

# The Joint Dynamics of Sovereign Ratings and Government Bond Yields

Makram El-Shagi  
Gregor von Schweinitz

March 2015

No. 4

Authors: *Makram El-Shagi*  
Henan University, Kaifeng  
HuangHe Street, Dingli International 11-2 # 18-2,  
475000 Kaifeng, Henan, China  
E-mail: [Makram.ElShagi@gmail.com](mailto:Makram.ElShagi@gmail.com)

*Gregor von Schweinitz*  
Martin Luther University Halle-Wittenberg, Chair of Macroeconomics  
Halle Institute for Economic Research (IWH)  
Department of Macroeconomics  
E-mail: [Gregorvon.Schweinitz@iwh-halle.de](mailto:Gregorvon.Schweinitz@iwh-halle.de)  
Phone: +49 345 7753 744

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of the IWH. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors.

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

Editor:

**Halle Institute for Economic Research (IWH) – Member of the Leibniz Association**

Address: Kleine Maerkerstrasse 8, D-06108 Halle (Saale), Germany  
Postal Address: P.O. Box 11 03 61, D-06017 Halle (Saale), Germany  
Phone: +49 345 7753 60  
Fax: +49 345 7753 820  
Internet: <http://www.iwh-halle.de>

ISSN 1860-5303 (Print)  
ISSN 2194-2188 (Online)

---

# The Joint Dynamics of Sovereign Ratings and Government Bond Yields\*

## Abstract

In the present paper, we build a bivariate semiparametric dynamic panel model to reproduce the joint dynamics of sovereign ratings and government bond yields. While the individual equations resemble Pesaran-type cointegration models, we allow for different long-run relationships in both equations, nonlinearities in the level effect of ratings, and asymmetric effects in changes of ratings and yields. We find that the interest rate equation and the rating equation imply significantly different long-run relationships. While the high persistence in both interest rates and ratings might lead to the misconception that they follow a unit root process, the joint analysis reveals that they converge slowly to a joint equilibrium. While this indicates that there is no vicious cycle driving countries into default, the persistence of ratings is high enough that a rating shock can have substantial costs. Generally, the interest rate adjusts rather quickly to the risk premium that is in line with the rating. For most ratings, this risk premium is only marginal. However, it becomes substantial when ratings are downgraded to highly speculative (a rating of B) or lower. Rating shocks that drive the rating below this threshold can increase the interest rate sharply, and for a long time. Yet, simulation studies based on our estimations show that it is highly improbable that rating agencies can be made responsible for the most dramatic spikes in interest rates.

Keywords: sovereign risk, rating agencies, semiparametric model, nonlinearities

JEL Classification: C14, C25, F34, G24

---

\* Parts of this research were conducted while Gregor von Schweinitz was visiting scholar at the Henan University, Kaifeng. Research of Gregor von Schweinitz was partly funded by the European Regional Development Fund through the programme „Investing in your Future“ and by the IWH Speed Project 2014/01. The authors are grateful for the comments by participants of the research seminar at the Wang Yanan Institute for Studies in Economics (WISE) at Xiamen University.

# Die gemeinsame Dynamik von Ratings und Zinsen auf Staatsanleihen\*

## Zusammenfassung

Das Paper beschäftigt sich mit der Frage, ob es nach einer Abwertung eines Staates zu einer Dynamik von steigenden Zinsen und weiter fallenden Ratings kommen kann, die unausweichlich in einem Staatsbankrott endet. Die hohe Persistenz von Ratings sowie die negativen Auswirkungen einer Abwertung auf Zinsen legt die Möglichkeit einer solchen Dynamik nahe. Wir zeigen, dass diese Schlussfolgerung zu kurz greift. In den Daten ist im Gegenteil ausschließlich eine sehr langsame Annäherung an ein langfristiges Gleichgewicht von guten Ratings und niedrigen Zinsen zu beobachten. Gleichzeitig ist die Persistenz von Ratings allerdings hoch genug, um nach einer Abwertung auf ein hochspekulatives Niveau (Rating von B oder schlechter) massive und langandauernde Zinsaufschläge zu erzeugen. Da eine solche Abwertung in der Realität allerdings äußerst selten erfolgt, müssen wir die Existenz des oben beschriebenen Teufelskreises verneinen. Eine negative Entwicklung wie zum Beispiel in Griechenland in den Jahren 2010 und 2011 lässt sich nicht aus der Dynamik von Ratings und Zinsen erklären.

Schlagwörter: Souveränrisiken, Ratingagenturen, semiparametrische Modelle, nicht-lineare Modelle

JEL-Klassifikation: C14, C25, F34, G24

---

\* Ein Teil der Forschungsarbeit wurde während eines Gastaufenthalts von Gregor von Schweinitz an der Henan University, Kaifeng, erstellt. Die Forschung von Gregor von Schweinitz wurde teilweise mit Mitteln der Europäischen Union, Europäischer Strukturfonds für regionale Entwicklung „Investition in Ihre Zukunft“ und durch das IWH Speed Project 2014/01 gefördert. Für die Kommentare von Teilnehmern des Forschungsseminars am Wang Yanan Institute for Studies in Economics (WISE) der Xiamen University sind die Autoren dankbar.

# 1 Introduction

Credit rating agencies – especially the so called Big Three: Moody’s, Standard and Poors (S&P) and Fitch IBCA – took heavy blame after the recent financial crisis and the subsequent (and partly still ongoing) great recession. When a crisis happens, it is almost tautological to note that rating agencies fail to predict it. After all, if rating agencies had foreseen the high risk of specific assets, whether those are senior tranches of asset backed securities, credit default swaps or just sovereign bonds, the over-investment in this specific asset class would probably have been avoided. However, in particular regarding government bonds, which are the main interest of this paper, the most frequently voiced concern is not the rating agencies’ failure to predict, but the possibility that bad ratings cause capital flight, driving the risk premium up, thereby causing further problems that are sanctioned with another rating downgrade. When S&P downgraded the French rating from triple to double A on August 11, 2013, the first response of the French government (through the minister of finance Pierre Moscovici) was to criticize the “critical and inexact” decision. It is obvious why politicians favor this view of a vicious cycle that can befall the best of us. However, this argument essentially builds on a result that was first put forward by Ferri, Liu & Stiglitz (1999) in their widely cited paper on the Asian Flu in the late 1990s. Yet, their result is far from uncontroversial. The original study has been harshly criticized by Mora (2006) and El-Shagi (2010). Similarly, more recent papers, who address the same question in the context of the European debt crisis, come to very different conclusions. While Baum, Karpava, Schäfer & Stephan (2014) find evidence for a substantial impact of ratings on capital allocation, De Vries & De Haan (2014) focus on the finding that the volatility following the rating was only temporarily.

The seeming contradiction between the latter two contributions highlights a key omission in the literature that the present paper aims to fill. A lot of the literature criticizing rating agencies focuses on the short term impact, or aims at showing that there is some arbitrariness to ratings (e.g Bolton, Freixas & Shapiro 2012). Nevertheless, as pointed out by El-Shagi & von Schweinitz (2015), neither of those is sufficient to establish empirical evidence for a vicious cycle between ratings and the risk premium that can drive a country from a good to a bad equilibrium. Even if a rating downgrade does increase the interest rate, the new high interest is merely paid on new and rolled over debt. That is, if the average maturity is not extremely short (which it usually only is in countries that have bad ratings to start with), the increase of the interest rate has to be sustained over a long time, to actually increase the fiscal burden. At the same time, the literature on the long run effect frequently merely estimates long run equations (e.g. De Vries & De Haan (2014) and Gärtner & Griesbach (2012)). This does not only cause identification issues, since the stationarity of ratings and yields is questionable, but more importantly misses the economic problem, that long lasting problems might impose high costs even if they are technically transitory.

In our paper, we augment a recently suggested approach by El-Shagi & von Schweinitz (2015), who simultaneously model the long run relationship and the short run dynamics in a model that explicitly allows for multiple equilibria. While the original paper pays little attention to the dynamics, mostly modeling them to allow robust inference on the long run relationship, the current paper focuses on the joint dynamics of ratings and yields. We augment the dynamic part of their model substantially, most importantly by allowing for asymmetric effects of both rating and yield changes.

We confirm their finding that there is strong evidence for a single good equilibrium. At

no point the typical risk premium associated with a rating is sufficient to justify further rating downgrades. Yet, we find that downgrades can come at a substantial cost. In the short run, rating changes tend to reinforce themselves, increasing the risk of further downgrades in the short run. If this drives the rating far below the investment thresholds, this decline in the risk assessment is associated with a substantial increase in the interest rate. Due to the extremely high persistence of ratings, the impact of a rating change can be felt for more than a decade. Yet, simulation studies based on downgrade episodes from the past decades show, that especially the more severe episodes cannot be explained through the common joint dynamics of ratings and yields, and are thus most likely driven by an actual change in the fundamentals (or a correction in their assessment).

The remainder of the paper is structured as follows. In Section 2 we briefly review the previous empirical findings, provide some introductory stylized facts on the dynamics of ratings and yields and discuss some of the associated measurement problems. In Section 3 we explain our econometric model and the methods employed. Section 4 presents our results on the long run and short run relation of ratings and yields, including the scenario simulations reproducing previous downgrade episodes using our model. Section 5 concludes.

## 2 Previous evidence and Stylized facts

In particular the critics of rating agencies believe them to hold – as Ferri et al. (1999) puts it — ”a tremendous power“. They claim that the major rating Moody’s, Standard and Poors (S&P) and Fitch agencies – frequently dubbed *the big three* – strongly shape the behavior of international investors, despite adding only limited information.

There is a long and extensive literature that aims to identify the factors that influence sovereign credit rating decisions. In general, debt sustainability measures, the degree of economic development and the default history (Cantor & Packer 1996, Gärtner, Griesbach & Jung 2011) as well as political stability and governance indicators (Mellios & Paget-Blanc 2006) are found to be important. One of the most general results in explaining ratings is that agencies react to past fundamentals rather than successfully predicting future shocks (Cantor & Packer 1996, Reisen & von Maltzan 1999). However, while providing only little information in normal times, rating agencies aim to reestablish their reputation after missing the emerging crisis, responding with overly restrictive downgrades (Ferri et al. 1999). This creates capital flight, increasing the interest rate and thus the fiscal burden of the government which is driven even closer to an actual default. In the case of sovereign ratings, this problem is particularly severe, since sovereign ratings also serve as so called sovereign ceiling, that is, a best-case rating for all but a few companies with headquarters in that country (Durbin & Ng 2005). Therefore, the capital flight can also hit the private sector, slowing down the economy and eroding the tax base. This problem is augmented by the reliance of regulators on ratings. These regulations often indicate that assets with ratings below a certain threshold are not considered ”investment“ but speculation and are thus strongly restricted or penalized (White 2010). Of course this argument of a self fulfilling prophecy is controversial. E.g. El-Shagi (2010) points out the inconsistency in simultaneously claiming that rating agencies obviously and systematically err, and that the majority of investors does not recognize this and follows the rating agencies despite their shortcomings.

Yet, the responsibility of rating agencies for financial market turbulences is a well estab-

lished narrative, in particular in politics. The ongoing debate in the European Union about the need of a big European rating agency is a textbook example illustrating the magnitude of the influence that is attributed to rating agencies by politicians. The fear of the big three – with headquarters all based in the US – became so severe that even major politicians in the European Union (such as Rainer Brüderle, at the head of the FDP (libertarian) group in the German Bundestag and the former federal minister of economic affairs) became worried that American institutions tried to deliberately exploit their influence to hurt the Euro (area).

The original example that fueled the debate about the dangers of rating agencies is the Asian Flu in the late 1990s. Ferri et al. (1999) argue that they played a significant role in accelerating the crisis. Their argument is mainly based on the first downgrades in Thailand (October 1997, from A- to BBB), Malaysia (December 1997, from A+ to A) and Indonesia (December 1997, from BBB to BB+). Their conclusion is challenged from two directions: Mora (2006) finds ratings to be sticky rather than procyclical. El-Shagi (2010) goes one step further and documents that there were many rating adjustments following these first downgrades, the last and most significant of them occurring only shortly before the end of the crisis and sometimes even after that. That is, there is at least as much evidence that rating agencies merely follow the market rather than triggering or worsening a crisis by downgrading a country.

The second prime example brought forward by critics of rating agencies relates to the ongoing European debt crisis. It was argued that actions of (US-based) agencies unduly increased market pressure on European periphery countries, increasing their government bond yields to unsustainable levels and thus triggering a public debt crisis with severe long run macroeconomic costs. Arezki, Candelon & Sy (2011) find that some downgrades in the Euro area, such as the one of Greece from A- to BBB+ by Fitch on December 8, 2009, had a systematic spillover effect to other European countries (see also Beirne & Fratzscher 2013). That is, the downgrade of Greece is found to have increased not only CDS-spreads of government bonds (as a measure of credit default risk) in Greece, but also in a number of other European countries. The authors claim that these spillovers alone may trigger further financial instability. However, their reported results are quite heterogeneous and may thus not be strong enough to support their claim in a more general setting. Afonso, Furceri & Gomes (2012) have a much more balanced view on the question at hand. First, they find a strong and asymmetric effect of rating changes. They also find, that these changes are mostly unanticipated one to two months ahead (with evidence of limited bidirectional Granger causality), concluding that there may be some limited evidence for market power of rating agencies. However, the reaction of markets is quick, which partly refutes the argument that ratings can actually drive a country into default. A similarly balanced view is taken by Gärtner et al. (2011), who use yearly data and a fundamental estimation of ratings as in Cantor & Packer (1996) to determine non-fundamentally justified parts of ratings. They argue that these parts of rating decisions also affect yield spreads. That is, an erroneous (arbitrary) downgrading decision might trigger yield increases, which would open the possibility of further downgrades in the future. This result is somehow challenged by the finding of De Vries & De Haan (2014), that credit ratings and yields have disentangled in the most recent past: after summer 2012, yield levels in European periphery countries dropped fast while ratings stayed at very low levels. The authors attribute this to either unconventional monetary policy or increasing conservativeness of credit rating agencies. However, their econometric model does not include short run effects and does allow only to a very small extent for different

effects of different rating levels. Therefore, their model could be misspecified, and the slow adjustment of ratings could simply be due to their general stickiness.

After briefly introducing our dataset, we will present a few stylized facts regarding the joint dynamics of ratings and yields to motivate our own econometric approach in the following subsections. In particular, we will argue that it is (a) necessary to consider short and long run effects in the model, that should (b) consider nonlinearities in the long run relation, and (c) asymmetries in the short run relation.

## 2.1 Measurement and sample selection

**Ratings:** To maximize data coverage, we use ratings of foreign-currency denominated government bonds as provided by Moody's, Standard and Poors (S&P) and Fitch.<sup>1</sup> Due to higher international demand, foreign-currency bonds have often been rated before the agencies considered domestic-currency denominated bonds. Yet, the ratings for different bonds issued by the same country are generally highly correlated.

The three agencies use different grades in order to assess the probability of a default over the medium- to long-term horizon, where better grades correspond to lower default probabilities. The names of grades differ across agencies. However, this difference is just notational (Cantor & Packer 1996). Therefore, grades can be easily compared and transformed to a numeric scale, as shown in table 1. In the rest of the paper, we will use the notation of S&P.

As market movements are often found to be strong around rating announcements (which provides new signals), the most important signal is probably provided by the first agency to adjust its rating. However, there seems to be some evidence for a specialization and leadership of agencies to specific markets (Hill & Faff 2010), which is why we should include information of all agencies rather than concentrating on a single agency. In addition to accounting for the timeliness of new information, average ratings provide an implicit safeguard against random judgment errors.

Where a rational representation of the rating is required, we use the mean rating of all three agencies (see also De Vries & De Haan 2014). However, for most of our analysis, we aim to maintain the ordinal nature of ratings, contrary to the majority of the literature considering rating levels. Our rating class dummies are generated by rounding the mean rating to the next integer and considering the joint rating as belonging to the corresponding rating class as defined by Table 1. It is fairly well-documented, that ratings of different agencies seldom differ by much, even in times of higher uncertainty, when rating agencies adjust their assessments at higher frequencies (Ferri et al. 1999).<sup>2</sup> Therefore this averaging mostly corresponds to the majority rating of the three agencies.

**Yields:** We measure the risk premium through the real government bond yield on sovereign bonds with a maturity of 5 to 10 years, denominated in domestic currency. Again this choice is mostly enforced by data availability and feasibility reasons. While real yields are strongly driven by default risk, they simultaneously capture other factors, such as the degree of market liquidity and global risk aversion (von Hagen, Schuknecht & Wolswijk 2011). Thus, they go slightly beyond the claim of rating agencies to consider solely the probability to default. The pure credit default risk component could also

---

<sup>1</sup>The data are collected from <http://countryeconomy.com/>

<sup>2</sup>Empirically, the standard deviation across agencies is 0.77 notches, the mean absolute difference to the average is 0.5 notches. That is, rating agencies are on average far less than one notch apart.



Table 1: Rating grades and transformation

Grade	Moody's	S&P	Fitch	Assigned Value
Prime	Aaa	AAA	AAA	24
High grade	Aa1	AA+	AA+	23
	Aa2	AA	AA	22
	Aa3	AA-	AA-	21
Upper medium grade	A1	A+	A+	20
	A2	A	A	19
	A3	A-	A-	18
Lower medium grade	Baa1	BBB+	BBB+	17
	Baa2	BBB	BBB	16
	Baa3	BBB-	BBB-	15
Non-investment grade speculative	Ba1	BB+	BB+	14
	Ba2	BB	BB	13
	Ba3	BB-	BB-	12
Highly speculative	B1	B+	B+	11
	B2	B	B	10
	B3	B-	B-	9
Substantial risks	Caa1	CCC+	CCC+	8
Extremely speculative	Caa2	CCC	CCC	7
	Caa3	CCC-	CCC-	6
In default with little prospect for recovery	Ca	CC	CC	5
		C	C	4
In default	C	D	DDD	3
			DD	2
			D	1

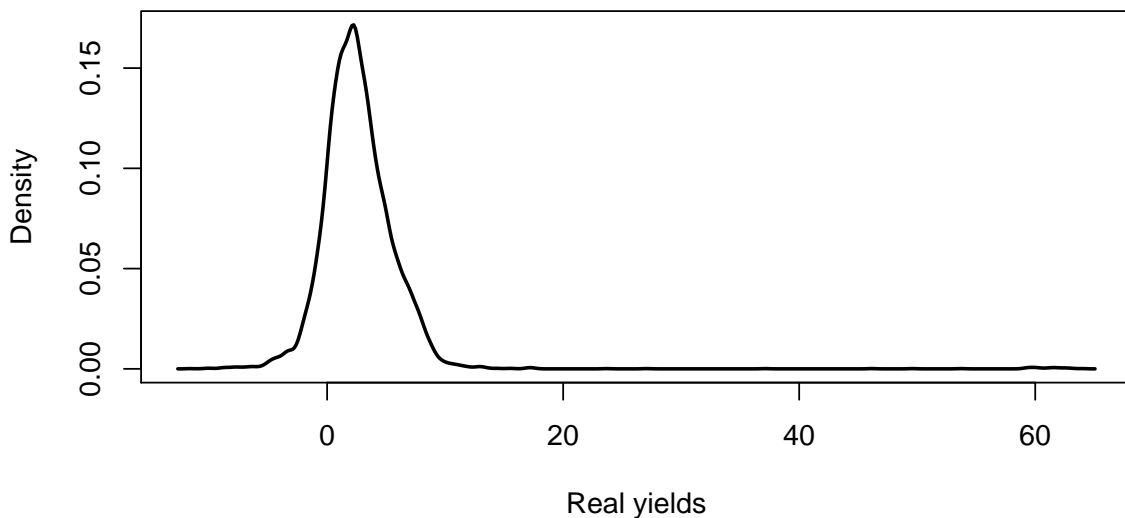


Figure 1: Density of yields

be captured by the prices of credit default swaps, which are essentially insurance contracts against the event of default. However, using CDS prices would reduce our sample significantly, especially reducing the number of observations with low ratings.

Contrarily, yields on domestic currency bonds are widely available for a broad range of countries for extended periods. The 5 to 10 year maturity mirrors the risk horizon of rating agencies. In order to make yields in different countries comparable we deflate yields using year-on-year inflation of the previous 12 month. While past inflation is not a perfect measure of inflation expectations that would be required to compute expected real returns, inflation time series are available for far more countries and periods than survey data or other more direct measures of expectations.

Figure 1 shows the distribution of those real yields. In around 85% of the periods yields are positive, and they have a mean and median of close to 2.5%. This result can be expected as investors would only be willing to accept yields below current inflation rates if they can be nearly certain that the latter will fall in the near future. Table 5 gives the impression that the distribution of yields is not too different between countries. Only in Greece, between October 2011 and January 2014, have real yields been higher than 20%.<sup>3</sup> In all other countries, high nominal yields were usually accompanied by high inflation rates, leading to a narrow distribution of real yields.

On average, we would expect developing countries to have higher credit risk and inflation volatility, contributing both to higher average real yields. However, these expectations are not fulfilled by the summary statistics in table 2.<sup>4</sup> There, advanced economies have on average higher real yields. This counterintuitive result can largely be attributed to Greece. If we exclude it, average yields are comparable between groups, with a much

<sup>3</sup>As a second exception, the nominal yield in Sri Lanka exceeded 450% in July 2010. This value is so unreasonably high that we exclude it as the only outlier in our estimation.

<sup>4</sup>The classification of countries follows (IMF 1997). A detailed list of countries with data availability can be found in table 4 the appendix.

lower standard deviation (i.e., volatility) in advanced economies.

As mentioned above, government bond yields are affected by more than one risk channel. Especially in times of turmoil on financial markets, market participants may shift their portfolio towards government bonds they deem safe (Vayanos 2004, Beber, Brandt & Kavaiecz 2009). This safe-haven effect (which is essentially herding behavior) may lead to self-fulfilling crises in other countries, when creditors with a risk slightly worse than “safe” (but far from default) are shunned by financial markets (De Grauwe & Ji 2013). Herding behavior also reinforces a tendency to treat superficially comparable countries similarly, without performing in-depth analyses of individual debtors. This, in turn, may lead to spillovers of risk from one country to the next (Beirne & Fratzscher 2013). That is, shocks to yields may be correlated across countries.

Table 2: Summary statistics of ratings and yields

Variable	IMF classification	mean	sd	min	max
ratings	total sample	20.75	3.93	4.50	24.00
	Advanced	22.51	2.37	4.50	24.00
	Developing	14.84	2.76	7.67	20.67
	Transition	18.20	1.96	13.67	22.00
yields	total sample	2.74	4.06	-11.59	64.00
	Advanced	2.83	4.35	-8.63	64.00
	Developing	2.53	3.27	-11.59	17.34
	Transition	2.28	2.07	-3.10	9.78

**Sample:** The afore defined indicators are available for an unbalanced monthly unbalanced panel of 46 countries from January 1980 to January 2014. It covers 27 advanced economies, as well as 4 eastern European transition economies and 15 developing economies. Especially transition and developing economies often have low data availability: there are several countries where yield data are only available after 2001. However, we are still able to work with roughly 9’100 observations in total.

## 2.2 Stylized Facts

**Nonlinearities:** As ratings measure “risk”, it is not very likely, that ratings (or their assigned values) can be used linearly. In the present context, this limitation holds even more, when the potential effect of an investment grade threshold and the possibly non-linear relationship between “risk” and sovereign yields is taken into account. Therefore, different authors used many different possibilities to transform ratings, when they wanted to explain (medium-run) yield movements.<sup>5</sup> Larraín, Reisen & von Maltzan (1997) test both a linear and a logistic transformation of ratings; Ferri et al. (1999) provide evidence for an exponential conversion, which is also used (together with linear and cubic conversions) by (Gärtner & Griesbach 2012).

Given that there is no agreement on how to transform ratings such that a linear relationship between transformed ratings and yields can be expected, we think that a more

<sup>5</sup>In event studies with windows of only a few days, the nonlinearity of ratings does not play such a huge role.

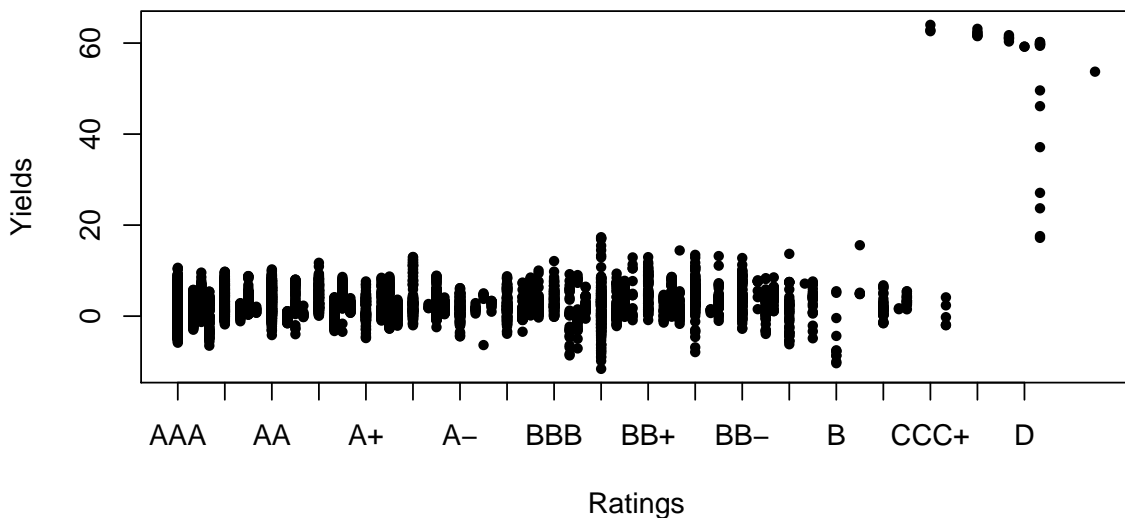


Figure 2: Scatterplot of ratings and yields  
*Note:* Rating scale (notation of S&P) inverted.

flexible transformation should be employed. This will be presented in the following section, together with our way of dealing with the short run and long run interaction of ratings and yields.

In Figure 2, it can be seen that the level relation between ratings and yields is basically flat for all but the lowest rating classes, even well below the investment grade threshold below BBB-. How can this “non-relation” be accommodated with the anecdotal evidence and the reasonable assumption, that (inverted) ratings and yields should be positively correlated, as they are both measuring the same thing? First, ratings are constant for long periods of time, in which yields may slowly adjust to new risk levels. Second, real yields are affected by many more factors in addition to sovereign risk. Rather than interpreting the flat slope as absence of a risk premium, it should be interpreted as a risk premium in an order of magnitude that is overshadowed by the general variance of interest rates. Only if risk becomes substantial, the risk premium starts to quickly increase. This is not purely related to the nonlinearity on risk measurement, but is actually in line with theory, that predicts the risk premium to go to infinity when the default probability becomes a certainty.

**Persistence:** Figure 3 shows the histogram of average ratings. Close to 60% of our ratings is high-grade, which can partly be explained by the greater data availability for advanced economies (table 4), which tend to have higher ratings. Ten of these countries have never (in our dataset) received a rating below AA, while only two industrialized countries (Greece and Israel) never achieved a rating above A+ (a “high grade”), see table 5. Transition and developing countries, on the other hand, tend to have much lower ratings. This is most visible in the mean rating reported in table 2. Among that group, only Slovenia achieved to get a “high grade” rating until they were downgraded in course of the European debt crisis in January 2012. Yet, the distribution features no bimodality

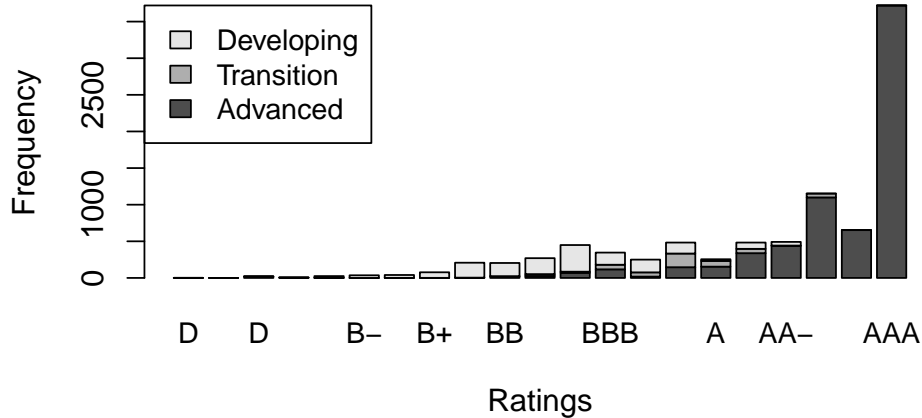


Figure 3: Histogram of ratings

that suggests that countries beneath a certain rating face consecutive downgrades until they eventually default.

While this is at least some indication for convergence to a good equilibrium, ratings feature an enormous persistence. The share of periods with a rating change is just 3.5%. In our whole sample, we observe 187 upgrades and 133 downgrades (slightly larger than upgrades on average). This low share of periods with rating changes is not driven by ratings already having reached the peak ratings. Even when excluding observations with a rating of AAA and AA+, the probability of a rating adjustment barely exceeds 5%. This degree of stickiness makes an analysis in a traditional AR framework difficult, even if the variables of interest might be technically stationary in the sense that they slowly return to a unique equilibrium (rather than an equilibrium curve as in a cointegration setting). Yet ignoring the short run dynamics would imply ignoring the shocks that drive ratings away from this long run equilibrium, which is why it is crucial to have a model that combines short and long run effects.

**Asymmetries:** Even if change happens, ratings move very slowly, see figure 4: if one or more agencies adjust their assessments, the average moves by at most one notch in more than 80% of downgrades and 95% of upgrades. Yet, another reading of those numbers is, that rating movements of more than one rating step are four times more likely for downgrades than they are for upgrades. That is, while we often observe a staggering of rating adjustments, with often one or two agencies moving first, and the third following the next month, the downwards dynamics seem to be much more intense than the recovery is. The strong downgrades occur most often in advanced economies (where there is admittedly more room for downgrades on average). This impression is confirmed if we look at longer horizons. There, we can see that an initial downgrade may trigger several more in advanced economies, which amounts to a large total downgrade: the six largest cumulative rating downgrades over one year, between 4.5 and 9.5 notches,

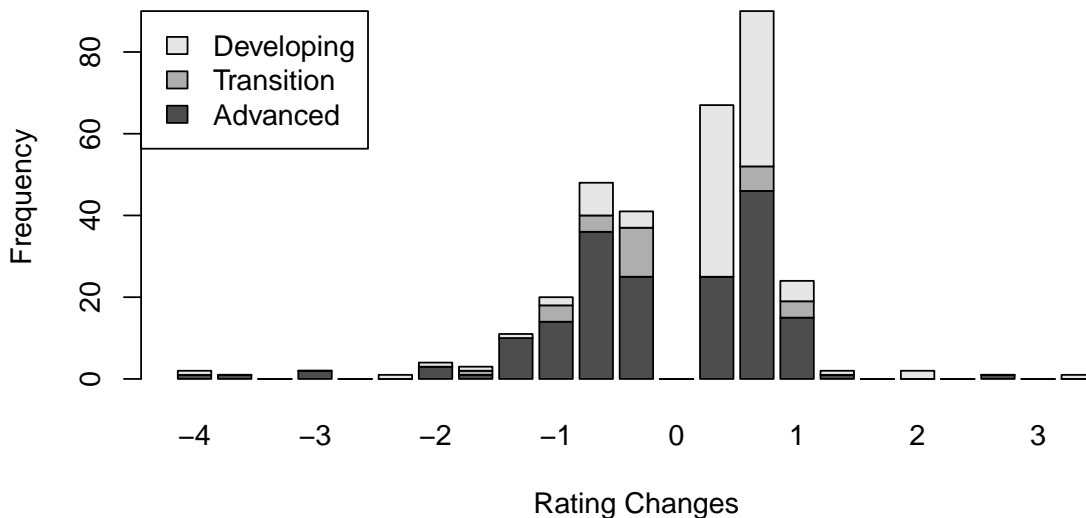


Figure 4: Histogram of rating changes

*Note:* Differences between bars are  $1/3$ , i.e., a rating change by one notch by one of the three agencies. Periods with no rating changes excluded

happened in advanced economies. On the other hand, cumulative rating improvements are more or less equally distributed over advanced, transition and developing economies. The asymmetry is not limited to the magnitude of change, but more importantly to the dynamics of change. Figure 5 shows the development of yields (normalized to 100 in the month of the rating change) at longer horizons of 12 months before and after a rating change. While this provides no conclusive evidence, the figure allows to roughly identify two stories. In periods of a downgrade, the rating change more often than not seems to happen jointly with peak in the interest rate. Yet, in times of upgrades, the rating changes are adjustments during an ongoing movement of declining risk premia.

While the general finding of an asymmetry is shared by most of the rich literature on event studies of rating changes, many of those studies imply much stronger interest rate dynamics. However, most of those studies (e.g. Ferri et al. (1999), Kiff, Nowak & Schumacher (2012), Afonso et al. (2012)) suffer from the fact that they use comparably short windows (between 2 weeks and 2 months) before and after a rating announcement. While the short run fluctuations they analyze may be highly relevant for speculation purposes, they seem to be negligible when it comes to a longer perspective.

In general, event studies usually find strong effects of rating announcements on yields during the days both before and after the event. It is sometimes argued, that this is a sign of an anticipation effect (Hill & Faff 2010). However, the argument that ratings are sticky (Mora 2006) or lagging (Reisen & von Maltzan 1999, El-Shagi 2010) seems to be more convincing. That is, we should expect Granger causality in both directions in general and not only for short windows around events.

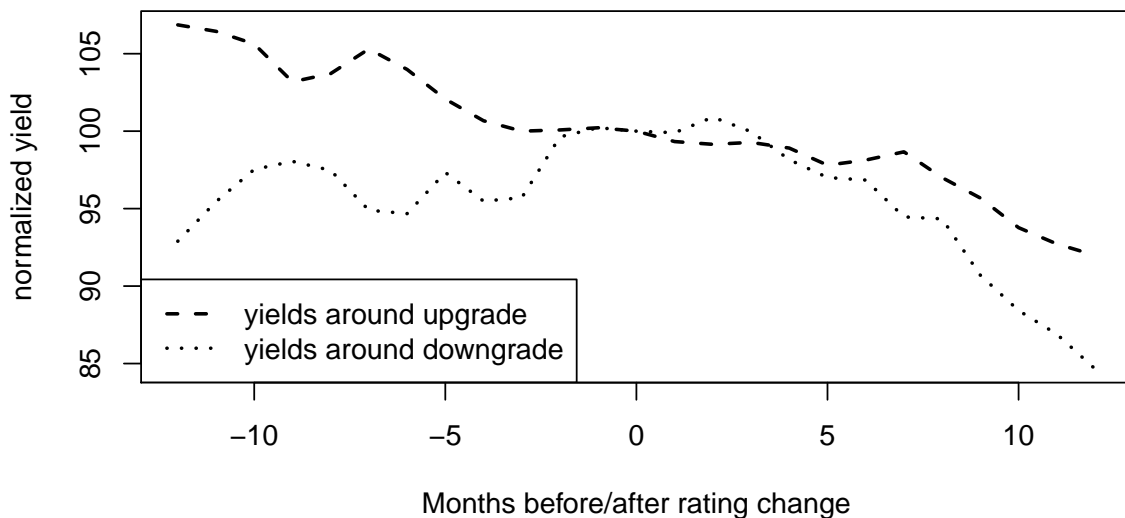


Figure 5: Development of yields in a two year window around rating changes

### 3 Model and estimation technique

#### 3.1 The model structure

In the following, we present and substantially extend the model by El-Shagi & von Schweinitz (2015). Our model is a simultaneously estimated bivariate two equation model consisting of a continuous yield equation and an ordered probit ratings equation. Each individual equation is inspired by the structure proposed by Pesaran, Shin & Smith (2001) in his seminal paper on the bounds cointegration test. That is, we explain the change of a variable through the lagged levels of the variables of interest (i.e. ratings and yields) and a range of lags of the first differences of those variables. The reason for this model setup is, that the equation structure estimated by Pesaran et al. provides valid inference (i.e. consistent and unbiased estimates) whether or not the included variables show unit root behavior or not.

This is particular important for us, since a key difference between our approach and typical cointegration approaches is that we allow for different long run coefficients in the two equations. I.e. rather than estimating *the* long run relation, we estimate the long run relation *implied by each* of the change variables (i.e. the first difference of interest rates, and the presence of up- or downgrades). Finding two individually significant, yet different long run relations in both equations implies that – while having enough persistence to differentiate between long and short run effects – the variables are technically not cointegrated in the traditional sense. Rather, they are either stationary or exhibit some kind of regime switching behavior.

Consider the exemplary long run relations featured in Figure 6 (a) and (b). Only in the first case (a), where *both* long run relations are the same, there is an infinite number of equilibria, in the sense that none of the points in the curve features a model intrinsic tendency for deviation (or no equilibrium in the sense that no single point acts as a per-

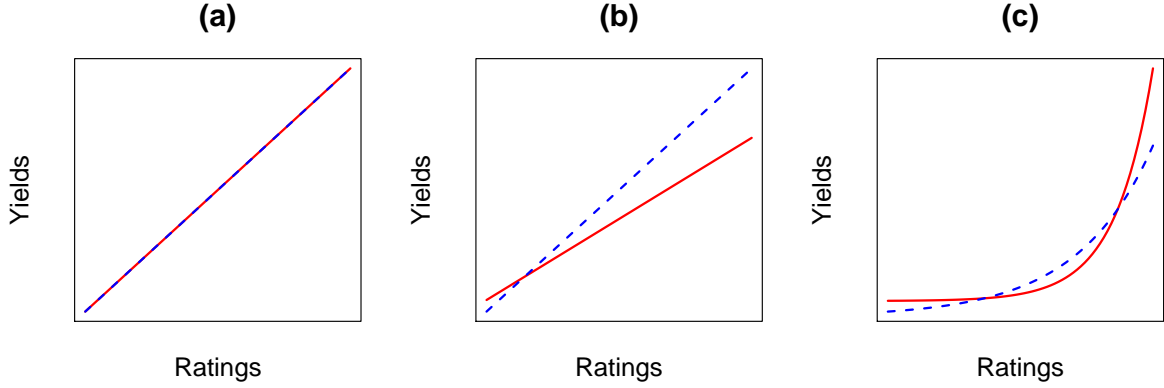


Figure 6: Exemplary long run relations of ratings and yields

manent center of gravity). That is, a shock can potentially have permanent impact, as in a nonstationary system. If the lines intersect once as in case (b), only this combination of yields and ratings is stable in the long run, and both yields and ratings will move towards that combination in the very long run. That is, despite high persistence as indicated by the very existence of a long run relationship, the system is essentially stationary. Since our model should host both, the traditional cointegration case (a) and the more complex case (b), it is essential for us to have a model that is still well identified whether or not there is stationarity.

In addition our model allows for nonlinearities in the long run relations, using a semiparametric approach to estimate a functional form over a set of rating dummies, rather than treating the (level) rating as continuous. Thereby, we allow the model to host situations with multiple equilibria. Thus our model is able to feature the frequently feared situation of a *good* and a *default* equilibrium (see Figure 6 (c)). Finally, our equations allow for asymmetric impacts of positive and negative changes of ratings and interest rates, thus allowing for the more complex dynamics previously documented in the literature.

**The yields model:** Our interest rate equation takes the form:

$$\begin{aligned}
 \Delta i_t = & \beta_0 + \beta_1 i_{t-1} + \sum_{c=7}^{24} \beta_c r_{c,t-1} \\
 & + \sum_{l=1}^{p_i} (\alpha_{l,p} 1_{\Delta i_{t-l} \geq 0} + \alpha_{l,n} 1_{\Delta i_{t-l} < 0}) \Delta i_{t-l} \\
 & + \sum_{l=0}^{p_i} (\gamma_{l,p} 1_{\Delta r_{t-l} \geq 0} + \gamma_{l,n} 1_{\Delta r_{t-l} < 0}) \Delta r_{t-l} + \varepsilon_t,
 \end{aligned} \tag{1}$$

where  $i$  is the interest rate,  $\Delta r \in \{-1, 0, 1\}$  the change of the rating,  $r_c$  a rating dummy,  $t$  the time index, and  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  an error term. We use up to  $p_i$  lags and the contemporaneous rating changes. The possibly asymmetric effect of past changes is captured by the different coefficients on positive ( $\alpha_{l,p}, \gamma_{l,p}$ ) and negative ( $\alpha_{l,n}, \gamma_{l,n}$ ) rating and yield changes. The rating dummies are defined such that  $r_c$  indicates that the rating is lower or equal than  $c$ . That is, for a rating of A the first 13 dummies (7-19) would be set to 1. For estimation purposes, we combine all default ratings (CCC- and below) in one



class, because only very few ratings at and below “Extremely Speculative” (CCC) are observed. This leaves us with 18 different rating dummies. Defining the rating dummies cumulative (i.e. a dummy equals 1, not only for countries with the respective rating, but also for countries with better ratings), simplifies assessing whether the difference of effects between adjacent rating classes is statistically meaningful.

**The rating model:** Exploiting the categorical nature of ratings, we estimate an ordered probit model for the change of the rating, again allowing for asymmetries:

$$\begin{aligned}
y_t^* = & \psi_1 i_{t-1} + \sum_{c=7}^{24} \psi_c r_{c,t-1} \\
& + \sum_{l=1}^{p_r+1} (\rho_{l,p} 1_{\Delta i_{t-l} \geq 0} + \rho_{l,n} 1_{\Delta i_{t-l} < 0}) \Delta i_{t-l} \\
& + \sum_{l=1}^{p_r} (\omega_{l,p} 1_{\Delta r_{t-l} \geq 0} + \omega_{l,n} 1_{\Delta r_{t-l} < 0}) \Delta r_{t-l} + \eta_t.
\end{aligned} \tag{2}$$

$$\begin{aligned}
\text{rating downgrade if: } & y_t^* < \mu_1, \\
\text{rating upgrade if: } & y_t^* > \mu_2
\end{aligned}$$

where  $y^*$  is the latent variable that is linearly dependent on the explanatory variables, and  $\eta_t \sim \mathcal{N}(0, 1)$  again an error term.

Our benchmark model uses monthly averages of the interest rate to avoid importing the daily variation into the monthly model, but rating levels from the last day of the month to preserve the categorical nature of ratings. Therefore, contrary to Pesaran et al. (2001) we do not use the contemporaneous first difference of the exogenous variable (i.e. the interest rate). Since our rating measure from the same period  $t$  is obtained slightly after the interest rate measure, we would implicitly explain ratings using interest rates from the (near) future. To account for this timing issue, we employ the lag  $p_r + 1$  instead.

**Inclusion of fixed effects:** Fixed effects are usually avoided in ordinal models, since they are no longer identified, when a cross sectional unit constantly is in one of the extreme groups. While this is true for the level of the rating (with several advanced economies always being rating AAA), this does not occur when using rating changes as dependent variable (since no country is permanently down- or upgraded). Thus, fixed effects could be easily included by adding (for  $N$  countries)  $N - 1$  fixed effects in every equation.

In our view, the inclusion of country fixed effects has both advantages and disadvantages. First, it accounts for possible differences between countries that may affect yields and ratings. Thus, we could implicitly capture effects on our two variables coming from other potentially explanatory variables. However, many rating classes are only observed in a small number of countries, and (conversely) only few countries experience a large range of different ratings. Therefore, fixed effects could also blur the relation between yields and ratings by capturing the effect of individual rating classes instead of the effect for individual countries. As an econometric distinction between these two very different channels is hardly possible, we present results without fixed effects as a baseline and use a fixed effects estimation in a robustness check.

### 3.2 Exploiting the ordinal nature of ratings to model nonlinearity

**The semiparametric estimation:** When estimating equations (1) and (2) through standard estimators, the representation of ratings through a series of dummies is problematic. Since some rating classes are merely observed in very few situations, time and country idiosyncrasies would drive the estimated coefficients rather than the actual impact of a rating of the corresponding class. While technically allowing “nonlinearities” in the impact of ratings (compared to treating ratings as pseudo continuous), this creates huge and economically unwarranted differences in the impact of very similar ratings. Essentially, although modeling ratings through class dummies, we would like to have a well behaved smooth function over rating classes, unless there is strong evidence (as in many data points) suggesting otherwise.

To achieve this objective we borrow from an approach suggested by Breitung, Roling & Elengikal (2013) for mixed frequency data sampling (MIDAS). In those MIDAS approaches, low frequency data (e.g. monthly inflation) is explained through high frequency data (e.g. daily oil price movements). Just estimating many coefficients for the high frequency lags usually yields substantial identification problems. The previous literature has dealt with this problem by restricting the coefficients on high frequency lags to follow a specific functional form that can be described by few parameters. However, Breitung et al. (2013) argues that this might be overly restrictive, and suggests a more flexible nonparametric approach. Instead of enforcing a specific functional form, Breitung et al. augments the objective function (such as the likelihood function) by a term that penalizes second differences between coefficients for lags of adjacent periods.

We employ the same strategy to enforce a smooth behavior of the impact of ratings. Since our ratings dummies are defined cumulatively, we do not restrict second but first differences between coefficient estimates. This is equivalent to minimizing second differences between mutually exclusive rating dummies, where the dummy equals 1 if and only if a country has the corresponding rating.

That is, instead of using the traditional likelihood functions for our models (denoted by  $LL_{model}$ ), we augment the likelihood functions as follows:

$$LL_{smooth} = \sum_{c=8}^{24} \ln(\phi(\sqrt{\lambda}(\beta_c - \beta_{c-1}), 0, \sigma_{model}^2)) \quad (3)$$

$$LL = LL_{model} + LL_{smooth},$$

where  $\phi(x, 0, \sigma_{model}^2)$  is the density of a normal distribution with mean zero and variance coming from the errors in the respective model ( $\sigma_\varepsilon^2$  or  $\sigma_\eta^2$ ), measured at point  $x$ .

By increasing the weight  $\lambda$  of the penalty  $LL_{smooth}$ , it is possible to enforce a smooth behavior of adjacent coefficients.<sup>6</sup> In the limit, when the penalty weight goes to infinity, the coefficients are forced to be identical, i.e. we would only estimate a single coefficient. If the weight of the penalty goes to zero, the results approach those of the unrestricted model, i.e. we would estimate 18 coefficients. Breitung et al. shows, that  $\lambda$  can generally be mapped on the effective loss of degrees of freedom. That is, we can use standard information criteria to select the degree of smoothing. This means, if high differences between the ratings are actually needed to explain the behavior of interest rates or ratings, we will chose low smoothing in the respective equation. On the other hand, if volatility

---

<sup>6</sup>The likelihood contains  $\sqrt{\lambda}$ , as equation (1) can be estimated by simple OLS with a quadratic penalty term (Breitung et al. 2013).

in those coefficients is merely driven by very few observations in specific classes, we will opt for smoothness.

In contrast to El-Shagi & von Schweinitz (2015) we estimate a single  $\lambda$  for both equations. This allows a more clear cut comparison to the model where the long run relation is restricted to be identical in both equations, which implicitly includes a single rather than individual  $\lambda$ 's. That is, a rejection of a single long run relationship can in our setup not be attributed to the influence of using different  $\lambda$ .

### 3.3 Dealing with heteroscedasticity

Especially the shocks to the interest rate behave far from the usual i.i.d assumptions. Rather, we find substantial heteroscedasticity, cross sectional correlation and heavy tails (as could, for example, be already inferred from the results of Arezki et al. (2011)). We take two measures to deal with those problems.

Firstly, we use a bootstrap to account for heteroscedasticity (both over time and cross countries) and cross sectional correlation. Since interest rates feature substantial heteroscedasticity both cross countries and over time, we use a bootstrap method accounting for both to produce confidence bands. Our approach is a model based bootstrap resampling residuals founded on the wild bootstrap originally proposed by Wu (1986). In the wild bootstrap, rather than resampling the original residuals, the bootstrap simulations are generated using error terms that are obtained by multiplying the original residual for the respective observation with a random multiplier  $v$  that is drawn from a distribution that makes sure that the expected value of  $\varepsilon v$  is 0, and the first few moments of the distribution of  $\varepsilon$  and  $\varepsilon v$  are identical or at least very close. In this paper, the random multiplier  $v$  is drawn from a 6 point distribution proposed by Webb (2013), which has been shown to have even more desirable properties than the traditionally used distributions suggested by Mammen (1993) and Davidson & Flachaire (2008). To reproduce the cross country correlation found in the original sample, we use the same multiplier for all countries at a given point in time, i.e.  $v_{i,t} = v_{j,t}$  for all pairs of  $i$  and  $j$ . This follows an approach suggested by Davidson & MacKinnon (2010) who used the same technique to reproduce the correlation between the residuals of several equations. For the lack of a better alternative, the probit equation is simulated using i.i.d. errors drawn from a standard normal distribution. The starting values for ratings and yields and the lagged differences are drawn from the empirical joint distribution. Since we need low ratings in the simulated sample to have identification of all coefficients, we combine the bootstrap with a reject accept algorithm, that discards those simulations, where extremely bad ratings - which are empirically also unlikely - do not occur. The reported confidence bounds are based on 1'000 accepted simulations of the entire two equation system. Tests with 5'000 simulations showed virtually no difference.

Secondly, we remove the most extreme outliers. To identify those, we rely on an initial maximum likelihood estimation of our system of equations. Those observations, where the residual in the interest equation of the ML estimated model has a probability of less than 1 in 100 Million, are removed from the sample. This results in dropping about 20 out of almost 10'000 observations, see table 6 in the appendix. The results reported in Section 4 are obtained from a second regression where those outliers have already been removed.

### 3.4 Assessing the long run relation

**Identifying the long run relation:** Including asymmetries in the model slightly changes how we have to treat the long run relation implied by each equation. Usually, we just set the first differences of all variables to zero on both sides of the equation, solving for the corresponding relationship of the level variables.<sup>7</sup> Of course, this approach relies on the idea that the expected value of the impact of a right hand side variable is zero, if the expected value of said variable is zero. However, this logic no longer holds in the case of asymmetric effects. Now, the impact of the lagged first differences of interest rates and the lagged rating changes is no longer zero on average, even in the event that we are in equilibrium and those deviations from zero are purely random. To approximately compensate the asymmetric effect of positive and negative deviations, we work with typical past changes of the interest rate. That is, rather than setting the past lagged change to zero, we replace all lagged changes by the standard deviation of changes observed in the data. Since rating changes are an extremely rare event, we keep the value of lagged rating changes at 0.

**Testing for the number of long run relationships:** To assess whether the model allowing for individual long run relationships in both equations outperforms the more traditional cointegration approach with identical long run relationships in both equations, we use a standard likelihood ratio test. To guarantee comparability, the likelihood ratio test is performed twice for pairwise identical  $\lambda$ . That is, we compare both models using the optimum  $\lambda$  as selected for the cointegration model, and we compare both models using the  $\lambda$  as selected for the model with individual long run relationships. The model comparison is based on the ML results for computational reasons.<sup>8</sup> This comes with a small caveat. Since the residuals are not normally distributed – which is the very reason for our bootstrapping procedure – the test results have to be interpreted with caution.

## 4 Results

### 4.1 The long run relation of ratings and yields

Our preferred model - allowing for different long run relations in both equations - strongly outperforms the more basic cointegration approach. As mentioned afore this comes with the caveat, that the LR test might not be reliable as a tool in the given circumstances with highly non normal residual. However, there is further evidence for the superiority of our benchmark model. Figure 7 plots the long run relationships as implied by our two equations including their confidence bounds. It is evident that the long run relations are not merely different for most rating levels, they differ economically and statistically significantly.

As theory predicts, both curves are significantly negatively sloped, implying that (a) ratings tend to deteriorate if the interest rate is high (dashed line, from equation (2)),

---

<sup>7</sup>As outlined in El-Shagi & von Schweinitz (2015) that corresponds to setting the left hand side of the ordered probit model equal to the mean of the thresholds, implying equal (and low) probabilities of up- and downgrades, i.e. an expected value of no adjustment at all.

<sup>8</sup>The clearly outperformed cointegration model takes an extremely long time to bootstrap. Since evidence from the model with multiple long run relations indicates that the coefficients only change slightly, we carry over the ML results to save running a fully fledged bootstrap of the cointegration system.

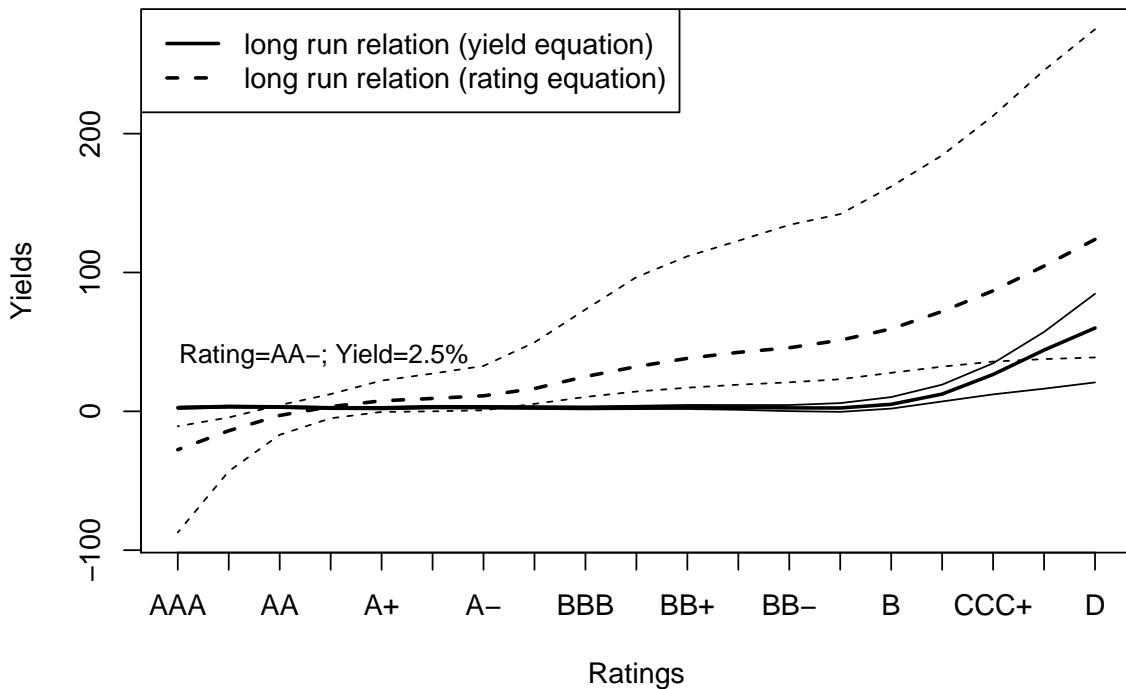


Figure 7: Long run relation of ratings and yields, accounting for asymmetric shocks  
*Note:* In addition to median long run relationship, we show the 16%- and 84% confidence bands from the bootstrap.

and that (b) the risk premium (measured through the interest rate) increases for countries with bad ratings (solid line, from equation (1)). However, as indicated by the stylized facts presented in Section 2, we only find a risk premium of noteworthy magnitude at extremely bad ratings. Only far below the investment grade threshold at and below rating B, the equilibrium interest rates as indicated by equation (1) start to pick up, quickly reaching extreme heights.

This behavior of the interest rate is one of the main reasons why our point estimates indicate only a single intersection of the long run relations which happens at top ratings and very low interest rates. When the rating is below AA- the interest rate implied by the current rating is so low, that it implies pressure for a rating upgrade. This is particularly true for bad ratings, which are thus highly transitory.

However, the confidence bands of the rating equation grow extremely wide for bad ratings. Yet, the impression of a possible second equilibrium is partly driven by the reporting of independent confidence bounds for both equations. When assessing the existence of a second equilibrium for each bootstrapped set of long run relationships individually, we see that we can still significantly reject the existence of a bad equilibrium. For 98% of the bootstrapped sets of long run relations we find that the equilibrium interest rate (according to the yield curve) creates upgrade pressure in the rating equation. Even in the unlikely case that a second equilibrium exists, the second equilibrium is unstable in the sense that there is no process driving the rating into that bad equilibrium. Essentially, a second intersection of the rating and yield curves below a rating of CCC would not

imply that a country can be driven into default by a vicious cycle of ratings and yields, but merely that a country needs a positive impulse to leave the default ratings. Since negative rating shocks are impossible at a D rating, the eventual positive shock initiating recovery would still be inevitable.

As a robustness test, we also run the model with country specific effects. Since this implies different intercepts in the long run relationships, each country now has a separate equilibrium. For all countries except Pakistan, this equilibrium corresponds to a yield of less than 7 percent; and just as in the baseline estimation, there is no evidence for a second intersection in any country. Regarding the dynamics, as outlined in the following subsection, the results are even more similar and are thus not discussed separately.

## 4.2 Short run dynamics and impulse-response functions

Although there is no gravitational pull towards a bad equilibrium, even starting from a relatively high risk level, our results indicate a substantial level of persistence, in particular for ratings. Despite this persistence the macroeconomic consequences of a downgrade are limited in most cases due to the afore mentioned little impact of ratings on interest rates, at most rating levels. If, however, a rating downgrade drives the rating below B grade, the impact is considerable. Due to the steep increase in the risk premium below this rating level, the interest rate quickly rises to partly extreme levels, and remains high over extended periods due to the afore mentioned persistence of ratings.

The development of rating and yield levels does not only depend on the weak pressure towards a stable long run equilibrium at high ratings and low yields, but also on the short run dynamics. In fact, the argument that an unduly large downgrade may unleash a vicious circle leading to default focuses almost exclusively on these short run dynamics. Therefore we need to analyze these dynamics after a negative rating shock to decide if empirical data indeed support such an argument. To do that, we simulate impulse-response-functions (IRF) for ten years (120 periods) after a large downgrade by two notches using the bootstrapped coefficients.

The standard way to calculate an IRF is to simulate a shock leading away from equilibrium. In our context this is, however, inadequate. Recovery from default ratings to the AA equilibrium takes a century in the median (see Table 3). Thus, it is apparently more than likely that a country is hit by a shock while still being far away from equilibrium. Since the nonlinearity in the risk premium causes shocks to have a pronouncedly different impact depending on the original rating, it is essential to consider shocks to countries with different risk assessment. Thus, rather than reporting a single impulse response function starting in the long run equilibrium, we report a range of impulse response functions for a negative 2 notch rating shock starting at the equilibrium recovery path but at different original rating levels and report the difference to the recovery path without the initial rating shock. The resulting IRF of yields is presented in figure 8, the IRF of ratings is given in figure 9.<sup>9</sup>

---

<sup>9</sup>While impulse response functions are usually computed deterministically, i.e. without further shock, this is apparently unfeasible in our case due to the probit equation. At each individual point in time, the most likely outcome is no rating adjustment. Yet, over time, the cumulative probability of a rating adjustment is increasing. Thus, the impulse responses are computed as the median of a range of simulations where both ratings and yields are subject to disturbances. By taking the median over many observations we remove the idiosyncratic effect of shocks. Rather than computing the difference between the median development with the initial rating shock, and the median development without said shock, we always run a simulation with and without the 2-notch shock with identical simulated disturbances.

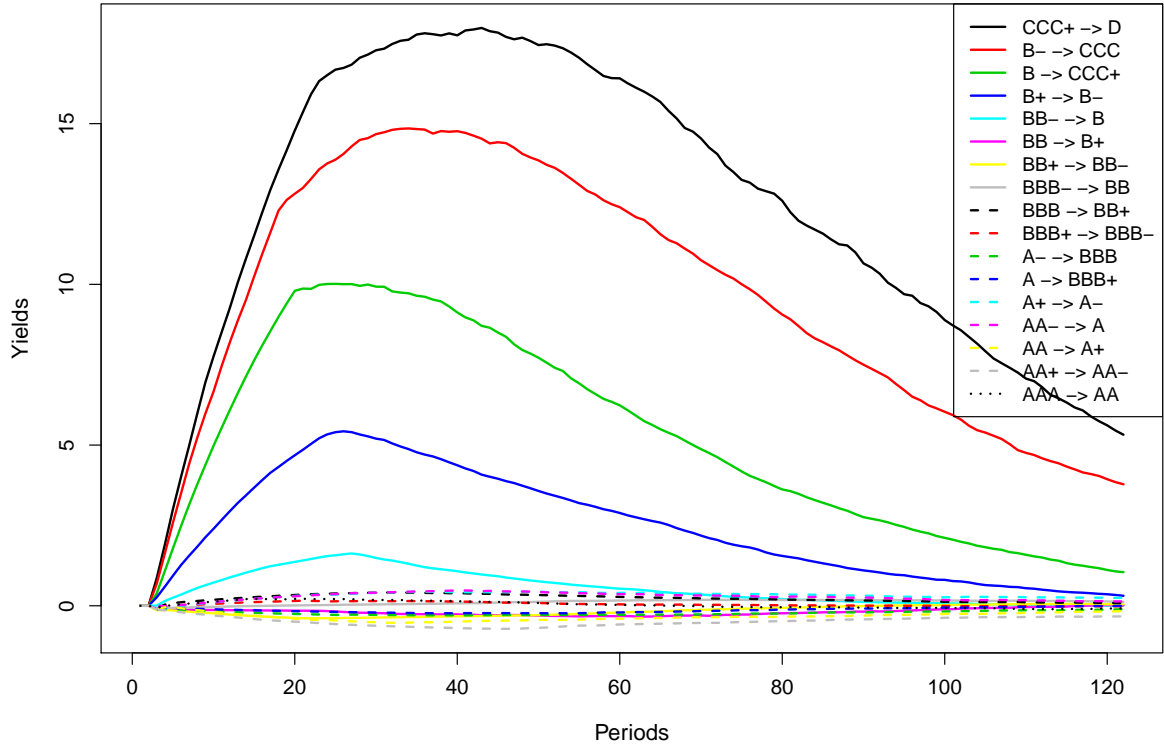


Figure 8: Median impulse-response function of yields after a rating shock of two notches

*Note:* Impulse-response functions are significant at the 5%-level for around 60 periods for all rating shocks occurring at a starting rating of BB- or below. For higher ratings, the whole IRF becomes insignificant.

Table 3: Median years until first occurrence of a rating after a D rating

Rating	D	CCC	CCC+B-	B	B+	BB-	BB	BB+BBB-	BBB	BBB+A-	A	A+	AA-	AA	AA	AAA			
Years	0	2	4	8	11	14	17	20	23	26	30	34	42	53	67	82	107	151	239

*Note:* The number of years until a rating is reached for the first time is calculated as the median of 5'000 simulations starting from a rating of 6 and a yield of 60% (as indicated by the yield equation). This estimation thus includes the (in this case beneficial) positive autocorrelation of rating changes. While the table includes all ratings, only the movement back to a level of AA is a recovery in the sense of a return to equilibrium. The further improvement to even higher ratings is purely driven by chance, which explains the apparent break in the recovery times.

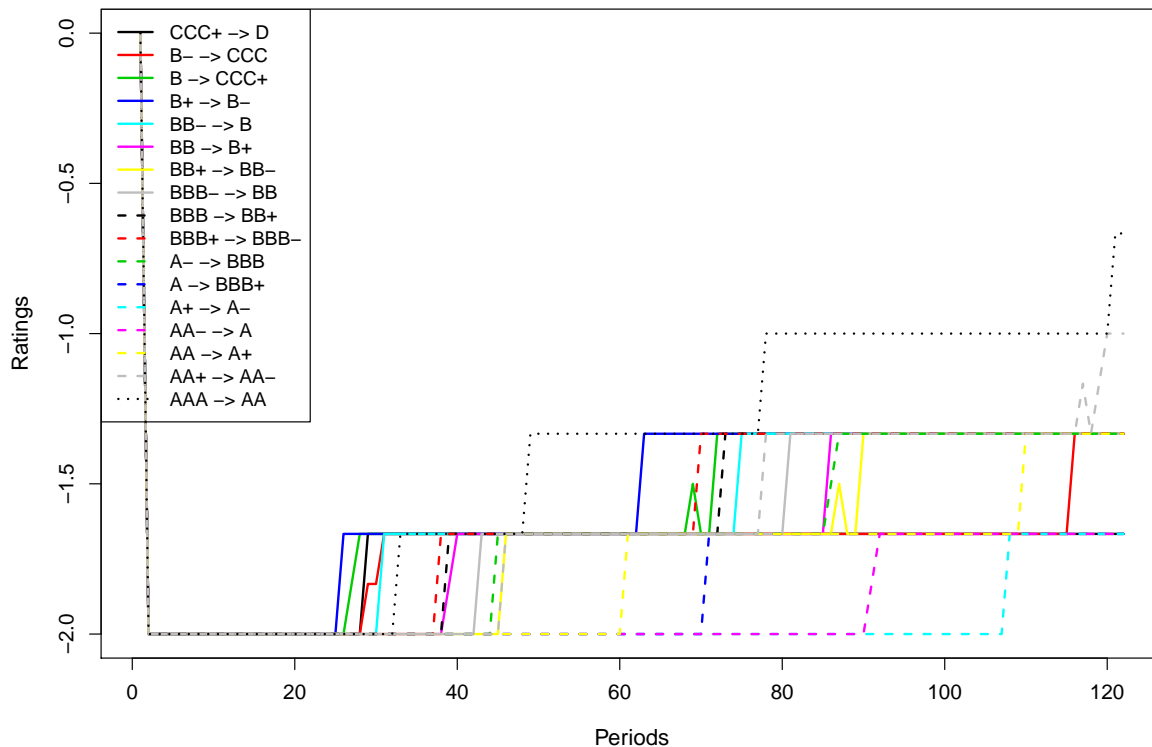


Figure 9: Median impulse-response function of ratings after a rating shock of two notches

*Note:* Impulse-response functions are significant at the 5%-level for around 40 periods for the lowest starting rating. For higher starting ratings, more periods are significant, and from BBB+ onwards, ratings stay significantly below the benchmark for the whole simulated period.

Only if the rating shock drives the rating below a rating of B, the impact of the rating shock on the interest rate is significant, see Figure 12. That is, only when the rating of a country is already (far) below the investment grade threshold, further downgrades have a significant increasing effect on yields. Yet, as soon as this happens, the effects are considerable. A rating shock driving the rating from B+ to B- causes a significant increase of the interest rate for more than six years, peaking at about 5% after two years. The impact of a shock driving the rating from B- to CCC remains significant for more than a decade, peaking at more than 15% after three years. For almost ten years, the interest rate is increased by 5% or more compared to the benchmark recovery path.

Concerning the dynamics of ratings, our results are mixed. On the one hand the already discussed persistence of ratings makes it almost impossible to catch up to the equilibrium recovery path after a substantial shock. Interestingly, this effect is visible the strongest when starting from good ratings, where the adjustment pressure on ratings is generally low. Starting from bad ratings, faster recovery is possible, due to the increasing insta-

---

For each pair we compute the difference individually, and then report the median of those differences. For the computation of confidence bounds, that should include parameter uncertainty, but not the idiosyncrasies of further disturbances, we repeat the afore mentioned procedure for every bootstrapped set of coefficients. That is, our confidence bounds are quantiles of those median IRFs for different possible coefficients.



bility of bad ratings. Yet, even after 10 years the ratings are still significantly below the original undisturbed recovery path for all starting ratings. This is partly due to the effect of past yield changes on rating changes. Both increasing and decreasing yields make rating downgrades more likely, see table 7 in the appendix. This seemingly contradictory result may be explained by the inability of rating agencies to differentiate between fundamentally justified yield movements and increased market volatility. As the second is a sign of financial distress, which may by itself affect the sustainability of government debt negatively, the probability of a downgrade increases (slightly).

While this confirms some of the worries concerning rating agencies, there is no evidence for a substantial self reinforcement. While our coefficient estimates do indicate a positive impact of lagged rating changes, this effect is so small, that further decreases of the rating caused by the initial shock are quite unlikely. At worst, our 90% confidence bands include delayed recovery of the rating compared to the benchmark. In no single case they include a vicious cycle augmenting the initial problems.

### 4.3 Scenario analysis

To provide some more conclusive evidence on the claim that the downward spiral of countries such as Greece was caused by the initial rating downgrade, this section presents some scenario analyses, where impulse response type developments are initialized with data from some episodes of major financial distress, and compared to the actually observed development.

The scenarios (i.e., the shock and the development thereafter) are depicted in Figures 10 and 11, together with impulse response functions and two sets of confidence bands. In addition to the confidence bounds reported for the IRFs, that only include parameter uncertainty, we also provide confidence bounds including the uncertainty concerning future shocks (not unlike the confidence bounds used in forecasting). This gives a more reasonable benchmark to assess the probability of an observed development, which does of course include an entire sequence of disturbances rather than merely the initial event. We report results for two events, the downgrade of Italy after the EMS crisis in the early 1990s, and the first sequence of downgrades of Greece during the ongoing debt crisis.

**Italy in 1992:** First, we want to look at Italy. As one of the founding members of the European exchange rate mechanism (ERM), Italy enjoyed a AAA-rating until July 1991, when Moody's first downgraded it by one notch to AA+ (or Aa1 in Moody's notation), making Italy the only G7 country with a rating below AAA at that time. A second downgrade, now by two notches to AA-, followed on August 13, 1992. The reason for these downgrades were the exceptionally high debt levels the Italian government had amassed in previous years (crossing 100% of GDP in 1992), together with large external imbalances. Harsh saving measures, privatization and laws aimed at reducing labor costs by the newly elected Italian government were considered to be positive by Moody's. However, the rating agency considered the measures to come too late. Therefore, they still predicted government debt to grow further during the 1990's (as it in fact did). Another fact that was thought to aggravate debt problems was the overvalued Italian currency, being restricted by the rules of the ERM. Again, the assessment of rating agencies was largely sound. Italy left the ERM one month later in September 1992 and devalued strongly (see also El-Shagi, Lindner & von Schweinitz 2014).

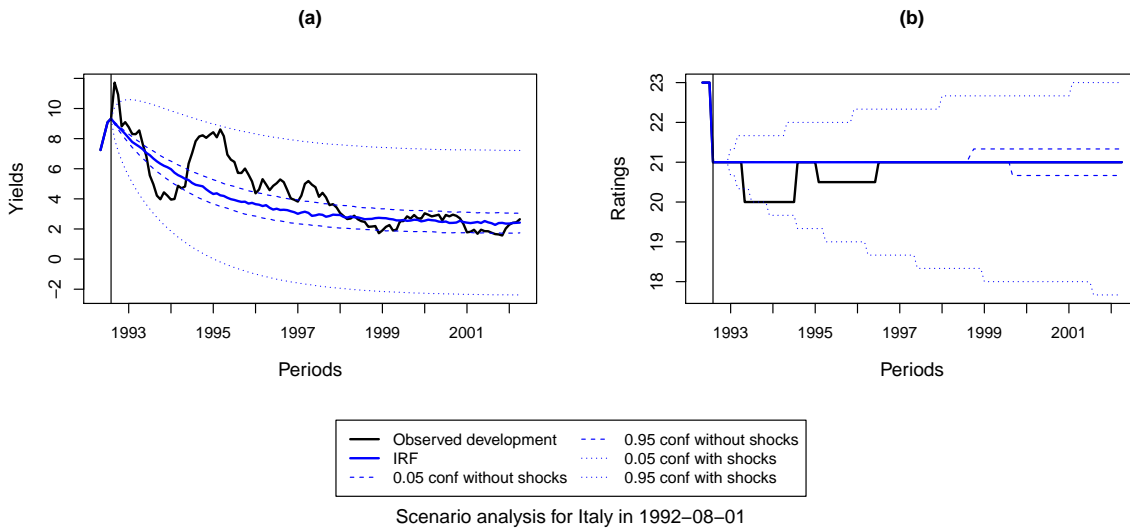


Figure 10: Observed development and impulse reaction functions in Italy after a downgrade in August 1992

It seems that the decision to downgrade the rating was largely a consequence of previous development. By judging reforms to come too late, Moody's itself acknowledged that there was basis enough to announce a downgrade earlier. Similarly, the European community voiced its concerns previously, citing the Maastricht criteria (among them, a maximum government debt level 60% of GDP) that were introduced in February 1992. Even the Italian minister of treasury was not too concerned by the downgrade. Therefore, it is not surprising that the reaction of markets was normal, see Figure 10. The observed development of yields is mostly quite close to the development of the simulated IRF, with all differences (except for the very first month) being inside the wider set of confidence bands. Observed ratings are identical to the simulated series for most of the time. Altogether, this is a typical case where the discrete rating change was a sign of an assessment of the sustainability of government debt that was mutually shared by the markets, which holds for the large majority of rating changes.

**Greece in 2010:** A completely different case is Greece in 2010. Mostly membership in the euro area had led to lower interest rates and improved ratings until October 20th, 2009, when the newly elected government opened their books and announced that the previous government had provided falsely low estimates of expected government deficit. Only two days later, Fitch downgraded Greek sovereign bonds to A-. While yields increased and the other members of the euro area (together with ECB and IMF) decided a first rescue package in March 2010 as well as the introduction of the European Financial Stability Facility), ratings further deteriorated. The starting point of our simulation is June 2010, when rating agencies had downgraded Greece to the investment grade threshold (Fitch just above, Moody's just below). Roughly until the end of 2010, observed ratings and yields are in line with our simulated results. However, at that time, recession and strong political opposition to reform led to renewed doubts regarding the ability and willingness of Greece to repay its debt. This in turn led to strongly increasing yields,

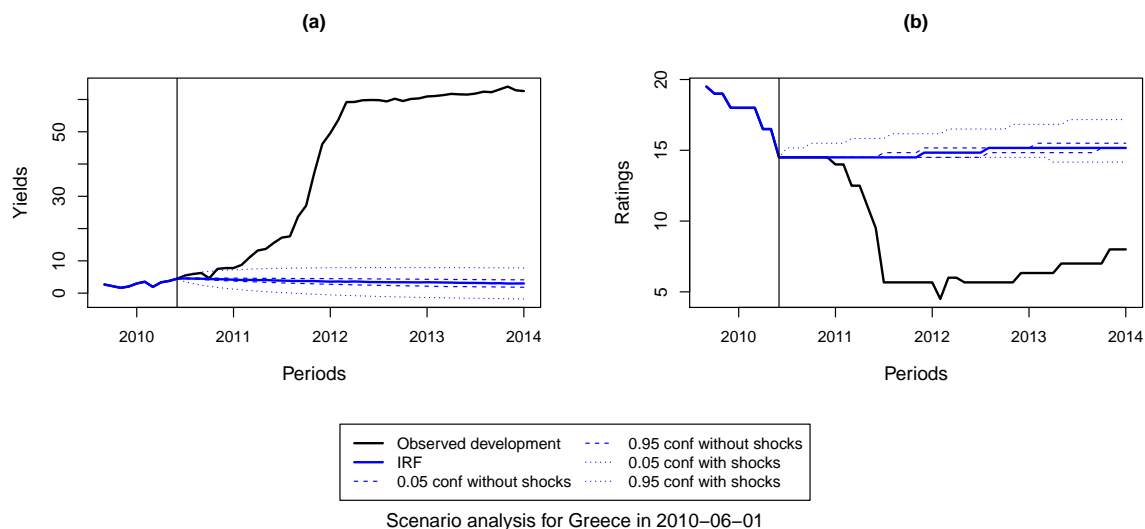


Figure 11: Observed development and impulse reaction functions in Greece after a series of downgrades until June 2010

further downgrades and in the end, when private investors became part of a second rescue package in July 2011, a last downgrade to CCC- (in default with little prospect to recovery). Taken together, it should be clear that not ratings, but political developments fueled the vicious cycle towards default in 2012. That is, this development cannot be attributed to the initial (weak) downgrades of rating agencies.

## 5 Conclusion

In conclusion, our evidence reconciles the two views prominently featured in the previous literature that partly hugely stresses the dangers of bad ratings, and partly denies the importance of ratings.

On the one hand, we find rather strong evidence against the theory of a vicious cycle. This is true for both the strong form of this theory that speculates the existence of a second – bad – equilibrium that might emerge after a rating is driven below moderate risk levels, and the weaker form of the theory that focuses on self reinforcing short run dynamics. Neither one is found in our data, at least not at a meaningful level. The usual interaction of ratings and yields fails to explain the downward spirals observed in a few cases. It seems as if the vicious cycle theory is mostly driven by misreading a few individual observations as typical.

On the other hand, there can be a substantial cost attached to rating downgrades. If a rating shock drives a country below a B rating, the risk premium can virtually explode. While the impact of ratings is negligible at better ratings, the increase is considerable for countries already starting at problematic levels before the downgrade and can easily reach a 2 digit magnitude. Due to the persistence of ratings, it is more than likely that the country will suffer this cost for extended periods.

## References

- Afonso, A., Furceri, D. & Gomes, P. (2012). Sovereign Credit Ratings and Financial Markets Linkages: Application to European Data, *Journal of International Money and Finance* **31**(3): 606–638.
- Arezki, R., Candelon, B. & Sy, A. N. R. (2011). Sovereign Rating News and Financial Markets Spillovers: Evidence from the European Debt Crisis, *IMF Working Papers* 11/68.
- Baum, C. F., Karpava, M., Schäfer, D. & Stephan, A. (2014). Credit Rating Agency Downgrades and the Eurozone Sovereign Debt Crises, *NBP Working Paper* 177.
- Beber, A., Brandt, M. W. & Kavajecz, K. A. (2009). Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market, *Review of Financial Studies* **22**(3): 925–957.
- Beirne, J. & Fratzscher, M. (2013). The Pricing of Sovereign Risk and Contagion During the European Sovereign Debt Crisis, *Journal of International Money and Finance* **34**: 60–82.
- Bolton, P., Freixas, X. & Shapiro, J. (2012). The Credit Ratings Game, *The Journal of Finance* **67**(1): 85–111.
- Breitung, J., Roling, C. & Elengikal, S. (2013). Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach, *mimeo*, University of Bonn.
- Cantor, R. & Packer, F. (1996). Determinants and Impact of Sovereign Credit Ratings, *The Journal of Fixed Income* **6**(3): 76–91.
- Davidson, R. & Flachaire, E. (2008). The Wild Bootstrap, Tamed At Last, *Journal of Econometrics* **146**(1): 162–169.
- Davidson, R. & MacKinnon, J. G. (2010). Wild Bootstrap Tests for IV Regression, *Journal of Business & Economic Statistics* **28**(1).
- De Grauwe, P. & Ji, Y. (2013). Self-fulfilling Crises in the Eurozone: An Empirical Test, *Journal of International Money and Finance* **34**: 15–36.
- De Vries, T. & De Haan, J. (2014). Credit Ratings and Bond Spreads of the GIIPS, *DNB Working Paper* 432.
- Durbin, E. & Ng, D. (2005). The Sovereign Ceiling and Emerging Market Corporate Bond Spreads, *Journal of International Money and Finance* **24**(4): 631–649.
- El-Shagi, M. (2010). The Role of Rating Agencies in Financial Crises: Event Studies from the Asian Flu, *Cambridge Journal of Economics* **34**(4): 671–685.
- El-Shagi, M., Lindner, A. & von Schweinitz, G. (2014). Real Effective Exchange Rate Misalignment in the Euro Area: A Counterfactual Analysis, *IWH-Discussionpaper* 6/2014.
- El-Shagi, M. & von Schweinitz, G. (2015). Risk and Return: Is there an Unholy Cycle of Ratings and Yields?, *Economics Letters* **129**: 49–51.

- Ferri, G., Liu, L.-G. & Stiglitz, J. E. (1999). The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis, *Economic Notes* **28**(3): 335–355.
- Gärtner, M. & Griesbach, B. (2012). Rating Agencies, Self-fulfilling Prophecy and Multiple Equilibria?: An Empirical Model of the European Sovereign Debt Crisis 2009–2011, *mimeo*, University of St. Gallen.
- Gärtner, M., Griesbach, B. & Jung, F. (2011). PIGS or Lambs? The European Sovereign Debt Crisis and the Role of Rating Agencies, *International Advances in Economic Research* **17**(3): 288–299.
- Hill, P. & Faff, R. (2010). The Market Impact of Relative Agency Activity in the Sovereign Ratings Market, *Journal of Business Finance & Accounting* **37**(9-10): 1309–1347.
- IMF (1997). World Economic Outlook May 1997 – Globalization: Opportunity and Challenges.
- Kiff, J., Nowak, S. B. & Schumacher, L. (2012). Are Rating Agencies Powerful? An Investigation Into the Impact and Accuracy of Sovereign Ratings, *IMF Working Paper 12/23*.
- Larraín, G., Reisen, H. & von Maltzan, J. (1997). Emerging Market Risk and Sovereign Credit Ratings, *OECD Development Centre Working Paper 124*.
- Mammen, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models, *The Annals of Statistics* pp. 255–285.
- Mellios, C. & Paget-Blanc, E. (2006). Which Factors Determine Sovereign Credit Ratings?, *The European Journal of Finance* **12**(4): 361–377.
- Mora, N. (2006). Sovereign Credit Ratings: Guilty Beyond Reasonable Doubt?, *Journal of Banking & Finance* **30**(7): 2041–2062.
- Pesaran, M. H., Shin, Y. & Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationships, *Journal of Applied Econometrics* **16**(3): 289–326.
- Reisen, H. & von Maltzan, J. (1999). Boom and Bust and Sovereign Ratings, *International Finance* **2**(2): 273–293.
- Vayanos, D. (2004). Flight to Quality, Flight to Liquidity, and the Pricing of Risk, *NBER Working Paper 10327*.
- von Hagen, J., Schuknecht, L. & Wolswijk, G. (2011). Government Bond Risk Premiums in the EU Revisited: The Impact of the Financial Crisis, *European Journal of Political Economy* **27**(1): 36–43.
- Webb, M. D. (2013). Reworking Wild Bootstrap Based Inference for Clustered Errors, *Working Papers 1315*, Queen’s University, Department of Economics.
- White, L. J. (2010). Markets: The Credit Rating Agencies, *The Journal of Economic Perspectives* **24**(2): 211–226.
- Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* **14**(4): 1261–1295.

# Appendix

Table 4: List of countries and data availability

Country	IMF classification	Data availability
Argentina	Developing	2009-10-01 – 2013-12-01
Australia	Advanced	1980-01-01 – 2014-01-01
Austria	Advanced	2003-02-01 – 2014-01-01
Belgium	Advanced	2005-01-01 – 2014-01-01
Brazil	Developing	2003-02-01 – 2014-01-01
Canada	Advanced	1986-05-01 – 2014-01-01
Chile	Developing	2009-01-01 – 2014-01-01
China	Developing	2002-06-01 – 2014-01-01
Colombia	Developing	2002-09-01 – 2014-01-01
Denmark	Advanced	1986-08-01 – 2014-01-01
Finland	Advanced	1992-01-01 – 2014-01-01
France	Advanced	1989-01-01 – 2014-01-01
Germany	Advanced	1986-01-01 – 2014-01-01
Greece	Advanced	1997-06-01 – 2014-01-01
Hong Kong	Advanced	1994-09-01 – 2014-01-01
Hungary	Transition	1997-02-01 – 2014-01-01
Iceland	Advanced	2003-08-01 – 2014-01-01
India	Developing	1993-10-01 – 2014-01-01
Indonesia	Developing	2003-05-01 – 2014-01-01
Ireland	Advanced	1992-01-01 – 2014-01-01
Israel	Advanced	2002-04-01 – 2014-01-01
Italy	Advanced	1988-11-01 – 2014-01-01
Japan	Advanced	1985-12-01 – 2014-01-01
Malaysia	Developing	2001-10-01 – 2014-01-01
Mexico	Developing	2001-08-01 – 2014-01-01
Netherlands	Advanced	1994-02-01 – 2014-01-01
New Zealand	Advanced	1994-03-01 – 2014-01-01
Norway	Advanced	1992-11-01 – 2014-01-01
Pakistan	Developing	2009-10-01 – 2014-01-01
Peru	Developing	2009-10-01 – 2014-01-01
Philippines	Developing	2001-02-01 – 2014-01-01
Poland	Transition	1999-03-01 – 2014-01-01
Portugal	Advanced	1994-12-01 – 2014-01-01
Singapore	Advanced	1990-06-01 – 2014-01-01
Slovakia	Transition	2007-04-01 – 2014-01-01
Slovenia	Transition	2007-04-01 – 2014-01-01
South Korea	Advanced	1997-12-01 – 2014-01-01
Spain	Advanced	1988-06-01 – 2014-01-01
Sri Lanka	Developing	2006-10-01 – 2014-01-01
Sweden	Advanced	1985-01-01 – 2014-01-01
Switzerland	Advanced	1994-01-01 – 2014-01-01
Taiwan	Advanced	1999-04-01 – 2014-01-01
Thailand	Developing	1999-09-01 – 2014-01-01
Turkey	Developing	2005-08-01 – 2014-01-01
United Kingdom	Advanced	1989-01-01 – 2014-01-01
United States	Advanced	1994-08-01 – 2014-01-01

Table 5: Summary statistics of ratings and yields

Country	Real Yields				Average Ratings			
	meany	sdv	miny	maxy	meanr	sdr	minr	maxr
Argentina	2.58	4.32	-6.18	7.88	10.77	1.47	7.67	12.50
Australia	3.99	2.29	-1.21	10.62	23.05	0.87	22.00	24.00
Austria	0.76	1.39	-2.19	3.43	23.94	0.13	23.67	24.00
Belgium	0.66	1.46	-1.59	4.46	22.63	0.55	21.67	23.00
Brazil	3.68	4.25	-10.34	9.35	13.71	2.02	10.00	16.00
Canada	3.22	2.23	-1.73	8.81	23.47	0.80	22.00	24.00
Chile	3.31	1.87	-0.95	8.43	20.17	0.40	19.33	20.67
China	0.42	1.90	-4.81	4.46	19.54	0.96	18.00	20.67
Colombia	4.63	1.58	2.38	11.73	13.81	0.85	13.00	15.67
Denmark	3.18	2.68	-2.42	8.65	23.47	0.46	23.00	24.00
Finland	2.83	2.81	-2.35	10.27	23.44	0.91	21.50	24.00
France	2.95	2.09	-1.29	7.41	23.94	0.21	22.67	24.00
Germany	2.76	2.04	-1.76	8.64	24.00	0.00	24.00	24.00
Greece	10.09	19.42	-1.28	64.00	16.45	4.71	4.50	20.00
Hong Kong	2.30	4.50	-6.49	13.00	20.63	1.60	19.00	23.33
Hungary	2.07	2.26	-3.10	9.78	16.92	1.78	13.67	19.00
Iceland	1.45	3.33	-8.63	8.30	18.64	3.63	14.67	22.50
India	1.45	3.56	-8.90	10.86	14.34	0.84	13.00	15.00
Indonesia	2.30	2.80	-5.43	7.61	12.49	1.57	9.50	14.67
Ireland	2.73	3.07	-2.20	12.10	22.15	2.67	16.00	24.00
Israel	2.79	2.52	-2.43	8.90	19.03	0.54	18.50	19.67
Italy	3.46	2.53	0.09	11.72	21.24	1.66	16.33	24.00
Japan	1.74	1.48	-1.41	5.66	23.14	0.98	20.67	24.00
Malaysia	1.31	1.63	-4.45	6.15	17.71	0.57	16.00	18.00
Mexico	3.22	1.52	-0.31	6.75	16.09	0.76	14.50	17.00
Netherlands	1.58	1.58	-2.39	4.88	24.00	0.04	23.67	24.00
New Zealand	3.69	1.72	-1.28	7.17	22.86	0.63	22.00	23.50
Norway	2.76	1.92	-1.38	7.27	23.91	0.18	23.50	24.00
Pakistan	2.11	1.72	-1.57	5.47	8.82	0.24	8.50	9.00
Peru	1.84	1.46	-0.13	6.32	15.53	0.58	14.50	16.67
Philippines	3.95	3.52	-2.78	11.42	12.99	0.91	12.00	15.00
Poland	3.36	1.71	0.42	8.55	17.88	0.62	16.00	18.50
Portugal	3.02	2.98	-0.94	14.43	20.41	2.99	13.00	22.00
Singapore	0.69	2.50	-5.84	6.37	23.22	1.08	21.00	24.00
Slovakia	0.78	1.45	-2.10	3.25	19.71	0.31	19.33	20.00
Slovenia	1.96	1.60	-2.17	4.53	20.61	2.13	16.33	22.00
South Korea	2.60	2.36	-0.97	9.02	18.24	1.88	11.50	20.67
Spain	2.94	2.58	-1.13	8.59	22.20	2.05	15.33	24.00
Sri Lanka*	3.28	5.80	-11.59	17.34	13.24	1.96	11.00	15.50
Sweden	3.70	2.41	-1.81	9.39	23.18	1.08	21.00	24.00
Switzerland	1.92	1.41	-0.53	5.40	24.00	0.00	24.00	24.00
Taiwan	1.26	2.16	-3.27	6.83	20.72	0.14	20.50	21.00
Thailand	1.23	1.97	-3.88	7.37	16.27	0.88	14.50	17.00
Turkey	4.26	3.77	-1.69	12.77	12.80	0.94	11.50	14.67
United Kingdom	2.84	2.52	-3.93	7.09	23.98	0.12	23.33	24.00
United States	1.53	1.90	-2.98	5.11	23.96	0.11	23.67	24.00

\*: excluding an extreme outlier in July 2010, when real yields shot up to more than 400%. Note: Real yields are the average monthly rating minus yoy-inflation in that month. Ratings are taken as the average rating at the end of the month.

Table 6: Outliers removed after initial estimation

Date	Country	$i_t$	$r_t$	$\varepsilon_t$
2010-03-01	Argentina	7.237	12	-4.004
2010-04-01	Argentina	3.803	12	-4.406
2013-03-01	Argentina	6.153	9	-8.132
2001-07-01	Australia	-0.536	22	3.812
2011-09-01	Greece	17.585	6	3.823
2011-11-01	Greece	27.090	6	7.382
2011-12-01	Greece	37.128	6	5.117
2008-12-01	Iceland	-3.430	17	-4.248
1998-12-01	India	-7.900	13	4.171
1999-01-01	India	-3.543	13	4.767
2005-10-01	Indonesia	5.874	11	-9.621
2006-10-01	Indonesia	-3.428	12	7.404
2011-08-01	Ireland	12.100	16	-4.785
2013-09-01	Peru	1.304	16	5.002
2013-10-01	Peru	6.322	16	-5.939
2008-01-01	Singapore	-1.332	24	-3.788
2008-10-01	Sri Lanka	-6.963	15	4.760
2008-11-01	Sri Lanka	-1.714	15	3.928
2009-11-01	Sri Lanka	13.332	15	-5.144
2008-10-01	Turkey	7.316	12	3.777
2008-12-01	Turkey	12.769	12	-3.645

*Note:* The standard deviation of estimation errors  $\sigma_\varepsilon$  in the initial estimation is 0.645.



Table 7: Median estimated coefficients from the yield and rating equation

coefficient	dyields0	dratings0
lyields	-0.032 ***	-0.011 *
facratings7	-0.46 ***	-0.184 **
facratings8	-0.458 ***	-0.177 **
facratings9	-0.391 ***	-0.162 **
facratings10	-0.234 ***	-0.133 *
facratings11	-0.085 **	-0.096 *
facratings12	-0.003	-0.063
facratings13	0.015	-0.046
facratings14	0.008	-0.049
facratings15	-0.009	-0.07 *
facratings16	-0.011	-0.093 **
facratings17	0.006	-0.105 ***
facratings18	0.008	-0.064 *
facratings19	0.002	-0.023
facratings20	-0.023	-0.02
facratings21	0.002	-0.056 *
facratings22	0.022	-0.083 **
facratings23	0.01	-0.136 ***
facratings24	-0.026	-0.17 ***
dyields0 <sub>p</sub>		-0.163 ***
dyields1 <sub>p</sub>	0.233 ***	-0.032
dyields2 <sub>p</sub>		-0.194 ***
dyields0 <sub>n</sub>		0.132 **
dyields1 <sub>n</sub>	0.111 ***	0.128 **
dyields2 <sub>n</sub>		0.137 ***
dratings1 <sub>p</sub>	-0.021	0.546 ***
dratings2 <sub>p</sub>	-0.004	0.115
dratings1 <sub>n</sub>	0.076	1.095 ***
dratings2 <sub>n</sub>	-0.043	1.058 ***
const	1.738 ***	
-1 0		-3.936 ***
0 1		0.635
λ	62.328	62.328
Country Fixed Effects	No	No
$LL_{model}$	3910.914	-1484.428
$LL_{smooth}$	-17.557	-16.068
$R^2$	0.06	0.064
$R^2_{adj}$	0.058	0.051
BIC	-7657.765	3178.469
AIC	-7785.823	3014.857

Note: *lyields* is the coefficient of lagged yields; *facratings* the dummy for ratings that are at most as good as this level; *dyields* and *dratings* denote lagged differences, with the asymmetric separation captured by the subindices *p* and *n*; *-1|0* and *0|1* are the thresholds of the ordered probit ratings model, the equivalent to the constant in the yields model. The smoothing coefficient  $\lambda$  is restricted to be equal in both equations. \*\*\*, \*\*, \* denote significance of a one sided test at the 1%, 5% and 10%-confidence levels.

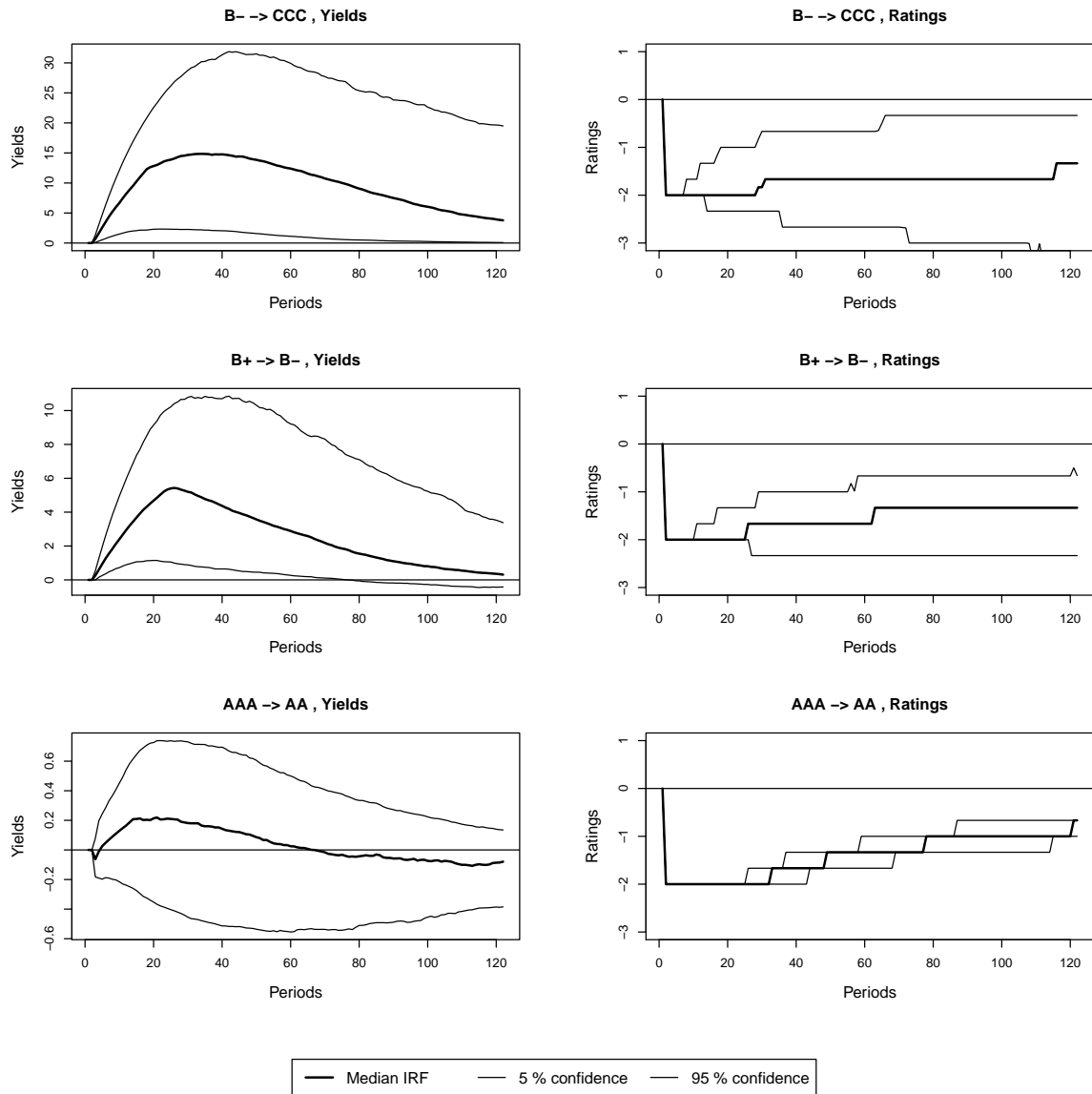


Figure 12: Selected impulse-response function after a rating shock with confidence bounds.

