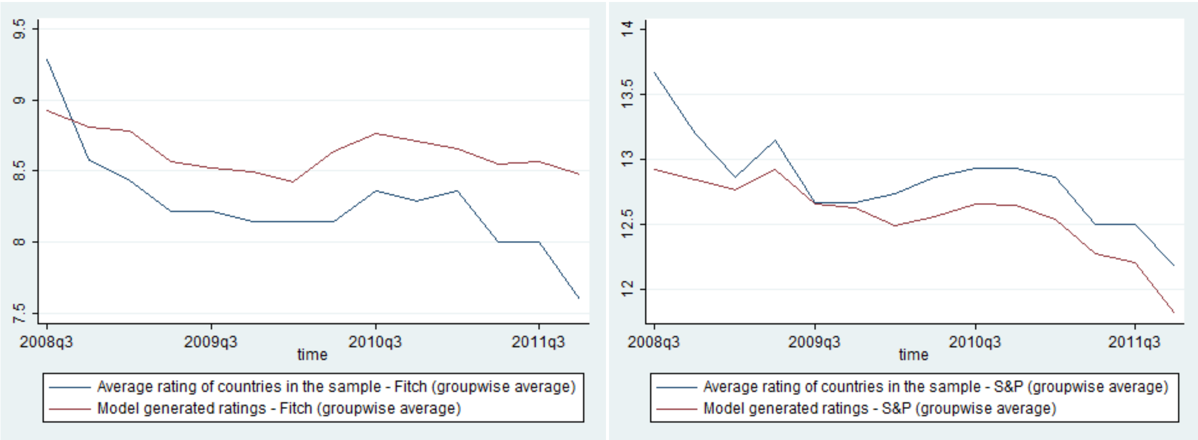

Figure A.11: Country group 9: post-socialist countries