The Joint Dynamics of Sovereign Ratings and Government Bond Yields*

Makram El-Shagi[1,2] and Gregor von Schweinitz[2,3]

[1]Center for Financial Development and Stability, School of Economics, Henan University, China
[2]Halle Institute for Economic Research, Germany
[3]Leipzig University, Germany

Abstract

Can a negative shock to sovereign ratings invoke a vicious cycle of increasing government bond yields and further downgrades, ultimately pushing a country toward default? The narratives of public and political discussions, as well as of some widely cited papers, suggest this possibility. In this paper, we will investigate the possible existence of such a vicious cycle. We find no evidence of a bad long-run equilibrium and cannot confirm a feedback loop leading into default as a transitory state for all but the very worst ratings. We use a bivariate semiparametric dynamic panel model to reproduce the joint dynamics of sovereign ratings and government bond yields. The individual equations resemble Pesaran-type cointegration models, which allow for valid interference regardless of whether the employed variables display unit-root behavior. To incorporate most of the empirical features previously documented (separately) in the literature, we allow for different long-run relationships in both equations, nonlinearities in the level effects of ratings, and asymmetric effects in changes of ratings and yields. Our finding of a single good equilibrium implies the slow convergence of ratings and yields toward this equilibrium. However, the persistence of ratings is sufficiently high that a rating shock can have substantial costs if it occurs at a highly speculative rating 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 Models, Nonlinearities

JEL classification: C14, C35, F34, G24.

*Corresponding author: Makram El-Shagi; Email elshagi@henu.edu.cn. Research of Gregor von Schweinitz was supported by DFG grant GZ:SCHW 1919/1-1. The authors thank the editor Carol Alexander and two anonymous referees for their constructive remarks. We are grateful to William Barnett, Manuel Buchholz, Stefan Eichler, Oliver Holtemöller, and Stephen Miller, as well as participants of the 2nd international workshop on Financial Markets and Nonlinear Dynamics, the WEAI annual conference 2015, and the CES 2015 annual conference for valuable comments and suggestions.

1 Introduction

Can a negative shock to sovereign ratings invoke a vicious cycle of increasing government bond yields and further downgrades that might ultimately push a country into default? In particular, the narratives of public and political discussions, supported by some widely cited papers such as Ferri, Liu, and Stiglitz (1999) and Bruneau, Delatte, and Fouquau (2014), suggest this possibility. In this paper, we propose a semiparametric bivariate framework to analyze the interaction between sovereign ratings and government bond yields to assess whether this narrative is empirically plausible.

Credit rating agencies – especially the so-called Big Three: Moody's, Standard and Poor's (S&P) and Fitch IBCA – took heavy blame for the recent financial crisis and subsequent (and partly still ongoing) Great Recession. It was argued that – like in previous major crises such as the Asian Flu and the collapse of the dot-com bubble – the agencies failed to predict the crisis, underestimated risk and thereby contributed to a bubble.[1]

However, regarding government bonds in particular, which are the main interest of this paper, the most frequently voiced concern is not the rating agencies' failure to predict crises, but the possibility that unfavorable rating changes cause capital flight, driving the risk premium up and thereby causing further problems that are sanctioned with another rating downgrade.

In the context of sovereign ratings this problem of ratings as self-fulfilling prophecies has first and most prominently been brought forward by Ferri et al. (1999). Building on that idea Carlson and Hale (2006) and Bruneau et al. (2014) propose explicit multiple equilibria models. Gärtner and Griesbach (2012) argue that the worse of the two equilibria would indeed lead to default (if not stopped by some intervention). In a similar but more elaborate model, Manso (2013) shows that, assuming rational behavior of competing rating agencies, rating downgrades can create feedback loops that ultimately can lead to default under endogenous default boundaries (the level of assets where firms decide to default). Although his model is originally meant for the analysis of corporate debt markets, it is of particular relevance for sovereign debt, where the debtors – as sovereigns – can indeed mostly choose whether to default.

Among politicians the narrative above is quite popular as it puts blame on rating agencies rather than bad policy. 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 then minister of finance, Pierre Moscovici) was to criticize the decision.[2] Based on a similar argument, a number of influential German politicians – backed by the European commission – started pushing for a European Rating agency after downgrades of Greece on Portugal.[3] It is

[1] Although rating agencies undeniably failed to predict the crisis, this very general criticism might be going too far. Following this logic, if rating agencies had foreseen that specific assets were highly risky, whether they are senior tranches of asset-backed securities, corporate bonds or sovereign bonds, then over-investment in these specific asset classes would probably have been avoided. Put differently, this criticism implicitly requires rating agencies to prevent crisis in general. However, this would apparently put too much of a burden on them.

[2] See, for example, http://www.bloomberg.com/news/articles/2013-11-08/france-credit-rating-cut-to-aa-by-s-p-on-weak-growth-prospects.

[3] See among others http://www.spiegel.de/international/europe/breaking-the-power-of-the-big-three-german-firm-wants-to-set-up-new-rating-agency-a-773549.html

obvious why politicians favor the view of a vicious cycle that can befall the best of us. Yet, the argument by Ferri et al. (1999) is not uncontroversial. The original study – focusing on the Asian Flu – has been criticized by Mora (2006) and El-Shagi (2010). Similarly, other recent papers that address the same question in the context of the European debt crisis reach very different conclusions. While Baum, Schäfer, and Stephan (2016) also find evidence of a substantial impact of ratings on capital allocation, De Vries and de Haan (2016) emphasize that the increased volatility following a rating downgrade was only temporary.

The seeming contradiction between the latter two contributions highlights a key omission in the literature that the present paper aims to fill. Much of the literature criticizing rating agencies focuses on their short-term impact or aims to show that there is some arbitrariness to ratings (e.g. Bolton, Freixas, and Shapiro, 2012). Nevertheless, as noted by El-Shagi and von Schweinitz (2015), neither of those effects provides sufficient empirical evidence of a vicious cycle between ratings and the risk premium that can push 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 usually occurs only in countries that have low ratings to begin with), the increase of the interest rate has to be sustained for a considerable length of time to actually increase the fiscal burden. To demonstrate the existence of a vicious cycle that inevitably leads to default unless the country is affected by subsequent positive shocks, it is necessary to prove the existence of explosive behavior in ratings and yields under particular conditions (such as a threshold for risk beyond which behavior becomes explosive). A weaker form of a vicious cycle – which is at least hypothetically still strong enough to cause default – is the combination of the strong self-reinforcing behavior of rating downgrades, combined with high persistence in rating levels in the absence of shocks and/or a lesser degree of self-reinforcing behavior for rating upgrades. Under those conditions, a negative rating shock might multiply before the system stabilizes (for a while) at a far lower rating, thus increasing government interest rates and the fiscal burden (Cantor and Packer, 1996). While technically not a second equilibrium, this kind of self-reinforcement with transitory stabilization and slow recovery would qualify as a vicious cycle in the sense used by the major critics of rating agencies. In other words, we need to focus on quantifying whether the impact of rating changes is substantial enough to be economically meaningful.

In our paper, we augment a recently suggested approach by El-Shagi and von Schweinitz (2015), who simultaneously model long-run relationship and short-run dynamics in a model that explicitly allows for multiple equilibria. Yet, the authors merely touch the subject of system dynamics, mostly including them to obtain valid inference on the long-run relationship. Our paper aims to dig deeper into the dynamic aspects, both in terms of a much richer econometric model that attempts to capture the stylized facts found in the data, and a more detailed analysis of the results including a range of simulation studies. By focusing on the dynamics, our paper fills a gap in a literature that was either concerned with long-run effects or immediate effects on impact.

In the abundant previous literature on ratings the dynamic part of the analysis usually takes the form of a simple event study. Rare exceptions (such as De Santis, 2012) derive impulse response functions based on VAR models that treat ratings as a continuous variable. Contrary to a standard VAR approach, we are able to account for the ordinal nature of ratings, include nonlinearities and asymmetries. Yet, this comes at a cost. The model we estimate is computationally demanding even in the bivariate form presented in this paper. Thus, we cannot account for the role of other macroeconomic variables. Since the interest rates react fairly quickly, and institutional variables are well covered by fixed effects, we do however believe that the benefits of our approach outweigh the losses.

We augment the dynamic part of the model of El-Shagi and von Schweinitz substantially, most importantly by allowing for asymmetric effects of both rating and yield changes. This modification allows for the type of vicious cycle that is driven by short-run dynamics rather than by convergence to a bad equilibrium. We confirm their finding that there is strong evidence of a single good equilibrium. At no point is the typical risk premium associated with a rating sufficient to justify further rating downgrades.

Yet, we find that downgrades can come at a substantial cost. Over the short run, rating changes tend to mildly reinforce themselves, slightly increasing the risk of further downgrades. Additionally, we do observe sharply increasing risk premia when ratings fall below the B+ level. Due to the high persistence of ratings, those interest premia can last many years, thus generating substantial macroeconomic costs without being vicious cycles. Yet, simulation studies based on downgrade episodes from the past decades show that unfavorable developments that have occasionally been observed after initial downgrades cannot be explained through the common joint dynamics of ratings and yields. They are thus most likely driven by an actual change in the fundamentals (or a correction in their assessment).

The remainder of this 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- and short-run relation between ratings and yields, including the scenario simulations reproducing previous downgrade episodes using our model. Section 5 concludes.

2 Previous evidence and stylized facts

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 and Packer, 1996; Gärtner, Griesbach, and Jung, 2011) as well as political stability and governance indicators (Mellios and Paget-Blanc, 2006) are found to be important. One of the most general results regarding ratings is that agencies react to past fundamentals rather than successfully predicting future shocks (Cantor and Packer, 1996; Reisen and von Maltzan, 1999). However, while providing only little new information in normal times, rating agencies may aim to reestablish their reputation after missing an emerging crisis, responding with overly restrictive downgrades (Ferri et al., 1999; White, 2010). In general, lower ratings induce higher borrowing costs and may thus reduce government expenditure and investment and lead to higher taxes (Cantor and Packer, 1996). They can impact macroeconomic fundamentals negatively, and thus have an additional indirect effect on government finances. Due to the so-called sovereign ceiling (an implicit rule whereby companies only rarely obtain a better rating than their home country), government downgrades will negatively affect ratings of companies, increasing interest payments and worsening the economic outlook (Durbin and Ng, 2005; Almeida, Cunha, Ferreira, and Restrepo, 2017). This is for example reflected in lower stock market returns (Kaminsky, Schmukler, et al., 2002), increasing capital outflows (Forbes and Warnock, 2012) and increasing funding constraints on banks' balance sheets (Bocola, 2016). This problem is exacerbated by the reliance of regulators on ratings. These regulations often indicate that assets with ratings below a certain threshold are not considered as "investment" but speculation; thus, these assets are strongly restricted or penalized (White, 2010).

Of course, the argument of a self-fulfilling prophecy is controversial, e.g., El-Shagi (2010) notes the inconsistency in simultaneously claiming that rating agencies obviously and systematically err and claiming that most investors do not recognize this and follow the rating agencies despite their alleged obvious shortcomings.

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 based on the first downgrades of 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.[4] 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 shortly before the end of the crisis and sometimes even after. 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 proposed by critics of rating agencies relates to the ongoing European debt crisis. It was argued that the actions of (US-based) agencies unduly increased market pressure on European periphery countries, increasing their government bond yields to unsustainable levels, thus triggering a public debt crisis with severe long-run macroeconomic costs. Arezki, Candelon, and 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 systematic spillover effects to other European countries (see also Beirne and Fratzscher, 2013). That is, the downgrading of Greece is found to have increased not only the CDS spreads of government bonds (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 results for spillover effects are quite heterogeneous and thus may not be strong enough to support their claim in a more general setting. The findings of Afonso, Furceri, and Gomes (2012) point to a much more balanced view of the question at hand. While not doing a fully-fledged dynamic analysis, the authors assess the impact of rating changes over several horizons. Although they find a moderate effect right after a rating announcement, this impact disappears within roughly six months. A similarly balanced position is taken by Gärtner et al. (2011), who use yearly data and a fundamental estimation of ratings as in Cantor and

[4] The stickiness of ratings may have several causes: first, they could be due to shortcomings in information processing; second, to a tendency of rating agencies to avoid rating reversals if default probabilities fluctuate near the boundary of two discrete rating classes (Löffler, 2005). Alternatively, stickiness may be explained by rational inattention when ratings are based on costly private information (Woodford, 2009; El-Shagi and von Schweinitz, 2017).

Packer (1996). They argue that non-fundamentally justified 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 partially challenged by the finding of De Vries and de Haan (2016) who observe that credit ratings and yields have recently become disentangled: after the summer of 2012, the yield levels of European periphery countries decreased quickly while ratings stayed at very low levels. The authors attribute this to either unconventional monetary policy or increasing conservativeness among credit rating agencies. However, their econometric model does not include short-run effects and allows for varying effects of different rating levels only to a very limited extent. Therefore, the econometric model of De Vries and de Haan may be misspecified, and the slow adjustment of ratings may simply be due to the general stickiness of ratings at certain levels.

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 necessary to consider (a) both short- and long-run effects in the model, (b) nonlinearities in the long-run relation, and (c) asymmetries in the short-run relation.

2.1 Measurement and sample selection

In this subsection, we discuss the data we use in this paper. Tables A1 and A2 in the Appendix report data sources and coverage by country, respectively.

Ratings: To maximize data coverage, we use average ratings of foreign currency denominated government bonds as provided by Moody's, S&P and Fitch. The three agencies use grades to assess the probability of a default over the medium- to long-term, where better grades correspond to lower default probabilities. The names of grades differ across agencies merely in notation (Cantor and Packer, 1996). Therefore, grades can be easily compared and transformed into the ordinal scale given in Table 1. Throughout the remainder of this paper, we will use the S&P notation.

As market movements are often found to be strong around rating announcements (which provide 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 specialization and leadership of the agencies in specific markets (Hill and Faff, 2010), which is why we should include information from all agencies rather than concentrate 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 and de Haan, 2016). However, for most of our analysis, we aim to maintain the ordinal nature of the ratings, contrary to the majority of the literature considering rating levels. The reason for this is the non-linear relationship between ratings and default probabilities. 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 defined in Table 1. It is fairly well documented that the ratings of different agencies seldom differ by much, even during times of higher uncertainty when

Table 1: Rating grades and transformation

Grade	Moody's	S&P	Fitch	Ass. 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
In default with little prospect for recovery	Caa3	CCC-	CCC-	6
	Ca	CC	CC	5
		C	C	4
In default	C	D	DDD	3
			DD	2
			D	1

rating agencies adjust their assessments more frequently (Ferri et al., 1999).[5] Therefore, averaging generally corresponds to the majority rating of the three agencies.
As a robustness check, we also employ the median of the three agencies. However, the differences in results (see Figures A4 and A5 in the appendix) are marginal.

Yields: Our interest rate variable is the the real government bond yield on sovereign bonds with a maturity of 5 years, denominated in domestic currency.[6] Again, this choice is mostly enforced by data availability and feasibility. Many studies that focus on emerging markets use the spread of dollar denominated bonds over the US treasury yield as a measure risk. However, our sample includes a large number of OECD countries whose debt is mostly (or in some cases entirely) denominated in the domestic currency.[7] For highly developed countries, the pure credit default risk component is often captured by the prices of CDS, which are essentially insurance contracts against the event of default. However, similar to using yield spreads, using CDS prices would reduce our sample significantly, in this case primarily by reducing the number of observations with low ratings. We therefore have to rely on real yields as a slightly more noisy, but also more widely available proxy of risk. 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, and Wolswijk, 2011). Thus, they go slightly beyond the claim of rating agencies to consider solely the probability of default. However, it seems plausible that the changes in real yields that are correlated to rating changes mostly reflect changes in the risk premium.
Conversely, 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. To produce comparable yields for different countries, we deflate the yields using the year-on-year inflation of the previous 12 months. While past inflation is not a perfect measure of inflation expectations, which 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 approximately 85% of the periods, yields are positive, and they have a mean and median near 2.5%. This is plausible, as investors would only be willing to accept yields below current inflation rates if they can be nearly certain that the latter will decrease in the near future. Table A3 suggests that the distribution of yields does not differ much among countries. Only in Greece between October 2011 and January 2014, real yields have been higher than 20%.[8] In all other

[5]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.

[6]Thus, there exists a currency mismatch between ratings and yields. However, the correlation between domestic and foreign ratings is very high (94% for rating levels, and 74% of domestic rating changes occur at the same time as foreign rating changes). The results are broadly robust to using domestic instead of foreign ratings or to using the emerging market bond index (EMBI, produced by JP Morgan) in US dollars. Both alternative measures for ratings or yields come at the cost of a significantly reduced dataset: domestic ratings are only available much later due to higher international demand mainly for foreign currency bonds, and the EMBI only covers emerging markets.

[7]Panizza and Presbitero (2014) show in a panel of 17 OECD countries from 1980 to 2007, roughly our sample period, that the share of foreign denominated debt is just 9%.

[8]In a second exception, the nominal yield in Sri Lanka exceeded 450% in July 2010. This value is so

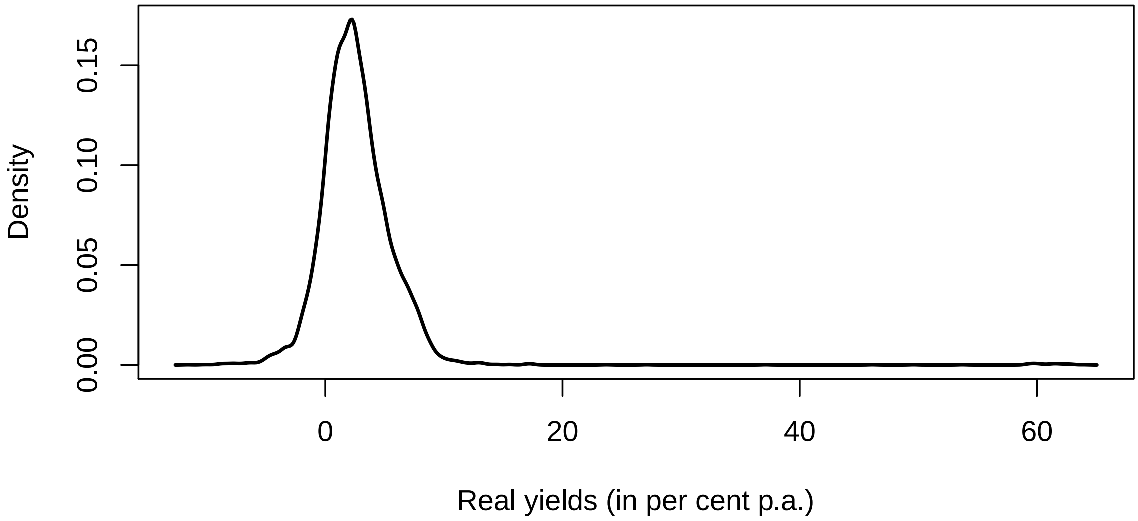

Figure 1: Density of yields.

countries, high nominal yields were usually accompanied by high inflation rates, leading to a narrow distribution of real yields.

On average, we expect developing countries to have higher credit risk and inflation volatility, both contributing to higher average real yields. However, these expectations are not fulfilled by the summary statistics in Table 2.[9] Advanced economies have, on average, higher real yields. This counterintuitive result can again largely be attributed to Greece. If we exclude it, average yields are comparable between groups, with a much lower standard deviation (i.e., volatility) in advanced economies.

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	2.36	4.50	24.00
	Developing	14.83	2.76	7.67	20.67
	Transition	18.21	1.97	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

unreasonably high that we exclude it from our estimation as an outlier. All other available observations are included in our sample.

[9] The classification of countries follows IMF (1997). A detailed list of countries with data availability can be found in Table A2 of the appendix.

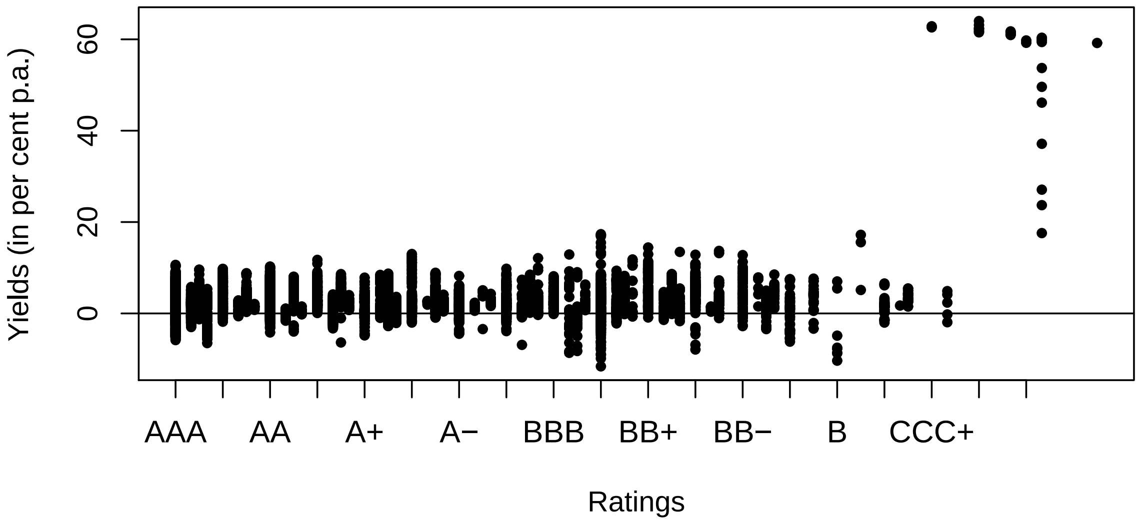

Figure 2: Scatterplot of monthly ratings and yields.

Note: The rating scale (notation of S&P) is inverted in this and all following graphs. That is, the x-axis displays increasing ratings (i.e., lower risk) when going from right to left.

Sample: The variables defined above are jointly available for an unbalanced monthly panel of 46 countries from January 1980 to January 2014. It covers 27 advanced economies as well as four Eastern European transition economies and 15 developing economies. Transition and developing economies often have low data availability: there are several countries for which yield data are only available after 2001. However, we still retain approximately 9,100 observations in total.

2.2 Stylized Facts

Nonlinearities: As ratings measure default probabilities in a nonlinear way (Löffler, 2005), it is not very likely that ratings (or their assigned values) can be used linearly. In the present context, this limitation especially holds when the potential effect of an investment grade threshold and the possibly nonlinear relationship between default probabilities and sovereign yields is taken into account. Therefore, different authors used various transformations of ratings when they seek to explain (medium-run) yield movements.[10] Larraín, Reisen, and von Maltzan (1997) test both a linear and a logistic transformation of ratings; Ferri et al. (1999) employ an exponential conversion, which is also used (along with linear and cubic conversions) by Gärtner and 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

[10]In event studies with windows of only a few days, the nonlinearity of ratings does not play such a large role.

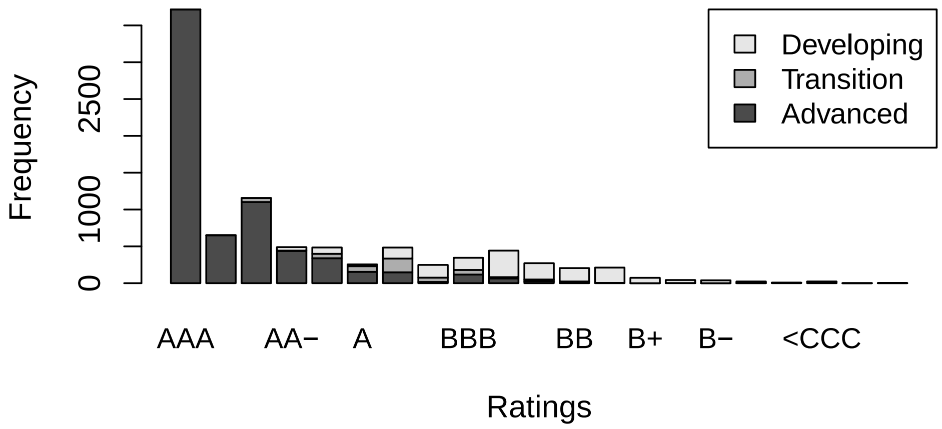

Figure 3: Histogram of monthly ratings.

flexible transformation should be employed. This transformation will be presented in the following section along with our method of addressing the short- and long-run interactions 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 between BBB- and BB+.[11] How can this "non-relation" be reconciled with the anecdotal evidence and the reasonable assumption that (inverted) ratings and yields should be positively correlated? First, ratings are constant for long periods of time during which yields may slowly adjust to new risk levels. Second, real yields are affected by many more factors orthogonal to sovereign risk. Rather than interpreting the flat slope as the absence of a risk premium, it should be interpreted as a risk premium of an order of magnitude that is overshadowed by the general variance of interest rates. The risk premium begins to quickly increase only if risk becomes substantial. This is not purely related to the nonlinearity of risk measurement; it is actually in line with theory, which predicts that the risk premium goes to infinity when the default probability approaches one.

Persistence: Figure 3 shows the histogram of average ratings. Nearly 60% of our ratings are high-grade, which can partly be explained by the greater data availability of advanced economies (Table A2 in the appendix), which tend to have higher ratings. In our dataset, ten of these countries have never received a rating below AA, while only two industrialized countries (Greece and Israel) have never achieved a rating above A+; see Table A3 in the appendix. Transition and developing countries, however, tend to have much lower ratings. Among that group, only Slovenia achieved a "high grade" rating

[11] Ghosh, Kim, Mendoza, Ostry, and Qureshi (2013) provide a theoretical explanation and empirical evidence for a very similar relationship between government debt and interest rates.

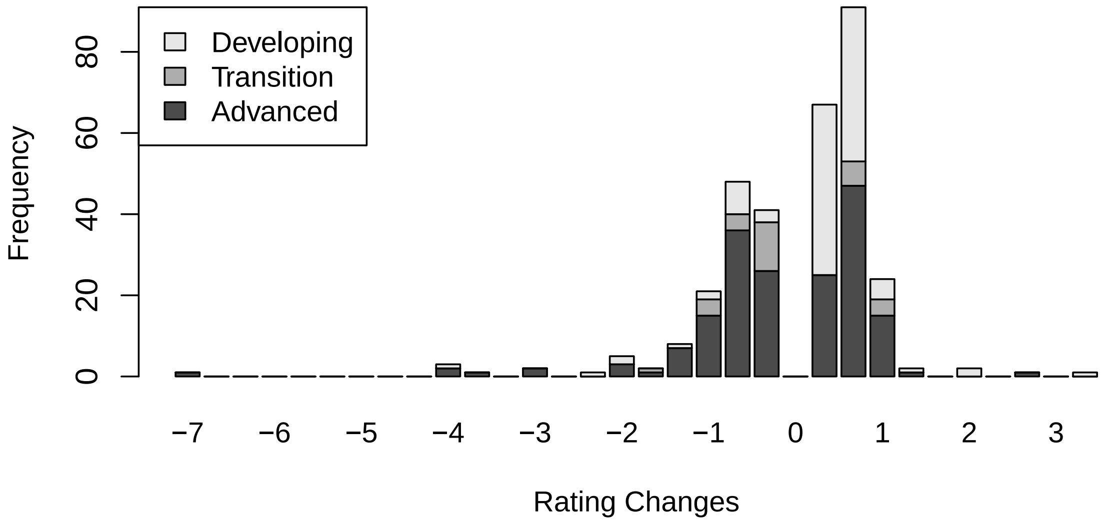

Figure 4: Histogram of monthly rating changes.

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

until they were downgraded during the European debt crisis in January 2012. Yet, the distribution features no bimodality that would suggest that countries below a certain rating face consecutive downgrades until they eventually default.

While this is at least some indication of existence of a unique, a good equilibrium, ratings are characterized by enormous persistence. The share of observations with a rating change is a mere 3.5%. In our entire sample, we observe 187 upgrades and 133 downgrades (which are on average slightly larger than upgrades). This low share of periods with rating changes is not driven by countries that have already achieved peak ratings. Even when excluding observations with ratings of AAA and AA+, the probability of a rating adjustment barely exceeds 5%.

This degree of stickiness makes conducting an analysis in a traditional AR framework difficult, even if the variables of interest are 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 use a model that combines short- and long-run effects.

Asymmetries: Figure 4 shows the distribution of monthly changes, excluding constant ratings. 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 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 one or two agencies moving first and the third following

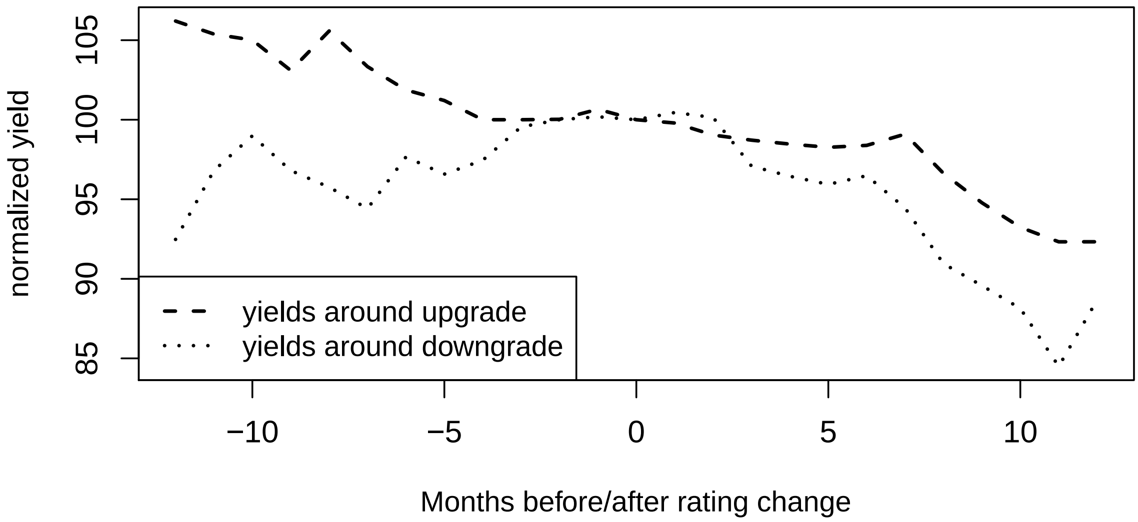

Figure 5: Development of yields in a two year window around rating changes.

the next month, downwards dynamics seem much more intense than the recovery. Such 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, occurred in advanced economies. However, cumulative rating improvements are more or less equally distributed over advanced, transition and developing economies.

Asymmetry is not limited to the magnitude of a 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 does not provide conclusive evidence, the figure roughly identifies two stories. During periods of downgrades, ratings changes, more often than not, seem to occur jointly with a peak in the interest rate. That is, the data confirm the finding of El-Shagi (2010) on a much broader level that rating downgrades occur late in an adjustment process, immediately before yields decrease again. Yet, during times of upgrades, rating changes are adjustments during an ongoing decline in risk premia.

While the general finding of 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, and Schumacher, 2012; Afonso et al., 2012) are limited by the fact that they use comparably short windows of approximately 14 days before and after a rating announcement. While the short-run fluctuations they analyze may be highly relevant for speculation purposes, they seem negligible given a longer macroeconomic perspective.

Event studies usually find strong effects of rating announcements on yields on the days

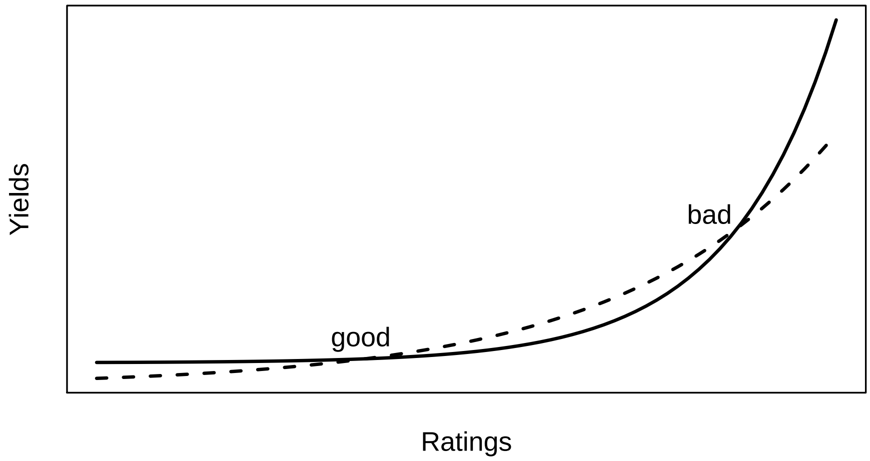

Figure 6: Exemplary long-run relations of ratings (depending on yields, dashed) and yields (depending on ratings, solid) with two equilibria.

before and after the event. It is sometimes argued that this is a sign of an anticipation effect (Hill and Faff, 2010). However, the argument that ratings are sticky (Mora, 2006) or lagging (Reisen and von Maltzan, 1999; El-Shagi, 2010) seems more convincing. That is, we should expect Granger causality in both directions in general rather than only for short windows around events.

3 Model and estimation technique

In this section, we present the econometric model we use to analyze dynamic developments. Due to the persistence of ratings, identifying (possibly asymmetric) short-run reactions of ratings and yields depends crucially on correct identification of the exact long-run relationship(s). Thus, our model is a simultaneously estimated bivariate two-equation model consisting of a continuous yield equation and an ordered probit ratings equation, which allows for much richer dynamics than the simpler model in El-Shagi and von Schweinitz (2015) (which focussed only on the long-run relationship). Each equation is inspired by the structure proposed by Pesaran, Shin, and Smith (2001) in their seminal paper on the bounds cointegration test.

A priori, there is no reason for a unique long-run equilibrium or a unique long-run relationship (implying an infinite number of equilibria along the relationship). Figure 6, inspired by the theoretical model of Gärtner and Griesbach (2012), shows a relationship with two equilibria: a good and stable equilibrium of low yields and high ratings, and a bad and unstable equilibrium of high yields and low ratings. If a country receives a rating below the bad equilibrium, this sets off the vicious cycle of rising yields and downgrades already described above.

As indicated by previous results and stylized facts, our model needs to account for the features described in section 2. In subsection 3.1, we explain how we perform a simultaneous estimation of the short- and long-run relations and how we perform the identification of (possibly many) long-run equilibria. In subsection 3.2, we explain how we incorporate nonlinearity in rating levels into the model while ensuring a certain degree of smoothness. Subsection 3.3 explains the bootstrap procedure needed to account for the time and cross-sectional heterogeneity of shocks, while subsection 3.4 describes how we interpret our results and simulate our impulse response functions. Some more specific details on long-run estimation and identification are provided in Appendix B.

3.1 The basic model

The model structure: Our model is a nonlinear extension of the model origininally proposed by El-Shagi and von Schweinitz (2015). It consists of an interest rate equation

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

and an ordered probit component explaining rating changes

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

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

where i is the interest rate, $\Delta r \in \{-1, 0, 1\}$ the change of the rating, r_c a rating dummy, r^* the latent variable governing the rating process, t the time index, and $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ $\eta t \sim \mathcal{N}(0, \sigma^2)$ are the respective error term. We use contemporaneous rating changes as well as p_i lags of interest rates and rating changes in the interest rate equation, and p_r lags of ratings and $p_r + 1$ lags of interest rates in the rating equation for reasons outlined below.

The main difference to the model proposed by El-Shagi and von Schweinitz (2015) is the inclusion of asymmetric effects of negative and positive changes of both ratings and yields, which is motivated by the discussion of Figures 4 and 5 in the previous Section. We test models with asymmetric effects against its symmetric counterpart at different numbers of lags and find that the symmetric model is always significantly outperformed by the asymmetric one.

Identification of long run relationships and the adjustment process: The individual equations of the model are inspired by the cointegration models popularized by Pesaran et al. (2001). The reason for adopting this model setup is that the equation structure estimated by Pesaran et al. allows valid inference (i.e., consistent and unbiased estimates) regardless of whether the two included variables exhibit unit root behavior. Usually those models only have a single equation, taking the form:

$$\Delta x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta x_{t-i} + \sum_{j=0}^{p} \eta_j \Delta y_{t-j} + \varepsilon_t, \tag{3}$$

and are used to obtain a single long-run relationship identified through the ratio of the coefficients on the level variables on the right hand side. However, we would like to estimate the long-run relationship implied by changes in ratings and yields separately, thereby identifying the equilibrium yield as a function of the current rating and the equilibrium rating as function of the current yield.

Structural identification: We require some assumptions to allow for the kind of semistructural identification implied by our lag selections. To this end, we borrow from the literature on structural VARs and impose restrictions on the contemporaneous effects.[12] We assume that ratings have a contemporaneous effect on yields, but not vice versa. This assumption is reasonable for two reasons. First, ratings are characterized by high persistence at monthly frequencies. Second, and more importantly, we measure ratings on the first day of the month and average yields over the month. That is, “contemporaneous” yields contain the information of up to a full month after the measurement of ratings. The identification of the contemporaneous effect of ratings on yields is achieved by including the contemporaneous rating change in the yield equation (following Pesaran et al. quite closely). This provides causal identification equivalent to the standard SVAR, which uses Cholesky decomposition of the covariance matrix. However, our approach comes at a considerable advantage. The ordered probit model used to explain (discrete) rating changes does not provide residuals. Therefore, a Cholesky decomposition could only be achieved by simultaneous estimation of the complete system of both equations and the covariance matrix. Our identification strategy, on the contrary, orthogonalizes residuals and thus (following the seminal argument by Sims, 1980) allows for block-wise estimation. Through the approach outlined above, 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

[12] In the original framework of Pesaran et al., the contemporaneous correlation is captured through η_0, i.e., the contemporaneous first difference of y without making any assumption as to whether this actually reflects an impact of y on x or vice versa. To identify a unique long-run relationship, this is not consequential. Equation 3 can easily be solved for Δy_t as dependent variable without changing the relation of the coefficients of the level variables β_1 and β_2.

not technically cointegrated in the traditional sense. Rather, they are either stationary or exhibit some type of regime-switching behavior.

In order to allow our model to identify the frequently feared situation of a *good* and a *default* equilibrium (see Figure 6), we need to allow for nonlinearities in the long-run relations. To this end, rather than including the rating as a continuous variable, we use a semiparametric approach to estimate a functional form over a set of rating dummies. This approach allows us to distinguish and test one, multiple, or an infinite number of equilibria (the last case being a cointegration relationship of ratings and yields).

In the results, we provide evidence that there is indeed only a single good long-run equilibrium towards which ratings and yields slowly converge.

Our model does not account for the - theoretically possible - situation that the long-run relationship (and thus the location of equilibria) changes over time. However, we would expect that time variation in the equilibrium creates (a) high uncertainty regarding the equilibrium location and (b) variation regarding the equilibrium between countries (which is accounted for in our fixed effects model). Since we find neither, it seems that time variation (although possible) is not driving our results.

Specification: Lags are selected in each equation separately by the Bayesian information criterion with a maximum of six lags. We adjust the dataset to estimate both equations with exactly the same observations.

Inclusion of fixed effects: Fixed effects are usually avoided in ordinal models because they are no longer identified when a cross-sectional unit is constantly in one of the extreme groups. While this is true for the level of the rating (with several advanced economies always being rated AAA, see the summary statistics by country in Table A3 in the Appendix), this does not occur when using rating changes as the dependent variable (because no country is permanently down- or upgraded). Thus, fixed effects for N countries could be easily included in the form of $N-1$ dummies in both equations.[13]

In our view, the inclusion of country fixed effects has both advantages and disadvantages. On the one hand, they account for possible differences between countries that may affect yields and ratings. Due to the structure of our econometric model, short- and long-run effects from other sources on our two variables of interest are simultaneously captured by fixed effects (if they are not already captured by the lag structures). On the other hand, many rating classes are only observed in a small number of countries, and (conversely) only few countries experience a large range of ratings. Therefore, fixed effects could also blur the relation between yields and ratings by capturing the effects of individual rating classes instead of the effects for individual countries. As an econometric distinction between these two very different channels is hardly possible, we present the results without fixed effects as a baseline and use fixed effects estimation as a robustness check.

3.2 Exploiting the ordinal ratings to model nonlinearity

Class dummies vs. continuous rating approximations: Contrary to the bulk of the literature that allows for nonlinearities, we do not merely choose a specific trans-

[13] Incidental parameters are sometimes a problem in ordered-choice models with panel data. However, this is of no concern as we have a large T dimension and relatively few countries N.

formation of a continuous representation of ratings. Instead, we use a nonparametric functional form based on (additive) dummies that represent the different rating classes to better account for the highly nonlinear relationship between ratings and yields indicated by Figure 2.

The rating dummies are defined such that $r_c = 1$ indicates that the rating is lower than or equal to 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 reference class (captured in the constant) because very few ratings at or below "Extremely Speculative" (CCC) are observed. This leaves us with 18 different rating dummies. Defining the rating dummies cumulatively (i.e., a dummy equals 1 not only for countries with the respective rating but also for countries with better ratings) simplifies the assessment of whether the difference in effects between adjacent rating classes is statistically meaningful.

The modeling of rating levels with dummies introduces a potentially strong degree of nonlinearity. Thus, we are also able to detect possible structural breaks, which could for example exist at or around the investment grade threshold (a rating of BB+ or below). In particular, our model allows to identify long-run relationships as depicted in Figure 6, where the behavior of the system can become explosive beyond a certain threshold and the country is driven automatically into default.

While rating dummies refer to the average rating class of the three rating agencies, rating changes $\Delta r \in \{-1, 0, 1\}$ capture every rating change (both smaller and larger than one notch). We do this because the rating decision of a single agency may have a short-run effect on yields even if the average rating class does not change.

The semiparametric estimation: When estimating equations (1) and (2) through standard estimators, the representation of the ratings through a series of dummies is problematic. Because some rating classes are 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.[14] While technically allowing "nonlinearities" in the impact of ratings (compared to treating ratings as pseudo-continuous), this creates considerable and economically unwarranted differences in the effects of very similar ratings. Essentially, despite modeling ratings through class dummies, we would like to obtain 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 and Roling (2015) for mixed frequency data sampling (MIDAS). In MIDAS approaches, low-frequency data (e.g., monthly inflation) are explained through high-frequency data (e.g., daily oil price movements). Merely estimating many coefficients for the high-frequency lags usually yields substantial identification problems. The existing literature has addressed 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 and Roling (2015) argue that this might be overly restrictive and suggests a more flexible nonparametric approach. Instead of enforcing a specific functional form, Breitung and Roling augment

[14]For example, highly-speculative ratings (and worse) are only observed in Argentina, Greece and Pakistan.

an 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 smooth behavior of the impact of ratings. Because our ratings dummies are defined cumulatively, we restrict not 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, we augment the traditional likelihood function of our models (denoted by LL_{model}) as follows:

$$\begin{aligned} LL_{smooth} =& \sum_{c=8}^{24} \ln(\phi(\sqrt{\lambda}(\zeta_c - \zeta_{c-1}), 0, \sigma^2_{model})) \\ LL =& LL_{model} + LL_{smooth}, \end{aligned} \tag{4}$$

where, depending on the model, ζ can be either β or ψ, and $\phi(\cdot, 0, \sigma^2_{model})$ is the density of a normal distribution with mean zero and variance drawn from the errors in the respective model (σ^2_{ε} or σ^2_{η}).
By increasing the weight λ of the penalty LL_{smooth}, it is possible to enforce the smooth behavior of adjacent coefficients.[15] 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 and Roling accordingly show that λ can generally be mapped onto the effective loss of degrees of freedom. That is, we can use standard information criteria to select the degree of smoothing. On the one hand, if high differences between the ratings are actually needed to explain the behavior of interest rates or ratings, we will choose a low degree of smoothing in the respective equation. On the other hand, if volatility in those coefficients is merely driven by very few observations in specific classes, we will opt for smoothness.
In contrast to El-Shagi and von Schweinitz (2015), we estimate a single λ for both equations. This allows a more clear-cut comparison of the model, where the long-run relation is restricted to be identical in both equations, which implicitly includes a single rather than individual λs. That is, in our setup, a rejection of a single long-run relationship cannot be attributed to the influence of using different λs.

3.3 Addressing heteroscedasticity

As mentioned in subsection 2.1, government bond yields are affected by more than one risk channel. Especially during times of turmoil in financial markets, market participants may shift their portfolios toward government bonds that they deem safe (Vayanos, 2004; Beber, Brandt, and Kavajecz, 2009). This safe-haven effect (which is essentially herding behavior) may lead to self-fulfilling crises in other countries when creditors with a risk that is slightly worse than “safe” (but far from being in default) are shunned by financial markets (De Grauwe and Ji, 2013). Herding behavior also reinforces a tendency to treat superficially comparable countries similarly without performing in-depth analyses of the

[15] The likelihood contains $\sqrt{\lambda}$, as equation (1) can be estimated by simple OLS with a quadratic penalty term (Breitung and Roling, 2015).

individual debtors. This, in turn, may lead to spillovers of risk from one country to the next (Beirne and Fratzscher, 2013), that is, shocks to yields may be correlated across countries.
Shocks to the interest rate may deviate from the usual i.i.d assumptions in our estimation. Indeed, we find substantial heteroscedasticity, cross-sectional correlation and heavy tails (as could, for example, be inferred from the results of Arezki et al., 2011). We use a bootstrap method to account for heteroscedasticity (both over time and across countries) and cross-sectional correlation in both equations. It is based on the wild bootstrap originally proposed by Wu (1986). In the wild bootstrap, rather than resampling the original residuals, the simulations are generated using error terms that are obtained by multiplying the original residual ε for the respective observation with a random multiplier v. The distribution of v ensures that the expected value of εv is 0 and that the first few moments of the distribution of ε and εv are identical or at least very similar. 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 and Flachaire (2008). To reproduce the cross-country correlations 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 and MacKinnon (2010), who used the same technique to reproduce the correlations among 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 the ratings and yields as well as the lagged differences are drawn from the empirical joint distribution. Because we need low ratings in the simulated sample to have identification of all coefficients, we combine the bootstrap with an acceptance-rejection algorithm that discards simulations where extremely low ratings do not occur. The reported confidence bounds are based on 1,000 accepted simulations of the entire two equation system.

3.4 Dynamic adjustments

For the interpretation of our results, we mostly focus on impulse response functions (IRF) and other visual representations. The reason is that the direct interpretation of the estimated coefficients is difficult for two reasons. First, we use 18 rating dummies to account for the nonlinear nature of the long-run relationship between ratings and yields. Second and as usual for VAR analyses, IRF are much more informative about the dynamics after a shock because all variables in the system are endogenous.
The standard method of calculating an impulse response function is to simulate the reaction after a shock leading away from equilibrium. In our context, however, this is inadequate. Recovery from default ratings to the long-run equilibrium takes considerable time. Because the nonlinearity in the risk premium causes shocks to have markedly different effects based on the original rating, it is essential to consider shocks at different risk assessments. Thus, rather than reporting a single impulse response function starting at the long-run equilibrium, we report a range of IRFs for a negative rating shock. Each IRF uses a different original rating level and the corresponding yield as given by the equilibrium yield curve implied by equation (1). We report the difference to the recovery path without the initial rating shock.

While impulse response functions are usually computed deterministically, i.e., without further shocks, this is unfeasible in our case due to the (ordered probit) rating equation. At each 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 (additional) shocks. Rather than computing the difference between the median development with the initial rating shock and the median development without said shock, we always conduct the simulations with and without the initial shock with identical simulated disturbances. For each pair, we compute the difference individually and then report the median of those differences. For the computation of confidence bounds, which should include parameter uncertainty but not the idiosyncrasies of further disturbances, we repeat the aforementioned procedure for every bootstrapped set of coefficients. That is, our confidence bounds are quantiles of those median IRFs for different possible coefficients.

4 Results

We base the discussion of the relationship between ratings and yields on coefficient estimates for our baseline model as reported in Table 3. The long run relation(s) can be obtained from the relative coefficients on the level variables (lagged yield and rating dummies). More precisely, because our rating dummies are defined additively, the relative coefficients define the slope of the equilibrium curve implied by the respective equation. This relation could in principle incorporate two different types of vicious cycles.

The first type is a vicious cycle, where an initial rating downgrade leads to increasing yields and further downgrades until the country defaults. This type of vicious cycle would imply at least two equilibria in the long-run relationship of ratings and yields: a good and stable equilibrium at high ratings and low yields, and an unstable equilibrium below which the vicious circle is set in motion. Our analysis confirms previous evidence of a single long run equilibrium El-Shagi and von Schweinitz (2015). A cointegration relationship of ratings and yields (an infinite number of equilibria) is rejected at the 1% level; similarly, the possibility of a second unstable equilibrium can be rejected at the 2% level. A more detailed analysis of the adjustment to the long-run equilibrium is given in Section 4.1.

The second type of vicious cycles combines three different elements, namely persistence of ratings, short-run reactions of yields and stronger effects of downgrades than upgrades. Under these circumstances, lower ratings and higher yields would persist for a long time after an initial downgrade. Moreover, the asymmetric (stronger) effect of downgrades would imply that an upgrade shock would not be able to offset the negative consequences of a similarly sized downgrade shock. As the second type is more strongly related to the short-run dynamics, it will be discussed in Section 4.2. We present evidence that, although upgrades have much weaker effects than downgrades, this second type of cycle is likely to exist only at very low ratings, when yields react very strongly. In Section 4.3 , we present some scenario analysis of downgrade episodes in our data to confirm the short-run dynamics implied by the impulse-response analyses.

Table 3: Median estimated coefficients from the yield and rating equation

	Δi_t	Δr_t
Lagged Yield	-0.033^{***}	-0.01^{*}
Rating CCC	-0.542^{***}	-0.187^{**}
Rating CCC+	-0.533^{***}	-0.179^{**}
Rating B-	-0.427^{***}	-0.16^{*}
Rating B	-0.239^{***}	-0.124^{*}
Rating B+	-0.09^{**}	-0.084^{*}
Rating BB-	-0.003	-0.054
Rating BB	0.022	-0.041
Rating BB+	0.016	-0.052
Rating BBB-	-0.015	-0.076^{**}
Rating BBB	-0.02	-0.097^{**}
Rating BBB+	0.003	-0.106^{**}
Rating A-	0.005	-0.066^{*}
Rating A	0.01	-0.017
Rating A+	-0.014	-0.016
Rating AA-	-0.005	-0.056^{*}
Rating AA	0.023	-0.084^{**}
Rating AA+	0.013	-0.133^{***}
Rating AAA	-0.028^{*}	-0.163^{***}
$\Delta i_{t-1,+}$	0.234^{***}	-0.151^{**}
$\Delta i_{t-2,+}$		-0.037
$\Delta i_{t-3,+}$		-0.189^{***}
$\Delta i_{t-1,-}$	0.109^{***}	0.153^{***}
$\Delta i_{t-2,-}$		0.116^{**}
$\Delta i_{t-3,-}$		0.118^{**}
$\Delta r_{t,+}$	-0.034	
$\Delta r_{t-1,+}$	0.002	0.586^{***}
$\Delta r_{t-2,+}$		0.12
$\Delta r_{t,-}$	0.06	
$\Delta r_{t-1,-}$	-0.06	1.038^{***}
$\Delta r_{t-2,-}$		1.09^{***}
const	1.949^{***}	
-1—0		-3.914^{***}
0—1		0.658
λ	55.142	55.142
# Effective Coefficients	18.305	23.264
Country Fixed Effects	No	No
LL(Data)	3938.304	-1481.072
LL(Smoothing)	-18.191	-16.036
R^2	0.062	0.063
R^2_{adj}	0.061	0.05
BIC	-7709.655	3174.08
AIC	-7839.997	3008.671

Note: Lagged yield is the coefficient of lagged yields; *Rating A* (and accordingly) the dummy for ratings that are at most as good as a rating of A; Δi and Δr denote lagged differences, with the asymmetric separation captured by the subindices + and $-$; *-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 λ is restricted to be equal in both equations. ***,**,* denote significance of a one sided test at the 1%, 5% and 10%-confidence levels.

4.1 Adjustment processes towards long-run equilibrium

Long-run relationship of ratings and yields: Here, we report the long-run relation as implied by the coefficient estimates in Table 3. We largely follow the strategy of El-Shagi and von Schweinitz (2015), adapting it in two respects to account for the additional nonlinearities in our model. First, to partially compensate for the asymmetric effects of positive and negative deviations, we use typical past changes in the interest rate.[16] That is, rather than setting the past lagged changes to zero, we replace all lagged changes by the standard deviation of the changes observed in the data. Because rating changes are extremely rare events, we retain a value of lagged rating changes of 0. Second, we carefully test for the number of long-run relationships, explicitly accounting for the potential role of the smoothing parameter λ. In several tests, we can reject the possibility of a cointegrating relationship with an infinite number of equilibria at the 1% level.[17]

There is further evidence for the superiority of our benchmark model as implied by Table 3. Figure 7 plots the long-run relationships 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 are economically and statistically significantly different.

As theory predicts, both curves are significantly positively sloped, implying that (a) ratings tend to deteriorate if the interest rate is high (dashed line, from equation (2)) and that (b) the risk premium (measured through the interest rate) increases for countries with low ratings (solid line, from equation (1)). However, as already indicated by the stylized facts presented in Section 2, we only find a risk premium of noteworthy magnitude at extremely low ratings. In the interest rate equation, only the rating class dummies far below the investment grade threshold of BB+ are significantly negative, implying an increase in the equilibrium interest rate which quickly reaches extreme heights.

This interest rate behavior is one of the main reasons why our point estimates indicate a single intersection of the long-run relations, which occurs at the highest ratings and at 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 low ratings, which are thus highly transitory.

The confidence bands of the rating equation grow extremely wide for low 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. The high significance of rejection in the presence of only a limited number of rating changes and – correspondingly – wide confidence bands gives further credence to our results. For 98% of the bootstrapped sets of long-run relations, we find that the equilibrium interest rate (according to the yield curve) creates upward pressure in the rating equation at all ratings below A-. Even in the unlikely case that a second equilibrium exists, that second equilibrium is unstable in the sense that there is no process driving the rating toward it. Essentially, a second intersection of the rating and yield curves below a rating of CCC would not imply that a

[16] Asymmetric effects of positive and negative deviations imply that random mean zero yield (or rating) changes have a non-zero impact on the left-hand side. Thus, the usual approach in error-correction models (set all changes to zero and solve the long-run relationship) needs to be adapted.

[17] More details are reported in Appendix B.

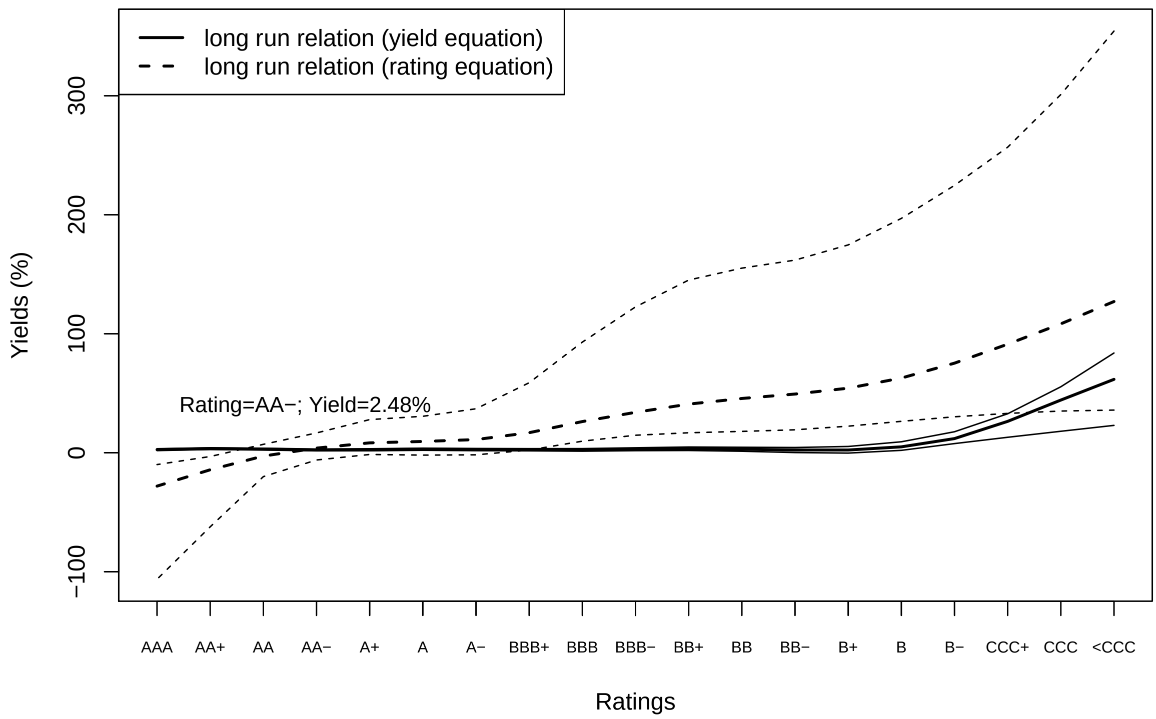

Figure 7: Long-run relation of ratings and yields, accounting for asymmetric shocks

Note: In addition to median long-run relationship, we show the 10%- and 90% confidence bands from the bootstrap.

country can be driven into default by a vicious cycle of ratings and yields.[18] This merely indicates that a country requires a positive impulse to escape default ratings. Because negative rating shocks are impossible at a D rating, a positive shock initiating recovery would be inevitable.

Recovery times towards equilibrium: While we find only evidence of a single long-run equilibrium of low interest rates and high ratings, the persistence of rating implies a very long adjustment process towards that equilibrium, as reported by Table 4.[19] If a country starts in default, first recovery steps are quick. The median duration until a non-default rating (CCC) without any additional shocks is only two years. However, following upgrades take successively longer: Reaching an investment grade takes 26 years in the median, and reaching the long-run equilibrium at a rating of AA- around 80 years.

Table 4: 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	7	11	14	17	19	22	26	30	34	42	53	66	81	105	148	231

Note: The number of years until a rating is reached for the first time is calculated as the median of 5,000 simulations beginning from a rating of D 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. Further improvement in ratings is driven purely by chance, which explains the apparent break in the recovery times.

4.2 Short-run dynamics and impulse response functions

Self-reinforcement of ratings in the short run: Like the analysis of the long-term relationship, the short-run dynamics do not provide evidence of a vicious cycle. While there is some autocorrelation in changes of ratings (especially for downgrades), it is far too little to imply the economically meaningful self-reinforcement of a negative shock, see Table 3. Autocorrelation coefficients need to be related to the large distance between the thresholds of the ordered probit equation. In equilibrium, a first downgrade has an unconditional probability of around 1.2%. Conditional on that, the coefficient on lagged rating changes, $\Delta r_{t-1,-}$ and $\Delta r_{t-2,-}$, imply a probability of 10% of a second downgrade up to 2 months later. Only in the rare case of two consecutive downgrades (having an unconditional probability of 0.1%), a third downgrade will happen with a probability of 44%.

Compared to the direct self-reinforcement of ratings, the feedback loop via higher yields is negligible at most rating levels. In the short run, yield changes do not significantly react to rating changes, and the long run coefficients reported in Table 3 only indicate a significant increase in yields for ratings below B. This increase in interest rates does

[18] This finding is robust with regard to the number of bootstrap draws. Tests with 5,000 simulations showed virtually no difference. Due to the high computational burden of more simulations, not all settings could be estimated with 5,000 simulations.

[19] Strictly speaking, ratings above equilibrium (AA+ and AAA) are in our simulation not reached as the natural outcome of long-run adjustment, but purely by chance.

in turn affect the downgrade probability. Although the effect is fairly small because the monthly adjustment of the interest rate is small (starting with the coefficient value for the respective rating dummy and declining from there until the new equilibrium is reached), the effect might matter because the trajectory of the interest rate is changed over an extended period of time. Thus, although the increase of the downgrade probability in any given month is below 1%, it is fairly persistent. In total, the probability of further downgrades being caused by an initial rating shock is strictly and substantially below 50% even for major shock at low rating level (see Figure 9).[20] Even the lower confidence bounds of our IRFs show one further downgrade only for shocks starting at rating levels of B+ and below (see Figure A1 in the Appendix).

Impulse response functions to rating shocks The weak self-reinforcing nature of rating downgrades does not completely eliminate the role of rating shocks in creating macroeconomic distress. As mentioned above, the impact on interest rates is considerable if a rating downgrade drives the rating below B. Due to the steep increase of the risk premium below this rating level, yields quickly rise to high levels and remain there over an extended period due to the high persistence of ratings. To illustrate this problem, we provide IRFs based on the bootstrapped coefficients after a relatively large rating shock, a downgrade of 2 notches (see Figures 8 and 9). A rating shock driving the rating from B+ to B- causes a significant increase in the interest rate for more than six years (see also Figure A1 in the Appendix), peaking at approximately 3% after one year. The impact of a shock driving the rating from B- to CCC remains significant for more than a decade, peaking at more than 12% after two years. For nearly five years, the interest rate is increased by 5% or more compared to the benchmark recovery path.

The duration of the interest premium suggests that there could be a substantial increase in the interest burden, even given moderate debt rollover during this time. For countries with a low average maturity – which might be the most likely to be hit by this type of shock (Arellano and Ramanarayanan, 2012) – the cost would be even higher. Such an effect will strain the government budget, thereby limiting public good provision substantially. At the same time, the sovereign ceiling function of the sovereign rating will cause similar rating changes in the private sector, most likely corresponding to similar interest premia, which hamper investment. Given the high path dependence of economic development, the consequences of such periods might be noticeable long after the interest rate has normalized.

Concerning the dynamics of ratings, our results are mixed. On the one hand, the already discussed persistence of ratings makes it nearly impossible to catch up to the equilibrium recovery path after a substantial shock. Even after eight years, ratings remain 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 3. 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 negatively affect the sustainability of government debt, the probability of a downgrade increases (slightly). However, ratings do

[20] As explained in Section 3.1, the timing structure of our data ensures that rating innovations in our estimation can be interpreted as structural exogenous shocks.

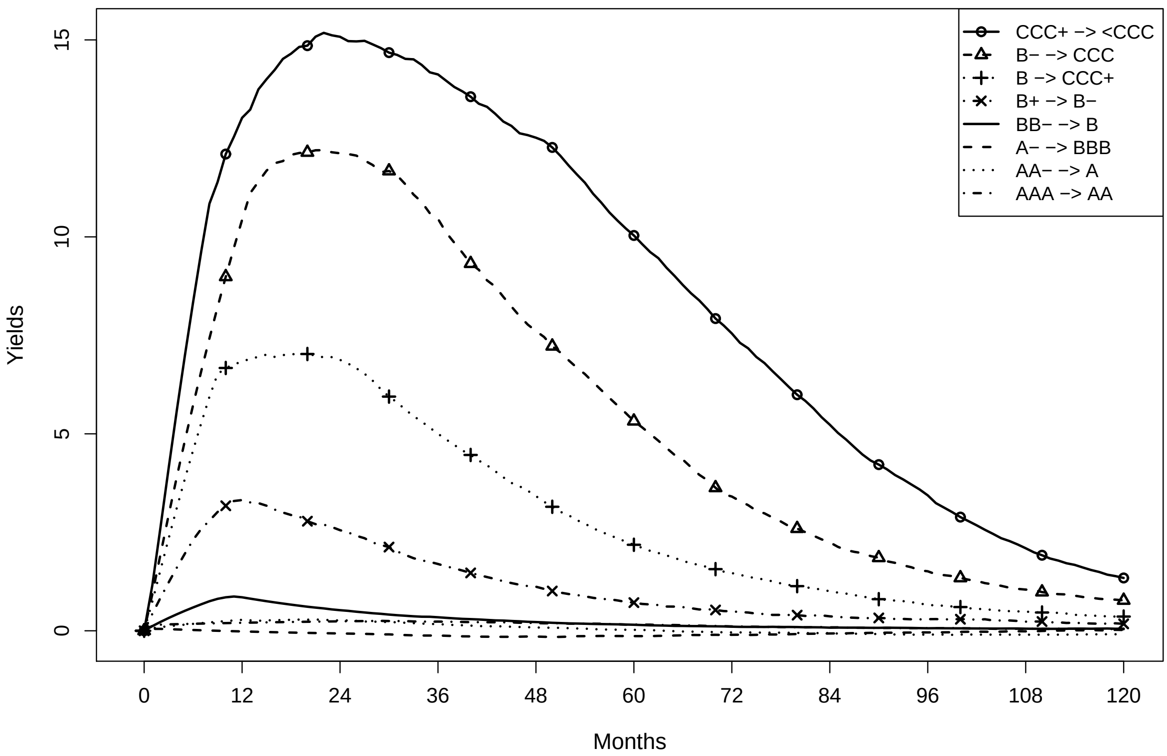

Figure 8: Median impulse-response function of yields after a rating shock of two notches, selected ratings

Note: Impulse response functions are significant at the 5% level for approximately 60 periods for all rating shocks occurring from an initial rating of BB- or below. For higher ratings, the entire IRF becomes insignificant. Selection of impulse response functions including confidence bands in Figure A1 in the appendix.

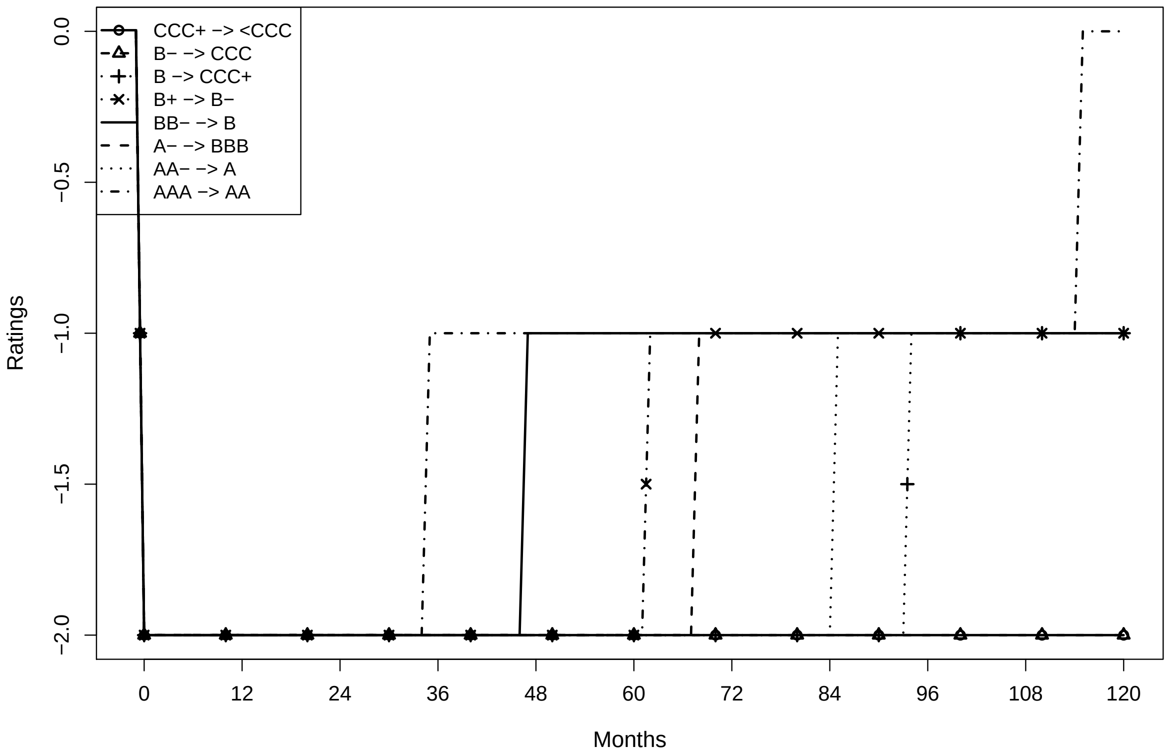

Figure 9: Median impulse-response function of ratings after a rating shock of two notches, selected ratings

Note: Impulse response functions are significant at the 5% level for approximately 40 periods for the lowest initial rating. For higher starting ratings, more periods are significant, and from BBB+ onwards, ratings remain significantly below the benchmark for the entire simulated period. Selection of impulse response functions including confidence bands in Figure A1 in the appendix.

not deteriorate further after an initial shock. Thus, we find no evidence of self-reinforcing rating dynamics.
The response to a corresponding upgrade is fairly similar in shape (although of opposite sign), as shown in Figure A2 in the Appendix. Moreover, the order of magnitudes of the developments after upgrade and downgrade shocks are comparable for initial rating levels of BB and above. However, downgrades at lower initial rating levels have a much stronger effect on yields than upgrades. For all starting rating of BB- or below, a downgrade of two notches has for all simulated periods a stronger effect than a similar upgrade in absolute terms. The negative effect of a downgrade at these low rating levels can be several times as large than the beneficial effect of an upgrade. There are two reasons for this. First, equilibrium yields increase strongly for ratings of B and below. Second, the strong asymmetric effects of past yield and rating changes in the rating equation imply that rating upgrades (along the recovery path) are much less likely in the downgrade scenario than in the upgrade scenario. That is, it is very likely that a weak vicious cycle in the sense of persistent increases of yields can occur at low rating levels.

Robustness checks As a robustness test of the results provided above, we test an alternative downgrade scenario, wherein we distribute the two-notch shock of the previous scenario over two consecutive months. As our rating equation employs two lags, this scenario tests the potential self-reinforcing behavior of downgrades. Without taking yield changes and level effects into account, two successive downgrades would increase the probability of a third downgrade from 1% to 44%. Therefore, these results should be interpreted as a worst-case response to a large rating shock. Both IRFs (yields and ratings) show larger differences to the benchmark adjustment cases, as shown in Figure A3 in the appendix. We can confirm that even transitory default after a negative rating shock is highly unlikely if the initial rating is not already close to default. Even for a starting rating of B, the probability of a default (as a transitory state) is only 2%, and it is almost zero for better ratings.
The results are also robust with regard to the case where different rating agencies have strongly differing opinions. Taking the median of the different rating decisions instead of its mean does not affect our results, as can be seen in Figures A4 and A5 in the appendix. As another robustness test, we also estimate the model that includes country-specific effects. Because this implies different intercepts in the long-run relationships, each country now has a separate long-run equilibrium. For all countries except Pakistan, this equilibrium corresponds to a real yield of less than 7 percent, and just as in the baseline estimation, there is no evidence of a second intersection for any country. Regarding the dynamics, the results are virtually even more similar.

4.3 Scenario analysis

To provide more conclusive evidence for the claim that the downward spirals of countries such as Greece were not caused by their initial rating downgrade, this section presents scenario analyses in which the impulse response developments are initialized with data from some episodes of major financial distress and compared to the observed development. The scenarios (i.e., the shock and the subsequent development) are depicted in Figures 10 and 11 along with the impulse response functions and the two sets of confidence

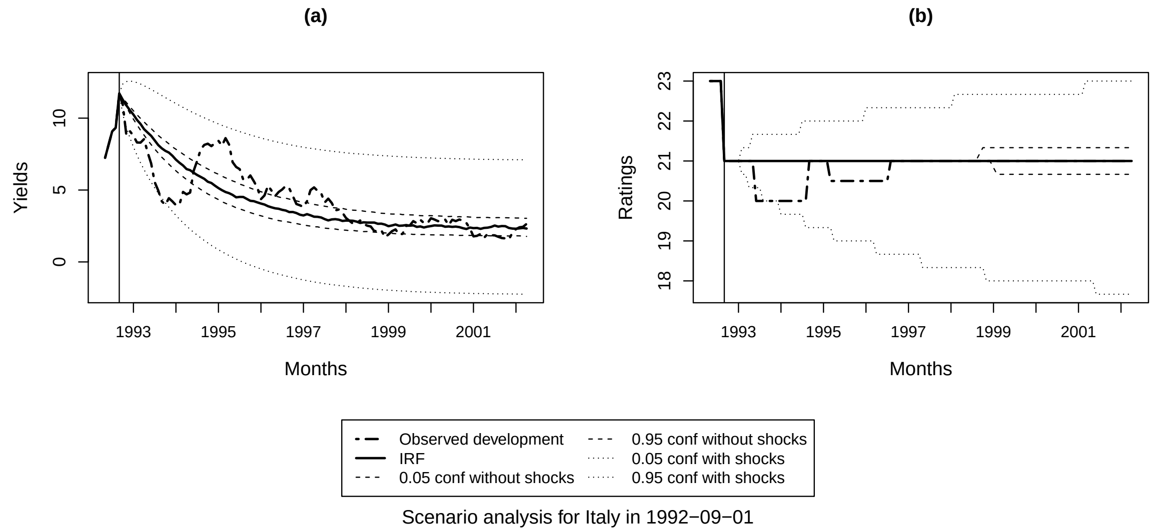

Figure 10: Observed development and impulse reaction functions in Italy after a downgrade in August 1992

bands. In addition to the confidence bounds reported for the IRFs, which only include parameter uncertainty, we provide confidence bounds including 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 the results for two events: First, the downgrade of Italy after the EMS crisis in the early 1990s as a case where dynamics are well predicted by our model. Second, the first sequence of downgrades of Greece during the ongoing debt crisis as an example where additional exogenous factors led to further consecutive downgrades and interest rate increases that have not been driven by the interaction of ratings and yields.

Italy in 1992: First, we examine 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+ (Aa1 in Moody's notation), making Italy the only G7 country with a rating below AAA at that time. A second downgrade, this time by two notches to AA-, followed on August 13, 1992. The reasons for these downgrades were the exceptionally high debt levels the Italian government had amassed in previous years (exceeding 100% of GDP in 1992) together with large external imbalances. Harsh austerity measures, privatization and laws aimed at reducing labor costs by the newly elected Italian government were considered positive by Moody's. However, the rating agency held that these measures had come too late. Therefore, the agency predicted that government debt would grow further during the 1990s (as it in fact did). Another fact that was thought to aggravate debt problems was the overvalued Italian currency, which was restricted by the rules of the ERM. Again, the rating agencies' assessment was largely sound. Italy left the ERM one month later in September 1992 and devalued

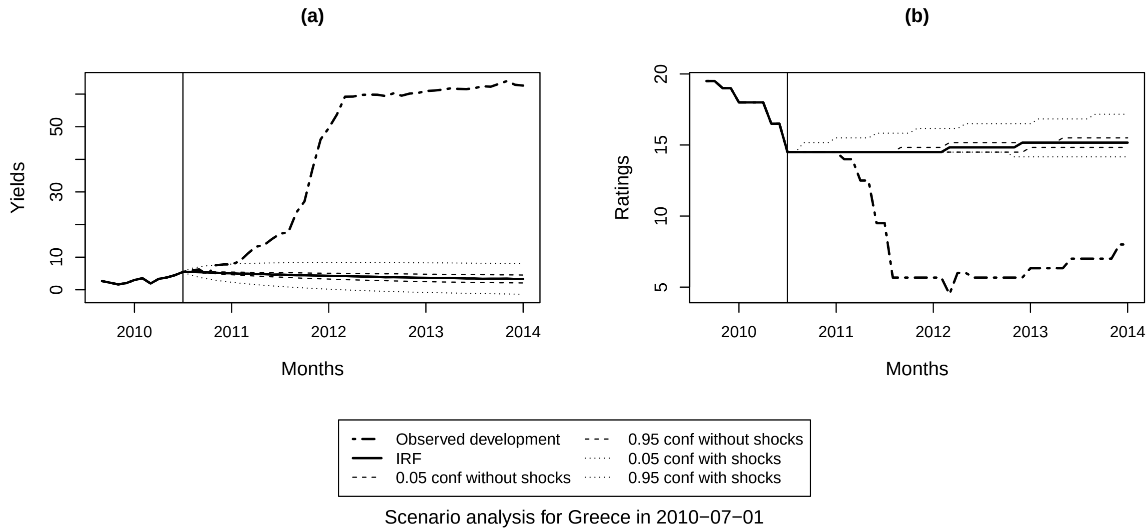

Figure 11: Observed development and impulse reaction functions in Greece after a series of downgrades until June 2010

strongly (see also El-Shagi, Lindner, and von Schweinitz, 2016).
It seems that the decision to downgrade the rating was largely a consequence of previous developments. Judging the reforms as coming too late, Moody's itself acknowledged that there had been enough to announce an earlier downgrade. Similarly, the European community had previously voiced its concerns, citing the Maastricht criteria (among them, a maximum government debt level of 60% of GDP) that were introduced in February 1992. Even the Italian minister of the treasury was not excessively concerned by the downgrade. Therefore, it is not surprising that the reaction of the markets was normal; see Figure 10. The observed development of yields is quite close to the development of the simulated IRF, with all differences (except for the first month) being within the wider set of confidence bands. The observed ratings are identical to the simulated series for most of the time. Overall, this is a typical case in which the discrete rating change reflected an assessment of the sustainability of government debt that was shared by the markets, which holds for the large majority of rating changes.

Greece in 2010: A completely different case is observed Greece in 2010. Membership in the Euro area had led to lower interest rates and improved ratings until October 20th, 2009, when the newly elected government opened its books and announced that the previous government had provided false low estimates of the expected government deficit. Only two days later, Fitch downgraded Greek sovereign bonds to A-. While yields increased and other members of the Euro area (together with the IMF, assisted by the ECB) issued a first rescue package in March 2010 and introduced the European Financial Stability Facility, ratings deteriorated further. The starting point of our simulation is June 2010, when rating agencies downgraded Greece to the investment grade threshold (Fitch just above, Moody's just below). Approximately until the end of 2010, the observed ratings and yields are in line with our simulated results. However, at that time, both 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 and further downgrades and, when private investors became part of a second rescue package in July 2011, a final downgrade to CCC- (in default with little prospect for recovery). Taken together, it should be clear that not ratings but political developments likely fueled the collapse toward default in 2012. That is, this development cannot be attributed to the initial (weak) downgrades of the rating agencies.

5 Conclusion

In conclusion, our evidence reconciles the two prominent conflicting views featured in the existing literature, which partly stresses the dangers of low 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 cycle is found in our data, at least not at a meaningful level. The usual interaction between ratings and yields fails to explain the downward spirals observed in a few cases. The vicious cycle theory seems driven by misinterpreting a few individual observations as typical.

On the other hand, there can be substantial costs 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 for better ratings, the increase is considerable for countries that begin 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 pay this cost for an extended period.

References

Afonso, A., D. Furceri, and P. Gomes (2012). Sovereign Credit Ratings and Financial Markets Linkages: Application to European Data. *Journal of International Money and Finance 31*(3), 606–638.

Almeida, H., I. Cunha, M. A. Ferreira, and F. Restrepo (2017). The real effects of credit ratings: The sovereign ceiling channel. *The Journal of Finance 72*(1), 249–290.

Arellano, C. and A. Ramanarayanan (2012). Default and the Maturity Structure in Sovereign Bonds. *Journal of Political Economy 120*(2), 187–232.

Arezki, R., B. Candelon, and A. N. R. Sy (2011). Sovereign Rating News and Financial Markets Spillovers: Evidence from the European Debt Crisis. IMF Working Papers 11/68.

Baum, C. F., D. Schäfer, and A. Stephan (2016). Credit rating agency downgrades and the eurozone sovereign debt crises. *Journal of Financial Stability 24*, 117–131.

Beber, A., M. W. Brandt, and K. A. Kavajecz (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. and M. Fratzscher (2013). The Pricing of Sovereign Risk and Contagion During the European Sovereign Debt Crisis. *Journal of International Money and Finance 34*, 60–82.

Bocola, L. (2016). The Pass-Through of Sovereign Risk. *Journal of Political Economy 124*(4), 879–926.

Bolton, P., X. Freixas, and J. Shapiro (2012). The Credit Ratings Game. *The Journal of Finance 67*(1), 85–111.

Breitung, J. and C. Roling (2015). Forecasting Inflation Rates Using Daily Data: A Nonparametric Midas Approach. *Journal of Forecasting 34*(7), 588–603.

Bruneau, C., A.-L. Delatte, and J. Fouquau (2014). Was the European Sovereign Crisis Self-Fulfilling? Empirical Evidence About the Drivers of Market Sentiments. *Journal of Macroeconomics 42*, 38–51.

Cantor, R. and F. Packer (1996). Determinants and Impact of Sovereign Credit Ratings. *The Journal of Fixed Income 6*(3), 76–91.

Carlson, M. and G. B. Hale (2006). Rating Agencies and Sovereign Debt Rollover. *Topics in Macroeconomics 6*(2).

Davidson, R. and E. Flachaire (2008). The Wild Bootstrap, Tamed At Last. *Journal of Econometrics 146*(1), 162–169.

Davidson, R. and J. G. MacKinnon (2010). Wild Bootstrap Tests for IV Regression. *Journal of Business & Economic Statistics 28*(1).

De Grauwe, P. and Y. Ji (2013). Self-fulfilling Crises in the Eurozone: An Empirical Test. *Journal of International Money and Finance 34*, 15–36.

De Santis, R. A. (2012). The Euro Area Sovereign Debt Crisis: Safe Haven, Credit Rating Agencies and the Spread of the Fever from Greece, Ireland and Portugal. ECB Working Paper 1419.

De Vries, T. and J. de Haan (2016). Credit Ratings and Bond Spreads of the GIIPS. *Applied Economics Letters 23*(2), 107–111.

Durbin, E. and D. Ng (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., A. Lindner, and G. von Schweinitz (2016). Real Effective Exchange Rate Misalignment in the Euro Area: A Counterfactual Analysis. *Review of International Economics 24*(1), 37–66.

El-Shagi, M. and G. von Schweinitz (2015). Risk and Return: Is there an Unholy Cycle of Ratings and Yields? *Economics Letters 129*, 49–51.

El-Shagi, M. and G. von Schweinitz (2017). Why They Keep Missing: An Empirical Investigation of Rational Inattention of Rating Agencies. IWH Discussion Papers 1/2017.

Ferri, G., L.-G. Liu, and J. E. Stiglitz (1999). The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes 28*(3), 335–355.

Forbes, K. J. and F. E. Warnock (2012). Capital flow waves: Surges, stops, flight, and retrenchment. *Journal of International Economics 88*(2), 235–251.

Gärtner, M. and B. Griesbach (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., B. Griesbach, and F. Jung (2011). PIGS or Lambs? The European Sovereign Debt Crisis and the Role of Rating Agencies. *International Advances in Economic Research 17*(3), 288–299.

Ghosh, A. R., J. I. Kim, E. G. Mendoza, J. D. Ostry, and M. S. Qureshi (2013). Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies. *The Economic Journal 123*(566), F4–F30.

Hill, P. and R. Faff (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, May). World Economic Outlook May 1997 – Globalization: Opportunity and Challenges.

Kaminsky, G., S. L. Schmukler, et al. (2002). Emerging Market Instability: Do Sovereign Ratings Affect Country Risk and Stock Returns? *The World Bank Economic Review 16*(2), 171–195.

Kiff, J., S. B. Nowak, and L. Schumacher (2012). Are Rating Agencies Powerful? An Investigation Into the Impact and Accuracy of Sovereign Ratings. IMF Working Paper 12/23.

Larraín, G., H. Reisen, and J. von Maltzan (1997). Emerging Market Risk and Sovereign Credit Ratings. OECD Development Centre Working Paper 124.

Löffler, G. (2005). Avoiding the rating bounce: Why rating agencies are slow to react to new information. *Journal of Economic Behavior & Organization 56*(3), 365–381.

Mammen, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *The Annals of Statistics*, 255–285.

Manso, G. (2013). Feedback Effects of Credit Ratings. *Journal of Financial Economics 109*(2), 535–548.

Mellios, C. and E. Paget-Blanc (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.

Panizza, U. and A. F. Presbitero (2014). Public Debt and Economic Growth: is there a Causal Effect? *Journal of Macroeconomics 41*, 21–41.

Pesaran, M. H., Y. Shin, and R. J. Smith (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics 16*(3), 289–326.

Reisen, H. and J. von Maltzan (1999). Boom and Bust and Sovereign Ratings. *International Finance 2*(2), 273–293.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica 48*(1), 1–48.

Vayanos, D. (2004). Flight to Quality, Flight to Liquidity, and the Pricing of Risk. NBER Working Paper 10327.

von Hagen, J., L. Schuknecht, and G. Wolswijk (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.

Woodford, M. (2009). Information-Constrained State-Dependent Pricing. *Journal of Monetary Economics 56*, S100–S124.

Wu, C.-F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics 14*(4), 1261–1295.

Appendix A: Tables and Figures

Table A1: Variables and sources

Variable	transformation	source
Ratings	Last day/month	www.countryeconomy.com
Yields (nom.)	mean, real	TR, 5Y benchmark bid
CPI	yoy growth	NSA

Note: The table contains the sources of all variables, with the list of countries being given in the following table A2. "TR" stands for Thomson Reuters, "NSA" for national statistical agencies

Table A2: 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	1986-09-01 – 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	1994-04-01 – 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	1994-09-01 – 2014-01-01

Note: Data availability refers to the main analysis based solely on ratings and (real) yields.

Table A3: Summary statistics of ratings and yields by country

	Real Yields				Average Ratings			
Country	meany	sdy	miny	maxy	meanr	sdr	minr	maxr
Argentina	2.58	4.32	-6.18	7.88	10.86	1.46	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	22.64	0.54	21.67	23.00
Brazil	3.68	4.25	-10.34	9.35	13.67	2.04	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	20.15	0.41	19.33	20.67
China	0.42	1.90	-4.81	4.46	19.52	0.96	18.00	20.67
Colombia	4.63	1.58	2.38	11.73	13.79	0.83	13.00	15.67
Denmark	3.17	2.67	-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	23.95	0.20	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	16.49	4.68	4.50	20.00
Hong Kong	2.30	4.50	-6.49	13.00	20.62	1.59	19.00	23.33
Hungary	2.07	2.26	-3.10	9.78	16.93	1.77	13.67	19.00
Iceland	1.45	3.33	-8.63	8.30	18.70	3.63	14.67	22.50
India	1.45	3.56	-8.90	10.86	14.34	0.83	13.00	15.00
Indonesia	2.30	2.80	-5.43	7.61	12.45	1.58	9.50	14.67
Ireland	2.73	3.07	-2.20	12.10	22.16	2.65	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	21.26	1.64	16.33	24.00
Japan	1.74	1.48	-1.41	5.66	23.15	0.98	20.67	24.00
Malaysia	1.31	1.63	-4.45	6.15	17.69	0.59	16.00	18.00
Mexico	3.22	1.52	-0.31	6.75	16.07	0.77	14.50	17.00
Netherlands	1.58	1.58	-2.39	4.88	24.00	0.03	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	23.90	0.18	23.50	24.00
Pakistan	2.11	1.72	-1.57	5.47	8.83	0.24	8.50	9.00
Peru	1.84	1.46	-0.13	6.32	15.48	0.57	14.50	16.67
Philippines	3.95	3.52	-2.78	11.42	12.98	0.90	12.00	15.00
Poland	3.36	1.71	0.42	8.55	17.87	0.64	16.00	18.50
Portugal	3.02	2.98	-0.94	14.43	20.44	2.95	13.00	22.00
Singapore	0.69	2.50	-5.84	6.37	23.21	1.09	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	20.68	2.08	16.33	22.00
South Korea	2.60	2.36	-0.97	9.02	18.23	1.88	11.50	20.67
Spain	2.94	2.58	-1.13	8.59	22.22	2.01	15.33	24.00
Sri Lanka	3.28	5.80	-11.59	17.34	13.24	1.95	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	16.25	0.89	14.50	17.00
Turkey	4.26	3.77	-1.69	12.77	12.77	0.93	11.50	14.67
United Kingdom	2.84	2.52	-3.93	7.09	23.98	0.11	23.33	24.00
United States	1.52	1.89	-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.

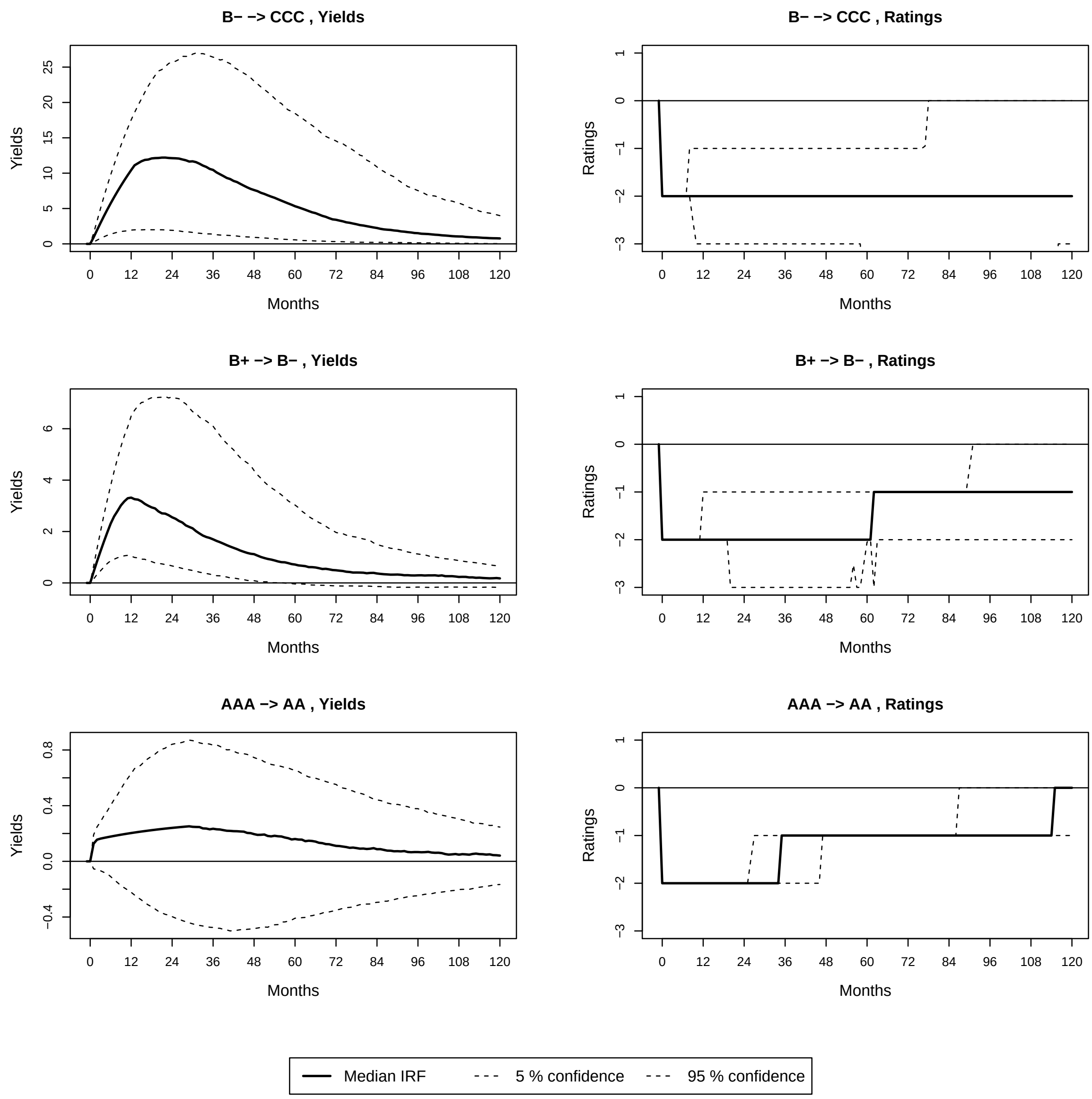

Figure A1: Selected impulse-response functions after a rating shock with confidence bounds.

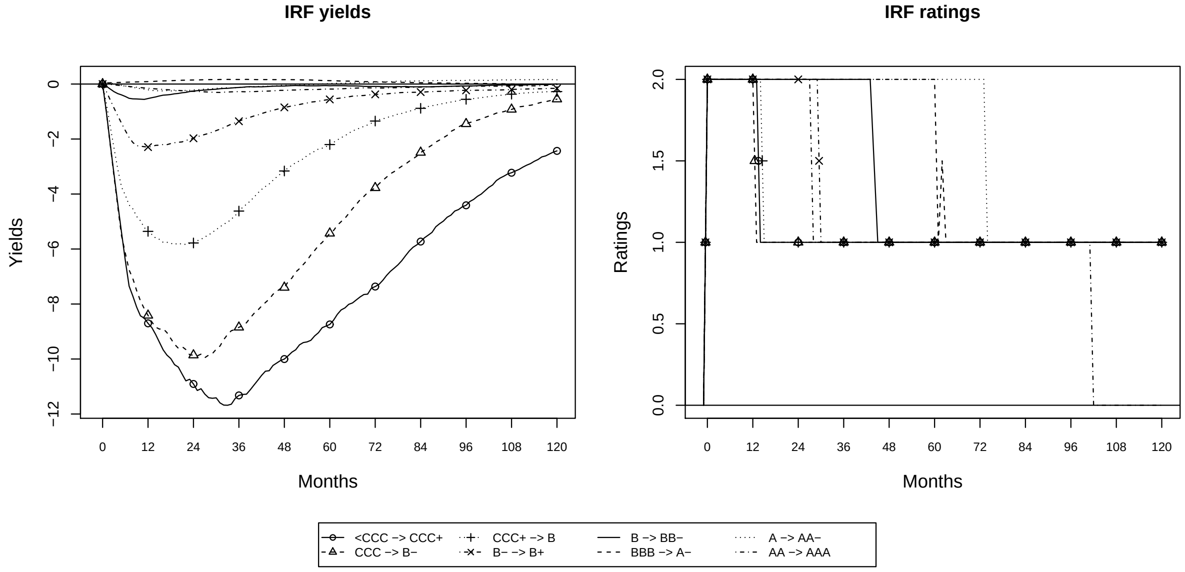

Figure A2: Selected impulse-response functions after a positive rating shock of two notches.

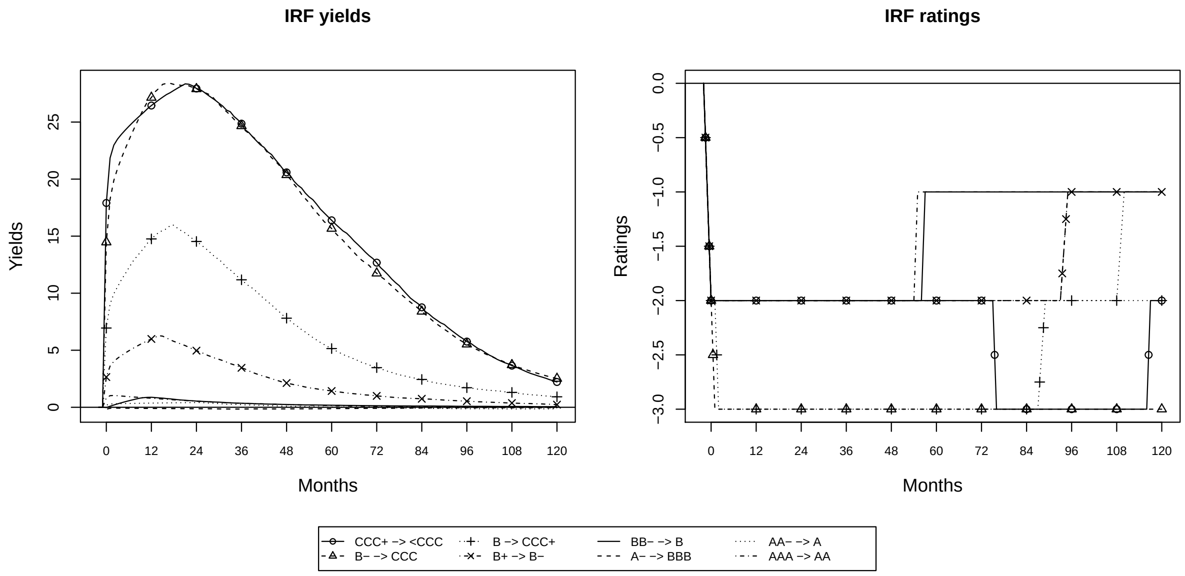

Figure A3: Selected impulse-response functions after a staggered rating shock.

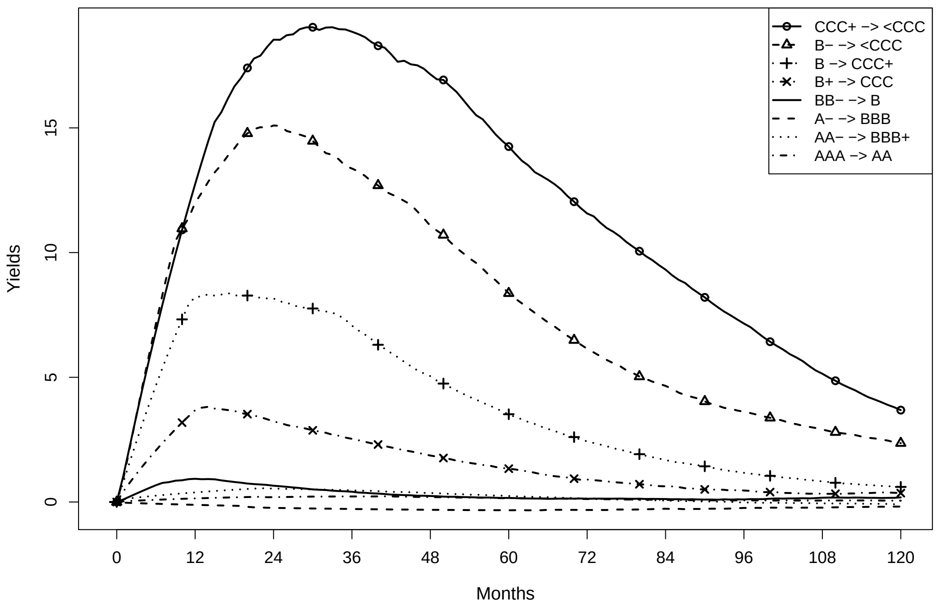

Figure A4: Robustness test with median ratings: Median impulse-response function of yields after a rating shock of two notches, selected ratings.

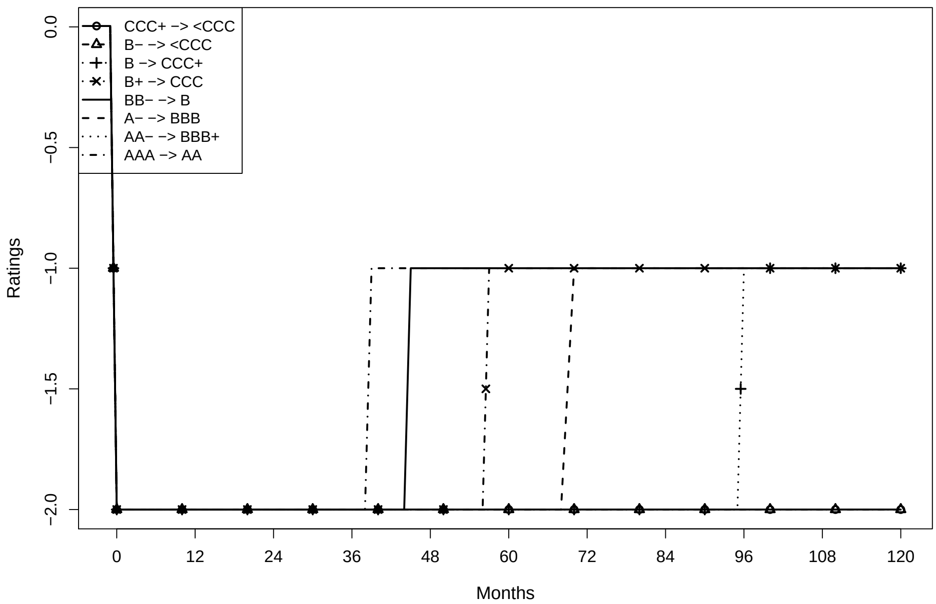

Figure A5: Robustness test with median ratings: Median impulse-response function of ratings after a rating shock of two notches, selected ratings.

Appendix B: 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. First, we estimate a restricted cointegration model where the long-run coefficients in both equations are equal ($\beta = \psi$). Second, we compare the restricted and the unrestricted models using model-specific optimal smoothing parameters λ. Third, to guarantee comparability, the likelihood ratio test is performed twice more for pairwise identical λ. That is, we compare both models using the optimum λ selected for the cointegration model, and we compare both models using the λ selected for the model with individual long-run relationships. The model comparison is based on the ML results for computational reasons.[21] This decision comes with a caveat: Because the residuals are not normally distributed – which is the very reason for our bootstrapping procedure – the test results have to be interpreted with caution. However, the three tests all return similar results. In all cases, the restricted cointegration model is rejected at the 1% level.

[21] The cointegration model takes an extremely long time to bootstrap. Because evidence from the model with multiple long-run relations indicates that the coefficients change only slightly, we carry over the ML results to avoid conducting a full-fledged bootstrap of the cointegration system.