Why they keep missing: An empirical investigation of sovereign bond ratings and their timing*

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

Two contradictory strands of the rating literature criticize that rating agencies merely follow the market on the one hand, and emphasizing that rating changes affect capital movements on the other hand. Both focus on explaining rating levels rather than the timing of rating announcements. Contrarily, we explicitly differentiate between a decision to assess a country and the actual rating decision. We show that this differentiation significantly improves the estimation of the rating function. The three major rating agencies treat economic fundamentals similarly, while differing in their response to other factors such as strategic considerations. This reconciles the conflicting literature.

*For valuable comments and suggestions, the authors thank anonymous referees, Geraldine Dany, Arno Hantzsche, Oliver Holtemöller, Alexander Kriwoluzky, Axel Lindner, Winfried Koeniger, Michael Koetter, Thomas Steger, Lena Tonzer and Mirko Wiederholt. The authors also thank participants at the EEA 2017, the 2017 IAAE Annual Conference, the CEF 2018, and seminar participants at the IWH. A previous version of this paper using averaged ratings across agencies is available as IWH discussion paper 1/2017. Replication data and code can be found on the authors websites.

†Corresponding author. Email: makram.elshagi@gmail.com

Keywords: Rating agencies; sovereign risk; decision timing

JEL-Classification: C14; C25; G24

1 Introduction

There is an abundance of papers showing that ratings merely follow financial markets, that is that interest rates (capturing the risk assessment of market participants) move before ratings are adjusted. During major economic crises, rating agencies often responded with downgrades after the downward spiral of worsening risk assessment captured in increasing sovereign bond yields had already begun.[1] Yet, there is an equally large literature providing evidence that rating changes have substantial effects on government bond yields and capital movements which has been interpreted both as evidence for informative ratings (Cantor & Packer 1996) and for markets merely following stale ratings (Ferri, Liu & Stiglitz 1999). When interpreting the effects of ratings on yields in the latter fashion, mutual (Granger) causality implies that there is a substantial risk of vicious cycles of downgrades and capital flight driving sound economies from a stable equilibrium to default as first argued by Ferri et al. (1999). However, this hypothesis was challenged by others who explicitly model the joint dynamics and find no evidence for multiple equilibria (Mora 2006). In this paper, we argue that a lot of these seemingly contradictory results in the previous literature on sovereign bond ratings are due to the lack of treating the decision to reassess a country's rating and the decision on the assigned rating separately.

We demonstrate that a lot of what is considered to be inaccuracy of the rating process, is indeed due to the natural inertia of ratings, driven by a reluctance to update. In a world where rating agencies are not the sole source of information on debtor quality and the other economic agents are aware of the rating process, the difference is quite meaningful. Based on private information, market participants can assess to some degree, whether ratings are merely being updated to the current general risk assessment (already known to the market) or whether they provide new information. Thus, it becomes apparent why ratings occasionally seem to lead markets, and follow them at other times. This situation, where the leading and following variables change over time, is distinctively different from the situation usually associated with mutual Granger causality, where two variables (such

[1] As an example, see e.g. El-Shagi (2010) for a case study of the Asian Flu. For corporate ratings, a lack in rating timeliness is e.g. shown by (Kou & Varotto 2008).

as risk and returns) always affect each other.

As the rating process combines private and public information with expert knowledge, rating agencies should in principle be able to provide valuable information. Indeed, in many instances there is evidence that interest rates respond to sovereign rating changes at least in the short run (e.g. Ferri et al. 1999, De Santis 2012). This response is heterogeneous in the degree of surprise of these changes, see Goh & Ederington (1993) (for the corporate bond market) and El-Shagi (2016). This nuanced reaction of markets makes it hard to believe that the adjustment in interest rates – when and where it occurs – merely reflects irrationality on the side of market participants who respond to stale ratings. Rather, it seems that markets are well aware that ratings are informative at times, while they fail to be so at other times. Therefore, it is unlikely that the frequent delay in ratings is caused by incompetence or a lack of understanding of the rated markets. The heterogeneous response of markets – and in turn the contradictory results of the literature – are much more in line with our interpretation that rating agencies often fail to do necessary evaluations of a country in time, but that they are fairly accurate if they do.

In this paper we explicity distinguish the decision of rating agencies' whether to evaluate a country and how to evaluate it. To do this, we make use of a novel data set which does not only contain rating levels but all rating announcements by the three biggest rating agencies (Moody's, S&P and Fitch) for sovereign ratings for 138 countries between 1974 and 2017. We propose a new *selection inflated ordered probit* model (SIOP). The SIOP is a system of two equations that separates (i) the selection of whether or not rating agencies opt to gather new information and update a rating, as described by a probit equation; and (ii) the decision how to update the rating, as described by an ordered probit equation. In spirit, the SIOP is thus closely related to a zero- or middle-inflated ordered probit model (MIOP), with a crucial difference in how a decision to not reevaluate a rating is separated from a reevaluation which confirms the previous rating: The MIOP-model treats both cases as observationally equivalent (as the observed rating level does not change), while the SIOP-model allows to differentiate between the two cases through the observation of announcement periods. We validate our assumption by showing that our new SIOP-model strongly outperforms both a MIOP-model and a simple ordered probit model (which would assume reevaluation every period).

So far, the decision whether to update a sovereign creditor rating at all has been underappreciated by the literature, partly because most of the literature aims to explain rating levels rather than actual rating decisions, and is thus unable to distinguish between deviations from the appropriate rating that are caused by misjudgment and those that are driven by lack of rating activity in general and the corresponding stickiness of ratings. While the stickiness of ratings is widely acknowledged and analyzed in the literature on corporate ratings, only few papers on sovereign ratings explicitly account for it. Moreover, the sources of stale ratings are typically not identified.

If rating agencies decide to update a rating, our analysis suggests that the direction of updates is strongly driven by macroeconomic fundamentals. The three big rating agencies treat fundamentals similarly. This is reassuring, as the credit default risk measured by sovereign ratings should be independent of the assessing agency. The reasons why rating agencies decide not to update a rating are manifold and differ between agencies. At least four possibilities should be mentioned: First, under *rational inattention* (Sims 1998) rating agencies weigh the cost of reassessing and collecting information against the benefit of a more accurate rating. Consistent with this channel, we find that both the time since the last announcement and the cumulative change of fundamentals since then affect announcement probabilities. Second, rating agencies may decide whether or not to act based on *strategic reasons*. In particular, we find that agency interaction affects announcement probabilities. Third, *rent seeking behavior* could guide their actions, as adjusting country ratings usually also requires to reassess a wide range of corporate rating of (paying) customers. However, if anything, we find that rating shopping (which would be an indication of rent seeking) seems not to be present at the level of sovereign ratings. Yet, despite controlling for those variables in the probit equation describing the decision to update a rating, we still find non-fundamental variables (such as lagged rating changes) to be important for the outcome of rating reassessments. This indicates, that the rating decision itself is subject to considerations beyond fundamental economic conditions. In particular, rating agencies seem to prefer a sequence of small changes over a major single rating change. Fourth, the common practice of *rating through the cycle* might cause situations that appear as if rating agencies respond too late. By comparing models with and without cyclical variation in explanatory variables, we show that this is not the predominant reason for rating agencies not to update a rating.

2 Literature review and institutional setup

The rating process The basics of the sovereign rating process are fairly well documented (Beers, Cavanaugh & Takahira 2002, Beers & Cavanaugh 2008, Fitch 2015). The rating decision is based on a wide selection of indicators capturing political risk, macroeconomic fundamentals, fiscal and monetary as well as external variables.[2] Importantly, not all necessary information is publicly available (in real-time). Instead, credit rating analysts need to be in close contact with ministries and other policy institutions. Thus, the decision to review a sovereign rating should come at significant costs to the credit rating agency.

Both bond and issuer ratings – such as sovereign ratings – are usually solicited ratings, i.e. paid for and requested by the issuer. Due to the business relation between issuers and raters, concerns of rating shopping and opportunistic behavior of rating agencies who do not want to lose customers have frequently been voiced, mostly with respect to corporate bonds.[3] For sovereign ratings, those issues seem to matter less, for at least three reasons. First, there is a non-negligible share of unsolicited sovereign ratings, with different degrees of cooperation (such as access to information) from the issuer. It is important to note that rating agencies traditionally keep updating originally solicited ratings for a while even after the solicitation stops (Kim & Wu 2011).[4] One reason for this behavior might be the

[2] A deeper discussion of these factors can be found further below.

[3] See, among others, Skreta & Veldkamp (2009), Sangiorgi, Sokobin & Spatt (2009), Griffin, Nickerson & Tang (2013), and Sangiorgi & Spatt (2016). Related to our research question, Cornaggia & Cornaggia (2013) document that issuer-paid ratings adjust slower (in particular downwards) than subscriber-paid ratings.

[4] Originally it was subject to the rating agencies' discretion when to review ratings. However, the

importance of reputation for the business model of rating agencies. As the accuracy of ratings, i.e. a correct determination of default probabilities, is hard to measure with rare events and comparably small samples (such as the number of sovereign issuers), a good reputation can be important for future income (see also the extensive discussion in Becker & Milbourn 2011). Thus, a second reason against unreasonably good sovereign ratings is a desire to signal market knowledge through high quality ratings in order to attract corporate clients. Third and last, rating shopping is based on the possibility to use the best rating issued (by accredited agencies) where regulation is concerned. Except in the past few years, when the ECB applied different haircuts to sovereign bonds based on their rating, regulation was of minor importance for home-country sovereign bonds that were often considered risk free under the Basel 2 framework. Moreover, most countries are rated by all the major rating companies and the ratings are freely available. That is, the negative signal of a downgrade is perceived by the market and frequently covered by the media whether or not other (possibly better) ratings exist. The public availability and visibility of sovereign ratings make opportunistic behavior less likely. Sovereign ratings are subject to immense scrutiny by politicians and media alike and thus the reputational dangers of assigning overly generous ratings are immense. To conclude, it seems to be in the best interest of credit rating agencies to deviate not too far from a rating that reflects the true credit default probability and adjust ratings regularly.

Fundamental determinants of sovereign ratings Over the past three decades, there has been an abundance of literature on ratings. Starting with Feder & Uy (1985), roughly 60 papers – to our knowledge – are concerned with the determinants of sovereign ratings. The first papers in this literature investigated Institutional Investor ratings and Euromoney ratings (see e.g. Brewer & Rivoli (1990), Cosset & Roy (1991) and Lee (1993)). However, starting with the seminal contribution of Cantor & Packer (1996), the

ECB introduced some regulation in 2013 for ratings issued in the European Union that requires biannual updates of sovereign ratings and – more importantly – requires the agencies to state the intended publication dates for ratings in the following year at the end of the year (EU Regulation No 462/2013). Due to the recent introduction of the regulation and its limited coverage, most of our sample is not affected. Additionally, there is some room for decisions of the rating agencies. If they feel that rating changes are required, updates that do not follow the calendar are generally permitted.

literature focused on the major credit rating agencies.

The core set of variables used in the current literature is still the one that has been established by Cantor & Packer (1996), who look at a combination of debt, the fiscal balance and a range of macroeconomic fundamentals, such as income per capita, inflation etc. A large number of additional indicators has been tested in later contributions for their potential impact on ratings. These extensions can be broadly grouped in two strands.

First, a fairly large range of papers look at the role of political and institutional factors for ratings (Brewer & Rivoli 1990). Depken, LaFountain, Butters et al. (2007) introduce corruption into the baseline model, which has been a staple variable in the later literature either as part of a wider index (Depken et al. 2007) or as a separate indicator (Amstad & Packer 2015). In a similar vein, Butler & Fauver (2006) add institutional quality and legal origin as indicators of the soundness of institutions. Haque, Mark & Mathieson (1998) look at a wide range of indicators of political stability, such as coup d'etats, strikes, demonstrations. While the previous papers found institutional quality to reflect positively on ratings, Block & Vaaler (2004) find a negative effect of elections in developing economies.

Second, a lot of authors test the assumption that different country groups are treated structurally different by rating agencies. Gültekin-Karakaş, Hisarcıklılar & Öztürk (2011) split their sample in emerging markets and developed economies, finding the latter to be favored slightly. Butler & Fauver (2006) split their sample by the level of debt. Recent papers by Fuchs & Gehring (2017) and Altdörfer, De las Salas, Guettler & Löffler (2016) find some evidence of a home bias of rating agencies.

Our paper takes a fairly wide approach, including – where available – all drivers that have been identified robustly in the previous literature. The key difference between our paper and the majority of the previous literature is that we explicitly account for the dynamics and persistence in rating decisions.

Ratings, persistence and timing A key criticism concerning rating agencies is the timeliness of sovereign ratings and the dynamic interaction of rating changes with the macroeconomy. Yet, the vast majority of papers digging deeper into determinants of sovereign ratings study rating levels (rather than changes) in a cross section of countries (Cantor & Packer 1996, Afonso 2003, Amstad & Packer 2015). Even where panel data is utilized, empirical strategies often aim to explain long-run rating levels in a nondynamic framework (see for example Ferri et al. 1999, Depken et al. 2007). In Table 1, we compare the actual rating levels to those predicted by a simple ordered probit model. In brackets, we add information on the number of observations where the model predictions implied a constant rating, although we observe a rating change in the data. We see that a large fraction of false rating level predictions occur in periods where agencies changed their rating. Even worse, there is no overlap between periods where a rating change occured, and periods where a rating change would have been predicted (i.e., where the predicted rating was different to the current one).

Table 1: Predicted versus actual ratings (ordered probit model)

Actual rating	Predicted rating							
	AAA	AA	A	BBB	BB	B	CCC	C
AAA	7618	16 (16)	0	0	0	0	0	0
AA	108 (17)	3662	10 (10)	0	0	0	0	0
A	0	12 (12)	3904	52 (21)	0	0	0	0
BBB	0	1 (1)	16 (16)	5409	23 (23)	0	0	0
BB	0	0	0	41 (22)	1906	23 (7)	0	6 (6)
B	0	0	1 (1)	0	24 (10)	1290	3 (3)	2 (2)
CCC	0	0	0	0	0	198 (14)	116	79 (8)
C	0	0	0	0	5 (5)	4 (4)	6 (6)	168

Note: Results of a simple ordered probit model of 8 rating classes on all variables employed in our baseline model, allowing for agency-specific thresholds between rating classes. The plain numbers show the fit of the model, while those in brackets count observations where an observed rating *change* was not predicted by the model.

In addition to this, the aforementioned strands of literature on sovereign ratings mostly fail to account for the dynamic aspects of ratings, such as persistence and speed of adjustment (as mentioned by Mora 2006, El-Shagi & von Schweinitz 2015). A more explicit treatment of persistence – although by no means explaining it – can be found in papers controlling for lagged rating levels (see e.g. Haque, Kumar, Mark & Mathieson 1996, Haque et al. 1998, Mulder & Monfort 2000). Some, such as Alsakka & ap Gwilym (2009), estimate models in first differences (accounting for persistence by construction), where they explicitly account for momentum in changes. Hu, Kiesel & Perraudin (2002) estimate transition matrices, which account for heterogeneity in rating persistence. Schumacher (2014) and El-Shagi & von Schweinitz (2015) estimate VAR models that jointly consider macroeconomic developments and ratings, thereby shedding more light on the dynamic aspects of rating decisions.

However, very few papers explicitly discuss persistence in depths (Dimitrakopoulos & Kolossiatis 2015, Hantzsche 2017). While the former estimates higher order AR models, the latter is a parallel paper to ours. It is to our knowledge the only other paper that explicitly tries to explain persistence, rather than just assuming persistence to be an exogenous feature of the data. We go beyond the model of Hantzsche (2017), because our paper exploits information on rating announcements whether or not the rating is actually changed, uses a much wider sample, and accounts for more indicators and potential nonlinearities.

Persistence, i.e. the absence of new rating assessments, might exist due to a range of factors, going far beyond what the literature suggested in the past. We are mostly interested in persistence that is caused by the decision of the rating agency to not evaluate a country or avoid rating changes it believes to be fundamentally appropriate. To do so, we can draw on the corporate rating literature where the dynamic properties of ratings have been a key issue for many decades. Mizen & Tsoukas (2012) is a similar paper to us in that they provide an in-depth discussion of state dependence (i.e. persistence) of corporate ratings and combine many rating-based variables from the literature that we consider to be important for sovereign ratings as well. They argue that lagged ratings are important determinants for current ratings, but also cover the time spent in the current rating group (as in Lando & Skødeberg 2002), momentum of rating upgrades and downgrades (Carty & Fons 1994) and cross-sectional waves of rating changes (Amato & Furfine 2004). These variables and others can be broadly categorized in four (partly overlapping) reasons for rating persistence.

First, what rating agencies do goes far beyond a simple econometric model with regularly observed variables. They need to conduct interviews, collect and assess the importance of unique factors such as newly introduced laws and regulation to make judgment based adjustments, etc. Thus, inattention is to some degree rational (Sims 2003, Woodford 2009, Maćkowiak & Wiederholt 2009), if there is no reason to believe that major changes occurred.

Second, rating agencies might have a preference to avoid adjustments and to prefer a sequence of small adjustments over larger ones for strategic reasons, partly to avoid upsetting financial markets (and thus draw more criticism), partly to avoid the reputational cost of admitting that they failed to reassess long enough to warrant such large changes. This will naturally give rise to rating momentum (Carty & Fons 1994). Those strategic decisions might be strongly influenced by competition which is believed to lower gains from reputation and thus to lower rating quality (Becker & Milbourn 2011). Applied to the decision to revise a rating, this might imply that rating agencies review less often (and potentially less thoroughly) in the face of stronger competition because the benefits of reviews are lower.

Third, there might be rent seeking behavior. Agencies might be reluctant to downgrade a paying customer or – in the case of sovereign ratings – even countries that do not solicit ratings (and thus do not pay) if they are sufficiently influential.[5]

Fourth, rating agencies often emphasize that they intend to provide ratings "through the cycle". This entails that they need to distinguish between cyclical variations of sovereign default probabilities and changes in the trend. At the current margin of the data, cyclical movements and sudden changes in the underlying trend are almost impossible to distinguish. If rating agencies' misjudge an event to be part of the cycle (i.e. transitory) although it turns out to be persistent, their response will follow the macroeconomic indicators with substantial delay. Since financial markets – in particular short term interest rates – often follow the cycle, this will result in ratings following interest rates rather than leading them. Whereas rating agencies have responded to the timing problems related to rating through the cycle by providing "point-in-time" ratings for many corporate borrowers, to the best of our knowledge those don't exist for sovereign

[5]Both strategic reasons and rent seeking behavior could be considered as special cases of rational inattention, because they essentially change the return of rating adjustments. In our classifications we mostly consider drivers to be "rational inattention" if they drive the cost side of ratings, rather than the benefit of issuing an accurate rating.

ratings. We therefore compare our baseline model to a model where we only include pseudo-real-time trend indicators.[6] We demonstrate that even then – i.e. in the absence of cyclical movements in potential rating drivers – we still observe excess persistence that our approach can explain.

3 How to model rating decisions

Rating agencies – as argued in the literature review above – do not necessarily reevaluate rating decisions continuously. In the majority of periods, the probability of coming to a new rating conclusion is insufficient to justify the cognitive and informational costs of a full reassessment. Instead, there may be long periods of time where agencies do not even consider a reevaluation. We capture this insight by a *selection-inflated ordered probit model* (SIOP) combining a probit equation for the probability of reevalution with an ordered probit equation for the direction of the rating decision, conditional on reevaluation.

In order to determine the drivers of the two decision problems of rating agencies, we model rating changes y as a combination of two processes y^d and $\tilde{y}$. The first process y^d describes the decision to reevaluate a rating. We assume that every reevalation is followed by an announcement of the rating agency, such that our observation of announcements gives us full knowledge about the reevaluation decision y^d. We show below that this assumption is supported by the data.

The second process is the direction of rating changes $\tilde{y}$ in case of reevaluation. There are three categories of rating changes, *downgrade* (-1), *no change* (0), or *upgrade* (+1), which can only be observed in periods where an actual reevaluation takes place. That is, only reevaluation periods ($y^d = 1$) are informative on the influence of explanatory variables on the direction of rating changes $\tilde{y}$.

These two models are combined to model the observed (directional) rating decision y, which now depends both on X (the determinants of rating reevaluations) and Z (the determinants of rating decisions in case of reevaluation). The probability of a rating *downgrade* is the joint probability of a rating reevaluation and a downgrade decision in

[6] This is inspired by the linear dynamic model of Koopman & Lucas (2005).

case of reevaluation ($Z\gamma \leq 0$). Similarly, the probability of a rating *upgrade* is the joint probability of a rating reevaluation and an upgrade decision in case of reevaluation.

Table 2: Observable states for rating decisions

reevaluation y^d \ change $\tilde{y}$	-1	0	1
0	not observable	$y = 0$	not observable
1	$y = -1$	$y = 0$	$y = 1$

In case of uncorrelated errors,[7] those joint probabilities can be obtained as products of the individual probabilities obtained from a probit model governing the behavior of y^d and an ordered probit model governing the behavior of $\tilde{y}$.[8] It needs to be noted, that the probability of *no change* (the remaining third case) combines two distinct observable cases: the case of no reevaluation ($Pr(y^d = 0|X)$) and the case of reevaluation with no rating change ($Pr(y^d = 1|X)Pr(\tilde{y} = 0|Z)$) yielding our full model:

$$\begin{aligned} P(y = -1|X, Z) &= P(y^d = 1|X)P(\tilde{y} = -1|Z) \\ P(y = 0|X, Z) &= P(y^d = 0|X) + P(y^d = 1|X)P(\tilde{y} = 0|Z) \\ P(y = 1|X, Z) &= P(y^d = 1|X)P(\tilde{y} = 1|Z) \end{aligned} \tag{1}$$

That is, we have de facto four different states jointly determined by y^d and $\tilde{y}$, see Table 2. The coefficients of the SIOP-model can be easily determined by likelihood maximization.

Statistically, our approach is very similar to a Heckman selection model, where we assess the direction of change in a limited dataset of observations where the rating has been assessed, and a "selection" equation determining when a country will be evaluated. Theoretically, those equations can be substituted in one another, to compute the total

[7] As a robustness check, we also allow for correlated errors. In this case, the probability of an observed rating decision y can be derived from a bivariate normal distribution instead of a multiplication of two normal distributions in equation (1). However, as the correlation is basically zero in the baseline SIOP model and is rejected by a likelihood ratio test, we do not discuss results separately.

[8] Since upgrades are impossible for a AAA rated country, and downgrades are impossible for a D rated country, we use a boundary adjusted ordered probit model. The equations for the adjusted ordered probit probabilities are found in the online appendix C.

effect of various indicators on the probability for upgrades and downgrades. Yet, contrary to selection models we are not only after estimating the joint effect, but actually interested in the individual equations, because both have an interesting story to tell about how rating dynamics work. Economically, we are thus much closer related to questions that have been assessed by midpoint inflated ordered probit (MIOP) models, where an ordered variable (such as rating changes) is modeled using two equations (Brooks, Harris, Spencer et al. 2007, Bagozzi & Mukherjee 2012).[9] The most prominent macroeconomic application of this method has probably been interest rate setting, but it has also been applied to sovereign bond ratings (Hantzsche 2017). Those models do, however, assume that the data does not allow to distinguish whether there is the deliberate decision to not change the variable of interest, or if change has not even been considered. We argue that this assumption is not warranted. Rating agencies frequently publish rating announcements confirming the current rating: around 8.5% of observations at the country-month-agency level in our baseline model contain announcements, of which less than 25% (2% of total observations) are rating changes. Due to the large share of confirmatory announcements, and the missing incentive for rating agencies to perform costly reevaluation exercises without sending a public signal about its actions, we can safely assume that we do indeed have the information on rating assessments. We confirm the superiority of our model when we test it against a MIOP model.

4 The determinants of ratings and their timing

Our analysis encompasses three different groups of variables: (a) announcements by rating agencies and non-fundamental variables derived from this information (most importantly, rating changes $\tilde{y}$ and reevaluation decisions y^d), explained in Subsections 4.1; (b) fundamental variables related to government default probabilities in Subsection 4.2; (c) home bias, political variables, and outlooks, which we only use in robustness checks in the online appendix B. In order to be able to describe the rating decision process at a more

[9]The MIOP was developed based on the zero-inflated ordered probit of Harris & Zhao (2007), which inflates the lowest instead of the middle category.

granular level, we will work with monthly data. This, however, forces us to interpolate some of the fundamental variables, especially for developing economies, as also explained in Subsection 4.4. Details on data treatment are provided in Table A.1, and summary statistics are given in Table A.2 in the online appendix A.

4.1 Rating agencies

Ratings Rating levels and announcement dates are drawn from the website http://www.countryeconomy.com. The website collects rating data for the three big rating agencies: Moody's, Standard & Poors and Fitch. Data on ratings for foreign-currency denominated loans currently span 138 countries partly going back as far as 1974.[10] As a rating stays constant from one announcement to the next, availability of rating data per country is solely determined by the first (reported) announcement. We consider the rating agencies separately. That is, each observation is time, country, agency specific. Figure 1 shows the number of countries for which rating data are available at each date. For the first part of the sample, our rating information is concentrated on OECD countries. Later, more and more countries are rated and data availability increases. In the last part of our data, certain explanatory variables are unavailable in some countries, reducing the scope of our analysis slightly.

In contrast to the wide literature on ratings that empirically assesses rating levels in the cross section, our panel approach requires to look at rating changes. We therefore opt for a simple ternary indicator of change as our dependent variable, only distinguishing upgrades (1), downgrades (-1) or unchanged ratings (0) by an individual agency within the current month.[11] Figure 2 shows the share of observations where there was an announcement or a rating change. While the frequency of announcements decreases only slightly for higher ratings, the share of rating changes drops towards zero very quickly.

The lagged rating level itself ($rating$) and its square ($rating^2$) are used as explanatory variable as in Mizen & Tsoukas (2012), giving some error correction interpretation to the

[10]We collected data on 2017-10-04. Due to the lower availability of explanatory variables, we are only able to work with data from 59 countries, starting in December 1996.

[11]In total, there are 114 observations, or 1.5% of all observations, where a rating agency changed the rating by more than one notch in one month. This number is too low to be empirically exploited.

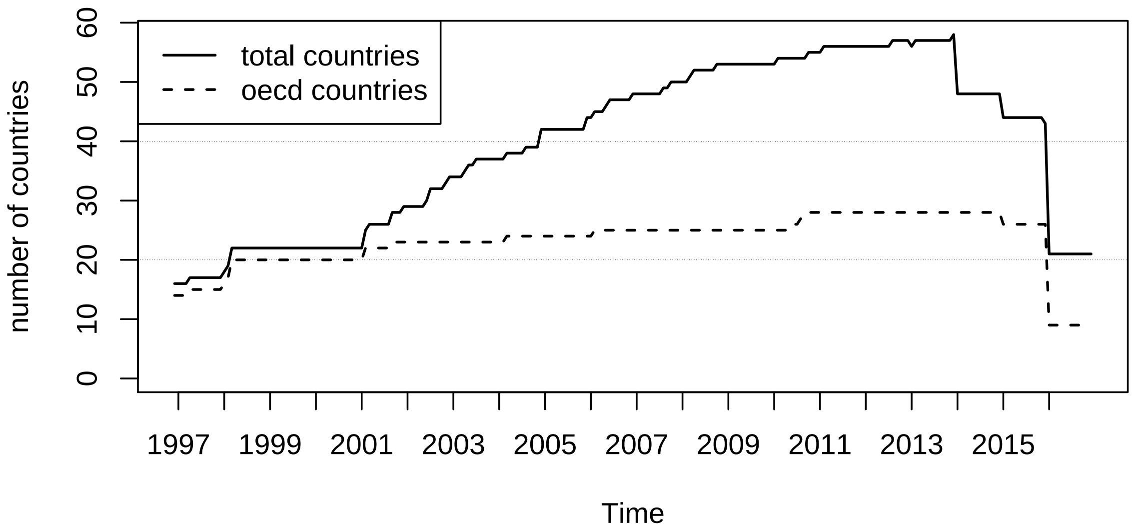

Figure 1: Number of observations over time

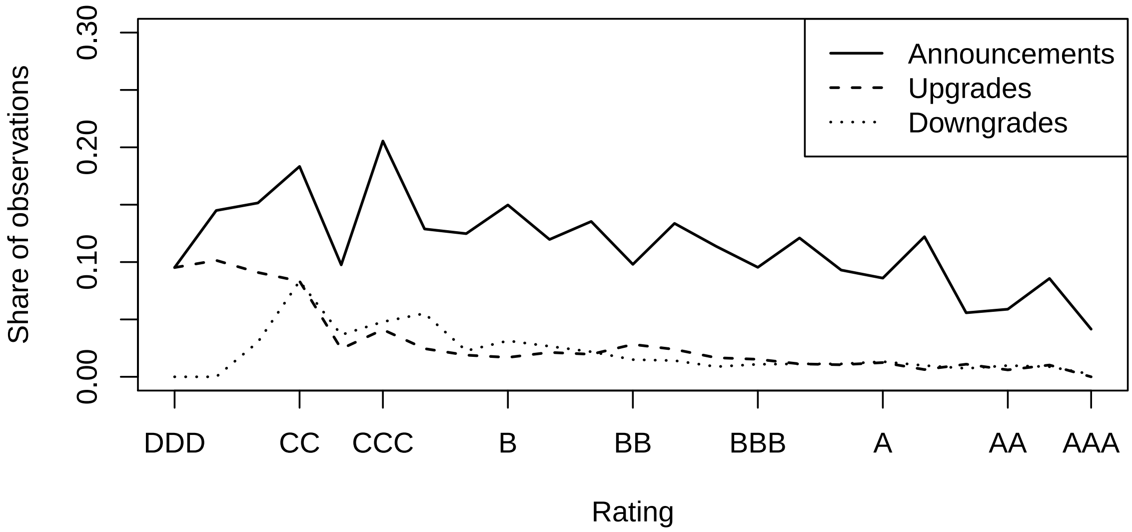

Figure 2: Rating announcements and changes for different rating classes

model, thus allowing to draw level conclusions from a first difference model. As in the previous literature, we code rating levels on a discrete scale where 24 is a AAA rating and where 0-3 denotes different default ratings.

Time and inattention The key innovation of this paper is to model ratings in a form that explicitly accounts for the often criticized – but possibly rational – inattention rating agencies seem to show. A main reason for inattention is that the fundamental reasons behind a rating change slowly. That is, unless there are specific reasons to look at a country at a certain time, a country will be only screened occasionally. We capture this idea in two ways. First, we look at the passing of time, measured in years since the last rating (*years*) as in Lando & Skødeberg (2002). Second, we assume that observing past economic fundamentals is largely cost-free, at least in comparison to judging future fiscal policy and other soft factors that enter sovereign ratings. Based on this logic, we generate an indicator of fundamental change since the last rating announcement at a country level (*changefund*). This indicator is based on all fundamental variables, as discussed in the next subsection. It measures the average squared total growth rate of all fundamentals since the last rating announcement relative to their usual month-on-month growth rates.[12] A large value of *changefund* indicates that there have been large changes of fundamental variables since the last announcement, warranting a reevaluation. As we take changes in squares and thus treat positive and negative changes equally, we expect this variable not to influence the direction of reevaluation.

By using time and structural change, we can capture fairly complex dynamics, that go beyond a simple linear model. In particular in conjunction with lagged rating changes –

[12]In addition to this, we winsorize growth rates at their 5% and 95% quantile to avoid that averages are driven by extreme growth rates (like hyperinflation). Let s be the time of the last rating announcement by rating agency a for country c. Then $(I_{j,c,t} - I_{j,c,s})/I_{j,c,s}$ is the percentage change of indicator I_j until time t. We normalize this percentage change by average monthly growth rates, take squares and winsorize (denoted by $w(\cdot)$) to get an indicator of fundamental change of indicator I_j:

$$changefund(j,c,t,a) = w\left[\left(\frac{(I_{j,c,t} - I_{j,c,s})/I_{j,c,s}}{N^{-1}\sum_{c=1}^{C}\sum_{\tau=1}^{T}\Delta I_{j,c,\tau}/I_{j,c,\tau-1}}\right)^2\right]. \quad (2)$$

The aggregate indicator $changefund(c,t,a)$ standard-normalizes $changefund(j,c,t,a)$ across all fundamental indicators.

measured through the total number of rating upgrades and rating downgrades of a country by the same agency within the past 12 months (*Up12* and *Down12*) – we can model staggering adjustment of ratings, both within and between the three rating agencies.

Agency interaction and competition To allow for interactions between rating agencies, we include the difference between the rating of the agency under consideration and the average of the two other agencies (if available) ($\Delta^{+}rating, \Delta^{-}rating$). This adds another level of error correction behavior by allowing agencies to converge to a common rating. Following the general spirit of our approach that strongly emphasizes potential nonlinearities and asymmetries, we allow for different reactions to positive and negative gaps to the competing agencies. Additionally, we consider the total number of rating upgrades and rating downgrades in a country by all other agencies within the past 12 months ($Up12^c$ and $Down12^c$) to capture rating momentum (Carty & Fons 1994). That is, we distinguish between the general tendency (or the lack thereof) to converge to a similar assessment, and rating adjustments following changes of other agencies. For both of those indicators, the economic interpretation is less straight forward. A deviation between an agency's ratings and others, and rating activity by other agencies might imply that relevant new information is available. This would trigger new ratings under a rational inattention hypothesis. However, it is similarly possible that strategic considerations make it unattractive to stand alone with a deviating rating. Erring as part of a group might overcompensate the potential benefit of being right when this comes at the risk of erring alone.

The above indicators don't capture the degree of competition, which may be important in itself (Becker & Milbourn 2011). In the initial years of our sample, many countries were only rated by one or two agencies (if they were rated at all). Contrarily, towards the end most countries are assessed by all of the big three. Exploiting the considerable cross-sectional variation at each point in time and over time, we include the total number of agencies assessing a country at any given time (N^c) to measure the degree of competition.

Cross-country spillovers We expect rating agencies to be more careful when they receive news that indicate the necessity of deeper investigation. To capture this, we use the shares of countries with rating upgrades and rating downgrades in other countries within the past 12 month (*UpAll12* and *DownAll12*), independent of the agency (Amato & Furfine 2004).[13] This captures (a) the possibility of cross-border contagion (Forbes & Rigobon 2001), (b) the possibility of raised awareness after having to re-rate several countries (if the general probability of change is affected by those indicators), and (c) the effect of foreign sovereign ratings on domestic capital markets (Gande & Parsley 2005).

4.2 Fundamental variables

Our paper includes a range of fundamentals, mostly suggested by the literature presented above. These fundamentals can be roughly divided into the subsets fiscal sustainability, macroeconomic environment and the institutional background.

Fiscal sustainability This group of variables encompasses variables of fiscal policy. That is, we consider both the debt to GDP ratio (*debt*), the fiscal balance (*fiscbal*) (relative to GDP) and real government bond yields (*yield*) as indicators of fiscal risk. While theoretically appealing to cover fiscal sustainability, the ratio of interest rate payments to revenues is only available for a very limited subset of countries. In particular, there are many gaps in the data, rendering this variable unsuitable for our dynamic approach. Finally, we include the growth rate of central bank reserves (*reserves*), to capture the possibility that the reserves of a country can counteract potential debt or crisis problems.

Macroeconomic environment This group of variables includes economic development and cyclical macroeconomic indicators. With respect to development, we use GNI per capita relative to the US (*gnipc*).[14] With respect to cyclical variables, our dataset covers industrial production growth (*growth*), inflation (*inflation*) and the real effective exchange rate growth (*reer*) as an indicator of competitiveness. These variables are more

[13] We use the share rather than the total number to account for changes in the cross-sectional dimension over time. Differentiating across agencies does not improve results.

[14] Normalizing local GNI by its US level is needed to overcome stationarity issues.

short-run in nature. However, large and persistent deviations from the level that is perceived as consistent with a given rating should still be taken into account by rating agencies.

Institutions The quality of institutions can and should have an effect on ratings, as they determine the degree of adaptability of a country and its government. To capture two dimensions of institutional quality, we include the measure of capital account openness by Chinn & Ito (2006), which has recently been updated until 2015 (*ka.open*), and the corruption perception index (*corrupt*).

In the cross sectional literature it has been established that the default history of a country has a major impact on ratings. In order to have information on the default history at the beginning of the sample, we use the Bank of Canada Database on Sovereign Defaults (Beers & Nadeau 2014), which starts in 1970. We use a dummy indicating whether a country has ever defaulted in the past (*default*).

In an alternative specification, we exploit the time dimension and set the dummy to one if a country had debt in default during the past 10 years (*recentdefault*). While our core results remain robust, we lose explanatory power compared to the baseline, indicating that rating agencies are surprisingly unforgiving when it comes to default.

4.3 Variables in robustness checks

In a series of robustness checks, we aim to capture (i) a potential bonus for rich countries, (ii) home bias and (cultural) proximity, (iii) political variables related to political stability and strategic behavior around elections, and (iv) information on rating outlooks. We find insignificant effects for the majority of these extensions. Only outlook changes have a significant effect, but the added information is largely orthogonal to the information in other variables, i.e. our main findings are not invalidated. Variable descriptions and estimation results are reported in the online appendix B.

4.4 Data treatment

The fundamental variables described in subsection 4.2 are often not available at a monthly frequency for a sufficiently large sample of observations. Instead, data from different sources are often only available for smaller subsamples at lower frequencies. Moreover, we have to account for stationarity issues and outliers. In order to deal with these issues, we perform the following transformations, see Table 3.

Table 3: Data treatment, fundamental variables

Variable	Sources	Comb rule	Measurement	Normalization	Orig freq	Lag
gnipc	WDI		% US GNI		Y	12
ip	NSO; NCB; IMF-IFS	a,c	yoy growth	sd	M/Q	1
reserves	NCB		yoy growth	win (99), sd	M	1
inf	IMF-IFS		yoy growth	win (99), sd	M	1
reer	IMF-IFS; BIS; NCB	b	yoy growth	win (1,99), sd	M	1
yield	TR; JPM	a,b	real rate	win (99), sd	M	1
debt	IMF-IFS; WDI	d	% GDP		Y	12
fiscbal	IMF-IFS; WDI	a	% GDP		Y	12
current	IMF-IFS; WDI	c	% GDP		Y	12
corrupt	TI		index		Y	12
ka_open	Chinn-Ito		index		Y	12

Note: Combination rules a-d are explained in the text. The data sources are: Bank for International Settlement (*BIS*); International Monetary Fund: International Financial Statistics (*IMF-IFS*); national statistical offices (*NSO*); national central banks (*NCB*); Transparency International (*TI*); Thomson Reuters (*TR*); World Bank: World Development Indicators (*WDI*)

First, we draw data from a variety of sources with different data coverage. To ensure the highest possible data availability at a similar data quality across countries, we use one of four combination rules, see the third column of Table 3. Under *rule (a)*, we use the country-specific source with the largest data availability. This rule applies to three cases. First, the fiscal balance, where we compare data from the IMF international financial statistics (IMF-IFS) and world development indicators (WDI). Second, we select monthly industrial production indices provided by national statistical offices (NSO) and national central banks (NCB).[15] Third, Thomson Reuters (TR) provides monthly nominal sovereign bond yields with different maturities. We apply the rule to the group of series with maturities between 5 and 10 years. Under *rule (b)*, we use a benchmark source

[15] Additionally, we remove seasonality using X13-ARIMA-Seats from the individual industrial production indices (and real effective exchange rates) before combination.

whenever possible, and turn to alternatives otherwise. This applies again to bond yields, where we use data on emerging market bond indices from JP Morgan when TR does not provide data. We also apply this rule to real effective exchange rates, where we prefer data by the IMF-IFS, but turn to statistics by the Bank for International Settlement (BIS) and NCBs if these data are not available. *Rules (c) and (d)* are used for data series where a benchmark source does not clearly dominate alternative sources in terms of coverage. For a prediction of the missing benchmark observations in y_{BM} based on available observations in an alternative source y_{alt}, we use coefficient estimates from a linear regression on the overlapping sample $y_{BM} = c + \beta y_{alt}$. We accept predictions only if the benchmark and alternative series have a correlation of at least 90%. In all cases, we adjust the constant c such that there is no structural break between predicted and officially reported series. The difference between rules (c) and (d) is the choice of benchmark series. Under rule (c), we use the same benchmark for all countries. For industrial production, our benchmark is the previously constructed monthly IP series, while alternatives are quarterly data from the IMF-IFS. For the current account balance, our benchmark are data from the IMF-IFS, which are extended using WDI-data. Under rule (d), the benchmark (and regression) is country-specific and chosen to be the longest available series. This rule applies to government debt, where we compare general government debt from the IMF-IFS to central government debt from the WDI.

The second set of transformations on our data relates to the necessity to make data comparable across countries. For GNI per capita, we remove the effect of long-run growth by expressing income per capita as a share of US income per capita in the same period. We make benchmark bond yields comparable to the EMBI by deflating the former using current CPI inflation. For industrial production, central bank reserves, consumer prices and real effective exchange rates, we calculate year-on-year growth rates. We winsorize these series, and yields, at the 99% level (REER additionally at the 1% level) to remove outliers. Moreover, to increase numeric stability, we standard-normalize winsorized series and industrial production growth. That is, all coefficients relate to a one standard-deviation increase.[16] For a similar reason, we express government debt, the fiscal balance and current account balances as share of GDP, such that coefficients describe an effect of a 100% increase of the explanatory on the latent variable. Similar to this treatment, we also divide lagged rating, and the rating differences to competitors (Δ^+ rating, Δ^- rating) by 24. This has the advantage that lagged rating and its square lie in the range $[0, 1]$.

The third set of transformations relates to the fact that many sources do not provide monthly data, but only quarterly or even annually, see Table A.1 in online appendix A. Yet, we want to avoid losing too many indicators that the cross-sectional literature has found to be important. Rather than dropping low frequency variables, we thus perform cubic interpolation of quarterly and yearly data to monthly frequency. Although not perfect, this approach can be justified with two arguments. Low-frequency fundamental data display a high persistence. The slow changes should affect the decisions of rating agencies, which claim to "rate through the cycle". Moreover, interpolation tries to mimic a continuous stream of news on fundamental developments. To avoid that the interpolation plays too big a role, we always lag the variables by one unit of their original frequency. This allows to identify Granger causality, which we interpret as causality in the face of low transition probabilities of ratings.

4.5 Pooling

Rating agencies do not necessarily rate countries in precisely the same way. In particular, when it comes to strategic interaction between agencies or non fundamental factors that affect ratings (and might be subject to more individual judgment), it does indeed seem unlikely that they do. Yet, estimating separate models for each agency, thereby losing efficiency by estimating separate coefficients where it is inappropriate, would go too far. We therefore run likelihood ratio tests to assess where pooling – and thus estimating a single coefficient for all agencies – is appropriate, and where it is not. As a rule of thumb, pooling rating-based variables is mostly rejected by the tests, with the exception of the

[16]The standard deviation of all data before standardization or normalization can be found in Table A.2 in online appendix A.

difference to competitors ratings $\Delta rating$ and the reaction to upgrades and downgrades by competitors, $Up12^c$ and $Down12^c$. Contrarily, pooling structural indicators is rarely rejected, with the exception being government bond yields and GNI per capita.[17] That is, rating agencies react more or less identically to those variables that they list as "official" determinants of ratings. On the other hand, they differ in how their rating process is driven by additional factors that do not (necessarily) reflect the pure credit risk of the rated sovereign.

5 Results

5.1 Is modeling inattention important?

Our baseline econometric model combines a probit model for the decision to reevaluate and an ordered probit model for the evaluation decision. Importantly, we estimate the ordered probit only based on those observations where rating agencies made an announcement. However, there are a couple of alternative models which could be employed to describe rating actions. Here, we present two sets of tests to show the superiority of our model over these alternatives. The key challenge is that our model differs quite fundamentally from the ordered probit models typically used in the literature. While we aim to explain the same economic phenomenon, the objective functions of the SIOP explains four different outcomes, while that of the ordered probit explains only three. This difference makes a direct comparison problematic. Additionally, the goodness of fit itself is of limited importance, as we are interested in establishing empirical evidence for an economic phenomenon, rather than being interested in prediction itself.

We therefore chose a two-step procedure. In the first step we compare a simple ordered probit model of rating changes (labeled *oprob* in Table 4) to the MIOP model (*MIOP*). The ordered probit model predicting rating changes uses the same likelihood and is nested in the MIOP, implicitly setting all slope coefficients of the probit equation to 0 and the threshold in the probit equation to negative infinity, thus implying a reevaluation

[17] More details on pooling tests can be found in Table A.3 in online appendix A. Variables employed only in robustness checks are always considered non-pooled.

Table 4: Model comparison

m1: reference	m2: alternative	#obs	$\#\text{coeffs}_{m1}$	$\#\text{coeffs}_{m2}$	LRstat	p(LR)	Hstat	p(H)
			Importance of inattention and announcement information					
MIOP	oprob	24703	99	51	259.74	0		
SIOP	MIOP	24703	99	99			-12.74	1.00
			Rating through the cycle and inattention					
MIOP trend	oprob trend	24703	99	51	228.64	0		
SIOP trend	MIOP trend	24703	99	99			9.21	1.00
SIOP	SIOP trend	24293	99	99	$0.09^{\dagger}$			
			Rating based variables and inattention					
SIOP	SIOP fundamentals	24703	99	29	1686.76	0	1472.25	0
SIOP	SIOP, no $rating^2$	24703	99	93	21.38	0	21.7	1
SIOP	SIOP, no dynamics	24703	99	75	181.86	0	140.5	0
SIOP	SIOP, no competitors	24703	99	85	188.19	0	216.23	0
SIOP	SIOP, no spillovers	24703	99	87	335.63	0	318.01	0

Notes: (i) The LR-test comparing the SIOP with the SIOP trend (marked with a †) relates to a Vuong-test. The critical value to reject the Null hypotheses of model equality at the 10% level is 1.66.
(ii) The Hausman test assumes coefficients of zero for all variables not contained in the nested model.
(iii) The model *SIOP, no* $rating^2$ drops $rating^2$; *SIOP, no dynamics* drops $rating$, $rating^2$, $Up12$ and $Down12$; *SIOP, no competitors* drops $\Delta^{+}rating$, $\Delta^{-}rating$, $Up12^c$, $Down12^c$ and N^c; *SIOP, no spillovers* drops $UpAll12$ and $DownAll12$.

probability of 100%. We can therefore apply a standard likelihood ratio test.

In the second step we compare the MIOP to our SIOP model (*SIOP*). While the models maximize different likelihoods and are thus not directly comparable through a likelihood ratio test, the two equations essentially model the same economic phenomenon using the same modeling framework. That is, the coefficients have identical interpretations in both models. The key difference is that we add the information on announcements without rating changes – making the SIOP model potentially more efficient – at the expense of the assumption that all evaluations eventually lead to announcements, which would render our model inconsistent if incorrect. This situation lends itself to a standard Hausman test.

We find that the MIOP significantly outperforms both ordered probit models in a likelihood ratio test. This implies that it is important for (in-sample) predictions of rating changes to model attention in an individual equation, see the first part of Table 4. The Hausman test does not reject the Null of identical coefficients in the MIOP and SIOP models, implying that our model is not inconsistent and – being the efficient choice – is thus preferable. Table 5 shows the average predicted probabilities of the four observed states for the SIOP (before the '/') and the MIOP model (after the '/').[18] The MIOP model creates a substantial bias in the probit equation. A comparison of the summary entries in the no-announcement column (prediction) and row (actual outcome) shows that the average predicted probability of no announcement is 83.2% in the MIOP against a share of 91.6% in the data. Correspondingly, the MIOP overestimates the share of announcements without rating change (14.6% instead of 6.4%). By using information on announcements, the SIOP does not incur this bias. The bias of the MIOP towards more announcement creates a slight advantage during upgrade and downgrade periods, where the overall predicted probability of a rating change is slightly higher than for the SIOP (7.4% vs. 5.1% during downgrade periods; 4.5% vs. 3.4% during upgrade

[18] The share of observations with rating changes (and therefore predicted probabilities for these events) is so small that an evaluation based on the most likely predicted event is not suitable. Instead, one would choose a lower signaling threshold for upgrade and downgrade predictions. This problem has been extensively discussed for early-warning models of financial crises (Kaminsky & Reinhart n.d., Alessi & Detken n.d., Sarlin & von Schweinitz n.d.).

periods). However, this is solely due to the higher probability of an announcement. The probabilities for rating changes conditional on an announcement are higher for the SIOP model than for the MIOP model.

Table 5: Predicted probabilities, SIOP vs MIOP model

Actual outcome		Predicted outcome				Frequency
probit		no ann	ann			
	oprob		-1	0	1	
no ann		92.8%/83.8%	0.8%/0.9%	5.4%/14.1%	0.9%/1.1%	22626 (91.6%)
	-1	84.7%/65.5%	5.1%/7.4%	9.7%/26.5%	0.5%/0.6%	243 (1.0%)
ann	0	75.6%/78.9%	2.0%/1.4%	20.8%/18.6%	1.6%/1.1%	1581 (6.4%)
	1	89.1%/71.0%	0.4%/0.5%	7.1%/24.0%	3.4%/4.5%	253 (1.0%)
Σ		91.6%/83.2%	0.9%/1.0%	6.5%/14.6%	1.0%/1.2%	24703 (100%)

Note: The table lists predicted outcomes for the four observed outcomes "no announcement" and "announcement with rating downgrade / no change / upgrade" for the SIOP model (first entry per cell) and the MIOP model (second entry). The last row contains sample means, the last column the sample frequency of outcome classes.

We want to make sure that the probit equation that we add to the SIOP model to predict rating activity does not merely compensate for the fact that we ignore rating through the cycle. We therefore look at a variations of the three models mentioned above where we use the trend component of the economic fundamentals rather than their level (*oprob trend, MIOP trend, SIOP trend*).[19] That is, the models do no longer include purely cyclical changes in fundamentals, which rating agencies should disregard if they truly rate through the cycle. Using the same approach that we use to establish the superiority of our SIOP opposed to a simple ordered probit model, we test if the SIOP with trend variables is preferable to the simple ordered probit, see the second part of Table 4. Again we find that the MIOP model significantly outperforms the simple ordered probit, and that the Hausman tests fails to reject the consistency of the SIOP. This strongly supports our original hypothesis that inattention goes beyond rating through the cycle (which could be captured by *oprob trend*). We then pitch the SIOP model with trend variables against the baseline SIOP. Due to the different explanatory variables, a classical likelihood ratio test is not feasible. Instead, we run the nondegenerate Vuong test proposed by Shi (2015). The test-statistic (reported as a LR-statistic in Table 4) shows that our baseline model

[19] We estimate country-specific trends for each variable using an HP-filter with the smoothing parameter adjusted to the original data frequency.

performs marginally and insignificantly better than model with trend data. This does not give us any evidence that supports disregarding the cyclical information. Moreover, the estimation of trends is not possible in countries where we do not have enough observations, resulting in a lower number of observations. Therefore, we use the baseline SIOP model with cyclical variables as benchmark model throughout the paper.

In further tests, we investigate if rating-based indicators add information to the model, or if economic fundamentals would be sufficient to explain rating dynamics in a SIOP model. In all cases, we compare the baseline SIOP model to nested smaller models using a LR-test and a Hausman test, as shown in the third part of Table 4. These tests confirm that the additional variables are highly relevant, as has previously been shown for corporate ratings (Mizen & Tsoukas 2012). First, we could disregard the information from the rating process and only focus on economic fundamentals (*SIOP fundamentals*). Such a model would be correct if (rational) inattention, strategic dimensions and rent seeking behavior did not play a role for rating decisions. We can strongly reject that this is indeed the case. The tests reported for the alternative model *SIOP fundamentals* very strongly indicate that such a model has a significantly lower explanatory power and yields inconsistent estimates. The main reason for this is that our baseline model does a much better job in predicting the outcome of announcements. The average probability for the case of an announcement that confirms the current rating level, $P(y^d = 1, \tilde{y} = 0)$, is 21% in the baseline model, while it is only 12% in the model based only on fundamentals. Similar differences apply for upgrade and downgrade announcements. We then use alternative specifications that include some but not all dimensions of rating-based variables. Specifically, we ask the question if we really need information on squared ratings (*SIOP, no rating*^{2}), rating dynamics (*SIOP, no dynamics*), competitor information (*SIOP, no competitors*) and information on cross-country rating spillovers (*SIOP, no spillover*). We find again that our baseline model increases explanatory power significantly. Moreover, the coefficient estimates of alternative models are only consistent in one case, namely if we only drop information on squared ratings. For the other alternative models, the Hausman test provides evidence of possible omitted variable bias.

5.2 The fundamentals behind ratings

For the most part, the coefficients on structural indicators have the expected direction (for the pooled variables see Table 6, for *gnipc* and *yield* see Table 7). In our discussion, we differentiate between the direct effects on reevaluation probabilities from the probit equation, the direct effect on reevaluation outcomes from the ordered probit equation, and the joint effect on rating changes. It is important to note, that (pooled) fundamental variables affect ratings mostly through the ordered probit part of the model, i.e., they influence agency behavior conditional on the decision to reassess a rating. We will see that this is different for rating variables, which have a much stronger influence on the reassessment decision.[20]

Fiscal sustainability It seems that the *fiscal balance* is much more important than the level of *debt*, which turns out to be nearly insignificant in both equations. This is in line with evidence from some highly indebted but stable developed economies that are consistently rated well. Ghosh, Kim, Mendoza, Ostry & Qureshi (2013) argue that ever higher primary balances are necessary to keep a country solvent under higher debt levels and interest payments. As the debt evolves slowly, it is mostly fiscal (and primary) balances that define the sustainability of sovereign debt. This also explains the asymmetric effect of fiscal balances. With a deteriorating fiscal balance, the probability for downgrades - conditional on being reevaluated - is increasing, while at the same time the probability for a reassessment is rising. In other words, a positive fiscal balance leads to a slowly improving rating, while a negative fiscal balance can quickly deteriorate the rating. Higher *central bank reserves* may act as a backstop against potential currency crises and thus decrease downgrade probabilities (Kaminsky & Reinhart n.d.). For *yields*, we find very different effects for Moody's and S&P on the one hand, and Fitch on the other, see Table 7. The first two agencies do not significantly react to past yields. Fitch, however, seems to put some weight on financial market behavior. We see that higher

[20] Following this argument and simplifying interpretation, we could split variables across equation – estimating the probit model only on non-fundamental variables and the ordered probit model only on economic fundamentals. However, such a split is strongly rejected by a likelihood ratio test.

yields significantly decrease announcement and upgrade probabilities. There are two possible explanations for this. First and related to the negative coefficient in the ordered probit equation, higher yields are in general a sign of lower fiscal sustainability, which leads to higher downgrade (and lower upgrade) probabilities. Second and with respect to the probit equation, periods with high yields may occur due to market volatility. In this case, the information content of yields with respect to fiscal sustainability is reduced, which should decrease announcement probabilities in general. It seems as if Fitch – as an agency with comparably many announcements – still wants to avoid wrong assessments that are driven by market volatility rather than fundamentals.

Macroeconomic environment The pooled variables related to the general macroeconomic environment do not significantly contribute to the decision to reassess a rating, see Table 6. However, they mostly play a strong role in the actual rating decision. We find that high *industrial production growth* and a positive *current account* all have beneficial effects in terms of higher upgrade probabilities. The same holds for an appreciating *real effective exchange rate*, which indicates economic strength seems to dominate possible future consequences on exports. Surprisingly, *inflation* has no impact on the rating when controlling for the other factors we include. The reason for that might be that our sample mostly coincides with the so called Great Moderation and the Global Financial Crisis where inflation was low in most countries with some even facing deflationary pressure.

GNI per capita enters non-pooled in the two equations. We find a mean-reverting tendency, whereby richer countries receive higher downgrade probabilities and the other way around. All agencies (Moody's significantly so) have on average a higher downgrade probability of richer countries in case of a reassessment. For Fitch, richer countries are additionally rated more often. However, it has to be kept in mind, that we control for a whole range of indicators. One might read that result as an increase in risk if GNI is too high compared to other factors such as sound fiscal policy, good institutions and the like.

Institutions High *corruption* (as indicated by a low corruption perception index) usually drives the rating down. This matches results from the previous literature. *Capital*

Table 6: Impact of pooled indicators on ratings

	Prob	OProb
Time and inattention		
$changefund$	0.07***	-0.058
	(0.017)	(0.046)
Competition		
$\Delta^{+}rating$	0.769*	4.083***
	(0.417)	(1.063)
$\Delta^{-}rating$	-0.37	4.823***
	(0.456)	(1.051)
$Up12^{c}$	0.031	0.4***
	(0.028)	(0.064)
$Down12^{c}$	0.117***	-0.202***
	(0.024)	(0.051)
Fiscal sustainability		
$debt$	0.076*	-0.068
	(0.045)	(0.114)
$fiscbal$	-1.896***	2.116*
	(0.420)	(1.088)
$reserves$	0.014	0.117***
	(0.017)	(0.041)
Macroeconomic environment		
$growth$	-0.02	0.19***
	(0.017)	(0.043)
$inflation$	0.048	-0.036
	(0.032)	(0.065)
$reer$	-0.015	0.176***
	(0.015)	(0.037)
$current$	0.236	5.177***
	(0.287)	(0.734)
Institutions		
$corrupt$	-0.003*	0.022***
	(0.001)	(0.004)
$ka.open$	0.165***	-0.019
	(0.058)	(0.133)
$default$	-0.065*	-0.293***
	(0.039)	(0.096)

Note: For the results on pooled coefficients see table 7.

account openness merely affects assessment probability. This might come from the fact that open capital markets – even when beneficial on average – make a country more subject to international fluctuations (Rancière, Tornell & Westermann 2008), thus requiring more monitoring from the agencies' side. Past defaults have a significantly negative and strong effect on rating changes conditional on reevaluation. While it has often been mentioned that countries can return to the capital markets quickly after a default (see e.g. the survey article by Panizza, Sturzenegger & Zettelmeyer 2009), it seems indeed as if something always "sticks", stigmatizing defaulting countries over extremely long periods.

5.3 Mean reversion, agency interaction and timing

Mean reversion Both Fitch and S&P show a considerable degree of mean reversion of ratings. Based on the estimated coefficients, we can look at monthly upgrade and downgrade probabilities conditional on all other variables being at their sample median. We find that downgrade probabilities peak at top ratings, while upgrade probabilities peak close to default ratings. That is, for both Fitch and S&P recovery from default happens fairly quickly. In the case of Moody's, announcement coefficients on *rating* and *ratings*^{2} are highly significant (and larger than for S&P or Fitch), but the effects largely balance each other. Announcement probabilities do not vary strongly over the range of different ratings (compared to the two other agencies). Similarly total upgrade and downgrade probabilities are low (usually below 1 percentage point) and move only moderately with the rating itself. This mirrors results for corporate ratings, which emphasize a strong state dependence of ratings (Mizen & Tsoukas 2012).

There is some indication on self-reinforcing behavior (as in the literature on corporate ratings Carty & Fons 1994), but with differences across agencies: Moody's shows a tendency for negative feedback loops: An initial downgrade by Moody's can lead to further downgrades (as indicated by the negative coefficient of *Down12* in the ordered probit equation), while an initial upgrade reduces the probability to reassess and does not affect actual rating decisions, see the negative coefficient of *Up12* in the probit equation. Another interpretation would be that Moody's staggers rating downgrades, but does not do the same for upgrades. S&P also shows a tendency for feedback loops. However, they only have positive feedback loops, i.e., they stagger rating upgrades. Opposed to these two, Fitch seems to dampen such loops, as past upgrades (conditional on reassessment) reduce the probability of further upgrades, while past downgrades the reduce the probability of all reassessments.

Cross-country spillovers Cross-country spillovers counter the feedback loops on own upgrades and downgrades we found above. In general, announcement probabilities increase substantially after rating changes in other countries ($UpAll12$ and $DownAll12$). With respect to the coefficients for reassessment outcomes, Moody's and S&P (which had feedback loops) dampen cross-country spillovers. Waves of downgrades/upgrades in other countries ceteris paribus lead to upward/downward rating pressure. This guarantees that a global recession does not trigger excessive downgrades in countries that are merely caught in the global business cycle. Fitch (which showed dampened behavior with respect to feedback loops in the same country) reacts to waves of downgrades/upgrades in the same direction. That is, rating changes in the own country and rating changes abroad may together result either in feedback loops (both positive and negative) or not. Which of the two dominates, depends on the share of countries that receive a rating change. This, in turn, should depend on the importance of international shocks in the past 12 months. In a broader sense, this is in line with the findings of Mora (2006), El-Shagi & von Schweinitz (2015) and El-Shagi (2010), who find evidence against vicious cycles in ratings.

Time and inattention We have two variables in this group, where *years* captures the time since the last rating announcement, and *changefund* the (normalized) change of economic fundamentals since then. Consistent with the idea of inattention, we find that large changes significantly increase the probability of a new assessment, while they are not informative for the actual rating decision. However, we do not find the same channel at work with respect to the time since the last announcement. On the opposite, S&P and Fitch significantly reduce reassessment probabilities as time goes by. This may be related to the fact that these two agencies make more use of outlooks than Moody's. Announcements with a confirmation of the current rating and an outlook to future assessments make up 70% of all announcements for Fitch, while Moody's has only 46% of these announcements (S&P is in the middle). The combination of $year$ and $changefund$ gives us the previously described U-shape of reassessment probabilities, with the degree of curvature depending on the speed of fundamental changes. This provides some evidence for (rational) inattention on the side of rating agencies, because both the time since the last announcement and the change of fundamentals since then can be measured at practically zero cost. In terms of rating decisions, we see no significant coefficient for $changefund$ in the ordered probit equation. However, Moody's and Fitch seem to lean towards rating upgrades the longer a rating has not been reevaluated, confirming previous evidence on rating ageing from the corporate rating literature (Mizen & Tsoukas 2012).

Agency interaction Interestingly, the agencies do not have a strongly significant increase in the reassessment probability when they deviate from the consensus in either direction (see coefficients for $\Delta^{+}rating$ and $\Delta^{-}rating$ in Table 6). However, conditional on reassessing (i.e. based on the ordered probit part of our model), they typically converge towards their competitors. There also is a tendency to follow up on upgrade and downgrades by other agencies (see $Up12^c$ and $Down12^c$ in Table 6). In theory, the effect could contribute to a snowball effect, where a downgrade in one agency triggers others to follow which does in turn increase the downgrade probability for the first agency. However, in practice it is too small to persist over a series of discrete rating steps, because upgrades by competitors only affect the outcome of reevaluations, while downgrades have a comparably small effect on reassessment outcomes.

While both of those results indicate some response to competing agencies, the "pure" effect of competition is highly heterogeneous across agencies, see coefficients for N^c in Table 7. As argued above, competition would suggest more activity and (in case of rating shopping) higher ratings. S&P and Fitch rate countries more often if they are also assessed by other agencies, yet there is no impact on the direction of rating changes (and thus no impact on the average rating). Contrary to its competitors, Moody's shows a lower

Table 7: Impact of non pooled indicators on ratings

	Moody		S&P		Fitch	
	Prob	OProb	Prob	OProb	Prob	OProb
Mean reversion						
$rating$	2.161***	0.369	0.215	-4.736	0.937	-3.774***
	(0.774)	(2.079)	(1.141)	(3.168)	(0.655)	(1.430)
$rating^2$	-2.115***	-0.931	-0.785	0.852	-1.516***	0.524
	(0.617)	(1.680)	(0.876)	(2.411)	(0.501)	(1.135)
$Up12$	-0.315***	-0.238	-0.097	0.559**	-0.081	-0.217*
	(0.077)	(0.209)	(0.098)	(0.245)	(0.053)	(0.126)
$Down12$	-0.019	-0.26**	-0.046	-0.068	-0.134***	-0.017
	(0.050)	(0.120)	(0.068)	(0.152)	(0.047)	(0.104)
Cross country spillovers						
$UpAll12$	10.247***	-10.064***	6.632***	-47.476***	15.354***	29.52***
	(3.674)	(0.070)	(0.125)	(0.113)	(3.001)	(0.066)
$DownAll12$	8.046***	8.036***	24.918***	14.278***	36.539***	-4.105***
	(2.702)	(0.035)	(4.420)	(0.121)	(2.298)	(0.019)
Time and inattention						
$years$	-0.009	0.127***	-0.027***	0.058	-0.129***	0.135**
	(0.006)	(0.035)	(0.010)	(0.036)	(0.016)	(0.057)
Agency interaction / competition						
N^c	-0.237***	-0.335***	0.406*	-0.097	0.194***	-0.006
	(0.058)	(0.119)	(0.246)	(0.684)	(0.051)	(0.116)
Fiscal sustainability and macroeconomic environment						
$yield$	-0.015	0.065	0.128	-0.375	-0.256***	-0.224*
	(0.063)	(0.139)	(0.103)	(0.298)	(0.054)	(0.125)
$gnipc$	-0.046	-1.129***	0.175	-0.121	0.436***	-0.43
	(0.141)	(0.400)	(0.184)	(0.536)	(0.101)	(0.315)
Constants						
$Intercept$	-1.98***		-3.096***		-2.261***	
	(0.285)		(0.579)		(0.269)	
$Thresh-1.0$		-1.587**		-3.656**		-3.172***
		(0.681)		(1.633)		(0.579)
$Thresh0.1$		0.603		-1.893		0.301
		(0.682)		(1.626)		(0.570)

Note: For the results on pooled coefficients see Table 6.

propensity to reassess and a significant negative response in the ordered probit equation, i.e. typically rates countries less generous if they are also assessed by other agencies. This indicates that the usual concerns, that competition increases ratings because it enables rating shopping can not be applied to sovereign ratings. In this case, the rating agencies seem to be more concerned about their reputation, rather than trying to incentivize governments to solicit a rating from them.

5.4 Robustness checks

To keep our baseline model as parsimonious as possible – given the complexity inherent to its nature – we address a range of other potential indicators from the earlier literature only in robustness checks. In particular, we check the following models: A model where the default history indicator is replaced by an indicator that is reset to zero ten years after the default. Then two models including an OECD dummy and an EU dummy, respectively. Two different models controlling for potential home bias of rating agencies, a model that controls for political variables like parliament majorities and election dates. Last, we control for outlooks by adding them either as separate control variables or adding them to rating levels. The specification tests are in all cases qualitatively similar to the ones for the baseline model, see Table 8: Modeling the separation between the decision to reassess and the actual rating decision in our two-equation model is important, as a single-equation ordered probit model leads to inconsistent estimates in all robustness checks. Moreover, the inclusion of rating-based variables is important to improve the efficiency and consistency of estimates.

Regarding the coefficient estimates, the coefficients on variables also included in the baseline model are in their majority robust. Coefficients on additional variables, however, are often inconclusive as all these variables enter non-pooled. We report detailed results in online appendix B.

Table 8: Specification tests for robustness checks

Baseline \ Alternative	oprob		fundamentals	
	p(LR-test)	p(Hausman)	p(LR-test)	p(Hausman)
recentdefault	1.000	$<$0.001	$<$0.001	$<$0.001
OECD	1.000	$<$0.001	$<$0.001	$<$0.001
EU	1.000	$<$0.001	$<$0.001	$<$0.001
home bias (small)	1.000	$<$0.001	$<$0.001	$<$0.001
home bias (large)	1.000	$<$0.001	$<$0.001	$<$0.001
Political variables	1.000	$<$0.001	$<$0.001	$<$0.001
Outlooks	1.000	$<$0.001	$<$0.001	$<$0.001
Outlooks in rating	$<$0.001	$<$0.001	$<$0.001	$<$0.001

5.5 Significance vs. Economic Impact

In our model, it is hard to gauge the economic significance of results from the estimated coefficients alone. Therefore, table A.4 in online appendix A presents the marginal effects in the baseline model. We report five different marginal effects, one for each probability implied by our model: announcement probability, upgrade/downgrade probability conditional on announcements and aggregate probabilities of rating changes. We evaluate marginal effects at median values of each variable, also reported in the table.

When looking at the marginal effect on aggregate up- and downgrade probabilities, at first glance it seems as if most of our results were rather small in magnitude. This is partly due to the monthly frequency of our data: Even when there are reasons to reevaluate a country, the probability that this happens in a specific month is rather low. This is augmented by the fact that we evaluate at the mean, i.e. in a situation where the general propensity to even consider reevaluation is low. The marginal effects increase by an order of magnitude, when looking at the directional equation only. That is, when considering a situation where rating reevaluations are generally likely (for example because fundamental variable changes have accumulated over time), a change in any specific variable is substantially more likely to actually trigger a rating change. Yet, even here, the marginal effects are typically clearly below one percentage point.

The exception to the small marginal effects come from rating-based variables, giving additional importance to the inclusion of these variables in our model. The effect of rating changes both by the same agency and other agencies of the same country within the past 12 months can be quite sizable. While the effect is clearly below the percentage mark at the mean, the probabilities in the ordered probit equations alone (i.e. conditional on other circumstances warranting a reevaluation) are typically several percentage points large. Even larger is the marginal effect of rating changes in other countries (*UpAll* / *DownAll*). However, scale has to be kept in mind when interpreting the marginal effect. A change of one unit in *UpAll* and *DownAll* would imply an increase from no rating changes (in other countries) at all to every single other country having its rating adjust in every single month over a year. When looking at a change of 0.01, i.e. rating changes happening with a probability of 1% more often than usual in the rest of the world, the reported marginal effects have to be multiplied by 0.01. However, the effects still have a meaningful order of magnitude. For example, a 1% increase in *DownAll* increases the aggregate downgrade probability by about 0.015% (Moody's) to 0.25% (S&P). Given that *UpAll* and *DownAll* are persistent by construction, it is not unlikely for a country to be caught in a reevaluation wave.

6 Conclusion

There is strong evidence that considering the reassessment decisions is highly important to understand how ratings contribute to macroeconomic dynamics. Ratings can be predicted substantially better when considering reassessment probability separately, as proposed in this paper. Most importantly, we find strong support for our hypothesis that deviations between the "appropriate" rating and the observed rating can frequently be explained by the impossibility or unwillingness of permanently monitoring a country. A better modeling of the rating process can help understanding why rating agencies seem to follow the market at some times, shocking it at others.

We find that there is considerable heterogeneity between agencies with respect to the importance of all variables that are *not* part of the official hard determinants of sovereign credit risk. However, regarding the fundamental determinants listed in official statements, we find evidence that their influence on the rating decision process is homogeneous across the Big Three agencies. Moreover, the direction of rating changes seems to be largely driven by said fundamentals, while the decision to reassess is mostly driven by other variables. In particular, we find evidence that inattention and a low degree of competition reduce the probability of an assessment.

References

Afonso, A. (2003). Understanding the Determinants of Sovereign Debt Ratings: Evidence for the two Leading Agencies, *Journal of Economics and Finance* **27**(1): 56–74.

Alessi, L. & Detken, C. (n.d.). Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity, **27**(3): 520–533.

Alsakka, R. & ap Gwilym, O. (2009). Heterogeneity of Sovereign Rating Migrations in Emerging Countries, *Emerging Markets Review* **10**(2): 151–165.

Altdörfer, M., De las Salas, C. A., Guettler, A. & Löffler, G. (2016). European versus Anglo-Saxon credit view: Evidence from the eurozone sovereign debt crisis, *IWH Discussionpaper 34/2016*.

Amato, J. D. & Furfine, C. H. (2004). Are credit ratings procyclical?, **28**(11): 2641–2677.

Amstad, M. & Packer, F. (2015). Sovereign Ratings of Advanced and Emerging Economies after the Crisis, *BIS Quarterly Review December* .

Bagozzi, B. E. & Mukherjee, B. (2012). A Mixture Model for Middle Category Inflation in Ordered Survey Responses, *Political Analysis* **20**(3): 369–386.

Becker, B. & Milbourn, T. (2011). How Did Increased Competition Affect Credit Ratings?, *Journal of Financial Economics* **101**(3): 493–514.

Beers, D. & Cavanaugh, M. (2008). Sovereign Credit Ratings: a Primer, *Technical report*, Standard & Poors.

Beers, D., Cavanaugh, M. & Takahira, O. (2002). Sovereign Credit Ratings: a Primer, *Technical report*, Standard & Poors.

Beers, D. T. & Nadeau, J.-S. (2014). Introducing a New Database of Sovereign Defaults, *Bank of Canada Technical Report 101*.

Block, S. A. & Vaaler, P. M. (2004). The Price of Democracy: Sovereign Risk Ratings, Bond Spreads and Political Business Cycles in Developing Countries, *Journal of International Money and Finance* **23**(6): 917–946.

Brewer, T. L. & Rivoli, P. (1990). Politics and Perceived Country Creditworthiness in International Banking, *Journal of Money, Credit and Banking* **22**(3): 357–369.

Brooks, R., Harris, M., Spencer, C. et al. (2007). An Inflated Ordered Probit Model of Monetary Policy: Evidence From MPC Voting Data, *MPRA Discussion Paper 8509*.

Butler, A. W. & Fauver, L. (2006). Institutional Environment and Sovereign Credit Ratings, *Financial Management* **35**(3): 53–79.

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

Carty, L. V. & Fons, J. S. (1994). Measuring changes in corporate credit quality, *The Journal of Fixed Income* **4**(1): 27–41.

Chinn, M. D. & Ito, H. (2006). What Matters for Financial Development? Capital Controls, Institutions, and Interactions, *Journal of Development Economics* **81**(1): 163–192.

Cornaggia, J. & Cornaggia, K. J. (2013). Estimating the costs of issuer-paid credit ratings, *The Review of Financial Studies* **26**(9): 2229–2269.

Cosset, J.-C. & Roy, J. (1991). The Determinants of Country Risk Ratings, *Journal of International Business Studies* **22**(1): 135–142.

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*.

Depken, C., LaFountain, C., Butters, R. et al. (2007). Corruption and Creditworthiness: Evidence from Sovereign Credit Ratings.

Dimitrakopoulos, S. & Kolossiatis, M. (2015). State Dependence and Stickiness of Sovereign Credit Ratings: Evidence from a Panel of Countries, *Journal of Applied Econometrics* .

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. (2016). Much Ado About Nothing: Sovereign Bond Ratings in the OECD, *IWH Discussion Paper 22/2016*.

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

Feder, G. & Uy, L. V. (1985). The Determinants of International Creditworthiness and their Policy Implications, *Journal of Policy Modeling* **7**(1): 133–156.

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

Fitch (2015). The Ratings Process: The Ratings Process Under Fitch Criteria, *Special Report*.

Forbes, K. & Rigobon, R. (2001). Measuring Contagion: Conceptual and Empirical Issues, *International Financial Contagion*, Springer, pp. 43–66.

Fuchs, A. & Gehring, K. (2017). The Home Bias in Sovereign Ratings, *Journal of the European Economic Association* **15**(6): 1386–1423.

Gande, A. & Parsley, D. C. (2005). News Spillovers in the Sovereign Debt Market, *Journal of Financial Economics* **75**(3): 691–734.

Ghosh, A. R., Kim, J. I., Mendoza, E. G., Ostry, J. D. & Qureshi, M. S. (2013). Fiscal fatigue, fiscal space and debt sustainability in advanced economies, *The Economic Journal* **123**(566): F4–F30.

Goh, J. C. & Ederington, L. H. (1993). Is a Bond Rating Downgrade Bad News, Good News, or No News for Stockholders?, *The Journal of Finance* **48**(5): 2001–2008.

Griffin, J. M., Nickerson, J. & Tang, D. Y. (2013). Rating Shopping or Catering? An Examination of the Response to Competitive Pressure for CDO Credit Ratings, *Review of Financial Studies* **26**(9): 2270–2310.

Gültekin-Karakaş, D., Hisarcıklılar, M. & Öztürk, H. (2011). Sovereign Risk Ratings: Biased Toward Developed Countries?, *Emerging Markets Finance and Trade* **47**(sup2): 69–87.

Hantzsche, A. (2017). An Estimation Strategy for Analysing the Determinants of Sovereign Credit Ratings. mimeo.

Haque, N. U., Kumar, M. S., Mark, N. & Mathieson, D. J. (1996). The Economic Content of Indicators of Developing Country Creditworthiness, *Staff Papers* **43**(4): 688–724.

Haque, N. U., Mark, N. C. & Mathieson, D. J. (1998). The Relative Importance of Political and Economic Variables in Creditworthiness Ratings, *Technical report*.

Harris, M. N. & Zhao, X. (2007). A Zero-Inflated Ordered Probit Model, With an Application to Modelling Tobacco Consumption, *Journal of Econometrics* **141**(2): 1073–1099.

Hu, Y.-T., Kiesel, R. & Perraudin, W. (2002). The Estimation of Transition Matrices for Sovereign Credit Ratings, *Journal of Banking & Finance* **26**(7): 1383–1406.

Kaminsky, G. L. & Reinhart, C. M. (n.d.). The Twin Crises: the Causes of Banking and Balance-of-Payments Problems, **89**(3): 473–500.

Kim, S.-j. & Wu, E. (2011). *Sovereign Debt: From Safety to Default*, Wiley, chapter Can sovereign credit ratings promote financial sector development and capital inflows to emerging markets.

Koopman, S. J. & Lucas, A. (2005). Business and default cycles for credit risk, *Journal of Applied Econometrics* **20**(2): 311–323.

Kou, J. & Varotto, S. (2008). Timeliness of spread implied ratings, *European Financial Management* **14**(3): 503–527.

Lando, D. & Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations, *Journal of banking & finance* **26**(2-3): 423–444.

Lee, S. H. (1993). Are the Credit Ratings Assigned by Bankers Based on the Willingness of LDC Borrowers to Repay?, *Journal of Development Economics* **40**(2): 349–359.

Maćkowiak, B. & Wiederholt, M. (2009). Optimal Sticky Prices under Rational Inattention, *American Economic Review* **99**(3): 769–803.

Mizen, P. & Tsoukas, S. (2012). Forecasting us bond default ratings allowing for previous and initial state dependence in an ordered probit model, *International Journal of Forecasting* **28**(1): 273–287.

Mora, N. (2006). Sovereign Credit Ratings: Guilty Beyond Reasonable Doubt?, *Journal of Banking & Finance* **30**(7): 2041–2062.

Mulder, C. B. & Monfort, B. (2000). Using Credit Ratings for Capital Requirements on Lending to Emerging Market Economies: Possible Impact of a New Basel Accord, *IMF Working Papers 69.*

Panizza, U., Sturzenegger, F. & Zettelmeyer, J. (2009). The Economics and Law of Sovereign Debt and Default, *Journal of Economic Literature* **47**(3): 651–698.

Rancière, R., Tornell, A. & Westermann, F. (2008). Systemic Crises and Growth, *Quarterly Journal of Economics* **123**(1): 359–406.

Sangiorgi, F., Sokobin, J. & Spatt, C. (2009). Credit-rating Shopping, Selection and the Equilibrium Structure of Ratings, *Working paper, carnegie mellon university.*

Sangiorgi, F. & Spatt, C. (2016). Opacity, Credit Rating Shopping, and Bias, *Management Science* **63**(12): 4016 – 4036.

Sarlin, P. & von Schweinitz, G. (n.d.). Optimizing policymakers' loss functions in crisis prediction: Before, within or after?, **25**(1): 100–123.

Schumacher, I. (2014). On the Self-fulfilling Prophecy of Changes in Sovereign Ratings, *Economic Modelling* **38**: 351–356.

Shi, X. (2015). A nondegenerate vuong test, **6**(1): 85–121.

Sims, C. A. (1998). Stickiness, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 49, Elsevier, pp. 317–356.

Sims, C. A. (2003). Implications of Rational Inattention, *Journal of Monetary Economics* **50**(3): 665–690.

Skreta, V. & Veldkamp, L. (2009). Ratings Shopping and Asset Complexity: A Theory of Ratings Inflation, *Journal of Monetary Economics* **56**(5): 678–695.

Vaaler, P. M., Schrage, B. N. & Block, S. A. (2006). Elections, Opportunism, Partisanship and Sovereign Ratings in Developing Countries, *Review of Development Economics* **10**(1): 154–170.

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

Appendix A: Data Description

Table A.1: Description of variables

Variable	Full name	Group	Source[a]	Freq.	Transform.	Comb. Rule	Normalization[b]
y^d	Rating announcement (Dummy)	Explained variable	CE	Daily	Sum		
$\tilde{y}$	Rating change (Ternary)	Explained variable	CE	Daily	Sum		
rating	Rating level	Ratings	CE	Daily	first day/mon		norm (24)
Δ^+rating	Pos. diff to other ratings	Ratings	CE	Monthly			norm(24)
Δ^-rating	Neg. diff to other ratings	Ratings	CE	Monthly			norm(24)
outlook$^+$	pos outlook (Dummy)	Ratings	CE	Daily	first day/mon		
outlook$^-$	neg outlook (Dummy)	Ratings	CE	Daily	first day/mon		
outlook$^{+,c}$	pos outlook by others (Dummy)	Ratings	CE	Daily	first day/mon		
outlook$^{-,c}$	neg outlook by others (Dummy)	Ratings	CE	Daily	first day/mon		
Up12	# Up in prev. year	Ratings	CE	Monthly			
Down12	# Down in prev. year	Ratings	CE	Monthly			
Up12^c	# Up in prev. year by others	Ratings	CE	Monthly			
Down12^c	# Down in prev. year by others	Ratings	CE	Monthly			
UpAll12	% foreign Up in prev. year	Ratings	CE	Monthly			
DownAll12	% foreign Down in prev. year	Ratings	CE	Monthly			
years	Years since last announcement	Ratings	CE	Monthly			norm (12)
N^c	# competitors	Ratings	CE	Monthly			

[a]Sources: Bank of Canada Database on Sovereign Defaults (*DSD*); Bank for International Settlement (*BIS*); countryeconomy.com (*CE*); Interamerican Development Bank: Database of Political Institutions 2015 (*DPI*); (Fuchs & Gehring 2017) (FG); International Monetary Fund: International Financial Statistics (*IMF-IFS*); national statistical offices (*NSO*); national central banks (*NCB*); Transparency International (*TI*); Thomson Reuters (*TR*); World Bank: World Development Indicators (*WDI*)

[b]Normalization Rules: *win* indicates windsorizing at the 99% (and 1%) level; *demean* indicates demeaned series; *sd* standardizes series to have mean zero and standard deviation one; *norm* indicates normalization to the range $[0,1]$ or by the given factor.

Table A.1: Description of variables, continued

Variable	Full name	Group	Source	Freq.	Transform.	Comb. Rule[a]	Normalization
gnipc	Gross national income per capita	Fundamentals	WDI	Yearly	rel. to US GNI		
growth	Growth of industrial production	Fundamentals	NSO	Monthly	growth (%yoy)	*LS*, first on monthly, then incl. quarterly	sd
			NCB	Monthly			
			NSO	Quarterly			
			IMF-IFS	Quarterly			
reserves	Central bank reserves	Fundamentals	NCB	Monthly	growth (%yoy)		win (top), sd
inf	Yearly change of consumer prices	Fundamentals	IMF-IFS	Monthly			win (top), sd
reer[b]	Real effective exchange rates	Fundamentals	IMF-IFS	Monthly	growth (%yoy)	*OC*	win (both), sd
			BIS	Monthly			
			NCB	Monthly			
yield	Benchmark bond yields (5-10 years)	Fundamentals	TR	Monthly	defl. by curr. inf.	*LS* of benchmark bonds, *OC* with EMBI	win (top), sd
	Emerging Markets Bond Index	Fundamentals	JPM	Monthly			
debt	General government debt	Fundamentals	IMF-IFS	Yearly		Country-specific *REG* on IMF-IFS series	norm (100)
	Central government debt		WDI	Yearly			
fiscbal	Fiscal balance	Fundamentals	IMF-IFS	Yearly		*LS*	norm (100)
	Fiscal balance		WDI	Yearly			
current	Current account balance	Fundamentals	IMF-IFS	Yearly		Full *REG* on IMF-IFS series	norm (100)
			WDI	Yearly			
corrupt	Corruption perception index	Fundamentals	TI	Yearly			
ka.open	Capital account openness	Fundamentals	Chinn-Ito	Yearly			
default	Dummy: default since 1970	Ratings	DSD	Yearly			
recentdefault	Dummy: default in last 10 years	Ratings	DSD	Yearly			

[a]Combination Rules: *LS* uses the longest available series in every country; *OC* uses series, if available, in the order given in the table; *REG* performs a regression of all series on a benchmark (country-specific or on the full sample), and employs the estimated coefficients for combination; the regression constant is adjusted in order to remove breakpoints.

[b]Series inverted for Chile, Costa Rica, Croatia, Hungary, Mexico, Mongolia, Sri Lanka , Sweden, Zambia

Table A.1: Description of variables, continued (robustness checks)

Variable	Full name	Group	Source	Freq.	Transform.	Comb. Rule	Normalization
dumoecd	OECD membership dummy	Institutional		Monthly			
dumeu	EU membership dummy	Institutional		Monthly			
yrsoffc	Exec. years in office	Institutional	DPI	Monthly			
maj	Majority of govt in parliament	Institutional	DPI	Monthly			
exelecpre	# exec. elec. in prev. year	Institutional	DPI	Monthly			
exelecpost	# exec. elec. in next year	Institutional	DPI	Monthly			
legelecpre	# leg. elec. in prev. year	Institutional	DPI	Monthly			
legelecpost	# leg. elec. in next year	Institutional	DPI	Monthly			
expshare	Export share of US	Home Bias	FG	Monthly			
inline	UN Voting overlap with US	Home Bias	FG	Monthly			
usmilaidshare	% US military aid	Home Bias	FG	Monthly			
comlang	Common language dummy	Home Bias	FG	Monthly			
delf.language	language distance to US	Home Bias	FG	Monthly			
delf.ethnic	ethnic distance to US	Home Bias	FG	Monthly			
bankexp	cross border bank exposure	Home Bias	FG	Monthly			

Table A.2: Summary statistics

	Moody						S&P						Fitch					
	min	median	mean	sd	max	N	min	median	mean	sd	max	N	min	median	mean	sd	max	N
Announcement y^d	0	0	0.048	0.214	1	9475	0	0	0.036	0.185	1	5888	0	0	0.151	0.358	1	9340
Rating change $\tilde{y}$	-1	0	0.001	0.141	1	9475	-1	0	-0.003	0.130	1	5888	-1	0	0.002	0.149	1	9340
rating	2	20	18.91	4.89	24	9475	3	21	19.51	4.82	24	5888	2	20	19.04	4.58	24	9340
Δ^+rating	0	0	0.3	0.94	13	9475	0	0	0.2	0.43	4.5	5888	0	0	0.32	0.8	8	9340
Δ^-rating	-7	0	-0.29	0.63	0	9475	-8	0	-0.33	0.91	0	5888	-7	0	-0.25	0.63	0	9340
outlook$^+$	0	0	0.1	0.3	1	9475	0	0	0.05	0.21	1	5888	0	0	0.08	0.27	1	9340
outlook$^-$	0	0	0.11	0.32	1	9475	0	0	0.15	0.35	1	5888	0	0	0.15	0.36	1	9340
outlook$^{+,c}$	0	0	0.1	0.3	1	9475	0	0	0.1	0.3	1	5888	0	0	0.11	0.32	1	9340
outlook$^{-,c}$	0	0	0.18	0.39	1	9475	0	0	0.21	0.41	1	5888	0	0	0.14	0.35	1	9340
Up12	0	0	0.13	0.35	2	9475	0	0	0.07	0.29	3	5888	0	0	0.14	0.37	3	9340
Down12	0	0	0.11	0.44	4	9475	0	0	0.12	0.42	4	5888	0	0	0.12	0.43	4	9340
Up12^c	0	0	0.18	0.46	4	9475	0	0	0.18	0.45	3	5888	0	0	0.18	0.45	3	9340
Down12^c	0	0	0.18	0.59	6	9475	0	0	0.24	0.74	5	5888	0	0	0.18	0.62	6	9340
UpAll12	0	0.02	0.02	0.01	0.04	9475	0	0.02	0.02	0.01	0.04	5888	0	0.02	0.02	0.01	0.04	9340
DownAll12	0	0.01	0.02	0.01	0.04	9475	0	0.02	0.02	0.01	0.04	5888	0	0.01	0.02	0.01	0.04	9340
years	0	1.34	3.29	6.14	54.74	9475	0	2.21	6.17	7.48	32.43	5888	0	0.62	1.63	2.28	11.03	9340
N^c	0	2	1.6	0.5	2	9475	1	2	1.97	0.16	2	5888	0	2	1.61	0.49	2	9340
gnipc	0.01	0.41	0.5	0.4	1.94	9475	0.02	0.58	0.57	0.42	1.94	5888	0.01	0.41	0.5	0.4	1.94	9340
growth	-34.7	2.16	2.11	7.27	59.76	9475	-32.56	1.8	1.8	6.98	59.76	5888	-34.7	2.18	2.12	7.24	59.76	9340
reserves	-99.16	5.81	8.09	23.79	131.84	9475	-99.16	3.8	5.02	22.53	131.84	5888	-99.16	5.91	8.2	23.83	131.84	9340
inf	-25.82	2.57	3.58	4.83	79.16	9475	-25.82	2.15	3.16	5.2	79.16	5888	-25.82	2.54	3.39	3.91	60.91	9340
reer	-26.51	0.2	0.22	6.69	26.39	9475	-26.51	-0.02	-0.25	6.34	26.39	5888	-26.51	0.17	0.19	6.7	26.39	9340
yield	-12.1	2.11	2.4	3.12	16.64	9475	-11.83	1.99	2.24	2.82	16.64	5888	-11.83	2.1	2.4	3.13	16.64	9340
debt	0.01	49.52	56.23	33.35	240.5	9475	0.01	52.19	57.31	31.89	240.5	5888	0.01	49.23	56.31	33.59	240.5	9340
fiscbal	-32.37	-2.55	-2.18	4.33	19.24	9475	-32.37	-2.54	-1.92	4.62	19.24	5888	-32.37	-2.56	-2.21	4.31	19.24	9340
current	-25.05	-0.26	0.42	6.3	27.13	9475	-25.05	0.65	1.59	6.41	27.13	5888	-25.05	-0.24	0.43	6.25	27.13	9340
corrupt	18.88	53.42	57.56	22.15	100.21	9475	20.96	62.62	61.75	22.61	100.03	5888	18.88	53.93	57.91	21.9	100.21	9340
ka.open	-0.03	1	0.74	0.33	1.03	9475	-0.03	1	0.77	0.32	1.02	5888	-0.03	1	0.75	0.32	1.03	9340
default	0	0	0.41	0.49	1	9475	0	0	0.4	0.49	1	5888	0	0	0.4	0.49	1	9340
recentdefault	0	0	0.28	0.45	1	9475	0	0	0.27	0.44	1	5888	0	0	0.27	0.45	1	9340
dumoecd	0	1	0.59	0.49	1	9475	0	1	0.6	0.49	1	5888	0	1	0.6	0.49	1	9340
dumeu	0	0	0.45	0.5	1	9475	0	0	0.49	0.5	1	5888	0	0	0.45	0.5	1	9340
yroffice	0	3.42	4.48	4.05	30.08	8109	0	3.42	4.34	3.73	30.08	4851	0	3.33	4.39	3.98	30.08	7975
maj	0.05	0.55	0.57	0.16	1	8109	0.05	0.57	0.59	0.16	1	4851	0.09	0.55	0.58	0.15	1	7975
exelecpre	0	0	0.09	0.29	2	8109	0	0	0.08	0.27	1	4851	0	0	0.09	0.3	2	7975
exelecpost	0	0	0.1	0.31	2	8109	0	0	0.08	0.28	2	4851	0	0	0.1	0.31	2	7975
legelecpre	0	0	0.26	0.44	2	8109	0	0	0.24	0.44	2	4851	0	0	0.26	0.44	2	7975
legelecpost	0	0	0.27	0.45	2	8109	0	0	0.27	0.45	2	4851	0	0	0.27	0.45	2	7975
expshare	0.01	0.52	1.96	4.14	22.8	5390	0.02	0.82	2.45	4.54	22.8	2694	0.01	0.53	1.97	4.17	22.8	5309
inline	13.03	41.82	38.84	15.46	85.6	5390	13.2	43.7	40	15.11	85.6	2694	13.03	41.99	39.01	15.5	85.6	5309
usmilaidshare	0	0.04	1.05	4.69	50.9	5390	0	0.01	0.36	1.83	18.15	2694	0	0.04	1	4.71	50.9	5309
comlang	0	0	0.21	0.41	1	5390	0	0	0.26	0.44	1	2694	0	0	0.19	0.39	1	5309
delf.language	48.47	95.02	90.16	11.25	99.92	5390	48.47	93.7	87.34	13.75	99.92	2694	48.47	95.02	90.06	11.3	99.92	5309
delf.ethnic	51.05	66.8	77.73	17.41	97.83	5390	51.05	65.25	75.48	18.11	97.83	2694	51.05	66.8	77.43	17.37	97.72	5309
bankexp	0	0.34	1.68	3.48	22.6	4022	0	0.44	2.23	4.22	22.6	2363	0	0.35	1.7	3.5	22.6	3967

Table A.3: Likelihood ratio pooling tests

	One pooled		One non-pooled	
	χ^2	p-value	χ^2	p-value
Mean reversion and convergence				
rating	19.5	0.012^{**}	25.7	0.001^{***}
UpDown12	21.3	0.006^{***}	30.5	0.000^{***}
UpDownAll	76.1	0.000^{***}	286.9	0.000^{***}
Time and inattention				
years	49.8	0.000^{***}	101.0	0.000^{***}
$changefund$	1.9	0.754	42.3	0.000^{***}
Competition				
$\Delta rating$	11.6	0.169	18.9	0.016^{**}
$UpDown12^c$	9.8	0.279	11.7	0.164
$\mathbf{N^c}$	39.3	0.000^{***}	196.4	0.000^{***}
Fiscal sustainability				
yield	16.7	0.002^{***}	27.1	0.000^{***}
$debt$	4.2	0.375	17.8	0.001^{***}
$fiscbal$	3.9	0.415	10.4	0.034^{**}
$reserves$	2.0	0.740	17.4	0.002^{***}
Macroeconomic environment				
gnipc	10.0	0.040^{**}	17.6	0.001^{***}
$growth$	3.3	0.508	5.2	0.270
inf	9.2	0.057^{*}	6.8	0.148
$reer$	1.8	0.771	7.1	0.131
$current$	7.9	0.095^{*}	20.0	0.000^{***}
Institutions				
$corrupt$	4.0	0.412	7.6	0.108
$ka.open$	4.3	0.368	4.4	0.349
$default$	7.1	0.129	5.0	0.288

Note: We combine variables based on one underlying variable in one test, i.e. we do not test separately for different "components" of a nonlinear specification of one variable: For example, $Up12$ and $Down12$ are considered together, see $UpDown12$.
In the columns labeled "One pooled", the unconstrained model has no pooled indicators and the constrained model only pools the variable indicated by the name of the row. In the columns labeled "One non-pooled", all variables except the one indicated are pooled in the unconstrained model, whereas all variables are pooled in the constrained model.
Stars (*** / ** / *) indicate significant rejections of the pooling tests at the 1% / 5% / 10% level. We consider variables unpooled (bold variable names) if both tests reject at the 5%-level.

Table A.4: Marginal effects (percentage points), baseline model

Variable	Value	Agency	Down	Up	Down (oprob)	Up (oprob)	Ann.
		Moody's	-0.11***	-0.45***	4.90	-5.53	-1.96***
$Up12$	0.00	S&P	-0.37***	0.19	-15.28***	14.87**	-0.44
		Fitch	0.13	-0.37***	1.54	-2.37**	-1.60*
		Moody's	0.18	-0.24***	5.43	-5.97**	-0.15
$Down12$	0.00	S&P	-0.02***	-0.05***	2.32	-1.30	-0.22
		Fitch	-0.05***	-0.19***	0.10*	-0.21	-2.56***
		Moody's	15.80	4.99	182.51	-264.29***	85.32***
$UpAll$	0.02	S&P	43.34	-15.40***	1617.09	-920.63***	33.32***
		Fitch	-13.36***	71.56	-163.83***	388.75***	320.73***
		Moody's	1.37	20.01	-141.65***	209.24***	66.87***
$DownAll$	0.02	S&P	25.04	21.57	-478.60***	287.70***	127.53***
		Fitch	21.99	44.25	23.64	-52.92***	772.47***
		Moody's	-0.01***	0.01	-0.19***	0.28***	-0.01*
$years$	13.63	S&P	-0.01***	0.00***	-0.16***	0.10*	-0.01***
		Fitch	-0.01***	0.00**	-0.06***	0.15***	-0.22***
		Moody's	0.10*	0.05**	1.03	-1.50	0.58***
$changefund$	-0.38	S&P	0.14	0.02**	1.93	-1.15	0.35***
		Fitch	0.08*	0.00***	0.33	-0.74	1.45***
		Moody's	-0.09***	0.23	-2.87***	4.61***	0.27**
$\Delta^{+}rating$	0.00	S&P	-0.07***	0.10	-5.57***	3.56***	0.16**
		Fitch	-0.10***	0.35	-0.90***	2.34***	0.67**
		Moody's	-0.16***	0.17	-3.82***	5.00***	-0.13
$\Delta^{-}rating$	0.00	S&P	-0.17***	0.07	-6.94***	3.77***	-0.08
		Fitch	-0.17***	0.29	-1.27***	2.40***	-0.32
		Moody's	-0.19***	0.54	-5.47***	12.27***	0.26
$Up12^{c}$	0.00	S&P	-0.22***	0.24	-11.69***	9.89***	0.16
		Fitch	-0.19***	0.95	-1.55***	6.86***	0.64
		Moody's	0.31	-0.04***	4.09	-4.79***	1.07***
$Down12^{c}$	0.00	S&P	0.38	-0.02***	7.17	-3.57***	0.65***
		Fitch	0.28	-0.17***	1.41	-2.24***	2.58***
		Moody's	0.05**	-1.05***	4.77	-10.01***	-2.41***
N^{c}	2.00	S&P	0.40	0.15	3.18	-2.06	1.35**
		Fitch	0.09*	0.23	0.04**	-0.08	3.58***
		Moody's	-0.01***	0.01**	-0.14***	0.21	-0.02
$yield$	2.07	S&P	0.05	-0.01***	1.54	-0.91	0.08
		Fitch	0.00**	-0.09***	0.16	-0.35**	-0.65***
		Moody's	0.00	0.00**	0.01	-0.02	0.01**
$debt$	49.78	S&P	0.00	0.00**	0.02	-0.01	0.00**
		Fitch	0.00*	0.00***	0.00	-0.01	0.02**
		Moody's	-0.03***	-0.01***	-0.38***	0.55**	-0.16***
$fiscbal$	-2.55	S&P	-0.04***	0.00***	-0.71***	0.42**	-0.09***
		Fitch	-0.02***	0.01	-0.12***	0.27**	-0.39***
		Moody's	0.00***	0.00	-0.07***	0.10***	0.00
$reserves$	5.36	S&P	0.00***	0.00*	-0.13***	0.08***	0.00
		Fitch	0.00***	0.01	-0.02***	0.05***	0.01
		Moody's	0.72	-1.19***	20.15	-29.47***	-0.38
$gnipc$	0.43	S&P	0.33	0.05*	4.09	-2.43	0.87
		Fitch	0.53	-0.10***	2.46	-5.54*	9.03***
		Moody's	-0.02***	0.02	-0.40***	0.58***	-0.02
$growth$	2.09	S&P	-0.02***	0.01*	-0.75***	0.45***	-0.01
		Fitch	-0.02***	0.03	-0.13***	0.29***	-0.05
		Moody's	0.01*	0.00**	0.06	-0.09	0.04*
inf	2.44	S&P	0.01*	0.00**	0.11	-0.07	0.02*
		Fitch	0.00*	0.00***	0.02	-0.04	0.09*
		Moody's	-0.02***	0.02	-0.40***	0.59***	-0.02
$reer$	0.14	S&P	-0.02***	0.01*	-0.76***	0.45***	-0.01
		Fitch	-0.02***	0.03	-0.13***	0.29***	-0.04
		Moody's	-0.03***	0.05	-0.92***	1.35***	0.02
$current$	-0.03	S&P	-0.03***	0.02	-1.74***	1.04***	0.01
		Fitch	-0.04***	0.09	-0.29***	0.67***	0.05
		Moody's	-0.02***	0.02**	-0.39***	0.57***	-0.02**
$corrupt$	56.07	S&P	-0.02***	0.01***	-0.73***	0.43***	-0.01**
		Fitch	-0.02***	0.03**	-0.12***	0.28***	-0.06**
		Moody's	0.15	0.22	0.34	-0.50	1.37***
$ka.open$	1.00	S&P	0.24	0.09*	0.64	-0.38	0.83***
		Fitch	0.10*	0.20	0.11	-0.25	3.43***
		Moody's	0.15	-0.31***	6.24	-6.62***	-0.51*
$default$	0.00	S&P	0.10	-0.12***	10.59	-4.88***	-0.30**
		Fitch	0.22	-0.43***	2.23	-3.02***	-1.30**

Appendix B: Robustness checks

In this appendix, we present additional data used in and results from robustness checks. In particular, we capture rich country bonusses, a potential home bias, the influence of political variables on ratings, and the effect of outlooks. Results on the first three extensions, which seem to have only marginal effects the rating process, are reported in Table B.1. Results on the extension using outlooks are reported in Table B.2. Pooling is rejected for all additional variables, so results are reported by agency. Although we do not repeat the results, all extension models also include all variables from the baseline model and yield mostly similar results.

Rich country bonus, home bias, or (cultural) proximity The previous literature has shown some evidence that rich countries, or those close to home countries of rating agencies, get preferential treatment. We address this possibility in robustness checks by adding a dummy variables indicating OECD (*dumoecd*) or EU membership (*dumeu*). To account for home bias and proximity to the home country, we also use the data from Fuchs & Gehring (2017), namely the share of exports from the host country going to the rated country (*expshare*), a measure of voting alignment with the US at the United Nations (*inline*), military aid from the US (*usmilaidshare*), the degree of cross-border exposure of home-country banks (*bankexp*)[21], a common language dummy (*comlang*) and indicators for cultural (*delf.ethnic*) and linguistic differences (*delf.language*).

Table B.1 shows that being a member of the OECD leads to either significantly less announcements by Moody's and an (not always significant) upward push of ratings in all three agencies. Given the positive feedback effects due to interactions between the three agencies, the positive push may create some self-reinforcing behavior, leading in general to a preferential treatment of OECD countries by all three agencies. The same treatment is, however, not extended to member states of the European Union (although there is a large overlap of the two groups). Instead, they receive more announcements by S&P, with other coefficients being insignificant. This may be explained by the fact

[21] This indicator has a much lower availability compared to the other variables in Fuchs & Gehring (2017). Thus we add it in a separate estimation.

Table B.1: Impact of preferential treatment and home bias on ratings

	Moody		S&P		Fitch	
	Prob	OProb	Prob	OProb	Prob	OProb
			Model with OECD dummy			
dumoecd	-0.12*	0.294*	0.074	0.35	-0.011	0.288**
	(0.065)	(0.170)	(0.095)	(0.237)	(0.049)	(0.117)
			Model with EU dummy			
dumeu	-0.041	-0.19	0.278***	0.112	0.043	-0.087
	(0.058)	(0.161)	(0.083)	(0.244)	(0.044)	(0.110)
			Model with home bias (without bankexp)			
expshare	-0.009	0.083*	-0.062*	0.065	0.001	0.051**
	(0.012)	(0.044)	(0.033)	(0.099)	(0.009)	(0.025)
inline	-0.002	0.019*	0.001	-0.022	0.005	0.011
	(0.004)	(0.010)	(0.006)	(0.019)	(0.003)	(0.007)
usmilaidshare	-0.006	-0.03	-0.011	-0.034	-0.012*	-0.034**
	(0.009)	(0.023)	(0.025)	(0.080)	(0.007)	(0.017)
comlang	-0.075	-0.28	-0.152	-0.284	0.018	0.155
	(0.090)	(0.238)	(0.155)	(0.419)	(0.077)	(0.194)
delf.language	0	-0.003	-0.003	-0.026	0.008**	0.002
	(0.005)	(0.013)	(0.007)	(0.022)	(0.004)	(0.010)
delf.ethnic	0.002	-0.002	0.006	-0.02	0.001	0.004
	(0.003)	(0.007)	(0.005)	(0.014)	(0.002)	(0.005)
			Model with home bias (with bankexp)			
bankexp	0.011	-0.028	0.001	0.143	0.028**	0.045
	(0.020)	(0.057)	(0.025)	(0.122)	(0.014)	(0.040)
			Model with political variables			
yroffice	0	0.092**	-0.016	-0.024	-0.008	0.003
	(0.017)	(0.045)	(0.025)	(0.066)	(0.014)	(0.033)
$yroffice^2$	0	-0.005**	0.001	0.001	0	0
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
maj	-0.8	4.715***	1.233	-4.105***	-0.492	4.298**
	(0.753)	(1.784)	(1.741)	(0.489)	(0.768)	(1.853)
maj^2	0.688	-3.753**	-1.563	6.169***	0.72	-3.053**
	(0.627)	(1.552)	(1.504)	(0.685)	(0.601)	(1.467)
exelecpre	0.208***	-0.004	0.132	0.209	0.093	-0.244
	(0.078)	(0.201)	(0.127)	(0.309)	(0.067)	(0.152)
exelecpost	0.061	-0.116	0.163	0.278	-0.09	0.265*
	(0.081)	(0.202)	(0.119)	(0.309)	(0.069)	(0.161)
legelecpre	-0.056	-0.275	0.049	-0.192	-0.038	0.06
	(0.063)	(0.176)	(0.099)	(0.251)	(0.050)	(0.115)
legelecpost	0.016	-0.1	0.073	-0.233	-0.031	-0.011
	(0.059)	(0.157)	(0.088)	(0.216)	(0.047)	(0.113)
baseline controls				YES		

Note: Every part reports estimates from different models that go beyond coefficients from the baseline model. For example, in the section "Model with OECD dummy", we add *dumoecd* to the baseline model. In case of the home-bias variables of Fuchs & Gehring (2017), we run two regressions, the first one with the variables *expshare* to *delf.ethnic*, the second one *adding bankexp* (which has much lower data availability) to the previous set of variables and the baseline model. Detailed results of all coefficients can be obtained from the authors.

that a large fraction of the EU sample is coinciding with the crisis and post crisis period. With binding commitments to bail out crises countries being introduced in that time, and some EU countries experiencing substantial problems, this might have created spillovers explaining the negative EU dummy.

The evidence using cultural proximity indicators of Fuchs & Gehring (2017) is largely inconclusive. Higher export shares seem to be positive for ratins, while receiving US military aid has a negative effect on the rating. This is explicable since the signal of instability that is related to receiving military aid might be more important than being considered an ally of the West.

To make sure that the insignificant results are not driven by parameter proliferation, rather than the absence of effect, we also construct a single proximity factor (the first principal component of the same indicators). However, this proximity factor does not have a significant effect on any agency.

Political variables In another robustness test, we include variables from the database of political institutions provided by the World Bank. First, we measure political stability, and second, we address strategical issues in the timing of rating announcements around elections. We utilize the availability of election dates in the data and assume that all political changes occur at election dates when we construct monthly variables. However, the database does not have the same coverage as our baseline data. Therefore, we loose some observations, as indicated in Table A.2.

With respect to political stability, we use the parliamentary majority of the government ($majority$) and the years in office of the executive ($yroffice$). A larger majority and longer time in office should indicate greater political stability. However, to be able to differentiate stable (democratic) countries from autocratic regimes, we also include both terms in squares ($majority^2$ and $yroffice^2$).

With respect to strategic actions around election dates, we include a range of indicators that extends the work by Block & Vaaler (2004) and Vaaler, Schrage & Block (2006) on the impact of elections and partisanship on ratings. Rather than just controlling for presidential elections in the current year, as their work does, we introduce separate dummies indicating the 12 months before and the 12 months after an election for the legislative (*legelecpre, legelecpost*) or executive (*exelecpre, exelecpost*) branch of government.

For all three agencies, mostly political stability seems to matter, while proximity to executive or legislative elections is nearly irrelevant, see Table B.1. We find that political stability mainly works through the magnitude of the government majority, which affects the direction of rating evaluations for all three agencies. Although the signs on $majority$ and $majority^2$ differ between agencies, the marginal effect of a more comfortable government majority is positive for the relevant order of magnitude between 40% and 70%. Beyond that – i.e. for a size of the majority that is rarely found outside of autocratic regimes – the marginal effect is slightly negative for both Moody's and Fitch. Years in office only matter for Moody's, with the expected hump-shaped effect. Around executive elections, there seems to be a slightly higher reevaluation probability (significantly so for Moody's). In case of a reevaluation after an executive election, Fitch has ceteris paribus a tendency to upgrade a rating.

Rating outlooks In an extension of our baseline specification, we also account for rating outlooks and their changes. In a first robustness check, we include the stance of the *outlook* (which can be positive, neutral or negative) as an explanatory variable. To account for possibly asymmetries, positive and negative outlooks both by the agency under observations and its two competitors are modeled as two dummy variables each ($outlook^+$, $outlook^-$, $outlook^{+,c}$ and $outlook^{-,c}$). We do not include these variables in our baseline specification to prevent obfuscating the view on the underlying rating process. If the reasons to change the rating correspond to reasons to change the outlook, this creates a multicollinearity problem. When explaining a rating agency's decision by its declared view that such a change is likely (i.e., the outlook), the outlook would capture effects that should be attributed to the fundamental indicators truly underlying the rating.

In this robustness check, explanatory power increases substantially.[22] Since outlooks

[22] With 24 additional coefficients (four variables in two equations for three agencies) being estimated, the likelihood increases by around 180 points.

reflect a first-stage assessment by rating agencies, and are meant to pave the way for upcoming rating changes, this is hardly surprising. The coefficients have the expected signs (see Table B.2). Outlooks increase announcement probabilities (significantly in six out of twelve coefficients). In case of an announcement, there is a very strong tendency to confirm the own outlook, and some pressure to follow the outlook by competitors.

Table B.2: Impact of outlooks on ratings

	Moody		S&P		Fitch	
	Prob	OProb	Prob	OProb	Prob	OProb
$outlook^{+}$	0.213***	0.407**	0.455***	1.739***	0.242***	1.569***
	(0.074)	(0.184)	(0.123)	(0.364)	(0.070)	(0.166)
$outlook^{-}$	0.08	-0.461***	0.329***	-1.273***	0.003	-1.016***
	(0.079)	(0.173)	(0.101)	(0.257)	(0.055)	(0.138)
$outlook^{+,c}$	0.038	0.175	0.088	0.783**	-0.013	0.182
	(0.077)	(0.198)	(0.117)	(0.346)	(0.057)	(0.144)
$outlook^{-,c}$	0.275***	-0.581***	0.124	-0.111	0.278***	-0.085
	(0.064)	(0.169)	(0.102)	(0.257)	(0.057)	(0.135)
baseline controls			YES			

Yet, the most striking result is that all other coefficients remain mostly robust (not shown in Table B.2). This is highly surprising. If positive and negative outlooks were explained by the same observable factors that drive ratings, their inclusion should replace the explanatory power of those indicators. In other words, if higher debt would always trigger agencies to first publish a negative outlook followed by a downgrade, then high debt should no longer play an additional role in the explanation of ratings. Yet, the coefficients on fundamentals are – for the most part – only marginally diminished. This strongly indicates that outlooks are mostly orthogonal to the indicators we include in our specification. As we already use an extremely general setup, covering a wide range of drivers of ratings identified in the previous literature, this indicates that the outlook often captures aspects of the rating decision that are hard to quantify and not easily available, such as institutional changes, new laws, etc. Economically this makes sense. If a country is downgraded because it has an apparently unsustainable level of debt, or is facing a crushing recession, the reasons to downgrade are fairly obvious. Yet, more subtle changes require deeper understanding and the corresponding analysis might take time. To still signal their attention to potential problems arising, changing the outlook might be a good strategy for rating agencies.

Since outlooks are obviously important to predict rating levels, and thus might be of equal interest to market participants as signals, we run a second robustness test where outlook changes are considered as part of the explained variable, i.e. as rating changes. This also changes the rating level (one of the explanatory variables): A negative (positive) outlook is treated as a 0.3 point deduction from (addition to) the numerical transformation of the rating. We find qualitatively very similar results in this setup. One of the reasons might be that quite a lot of outlook changes do indeed coincide with rating changes. As mentioned before, outlooks often turn negative on the initial downgrade in a sequence of downgrade and return to neutral on the last downgrade of the same sequence. Correspondingly, the actual impact on our estimation is fairly small.

Appendix C: Technical remarks

Boundary adjusted ordered probit This directional rating decision is modeled as an ordered probit model with explanatory variables Z (including an intercept) and two thresholds $\mu = (\mu_1, \mu_2)$, separating the three possible categories. We further account for the bounded nature of rating levels. For the highest (lowest) rating classes, further upgrades (downgrades) are impossible and should therefore have a probability of zero. As an adjustment, we treat predictions to upgrade (downgrade) at the highest (lowest) rating class as if they were predicting no change, effectively adding the corresponding probabilities to the probability of no change (Hantzsche 2017). Without this correction, the impact of the rating level might be overestimated, because the estimator would try to fit the zero probability to downgrade at low ratings or upgrade at high rating levels. Denoting with r the rating level, we introduce two dummy variables that are one for observations with boundary rating levels, $D^{AAA} = 1_{r=AAA}$ and $D^{D} = 1_{r=D}$. Using $1 - \Phi(x) = \Phi(-x)$, we thus model the directional decision $\tilde{y}$ as follows:

$$
\begin{aligned}
P(\tilde{y} = -1|Z) &= (1 - D^{D})\Phi(-Z\gamma + \mu_1) \\
P(\tilde{y} = 0|Z) &= (\Phi(Z\gamma - \mu_1) - \Phi(Z\gamma - \mu_2)) + \left(D^{D}\Phi(-Z\gamma + \mu_1) + D^{AAA}\Phi(Z\gamma - \mu_2)\right) \\
P(\tilde{y} = 1|Z) &= (1 - D^{AAA})\Phi(Z\gamma - \mu_2)
\end{aligned}
$$