ECONOMICS – WORKING PAPERS 2022/05

How much is too much?

Assessing the non-linear relationship between debt and sovereign creditworthiness

European Investment Bank
The EU bank

How much is too much?

Assessing the non-linear relationship between debt and sovereign creditworthiness

March 2022

European Investment Bank The EU bank

How much is too much? Assessing the non-linear relationship between debt and sovereign creditworthiness

EIB Working Paper 2022/05
April 2022

This is a publication of the EIB Economics Department.
economics@eib.org
www.eib.org/economics

About the Economics Department
The mission of the EIB Economics Department is to provide economic analyses and studies to support the Bank in its operations and in the definition of its positioning, strategy and policy. The department, a team of 40 economists, is headed by Director Debora Revoltella.

Authors
Sanne Zwart (EIB)

Disclaimer
The views expressed in this publication are those of the authors and do not necessarily reflect the position of the European Investment Bank.

EIB working papers are designed to facilitate the timely exchange of research findings. They are not subject to standard EIB copyediting or proofreading.

For further information on the EIB's activities, please consult our website, www.eib.org. You can also contact our InfoDesk, info@eib.org.

Published by the European Investment Bank.
Printed on FSC® Paper.

pdf: QH-BK-22-005-EN-N ISBN 978-92-861-5236-8 ISSN 2599-736X doi: 10.2867/961968

How much is too much? Assessing the non-linear relationship between debt and sovereign creditworthiness

9 March 2022

Sanne Zwart[1, 2]
European Investment Bank

Public debt is a very weak predictor of a country's credit rating if a country's other features are not taken into account. However, everything else equal, more public debt is associated with worse ratings. This paper explores the relationship between debt and sovereign creditworthiness by explicitly modelling the debt thresholds associated with rating changes. It finds that the impact of an increase in public debt is highly non-linear and crucially depends on a country's economic situation. In particular, low levels of GDP per capita are associated with a smaller range of possible ratings than higher levels. Hence, for countries with a higher GDP per capita, a change in debt levels is thus more likely to result in a rating change. Overall, the non-linear relationship between debt and creditworthiness is substantial, and accounting for it improves the performance of sovereign credit rating models significantly.

[1] The views expressed in this publication are those of the author and do not necessarily reflect the position of the European Investment Bank.

[2] The author would like to thank Aitor Erce, Frank Betz and Joana Conde for useful comments and suggestions, and Tomasz Olejnik for statistical support.

1 Introduction

Public debt is a major driver of sovereign creditworthiness despite the correlation with credit ratings being very low. These two seemingly contradicting statements can be reconciled by analysing public debt *conditional on a country's other characteristics*. In other words, a country's situation sets thresholds for public debt that drive the assessment of its creditworthiness and hence of its credit rating. A straightforward approach to derive the thresholds would be to assume a regular linear structure, which implies that an increase in public debt will always have the same impact on ratings. While appealing due to its simplicity, this is a strong assumption. A full understanding of the link between ratings, public debt and a country's conditions can only be obtained when no linear structure is imposed *a priori*.

Indeed, the methodologies of credit rating agencies (CRAs) are considerably more sophisticated than directly transforming a weighted average of economic variables, including public debt, in a linear way into ratings. Some, though, do utilise a linear transformation at an intermediate stage of the rating process. However, cognisant that no quantitative model can capture all relevant factors, CRAs methodologies allow for a qualitative overlay at various stages. The overlay needs to be justified and is sometimes limited in terms of notches. Nevertheless, the overlays provide considerable scope for ratings to be non-linear in the input variables. Besides a direct overlay, qualitative assessments and expectations are all subjective elements which, if applied in a consistent way, could introduce additional non-linearities. For instance, if for lower rated countries fiscal consolidation were considered less likely to be successful than for better rated countries, a fiscal deficit would be a greater drag on the rating.

In this paper, ratings of the three major CRAs are used to analyse the relationship between sovereign creditworthiness and public debt. In particular, the assumption that an increase in public debt affects a rating independently of a country's conditions is replaced by debt thresholds that are a function of the other economic variables. Taken together, these thresholds define areas associated with each of the ratings. Importantly, even if the borders of these areas are linear, they can imply areas that have different shapes as long as the borders are not having the same slope or the same distance to each other. Hence, estimating the borders independently allows for a non-linear relationship between public debt and ratings.

Using different estimation methods, this paper explores the effect of the standard assumption that borders have equal slopes and are equidistant. The analysis considers annual ratings and data for the period 2010-19. The various variables included in the analysis are those typically used by CRAs and reflect public finances, economic performance, external performance and institutions. In addition, GDP per capita is included as a proxy of economic development. To simplify the analysis, ratings are aggregated by band, *i.e.* C, B, Ba, …, Aaa, which is shown to be sufficient for many purposes (even if those involve individual ratings), while the analysis can also be easily extended to cover individual ratings. Compared to OLS, ordered logit relaxes the conditions on equal distances between borders. Both OLS and ordered logit utilise the inherent structure provided by the ratings. On the other hand, multinomial logit ignores the logical order of ratings, and by also allowing the borders to have different slopes, does not impose any *a priori* structure. In between these estimation methods is "sequential logit", which utilises the inherent rating structure by reducing the importance of distant ratings when estimating a border. This is either done by estimating each border independently using all ratings (sequential global logit), or by only considering ratings adjacent to a border (sequential local logit). While linear estimations are sufficient to derive non-linear relationships, quadratic specifications are also considered to allow for richer forms.

Most studies assume a regular and linear relation between debt and ratings, and the few that allow for more complex relationships typically still assume identical coefficients. Linear models are often used to derive conclusions about the drivers of sovereign risk in order to analyse the judgement applied by CRAs (*e.g.* D'Agostino & Lennkh (2016)), to support critique of CRAs (*e.g.* Haspolat (2015)) or to gauge the reaction of CRAs to the COVID-19 crisis (Tran et al. (2021)). Other studies rely on the inherent order of the ratings to relax the imposed structure. Afonso et al (2009) consider ordered logit and probit models to study the determinants of sovereign ratings, while Teker et al. (2014) employ ordered probit in combination with factor analysis. Broto & Molina (2014) also employ ordered logit to analyse the drivers of rating upgrades and downgrades separately. Reusens & Croux (2017) use an ordered probit model to assess whether CRAs changed their sovereign credit rating assessment after the start of the European debt crisis. The non-linear relationship between public debt and sovereign credit ratings is explicitly analysed in Hadzi-Vaskov & Ricci (2019). Their findings, employing a rolling ordered probit regression, confirm that the debt-ratings relationship is non-linear and depends on the rating. Interestingly, this study allows for coefficients to vary across ratings. However, as the focus is on differences between advanced and emerging countries, it does not explore the assignment of ratings in more detail.

This paper finds that allowing for a non-linear relation between ratings and debt significantly improves the model performance. Here, performance is primarily measured by the number of correct predictions (hits), but also, *e.g.*, by the distribution of the prediction error. Importantly, none of the methods directly maximizes the number of hits, so that this indicator is suitable for comparisons. Moving from OLS to ordered logit, so no longer imposing that all ratings have equally-sized areas, increases the share of hits from 52% to 55%. Logistic estimation allows for borders with different slopes and further increases this share to 62-65%, depending on the method followed. When including quadratic and interaction terms for public finance variables, 69% can be achieved compared to 55% for OLS. Overall, the considerable improvement in performance when accounting for non-linearity appears to justify the additional parameters needed for the logistic regressions.

Further analysis supports the use of rating bands instead of individual ratings for the estimations: solely including ratings directly adjacent to a border is found to not be a fruitful approach as these ratings do not sufficient information about the shape and location of the border. Hence, for the estimations a broader range of ratings needs to be used. However, it is shown that aggregating ratings into bands and estimating the smaller number of implied borders allows to assign individual ratings with only a relatively small cost in terms of performance when compared to estimating all borders between individual ratings.

The validity of the results is confirmed by robustness checks and an assessment of the scope of overfitting. The good performance of multinomial logit turns out to depend crucially on the aggregation of ratings in bands. When considering individual ratings, the performance drops dramatically due to the much larger number of classes and the lack of structure. In contrast, the sequential logits continue to significantly outperform OLS. Overall, cross-fitting yields comparable performance results as the full estimations, which indicates that the sample is large enough relative to the number of parameters. Finally, the relatively poor performance of all models on C-ratings is mostly due to the unbalanced number of observations. This follows from analysing a private dataset with the European Investment Bank's internal sovereign ratings. This sample includes more C-rated countries than the CRA sample, and when conducting the same estimations, the share of correct predictions in the C-band is now comparable to that of other bands.

The non-linear relationship between debt and ratings can be further inspected through a graphical representation. The borders are then shown as a function of GDP per capita, as this variable is a key driver of the ratings. The B-ratings form a vast region, indicating that it is difficult for developing countries to reach Ba-status. Similarly, the Baa-area is also rather broad, signalling that once investment-grade status has been accomplished, a considerable fall in public debt in combination with an increase in GDP per capita is needed before the A-region is reached. The graphical representation confirms that for a country to be in the Aaa-region, its overall economic conditions should be very strong as otherwise the required GDP per capita level is almost forbiddingly high. The C-ratings are exceptional in that they require high levels of debt, which might reflect that there are relatively few observations. From moderate GDP per capita levels onwards and for higher debt levels, the areas are closer to each other, implying that in particular A- and Ba-ratings can change relatively easily

An important implication of the observed non-linearity is that general questions about rating drivers do not have simple answers. Analysing rating drivers by using an OLS estimation is appealing, but just because it gives a single set of coefficients does not mean it is the right approach. In fact, this paper indicates that the rating drivers are dependent on the rating. Hence, the effect of changes in economic variables on ratings can still be analysed, but to do so the sample should be restricted to a subset of similar ratings.

The findings also have important implications for the modelling of sovereign risks. Again, because of its simplicity it is tempting to use an OLS specification. However, there is a significant cost in terms of performance. An off-the shelve estimation method as multinomial logit would improve the performance when only considering a smaller number of rating classes, but breaks down for larger numbers (which probably explains why this method is not used in academic studies). When striving for the best possible results, sequential logit should be considered, as the improved quality of the credit assessments is likely to be worth the additional cost in terms of model design, computing and tractability. Regardless of the modelling approach, but in particular for OLS, it should be checked whether model performance can be improved by treating C-ratings separately and excluding them from the estimations.

The outline of the paper is as follows. Section 2 discusses the methodologies of CRAs. Section 3 covers the sample and discusses the stylised relation between debt and ratings. Section 4 discusses the modelling approach and introduces the borders between ratings. It continues by briefly explaining the estimation methods and the criteria used to assess their performance. The estimations are presented in Section 5. Building on the analysis of the border between speculative and investment grade, first linear estimations are conducted before second-order terms for public finance variables are included. This section also covers various robustness checks. Section 6 concludes.

2 Credit rating agencies' approaches to sovereign ratings

The assumption that ratings are a weighted sum of economic variables is one frequently made by both scholars and modellers. However, the rating methodologies of the three major CRAs are much more sophisticated than a linear aggregation, and hence it is *a priori* not evident that the relationship can be considered linear.[3]

All three major CRAs publish their rating methodologies (see Fitch Ratings (2021b), Moody's Investor Service (2019) and S&P (2017)). Broadly speaking, all use some aggregate of economic variables, sometimes involving several stages, to derive scores and consider a linear or almost linear scale to translate these scores into ratings. However, in practice their approaches allow for substantial deviations between the ratings implied by the aggregated economic variables and the ratings ultimately assigned, as adjustments can be introduced at various points during the rating process. Beyond adjustments related to factors not included in the methodologies, discretion can be introduced based on qualitative assessments and expectations about future developments.

2.1 Fitch

The Sovereign Rating Model is a multiple regression model, consisting of 18 variables which are structured around four pillars: i) Structural Features; ii) Macroeconomic Performance, Policies and Prospects; iii) Public Finances; and iv) External Finances. In case of volatile variables, a centered three-year average is taken. The coefficients of the variables are derived from an OLS regression, and there is no subjective judgement involved at this stage. The OLS regression is re-estimated annually. The model output is a score that is calibrated to the sovereign ratings. Crucially, this rating scale is linear in the model output. The lowest model output is associated with a CCC+ rating. For sovereigns with this rating, the model is not used to assign the ratings, as the rating definitions will be the determining factor.

For each of the four pillars, a justified quantitative overlay up to two notches can be applied, with the overall adjustment capped at three notches except for certain circumstances. Overall, this allows for a typical interval of six notches around the model rating.

2.2 Moody's

The rating methodology is based on a scorecard approach with four factors: i) Economic Strength, ii) Institutions and Governance Strength, iii) Fiscal Strength; and iv) Susceptibility to Event Risk. Underlying the factors are multiple variables as well as qualitative assessment. The weights of variables and factors are calibrated to reflect the agency's analytical thinking, with adjustments made every couple of years. Combining the first two factors with equal weights forms the Economic Resilience score. Subsequently, combining this score with that of Fiscal Strength using dynamic weights yields the Government Financial Strength score. Finally, the Susceptibility to Event Risk can lower this score by up to five notches. Throughout, the scores of the factors are associated with ratings linearly, covering Aaa- through C-ratings.

For the Economic Strength and the Fiscal Strength Factors, an adjustment of up to nine notches can be applied, while for the Institutions and Governance Strength an adjustment up to six notches is possible, and a similar overlay can be for applied to Susceptibility to Event Risk (two steps out of the eight scoring categories for this factor). In general, "the scorecard-indicated outcome is not expected to match the actual rating for each issuer," and several examples are given why they can be different. No limitation in terms of notches for this final adjustment is provided in the rating guidelines. When defaults are becoming likely, the ratings assigned reflect expected recovery rates.

[3] The probability of default is also non-linear in ratings, as is its logarithm.

2.3 S&P

The methodology relies on five pillars: i) Institutional Assessment; ii) Economic Assessment; iii) External Assessment; iv) Fiscal Assessment; and v) Monetary Assessment. Each of these pillars is assessed on a six-point scale, including both quantitative and qualitative factors. The average of the first two pillars forms the Institutional and Economic Profile, the average of the latter three the Flexibility and Performance Profile. These two profiles are combined in an almost linear way to form the indicative rating.

Adjustments can be made to each of the five pillars, with the maximum adjustment differing across pillars, but generally about one-third of the scoring range. The assigned rating is “most likely” within one notch of the indicative rating, but larger deviations can occur under certain conditions without an overall limitation. Ratings of CCC+ and below are based on a specific methodology.

3 Data

3.1 Sample

The analysis considers the period 2010-2019. The sample has an annual structure, reflecting that macroeconomic data is typically reported by calendar year. Due care is given to ensure that the data used for the estimations reflect the information available at previous points in time, *i.e.* for the assessment of the 2013 end-of-year ratings, the data available at the end of that year is used. This implies that, for example, the 2013 GDP real growth figure is an estimate and the one-year ahead GDP real growth figure is a forecast instead of being the actual 2013 and 2014 figures respectively.

The country selection is taken as widely as possible by including all countries which, for one or more years during 2010-2019, have public debt data and are rated by at least one of the thee CRAs (for the extended sample used in the robustness analysis, also countries rated by the EIB are included). Overseas territories, dependent areas and similar entities are excluded if they are consistently rated below the parent state. For these countries, (potential) support from the mainland can be expected to be a main driver of the rating, but quantifying this effect is beyond the scope of this analysis. The final selection consist of 139 countries.

Ratings are the long-term foreign-currency issuer ratings at 31 December of each year during 2010-2019. They are translated into the Moody's rating scale, using the standard conversion table[4] to allow for aggregation across agencies. The consistency between the rating scales is generally considered very high, which allows comparison of the CRAs' ratings of a country. Including the three CRAs increases the number of data points and thus allows for richer specifications. As the focus of this paper is on the relation between debt and ratings, any differences between CRAs is not further explored. During the time period, rating agencies have generally revised their methodologies, for instance the variables used and their weights, which could introduce differences. However, the definition of the ratings has been stable, so any differences could be absorbed by quantitative overlays. In general, the CRA methodologies are sufficiently flexible to address considerations beyond the direct drivers of creditworthiness, such as consistency with ratings of other countries or with previous assessments, or even the positioning of the rating relative to those of other CRAs. These factors are also ignored in this study.

For estimations, ratings are usually grouped into bands to have more observations.[5] The bands typically consist of the three ratings with identical letters, *e.g.* the Ba-band consists of Ba1, Ba2 and Ba3.[6] Aaa-ratings are considered separately, and all ratings of Caa1 and below, including ratings indicating types of default, are mapped into a single C-band, which eliminates any effect that the various approaches of the CRAs for these ratings may have. The rating with the largest number of observations is Aaa, followed by B1 and B2 (see Table 1). C-ratings occur rather infrequently.

[4] Ratings at the lower end, starting from Caa1, are not directly comparable across CRAs due to different definitions. For the data description, the standard mapping is used where Fitch and S&P ratings indicating Default, Restricted Default and Selected Default are mapped into Moody's rating C, while Fitch and S&P ratings C and Ca are mapped into Moody's rating Ca. For the estimations, all ratings of Caa1 and below are mapped into a single C-group.

[5] Individual ratings are analysed in Section 5.4, while Appendix A discusses how even an analysis of the bands allows for assigning individual ratings by utilising the probabilistic nature of the logistic regressions.

[6] This grouping is in line with industry standards, as all three CRAs describe the ratings by rating band (see *e.g.* Fitch Ratings (2021a), Moody's Investors Service (2021a) and S&P Global Ratings (2001)).

Table 1: Ratings by agency (2010-2019)

Rating	Fitch	Moody's	S&P	Total	Share	Band	Total	Share
Aaa	126	137	113	376	11.1%	Aaa	376	11.1%
Aa1	18	17	28	63	1.9%			
Aa2	40	43	52	135	4.0%	Aa	341	10.1%
Aa3	32	51	60	143	4.2%			
A1	69	58	38	165	4.9%			
A2	29	35	36	100	3.0%	A	405	12.0%
A3	41	48	51	140	4.1%			
Baa1	50	53	50	153	4.5%			
Baa2	74	77	67	218	6.5%	Baa	657	19.4%
Baa3	102	97	87	286	8.5%			
Ba1	61	70	46	177	5.2%			
Ba2	36	36	66	138	4.1%	Ba	588	17.4%
Ba3	94	83	96	273	8.1%			
B1	93	125	119	337	10.0%			
B2	103	84	122	309	9.1%	B	887	26.3%
B3	41	108	92	241	7.1%			
Caa1	0	28	13	41	1.2%			
Caa2	14	15	3	32	0.9%			
Caa3	0	18	1	19	0.6%	C	124	3.7%
Ca	6	2	4	12	0.4%			
C	7	3	10	20	0.6%			
Total	1036	1188	1154	3378	100%	Total	3378	100%

Note: Ratings of Baa3 and above are referred to as investment grade and those of Ba1 and below as speculative grade.

Data is taken from the International Monetary Fund's (IMF) World Economic Outlook (WEO) database where possible. This database covers almost all countries and a very wide range of variables, which ensures comparability of data and consistency across variables. In addition, the dataset includes forecasts, typically for the next five years. Finally, the database is updated twice a year, namely in April and October, which allows to base the analysis on the data (or estimates and forecasts) that were available at the time of the rating analysis. As the focus is on the year-end situation, the October editions are used here. As various editions of the dataset are used, contemporaneous and future values of economic variables are necessarily estimates or predictions of the IMF. At the time of their assessment, CRAs may of course have augmented these values with their private views. It is therefore implicitly assumed that the assessments of CRAs are not structurally different from that of the IMF. Besides replacing zeros in the data that obviously represented missing data, no data adjustment or manipulations were made.

The choice of variables closely follows that of the CRAs, while also aiming to keep the sample as large as possible (see Table 2 for an overview of the sample statistics). Public finances are covered by general government gross debt (in per cent of GDP), interest expenditure as share of government revenues and the primary balance as share of GDP. For many countries, the primary balance was not covered in the WEO before 2015. In case of missing data for this variable, the historical figures are taken from the first edition where the primary data is included.

Table 2: Sample statistics (2010-2019)

		Minumum	10th-percentile	Mean	Median	90th-percentile	Maximum	Standard deviation	Number of observa-tions
Fitch	(notches)	C	B2	Baa2	Baa3	Aaa	Aaa	5.0	1,036
Moody's	(notches)	C	B3	Baa3	Baa3	Aaa	Aaa	5.2	1,188
S&P	(notches)	C	B3	Baa3	Baa3	Aa1	Aaa	5.1	1,154
Debt	% of GDP	2.6	19.8	54.5	46.5	96.5	250.4	34.2	1,364
Interest	% of revenues	0.0	1.7	8.2	5.9	16.5	53.3	7.8	1,364
Primary balance	% of GDP	-29.8	-5.0	-1.2	-1.1	2.5	24.7	4.2	1,374
Growth	%	-35.0	0.5	3.2	3.2	6.5	18.7	2.9	1,387
Current account	% of GDP	-58.0	-11.8	-2.7	-2.7	7.0	44.1	9.0	1,375
Free-floating	dummy	0.0	0.0	0.2	0.0	1.0	1.0	0.4	1,390
Reserves	months of imports	-0.6	0.0	5.0	4.1	10.3	38.8	5.0	1,379
Governance	points	-1.9	-0.9	0.1	0.0	1.5	1.9	0.9	1,390
GDP p.c.	USD 2017 PPP	746	2,960	23,107	15,805	51,374	116,493	20,925	1,390
ln(GDP p.c.)	X	6.6	8.0	9.6	9.7	10.8	11.7	1.1	1,390

Note: To obtain the mean and the standard deviation of the ratings, ratings C to Aaa are associated with the numbers 1 to 21.

Economic performance and prospects are covered by real GDP growth (in per cent). External performance is captured through the current account balance (in percent of GDP). In addition, a dummy captures the *de facto* classification of a country's currency as free-floating at the end of the year according to the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions. The contemporaneous and future import cover of reserves (in number of months) is included for countries with other exchange rate regimes, and is obtained by dividing reserves (in US dollars) by imports (in US dollars) over 12. Reserves here are measured according to the BoP/IIP manual and are only available since the October 2015 edition. Hence, for earlier years, the historical values of the October 2015 WEO are taken.

A measure for governance is obtained from the World Bank World Governance Indicators. It is constructed as the simple average of the six factors (Voice and Accountability, Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption). When releasing a new wave, the World Bank labels it with the preceding year to reflect the time of the underlying data. Here, however, the simple average is associated with the year in which the underlying data was released (as this was the best possible assessment available at the end of the year). No forecasts are available.

Finally, GDP per capita (in PPP-adjusted constant dollars), a rather slow-moving variable, is included as a proxy for development. Although the link with sovereign creditworthiness is perhaps less direct than for variables related to public finances or external performance, it is typically found to be an important driver of ratings. Moreover, its very strong correlation with ratings makes it a key variable, and indeed it will be the main explaining variable in Section 4. As is standard, the logarithm of GPD per capita is used in estimations. In October 2014, the WEO's estimates of purchasing-power-parity weights were updated, following the release of the 2011 International Comparison Program (ICP) survey for new purchasing-power-parity benchmarks. As a result, GDP valued at purchasing power parity were updated as well, and figures were no longer directly comparable with those of previous editions. Hence, to ensure consistency across years for this variable, data for all years are taken from a single WEO edition, namely that of October 2020 (which is based on the 2017 ICP).

While these variables capture the main factors considered by CRAs, this list is not exhaustive. For instance, information related to the debt structure, interest rates, of balance-sheet liabilities related to state-owned enterprises, an indicator of political stability or the default history would be candidates for inclusion. However, the lack of single databases covering all countries is a major bottleneck for including these variables. Especially the lower rated countries are typically less well covered. Here, priority is given to keeping as many countries as possible in the sample so as to have the most complete picture related to the ratings.

During 2010-2019, ratings of Fitch and S&P deteriorated by half a notch, and those of Moody's by almost a full notch (see Table 3). The deteriorations most likely reflect that public debt increased by 19% (9.5 percentage points of GDP) during the same period. On the other hand, GDP per capita rose by 14% which is likely to be supportive of sovereign creditworthiness. Governance remained broadly stable.

Table 3: Averages of ratings and selected economic variables over time

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2019 vs 2010
Fitch	13.5	13.4	13.1	13.1	13.1	13.0	13.0	13.0	13.0	13.0	-3.7%
Moody's	13.5	13.1	12.9	12.7	12.8	12.8	12.6	12.6	12.6	12.6	-6.4%
S&P	13.0	12.9	12.6	12.6	12.5	12.5	12.4	12.3	12.3	12.5	-4.1%
Debt	49.1	49.0	50.2	52.0	53.1	56.4	58.9	59.0	58.6	58.6	19.3%
Interest	7.9	7.6	7.6	7.6	7.8	8.2	8.5	9.1	9.0	9.1	14.9%
Governance	0.12	0.12	0.12	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.6%
GDP p.c.	21,631	22,120	22,255	22,493	22,806	23,217	23,507	23,946	24,403	24,696	14.2%

Note: To obtain average ratings, ratings C to Aaa are associated with the numbers 1 to 21 (13 is Baa2). For governance, the change in per cent takes into account that the variable ranges between -2.5 and +2.5. For each variable, only countries with data in all periods are included.

Unsurprisingly, the ratings of the three are CRAs are closely correlated (see Table 4). Ratings also have a high correlation with governance and with GDP per capita. On the other hand, debt has a correlation close to zero, although with the expected sign – this relationship will be looked into closer in the next subsection.[7]

Table 4: Contemporaneous correlations of ratings and variables (2010-2019)

	Fitch	Moody's	S&P	Debt	Interest	Pr. Bal.	Growth	Current ac.	Free fl.	Reserves	Govern.	ln(GDP p.c.)
Fitch	1.00	0.93	0.95	-0.04	-0.32	0.09	-0.15	0.38	0.47	-0.12	0.55	0.61
Moody's	0.93	1.00	0.93	-0.09	-0.30	0.06	-0.08	0.38	0.43	-0.06	0.53	0.59
S&P	0.95	0.93	1.00	-0.06	-0.31	0.08	-0.11	0.40	0.46	-0.07	0.54	0.62
Debt	-0.04	-0.09	-0.06	1.00	0.41	-0.05	-0.34	-0.14	0.32	-0.25	0.30	0.20
Interest	-0.32	-0.30	-0.31	0.41	1.00	-0.01	-0.02	-0.22	-0.18	0.02	-0.12	-0.15
Pr. balance	0.09	0.06	0.08	-0.05	-0.01	1.00	0.16	0.32	0.03	0.06	0.07	0.14
Growth	-0.15	-0.08	-0.11	-0.34	-0.02	0.16	1.00	-0.06	-0.25	0.11	-0.25	-0.36
Current ac.	0.38	0.38	0.40	-0.14	-0.22	0.32	-0.06	1.00	0.20	0.19	0.22	0.47
Free fl.	0.47	0.43	0.46	0.32	-0.18	0.03	-0.25	0.20	1.00	-0.52	0.64	0.50
Reserves	-0.12	-0.06	-0.07	-0.25	0.02	0.06	0.11	0.19	-0.52	1.00	-0.25	-0.03
Governance	0.55	0.53	0.54	0.30	-0.12	0.07	-0.25	0.22	0.64	-0.25	1.00	0.69
ln(GDP p.c.)	0.61	0.59	0.62	0.20	-0.15	0.14	-0.36	0.47	0.50	-0.03	0.69	1.00

Note: The shown correlations are Kendall tau correlations if they involve ratings, and Pearson correlations otherwise.

Interest has a more pronounced negative correlation with ratings, and the primary balance a small positive one. Growth has a low and negative correlation, probably reflecting that many low-rated

[7] Analysing the univariate relationship is typically a regulatory requirement for credit rating model design, see *e.g.* article 32.5.f of the EBA/RTS/2016/03 (European Banking Authority, 2016) which states that "… competent authorities shall verify that the documentation includes … the univariate analysis of the variables considered and respective criteria for variable selection". When designing a sovereign credit risk model, the weak correlation of debt and ratings indicates that the univariate analysis should not be driving the variable selection.

countries are catching up and thus have high growth rates. The correlation of the current account is relatively high, which likely reflects the trade deficit of many low-rated countries. As mostly highly rated countries have a free-floating exchange rate, the correlation is positive. The negative correlation of reserves could indicate that lower-rated countries have to maintain higher reserves. When comparing economic variables with each other, the high correlation between GDP per capita and governance stands out, followed by that of governance and the free-floating dummy.

3.2 Debt and ratings

The low correlation between ratings and debt does not disappear when assuming that ratings are based on forward looking debt developments, as, if anything, estimations of the future debt stock only have a somewhat stronger correlation (see Table 5).

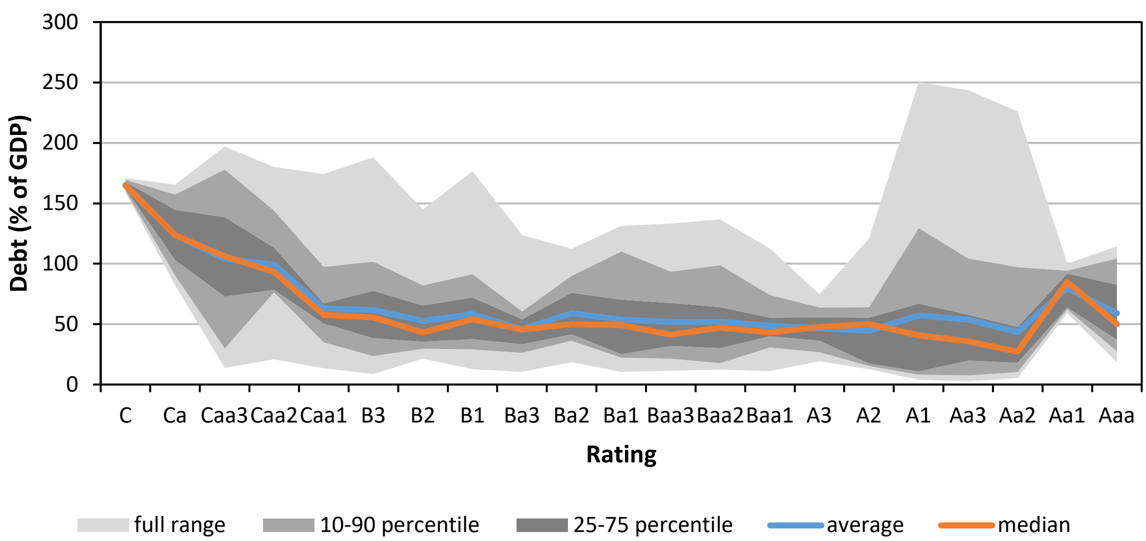
Table 5: Contemporaneous and forward correlations of ratings and debt (2010-2019)

	$Debt_t$	$Debt_{t+1}$	$Debt_{t+3}$	$Debt_{t+5}$
Fitch	-0.04	-0.05	-0.05	-0.05
Moody's	-0.09	-0.10	-0.10	-0.09
S&P	-0.06	-0.07	-0.07	-0.07
Combined	-0.07	-0.07	-0.07	-0.07

Note: The shown correlations are Kendall tau correlations.

Further insight into the relation between ratings and debt is obtained by looking at the average debt levels associated with ratings (see Figure 1 for Moody's; the figures for Fitch and S&P are similar). Low ratings (the C's) have higher average debt levels, although the dispersion for Caa -ratings is higher. However, from B3 to A2, median and average debt levels are broadly between 45% and 55% of GDP with a very modest decline. Strikingly, the average debt level for B1, B2 and B3 is identical to that of A1 (57% of GDP). Several countries with A1-, Aa3- and Aa2-ratings have rather high debt levels, creating a wedge between the median and the average levels for these ratings. Aa1-ratings stand out as having a considerably higher average debt level (80% of GDP) than other non-C ratings, while for Aaa-ratings it is again close to 60%. Overall, the weak negative bilateral relation between ratings and debt levels seems thus mostly driven by low ratings.

Figure 1: Debt level by rating (Moody's, 2010-2019)



Regressions of ratings on a single economic variable (and a constant) confirm that debt is not a very good predictor of ratings (see Table 6). The coefficients of all variables have the same signs as the respective correlation, but growth and reserves are not statistically significant. Although the OLS-coefficient of debt is highly significant, the estimated ratings are mostly Baa3 or Baa2, and the low R-squared of 0.007 confirms the poor performance in explaining rating levels. GDP per capita and governance have a much wider range of estimates and an R-squared above 0.5. The current account also has a wide range, but its R-squared is markedly lower.

Table 6: Bivariate OLS-regressions of ratings and estimated ratings (2010-2019)

	Estimations			Estimated ratings					
	Coefficient		R-squared	Minumum	10th-percentile	Mean	Median	90th-percentile	Maximum
Debt	-0.009	***	0.004	Ba1	Baa3	Baa3	Baa3	Baa2	Baa2
Interest	-0.261	***	0.177	C	Ba2	Baa2	Baa3	Baa1	A3
Primary balance	0.132	***	0.011	Ba3	Baa3	Baa3	Baa3	Baa2	A3
Growth	-0.059	*	0.001	Ba1	Baa3	Baa3	Baa3	Baa2	A3
Current account	0.300	***	0.241	C	Ba2	Baa3	Baa3	A3	Aaa
Free-floating	6.442	***	0.306	Ba1	Ba1	Baa3	Ba1	A1	A1
Reserves	-0.024		0.001	Baa3	Baa3	Baa3	Baa3	Baa3	Baa3
Governance	4.119	***	0.515	Caa3	B1	Baa3	Baa3	Aa3	Aa2
ln(GDP p.c.)	3.990	***	0.561	C	B2	Baa3	Baa2	A1	Aa1

Note: Ratings C to Aaa are associated with the numbers 1 to 21. ***, ** and * indicate statistical significance at the 1%-, 5%- and 10%-level respectively. Standard deviations are derived without accounting for potential dependence across time periods or CRAs.

Although the debt level appears not to play an important role for the rating level, it is the main driver of rating changes. When the same bivariate OLS-estimations are carried out with a country dummy, the coefficient of debt increases considerably (see Table 7). For 10% of the countries, the debt level changes by 55.9 percentage points of GDP or more during 2010-2019, which translates into a change of 3.0 notches. Its maximum range is even 7 notches. Interest, governance, growth and GDP per capita also have a relatively large impact in terms of notches. On the other hand, the effect of the current account is much smaller, and the effect of the free-floating dummy, reserves and the primary balance is still small.

Table 7: Bivariate OLS-regressions of ratings and estimated ratings with a country dummy

	Estimations			90th-percentile range		Maximum range	
	Coefficient		R-squared	Variable	Predicted rating	Variable	Predicted rating
Debt	-0.054	***	0.962	55.9	3.0	136.1	7.3
Interest	-0.155	***	0.955	14.9	2.3	26.2	4.0
Primary balance	0.021	***	0.948	13.4	0.3	38.8	0.8
Growth	0.165	***	0.953	9.6	1.6	40.7	6.7
Current account	0.011	**	0.948	20.3	0.2	51.8	0.5
Free-floating	0.499	***	0.948	0.0	0.0	1.0	0.5
Reserves	0.053	***	0.948	6.9	0.4	18.7	1.0
Governance	3.501	***	0.951	0.4	1.4	0.9	3.3
ln(GDP p.c.)	3.302	***	0.951	0.4	1.3	1.0	3.4

Note: Ratings C to Aaa are associated with the numbers 1 to 21. ***, ** and * indicate statistical significance at the 1%-, 5%- and 10%-level respectively. Standard deviations are derived without accounting for potential dependence across time periods or CRAs. The range of a variable for a country is the maximum value minus the minimum value over the estimation period. The table shows the 90th-percentile and the maximum across countries. The range of the predicted rating for a country is the range of the variable times the absolute value of the coefficient, and thus shows the impact in notches.

4 Methodology

4.1 Model

The above preliminary analysis suggest that debt is an important driver of rating changes, but not of rating levels. This in turn suggests the following approach to establish the link between debt and ratings: first derive critical debt levels, *i.e.* the thresholds between ratings (bands), based on other economic fundamentals, and then assign the rating (band) based on the particular debt level. This approach is made precise below.

Assume the general case of the rating being a function of *debt* and a vector $x = (x_1, \ldots, x_n)$ of other variables

$$rating = F(debt, x), \tag{1}$$

with the possible ratings 1, …, R ordered from worst to best. Loosely speaking, the aim is to derive, given x, the debt level that forms the border between ratings r and r+1, and to do so in such a way that the border can be estimated empirically.

A form widely used in the literature assumes that F assigns ratings based on a linear function f and intervals $[L_r, U_r)$ such that

$$f(debt, x) = \alpha + \beta \cdot debt + \gamma^T \cdot x \in [L_r, U_r) \quad \Rightarrow \quad rating = r. \tag{2}$$

Requiring that the intervals $[L_r, U_r)$ are disjoint and together span the codomain of f implies that the upper bound of an interval coincides with the lower bound of the next, so $U_r = L_{r+1}$, for $r \in \{1, \ldots, R\text{-}1\}$ (L_1 and U_R could potentially be required to be $-\infty$ and $+\infty$ respectively).

In practice, the bounds $U_1, \ldots, U_{R-1}$ are unknown. Often, they are considered to be equally spaced on some specific interval. This strong regularity assumption is typically made because of its appealing simplicity and not for empirical reasons. Importantly, making such an assumption on the bounds in the rating space has direct implications on the distance between the borders as functions of debt. In a similar fashion, the parameters and shape of the function f affect all the ratings and therefore all borders between them. Hence, imposing a certain structure on the bounds and assigning ratings based on a single function have the unwanted effect of pushing the analysis towards certain results.

Instead, here a more flexible form for the assignment of ratings underlying Equation (1) is proposed. Assume that the probability of assigning rating r is a function of the level of debt and the vector x of other variables

$$\mathbb{P}(\text{rating} = r) = p^r(debt, x). \tag{3}$$

The simplest way to assign ratings based on these probabilities is to assign the rating with the highest probability[8,9]

$$rating = \text{argmax}_r \, p^r(debt, x). \tag{4}$$

In a way, this method replaces the R-1 unknown bounds U_r with R-1 unknown functions p^r (as the R^{th} function follows due to the condition that the probabilities should sum up to 1). The additional parameters to be estimated can be seen as the "price" of imposing less structure on the bounds. In other words, prescribing the shape of the functions p^r imposes some structure on the borders, but

[8] In case ratings have equal probabilities, one will have to be selected randomly.
[9] See also Appendix A for an alternative way to allocate ratings.

due to the additional parameters this is less stringent than the structure imposed by the shape and single set of parameters of function *f* in Equation (2).

Rather than directly imposing a form on p^r, it is instructive to look at a pair of ratings *r* and *s* and define the set $A_{r,s}$ of inputs where rating *s* is at least as likely as rating *r*

$$A_{r,s} = \{ (debt, x) \mid p^s(debt, x) \geq p^r(debt, x)\}. \tag{5}$$

Reordering of the condition gives a condition on the odds ratio

$$\frac{p^s(debt, x)}{p^r(debt, x)} \geq 1. \tag{6}$$

This suggests to impose a functional form on the odds ratio instead of the probabilities themselves. Clearly, *R*-1 independent odds ratios pin down all the other odds ratios, while the condition that the sum of the probabilities equals 1 allows to obtain the probabilities. The natural choice for the functional form is $e^{q(debt,\, x;\theta)}$, where *q* is a polynomial with coefficients θ. This gives

$$A_{r,s} = \{ (debt, x) \mid e^{q(debt,\, x;\theta_{r,s})} \geq 1\} = \{ (debt, x) \mid q(debt, x; \theta_{r,s}) \geq 0\}, \tag{7}$$

The specific functional form for the odds ratio implies that the approach is equivalent to the logit model, which is hence the obvious regression method to be used to obtain estimates of the parameters $\theta_{r,s}$. Of course, in practice one would focus on estimating the parameters of the border for two subsequent ratings *r* and *r*+1.[10]

4.2 Borders

The border where the ratings *r* and *s* have equal probabilities is defined by

$$B_{r,s} = \{ (debt, x) \mid q(debt, x; \theta_{r,s}) = 0\}. \tag{8}$$

For a particular value of *x*, this set defines the debt values for which the probabilities of ratings *r* and *s* are equal

$$debt^{r,s}(x) = \{debt \mid q(debt, x; \theta_{r,s}) = 0\}. \tag{9}$$

In general, the border $debt^{r,s}(x)$ can be empty, contain a single value, or multiple values. Suppose the polynomial *q* is of the order 1, so

$$q(debt, x; \theta_{r,s}) = \alpha_{r,s} + \beta_{r,s}\cdot debt + \gamma_{r,s}^T\cdot x, \tag{10}$$

where $\alpha_{r,s}$, $\beta_{r,s} \in \mathbb{R}$ and $\gamma_{r,s} \in \mathbb{R}^n$. In this case, the set $debt^{r,s}$ contains exactly one number

$$debt^{r,s}(x) = (-\alpha_{r,s} - \gamma_{r,s}^T\cdot x) / \beta_{r,s}. \tag{11}$$

The set of debt values that constitute the border between rating *r* and rating *s* can thus be described by a linear function of the variables $x_1, \ldots, x_n$. Clearly, without any restrictions on the vector of variables x, $debt^{r,s}(x)$ can be negative or unrealistically high.

[10] At first sight, it seems that when analysing two subsequent ratings *r* and *r*+1, a linear probability model can also be used to estimate Equation (2). However, on top of its methodological drawbacks, this would not allow for combining the found probabilities into a full probability distribution over the ratings. In contrast, in the model characterized by Equation (4) there are *R*-1 independent odd ratios which can be estimated either separately or simultaneously and which yield the full probability distribution.

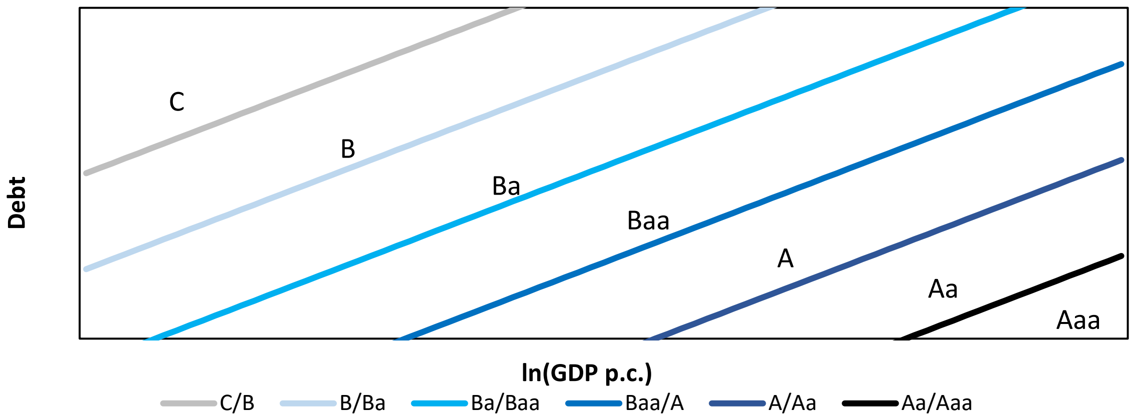
When $q^{r,s}$ is quadratic in *debt*, it is straightforward to derive that $debt^{r,s}(x)$ contains up to two values. If x_i enters f linearly or quadratically (possibly with an interaction term), the border between ratings r and s evaluated at specific values of the other control variables is a conic section, so an ellipse, parabola or hyperbola.

As the parameters of the *R-1* odds ratios are obtained without imposing any restrictions, there is no guarantee that a continuous path in the (*debt*, *x*)-space would traverse through the ratings in a sequential way. To make this more precise, consider the set S^r with inputs that yield rating r or create a tie with other ratings

$$S^r = \{ (debt, x) \mid r \in \text{argmax}_s\, p^s(debt, x) \}. \tag{12}$$

The sets S^r cover the entire input space, and the intersection between S^r and S^s is a subset of $B^{r,s}$. One may expect that $S^r \cap S^s = \varnothing$ if r and s are not adjacent. However, this would only hold in specific cases, for example in the linear case where $\gamma^T_{r,r+1} / \beta_{r,r+1}$ is the same for all ratings r and where $-\alpha_{r,r+1} / \beta_{r,r+1}$ is strictly monotonous (preferably decreasing) in r (see Figure 2). Obviously, these conditions do not hold in practice. However, the data often imply that if S^r are S^s are overlapping for non-adjacent r and s, this happens for very high values of the variables. Alternatively, the parameters $\gamma^T_{r,r+1} / \beta_{r,r+1}$ can be constrained to be identical, which yields a model close to that of Equation (2) with the crucial difference that the bounds are not imposed but estimated through the unconstrained intercepts. In fact, imposing bounds such that they imply equal distances between the borders, yields exactly the specification of Equation (2), so that the probability-based model of Equation (4) can be seen as a generalisation.[11]

Figure 2: Illustration of borders implying a gradual change in ratings



Note: The six borders constitute debt thresholds as function of GDP per capita. When a country's debt is below the threshold implied by its GDP per capita and the "Aa/Aaa"-border, the associated rating is Aaa. As the borders have the same slope, they do not cross, and when the debt level changes gradually, no rating band can be skipped. When borders do not have the same slope, they can still imply a gradual change in ratings as long as they do not cross for values of GDP per capita and debt that are economically meaningful.

4.3 Estimation method

As mentioned above, the model design is such that logistic regression is the natural and most flexible estimation approach.[12] However, other methods will be explored as well to gauge the impact of the various linearity assumptions. Ordered logit assumes that the coefficients are identical across bands (the proportional odds assumption which implies parallel borders in Figure 2), but that the cut-off

[11] In this case, the parameters are determined up to a factor $\beta_{r,r+1}$, which determines how fast the odds ratio is changing when moving away from the border. Also note that Equation (2) is typically estimated with OLS assuming normally distributed disturbances, and the probability-based approach with logit, which implies different results.

[12] Alternatively, linear discriminant analysis can be used, but this would require additional assumptions and is in general not preferred. See Hastie et al. (2017) for details.

points are not identical, allowing for different widths of the bands. OLS imposes equidistant borders, ensuring that the bands have equal width. All estimation methods can allow for curved borders by simply including higher-order terms.

Three types of logistic regression are considered. Multinomial logit maximises the likelihood of obtaining the observed ratings simultaneously. As a result, the estimation of, *e.g.*, the Ba/Baa-border takes into account the likelihood of all ratings. Importantly, for this approach the order of the ratings is not relevant, and it hence does not benefit from the inherent structure of the ratings. Secondly, the estimation of borders between subsequent ratings can be done independently (meaning that six regressions will have to be carried out). Note that the structure of the ratings is already used here to single out the most relevant borders. To estimate the Ba/Baa border, all Ba-ratings can be assigned a 0 and all Baa-ratings a 1 (sequential local logit). The structure is used again here to narrow down the area where the Ba/Baa border is supposed to be and to not let the likelihood of remote ratings interfere with the estimation of this particular border. Thirdly, the independent regressions can be carried out by assigning 0's to all ratings of Ba and below and 1's to all ratings of Baa and above (sequential local logit). Clearly, this approach further utilises the inherent order of the ratings to specify which ones are lower and which ones are higher. In this way, the information contained in each data point can be utilised for the estimation of all borders. Whether the local or global variant performs better thus depends on whether the ratings away from a particular border contain useful information or are merely introducing a bias.

Logistic regression maximises the likelihood of obtaining the actual ratings given the observed debt and vector of economic variables *x.* Importantly, a logit regression does not directly maximise the number of correct predictions, although in practice this aim is closely related to the maximisation of the likelihood. To avoid overfitting, ratings are typically grouped in bands of three, *e.g.* Baa1, Baa2 and Baa3, so that there are more observations. This again reflects that also Baa1 ratings might have some information regarding the border between Baa- and Ba-ratings. However, observations close to the border will have a larger impact on the parameters, as the probabilities will be more sensitive to the parameters than for observations further away from the border. If a country is rated by multiple CRAs, its characteristics (*debt, x*) can be associated with multiple ratings or can have multiple instances of the same ratings. All cases are included in the estimation as each of them contains information regarding the specific point: identical ratings indicate a higher reliability of the particular rating, while different ratings indicate that this country is closer to the border. Indeed, there is evidence that the average rating of all agencies is a better indicator of an impending default than a single rating of individual agencies (Kraemer, 2021).

One of the variables, say x_1, might be of particular relevance, and one might want to express or show the debt border as function of this variable. Here, ln(GDP p.c) would be the obvious candidate given its relatively high correlation with ratings and its relatively low correlation with debt (see Table 4), its economic relevance and its low volatility. However, the level and potentially the shape of the border depend on the values of the other (control) variables as well, as mathematically the border is an affine n-dimensional subspace in the (n+1)-dimensional space spanned by debt and the control variables. Perhaps the most relevant border is the one where all variables are at their means. It is thus convenient to do the logit regression with respect to the demeaned (centred) variables $\hat{x}_2, \ldots, \hat{x}_n$, which does affect the intercept but not the coefficients. It then directly follows that for the most relevant border, *i.e.* the one obtained with these variables set to 0, the intercept and the coefficients of *debt* and x_1 (including possible higher order and interaction terms) provide the full description.

In case of linear estimations, the debt levels of all countries can easily be adjusted to account for the actual values of the control variables $x_2, \ldots, x_n$ to facilitate a graphical representation. The adjustment accounts for the unexplained part of the control variables and follows directly from Equation (10),

$$\overline{debt} = debt + (\gamma_2\hat{x}_2 + \ldots + \gamma_n\hat{x}_n)/\beta, \tag{13}$$

where redundant sub- and superscripts are dropped for simplicity. In essence, the adjustment reflects that if a country performs better than average on a particular variable, say, growth, it can have a higher debt level before crossing the border towards a worse rating (assuming that β is negative and the coefficient of growth positive). Clearly, the adjustment can result in negative or unlikely high debt levels. Note that the adjustment depends on the particular border as the related coefficients are used.[13] Alternatively, for a single country in a particular year, the border can be adjusted to absorb the impact of the control variables. In this way, all borders of a particular country in a certain year can be showed in a single chart.

The above methods allows to analyse the effect on the rating of a change in *debt* or in x_1 while keeping the control variables constant. This would give an answer to the question how much debt can grow before the rating would be lowered. However, when comparing countries with different levels of *debt* and x_1, it is unlikely that the control variables are identical, as they might be related to *debt* and x_1. Hence, when comparing different combinations of *debt* and x_1, the levels of the control variables would have to be consistent. This can be accomplished by regressing the variables x_2, …, x_n on a constant, debt and x_1. The residuals $\tilde{x}_2$, …, $\tilde{x}_n$ are then used for the logit regression (debt and ln(GDP p.c.) are still centred to reduce collinearity in the quadratic estimations). Overall, the relation between variables on the one hand and debt and ln(GDP p.c.) on the other is in line with economic reasoning (see Table 8). For example, interest expenditure as share of revenues increases with debt and falls with economic development. However, for some variables the causality is reversed, *e.g.* countries with better governance typically have higher debt levels.

Table 8: Regressions of variables on debt and ln(GDP p.c.)

	Interest	Primary balance	Growth	Current account	Free-floating	Reserves	Governance
(intercept)	16.900	-6.877	12.357	-40.547	-1.605	6.240	-5.302
Debt	0.144	-0.006	-0.025	-0.094	0.002	-0.039	0.005
ln(GDP p.c.)	-1.710	0.636	-0.816	4.479	0.177	0.092	0.540
Values for combinations of debt and GDP p.c.							
(0, 2000)	6.1	-1.9	5.3	-6.6	-0.1	6.2	-0.9
(0, 64,000)	1.3	-0.1	3.0	6.0	0.4	6.4	0.6
(55, 25,000)	7.6	-0.8	2.8	-0.7	0.3	5.0	0.4
(100, 2000)	17.6	-2.4	3.3	-14.1	0.0	3.0	-0.5
(100, 64,000)	12.8	-0.6	1.0	-1.5	0.5	3.3	1.0

NB: The regressions exclude Greece and Japan, see section 4.4.

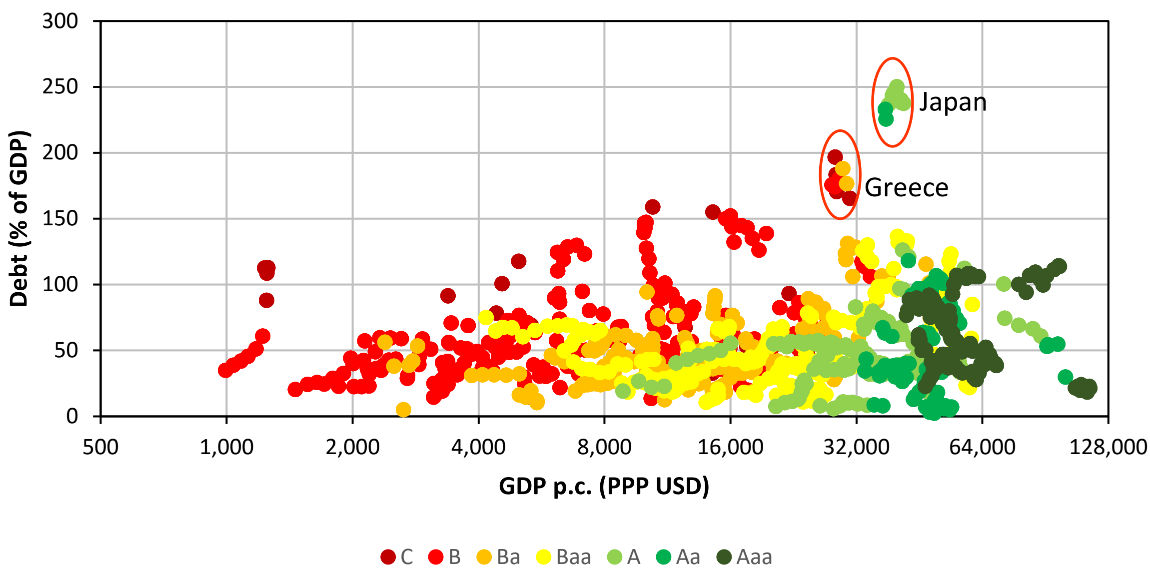
Mathematically, regressing the control variables on debt and ln(GDP p.c.) is a projection which ensures that the space spanned by *debt* and x_1 is orthogonal to the space spanned by the control variables. The new basis singles out a particular two-dimensional subspace of the border, namely the one where the values for the other values are all zero. In case of a linear specification, the projection will not affect the coefficients of these variables. The coefficients of *debt* and x_1, as well as the intercept, will be different though, as they now also capture the indirect effects through the control variables. Of course, the border will not be affected, it is merely expressed relative to a different basis. The debt can be adjusted to account for the actual value of the control variables in the same way as discussed above. Similarly, for quadratic estimations the regressed control variables can be used if all interaction terms are included.

[13] For the quadratic specification, the unexplained part of the auxiliary variables cannot always be absorbed in the debt level. The adjusted debt level satisfies $\beta\overline{debt} + \delta\overline{debt}^2 + \varphi\overline{debt} \cdot x_1 = \beta debt + \delta debt^2 + \varphi debt \cdot x_1 + g(x_2, \ldots, x_n)$. In case this quadratic equation in $\overline{debt}$ has two solutions, the highest one is the relevant one as it is increasing in *debt*. In case there is no solution, the non-linear relationship between the rating and *debt* prevents the debt level to absorb the effect of the variables x_2, …, x_n, for example because β and δ are positive and the effect of the auxiliary variables is very negative. Given that it is not always possible to adjust the debt level and that the higher order terms prevent a straightforward economic interpretation, this route is not pursued here.

4.4 Relevant area and performance criteria

The obtained borders are stretching well beyond economically relevant values. However, even for plausible values, the border can be an extrapolation if derived from remote data points. Most of the countries in the sample have a GDP between USD 2,000 and USD 64,000 and public debt below 130% of GDP (Figure 3). Beyond these values, the border are extrapolations and not necessarily meaningful. For this reason, Greece and Japan are excluded from the analysis. In addition, the number of observations in the C-band is relatively low and may not be sufficient to represent their distribution adequately. This issue is explored in more detail in Section 5.4.

Figure 3: Distribution of countries



Ex ante, rating bands could be expected to exhibit the three following characteristics: 1) higher debt leads to lower ratings; 2) higher GDP per capita leads to higher ratings; 3) small changes in variables result in at most one rating band difference (no jumps).[14] Together these characteristics imply upward sloping borders that are not crossing for economically-relevant values. Linear OLS is consistent with these requirements if the coefficient of debt is negative and that of ln(GDP p.c.) positive. For estimation methods imposing less structure, these requirements do not necessarily hold.

A full set of borders implies a classification of countries. Hence, a measure for the performance of the borders can be obtained by comparing the predicted classification with the true classification. Importantly, none of the methods considered here directly maximises the number of correct predictions (hits): OLS is minimising the squared distance between the predicted and true bands while logistic regressions are maximising the likelihood of finding the true bands. Both methods put emphasis on further out observations that are not well predicted, but OLS to a larger extent. This indicates that beyond hits, the share of large deviations or the standard error of the difference between predicted bands should be inspected. Also, some degree of symmetry in this difference could be wanted. Throughout, there is the risk of overfitting the models to the data. Cross-validation could be used to assess whether this is indeed an issue.

[14] The no-jump condition is very strong as borders with different slopes will always cross. While the logistic regression does not ensure that the band with maximum probability is changing in a gradual way, the changes in the underlying probabilities are continuous. Hence, when instead of the band with the maximum probability, the one associated with the 50% cumulative probability is chosen, there would not be any jumps. This approach made precise in Appendix A.

5 Estimations

5.1 The border between speculative and investment grade

The most important border is probably the one between speculative and investment grade, so between Ba- and Baa-bands. There are two conceptually different ways of understanding this border: it can be considered in a local sense as distinguishing between Ba- and Baa-bands or in a global sense as distinguishing between all speculative grade bands on the one hand and all investment grade bands on the other. In the local case, the regression would either include the Ba- and Baa-bands. In the global case, all bands are included with all speculative grade ratings associated with 0 and all investment grade ratings with 1. This estimation assumes that observations further away from the border still provide useful information about its location and, for quadratic estimations, its shape. Both approaches are followed here (see the local and global logit regressions in Table 9).

All coefficients of the linear estimations (1)-(4) have the expected sign, except for interest which has a positive relation with rating bands. Again with the exception of interest, all variables are also highly statistically significant. Regressions with the centred control variables are shown in (1) and (3) and those with the regressed control variables in (2) and (4). A comparison between the coefficients of the linear centred and regressed regressions (so between regressions (1) and (2), or between (3) and (4)) confirms their equivalence. An obvious potential improvement would be to include forward looking variables. Appendix B provides evidence that indeed the assessment can be improved somewhat, mostly by including future real GDP growth. A full analysis would have to be done after model selection and is considered outside the scope of this study.

Table 9: Regressions for the border between speculative and investment grade

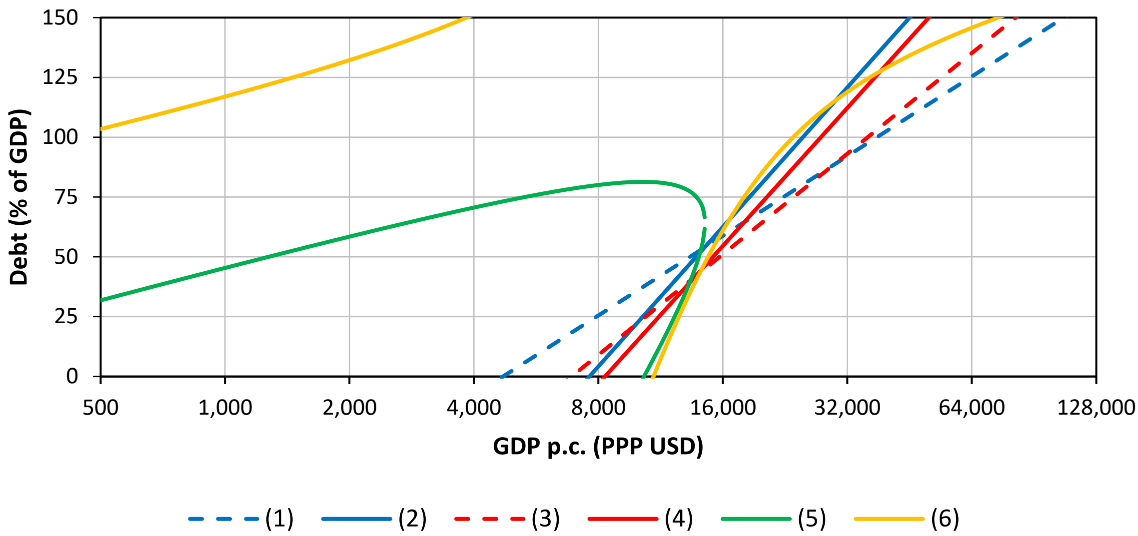
	(1) Local logit centred	(2) Local logit regressed	(3) Global logit centred	(4) Global logit regressed	(5) Local logit regressed	(6) Global logit regressed
Intercept	-0.025	-0.036	-0.363 ***	-0.381 ***	0.013	-0.246 *
	(0.081)	(0.080)	(0.073)	(0.073)	(0.151)	(0.131)
Debt	-0.024 ***	-0.026 ***	-0.036 ***	-0.042 ***	-0.008 *	-0.027 ***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)
Debt^2					0.000 ***	0.000 *
					(0.000)	(0.000)
Interest	0.008	0.008	0.008	0.008	0.201 ***	0.177 ***
	(0.013)	(0.013)	(0.011)	(0.011)	(0.031)	(0.025)
Interest^2					-0.025 ***	-0.021 ***
					(0.003)	(0.003)
Primary balance	-0.056 **	-0.056 **	-0.064 ***	-0.064 ***	-0.038	-0.061 **
	(0.024)	(0.024)	(0.020)	(0.020)	(0.04)	(0.029)
Primary balance^2					0.004	-0.006 *
					(0.004)	(0.003)
Growth	0.193 ***	0.193 ***	0.262 ***	0.262 ***	0.187 ***	0.249 ***
	(0.034)	(0.034)	(0.029)	(0.029)	(0.04)	(0.032)
CA	0.040 ***	0.040 ***	0.059 ***	0.059 ***	0.054 ***	0.053 ***
	(0.013)	(0.013)	(0.011)	(0.011)	(0.018)	(0.014)
Free-floating	1.998 ***	1.998 ***	3.039 ***	3.039 ***	2.504 ***	3.490 ***
	(0.313)	(0.313)	(0.272)	(0.272)	(0.403)	(0.332)
Reserves	0.159 ***	0.159 ***	0.230 ***	0.230 ***	0.200 ***	0.287 ***
	(0.021)	(0.021)	(0.019)	(0.019)	(0.029)	(0.022)
Governance	1.323 ***	1.323 ***	1.373 ***	1.373 ***	1.608 ***	1.531 ***
	(0.192)	(0.192)	(0.116)	(0.116)	(0.24)	(0.146)
ln(GDP p.c.)	1.158 ***	2.209 ***	2.165 ***	3.459 ***	2.351 ***	3.924 ***

	(0.164)	(0.159)	(0.131)	(0.134)	(0.197)	(0.166)
ln(GDP p.c.)^2					1.074 ***	0.643 ***
					(0.26)	(0.143)
Debt x Interest					-0.001	-0.001 **
					(0.001)	(0.001)
Debt x Primary b.					0.002 *	0.000
					(0.001)	(0.001)
Debt x ln(GDP p.c.)					-0.061 ***	-0.043 ***
					(0.008)	(0.005)
Interest x Primary b.					0.000	0.002
					(0.008)	(0.006)
Interest x ln (GDP p.c.)					-0.109 ***	-0.183 ***
					(0.032)	(0.026)
Primary b. x ln (GDP p.c.)					0.064	0.030
					(0.063)	(0.04)
Observations	1216	1216	3215	3215	1216	3215
AIC	1394.3	1394.3	1797.8	1797.8	1215.9	1567.3
DF	1206	1206	3205	3205	1196	3195
Adjacent bands						
Observations	1216	1216	1216	1216	1216	1216
% correct	72.2%	72.2%	70.6%	70.6%	76.5%	75.7%
% higher band correct	79.9%	79.9%	72.8%	72.8%	80.6%	76.2%
% lower band correct	63.4%	63.4%	68.1%	68.1%	71.8%	75.2%
All bands						
Observations	3215	3215	3215	3215	3215	3215
% correct	87.2%	87.2%	87.0%	87.0%	85.9%	89.8%
% higher bands correct	90.5%	90.5%	88.2%	88.2%	92.0%	90.8%
% lower bands correct	83.5%	83.5%	85.7%	85.7%	79.1%	88.7%

Note: ***, ** and * indicate statistical significance at the 1%-, 5%- and 10%-level respectively. Standard deviations are derived without accounting for potential dependence across time periods or CRAs.

Not surprisingly, the local regression performs somewhat better in terms of hits when only considering the adjacent bands (72.2% versus 70.6%). However, when considering all bands it still performs marginally better (87.2% versus 87.0%), indicating that the information contained by the non-adjacent bands is not very relevant when only considering the number of hits. The borders obtained by local and global logit are rather similar (see Figure 4). The border for the estimations with regressed variables is steeper than for the centred variables. This reflects that higher debt goes hand-in-hand with a deterioration in control variables, which is ignored by the centred regression as the level of the control variables is kept at their average.

Figure 4: The border between speculative and investment grade based on GDP per capita and debt, based on the regressions of Table 9



When considering the quadratic estimations (5) and (6) and all bands, the global logit clearly outperforms the local logit, which even performs worse that the linear estimation. The reason for the bad performance of quadratic local logit is the overfitting of the bound (see Figure 4): due to the high curvature, the border passes right through an area with numerous observations from other bands. Clearly, for quadratic estimations, the wider range of shapes[15] for the bounds should be balanced by more information. In this case, the more remote observations in the lower left corner contain the important message that the border should not be passing through that area – but this message is not used by the local logit. Note that the border of the global logit in the left upper corner does not affect the performance in terms of hits as there are no observations in that area. For quadratic logit, the Akaike criterion indicates that the additional parameters improve the model compared to the linear regression. Importantly, the borders are sensitive enough to the variables to predict different bands for countries over the sample period: of the 60 countries with ratings in the Ba- and Baa-bands, both local and global logit predict different bands over time for 18 countries, while in reality 21 experienced a change.

Additional insight in the methodology can be obtained by comparing the regression on only debt and ln(GDP p.c.) and the regression with additional variables, and looking at the different implications for the border and the positioning of observations of the adjacent bands (see Figure 5 and Figure 6 respectively). Without additional variables, the border is crossing a dense cloud of observations, with multiple observations on the wrong side and the slope not directly obvious. When also including control variables and considering the adjusted debt (see section 4.3), the cloud is pulled apart, and even visually it is clear that the border is doing a better job in separating the two bands.

[15] As discussed in Section 4.2, the possible shapes are all cross-sections of cones, reflecting the quadratic dependence on both debt and ln(GDP p.c.).

Figure 5: The border between speculative and investment grade based on only GDP per capita and debt

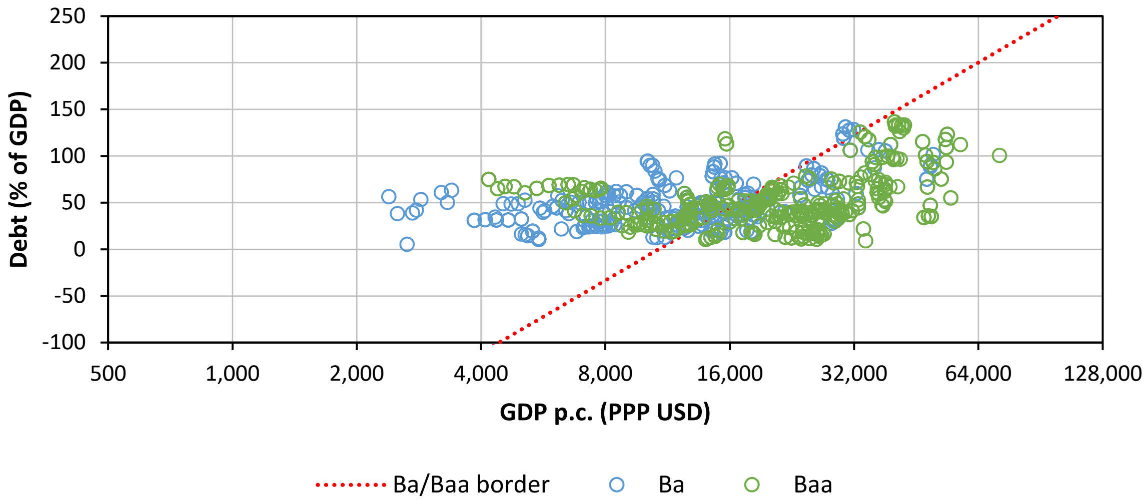
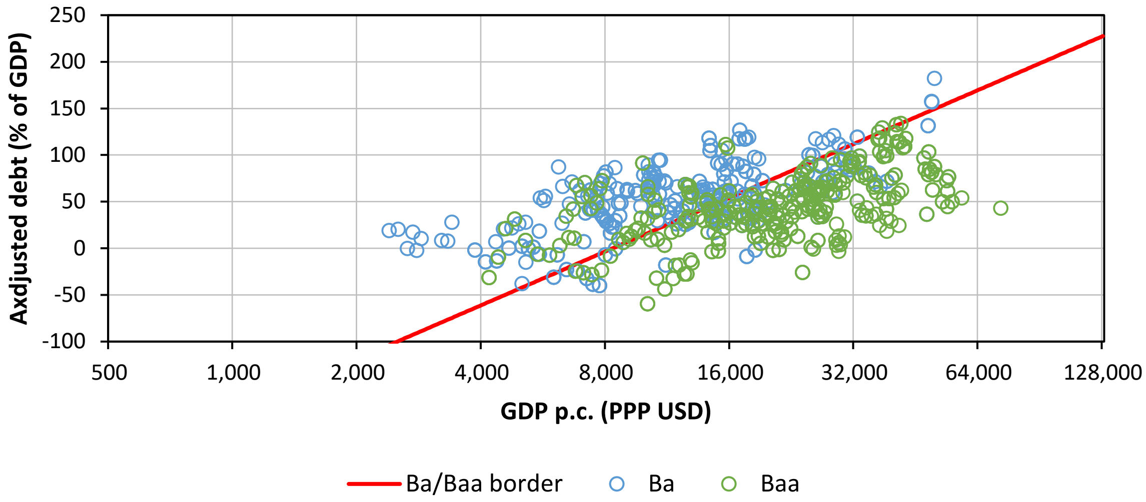
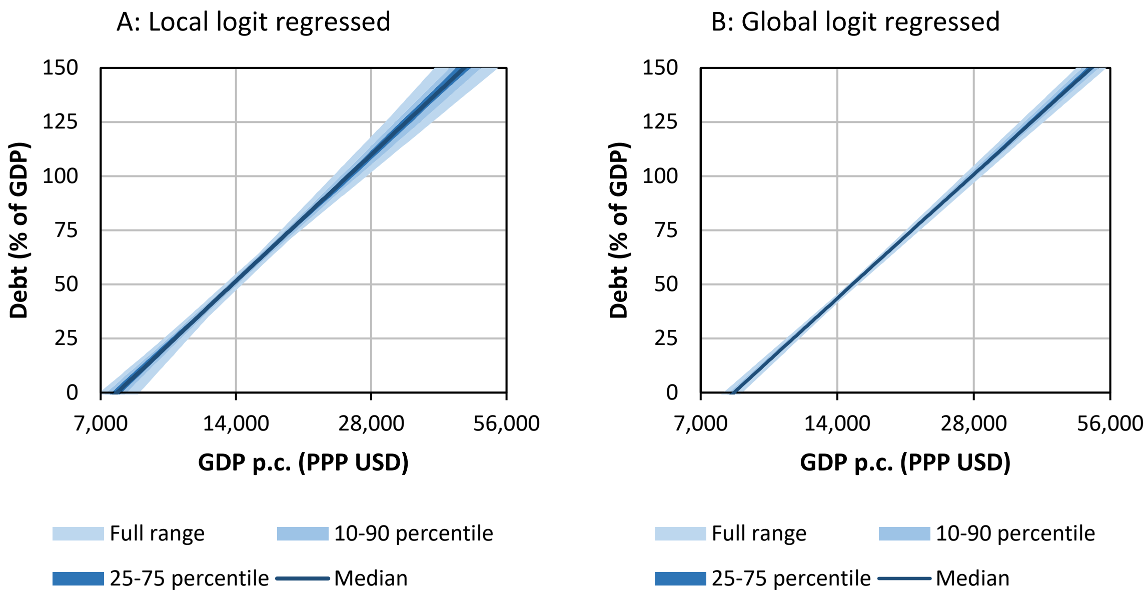


Figure 6: The border between speculative and investment grade based on Equation (4)



Global logit yields a border with a higher degree of stability than local logit (see Figure 7). Clearly, the higher number of observations results in better estimates of the parameters – especially as many observations in other bands are still close to the border.

Figure 7: The variation in the border based on subsampling (Equations (2) and (4); 100 runs; each 80% of sample)



The border differentiates between conditions leading to higher and lower ratings, but it might not be relevant in its entirety as a boundary between its adjacent bands. For example, it could be the case that for very high debt levels, it crosses another border, which would imply that only for lower debt levels the border is a real boundary, as for higher values, one of the adjacent bands is no longer assigned (skipped).

5.2 Credit ratings and debt

A complete understanding of the relation between debt, GDP per capita and ratings can be obtained by estimating all borders. Starting from OLS, which imposes the most structure, results are shown for ordered logit, which allows for different widths of the bands, sequential global and local logit, which estimate all borders individually, and finally multinomial logit, which imposes no structure (Table 10). While OLS and ordered logit have a single underlying regression, the logit regressions have six (see Table 11 for those of sequential local logit).

Table 10: Performance of the linear regressions

	OLS	Ordered logit	Sequential global logit	Sequential local logit	Multinomial logit
Hits	51.9%	55.1%	62.9%	64.9%	64.0%
Hits +/- 1 band	93.9%	93.2%	94.4%	92.4%	94.0%
>= 3 bands away	1.0%	1.0%	0.7%	1.8%	1.0%
Too high	23.8%	21.5%	19.8%	19.0%	20.6%
Too low	24.3%	23.4%	17.4%	16.1%	15.4%
Standard error	0.85	0.84	0.76	0.83	0.78
Hits by band					
C	42.6%	26.9%	33.3%	29.6%	29.6%
B	55.6%	74.8%	77.5%	82.7%	84.0%
Ba	56.9%	41.9%	41.5%	32.9%	31.0%
Baa	58.8%	58.2%	58.3%	66.2%	65.1%
A	56.1%	47.4%	58.2%	57.7%	54.8%
Aa	37.3%	31.7%	62.1%	60.8%	54.2%
Aaa	34.1%	61.4%	84.6%	92.4%	94.9%
Observations	3215	3215	3215	3215	3215

Note: The underlying regressions include a constant, debt, interest, primary balance, growth, current account, free-floating dummy, reserves, governance and ln(GDP p.c.).

Table 11: Sequential local logit regression (regressed control variables)

	C/B border		B/Ba border		Ba/Baa border		Baa/A border		A/Aa border		Aa/Aaa border	
Intercept	4.46	***	1.2	***	-0.036		-3.508	***	-3.635	***	-21.789	***
	(0.351)		(0.121)		(0.08)		(0.253)		(0.374)		(3.07)	
Debt	-0.043	***	-0.048	***	-0.026	***	-0.054	***	-0.039	***	0.070	***
	(0.006)		(0.004)		(0.003)		(0.005)		(0.006)		(0.017)	
Interest	0.042	**	-0.022	**	0.008		0.042		-0.655	***	0.600	***
	(0.016)		(0.011)		(0.013)		(0.034)		(0.067)		(0.114)	
Primary balance	0.003		-0.031		-0.056	**	-0.145	***	0.078	***	-0.397	***
	(0.028)		(0.022)		(0.024)		(0.03)		(0.028)		(0.073)	
Growth	0.323	***	0.129	***	0.193	***	0.444	***	-0.029		-0.215	**
	(0.045)		(0.032)		(0.034)		(0.057)		(0.05)		(0.09)	
CA	-0.004		0.066	***	0.04	***	0.087	***	-0.048	**	0.282	***
	(0.016)		(0.013)		(0.013)		(0.019)		(0.022)		(0.051)	
Free-floating	2.429	**	1.571	***	1.998	***	3.967	***	-1.172	***	1.048	
	(1.064)		(0.513)		(0.313)		(0.355)		(0.323)		(1.607)	
Reserves	0.281	***	0.209	***	0.159	***	0.197	***	0.007		0.118	
	(0.06)		(0.027)		(0.021)		(0.028)		(0.018)		(0.301)	
Governance	1.23	***	1.863	***	1.323	***	1.351	***	-0.079		8.468	***
	(0.282)		(0.19)		(0.192)		(0.205)		(0.142)		(1.044)	
ln(GDP p.c.)	1.075	***	2.913	***	2.209	***	3.773	***	2.638	***	14.026	***
	(0.226)		(0.162)		(0.159)		(0.315)		(0.289)		(1.983)	
Observations	945		1405		1216		1026		684		676	
AIC	480.1		1274.8		1394.3		789.5		662.4		354.0	
DF	935		1395		1206		1016		674		666	
Adjacent bands												
Observations	945		1405		1216		1026		684		676	
% correct	90.7%		78.2%		72.2%		83.0%		80.4%		91.3%	
% higher band correct	98.6%		69.5%		79.9%		75.1%		75.2%		94.9%	
% lower band correct	29.6%		84.1%		63.4%		87.7%		84.7%		86.9%	
All bands												
Observations	3215		3215		3215		3215		3215		3215	
% correct	97.3%		87.9%		87.2%		92.3%		89.1%		96.5%	
% higher bands correct	99.6%		89.3%		90.5%		85.4%		77.7%		94.9%	
% lower bands correct	29.6%		84.6%		83.5%		95.7%		92.2%		96.7%	

Note: ***, ** and * indicate statistical significance at the 1%-, 5%- and 10%-level respectively. Standard deviations are derived without accounting for potential dependence across time periods or CRAs.

A first assessment of the estimation methods would focus on the share of hits. It is insightful to consider that without any analysis, it is possible to reach a share of 26.0% by always choosing the band with the largest share (B). Also, when only using GDP per capita (the variable with the highest correlation with ratings), and setting six thresholds (one-dimensional borders), a share of 49.4% can be reached.[16] Finally, for a country to be allocated correctly, all borders, but at least its direct borders should assign it correctly, implying that based on the sequential local logit estimates (see Table 11), 90% x 90% = 81% is a rough estimate for the maximum attainable share.

[16] When using debt to differentiate between countries' creditworthiness, not all bands are assigned as the six thresholds do not monotonically increase due to the low correlation with ratings.

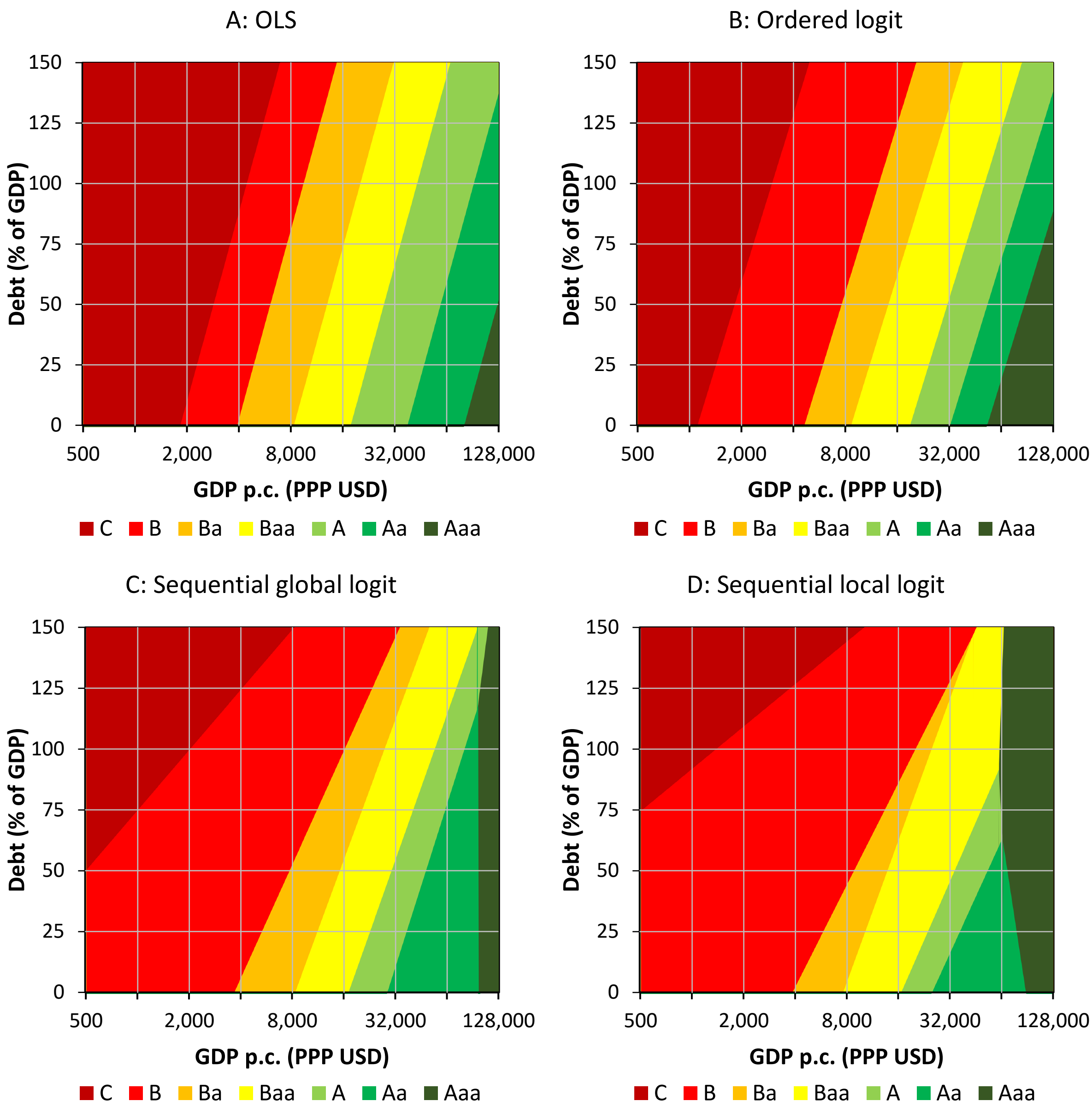
With 51.9% hits, OLS is the worst performer of the five models and only marginally better than using only six thresholds on GDP per capita. Ordered logit still has a single set of coefficients, but the cut-off points allow the distance between borders to vary, and the performance rises to 55.1%. This is a first confirmation of the non-linear relationship between debt and ratings. Allowing the coefficients for each border to vary, yields shares of around 63-65%. While sequential global logit has the lowest performance among these three regressions, the additional structure imposed reduces the number of predictions that are further off, as evidenced by its high share of predictions within 1 band from the true value, the lowest share of predictions 3 or more bands off, and the lowest standard error.

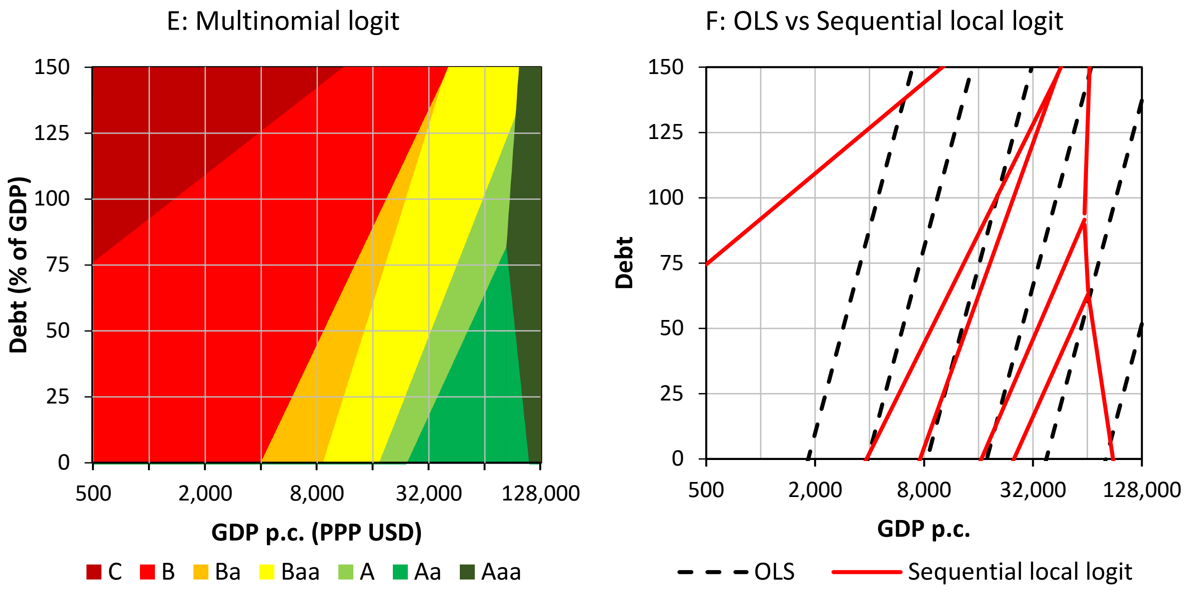
The sequential logistic regressions, and in particular the multinomial regression, are skewed towards assigning too high bands. They also perform poorly on the C-band, predicting only around 30% of the observations correctly. This probably reflects the low number of observations in this band (108 compared to 837 in the adjacent B-band), and this is analysed in more detail in Section 5.4. On the other hand, the correct share in the Aaa-band is very high, and especially for multinomial logit seems to come at the cost of a worse performance on the Aa-band. Indeed, the coefficient of debt has a positive sign in the sequential local logit regression (see Table 11), indicating that it might suffer from overfitting. OLS is performing relatively badly on the exterior bands (C and Aaa) compared to the interior bands. While sequential local logit performs best overall, its performance is less uniform over the bands than sequential global logit.

The regressions imply different borders and hence regions associated with each of the bands (see Figure 8, panels A-E). As the borders are affine subspaces with more than two dimensions, a particular slice needs to be selected for a graphical representation. As discussed in Section 4.3, the selection assumes that the control variables assume their typical values, found by regressing on debt and ln(GDP p.c.). Clearly, for countries that have better economic fundamentals, the borders will be higher in terms of debt, resulting in higher ratings. Importantly, the shown borders follow directly from the regression results, unless they are not between adjacent bands. In that case, the mathematical representation of a border can be found by summing the coefficients of all intermediate borders for each explaining variable and equating the obtained formula to zero.

For the OLS regression, the interior bands are equal-sized, while the ordered logit drops this restriction. For the two sequential logits and the multinomial logit the slope of the various borders can also be different as well. These three regressions yield comparable results though, with similar implications, leading to the following observations. The C-ratings are exceptional in that they require high levels of debt. The B-ratings form a vast region, especially compared to the Ba-region. The large area can simply reflect that the associated probability of default (PD) has a large basin. However, it is also consistent with a conservative approach, in the sense that CRAs are cautious in assigning Ba-ratings for at least a subset of countries. In this case, the observed PD for B-rated countries would either be below the expected value, or for a group of countries the observed PD would be significantly lower (*e.g.* for those with improving as opposed to deteriorating fundamentals, or for those from a particular region).

Figure 8: The relation between debt, GDP per capita and rating bands – linear regressions (regressed control variables)



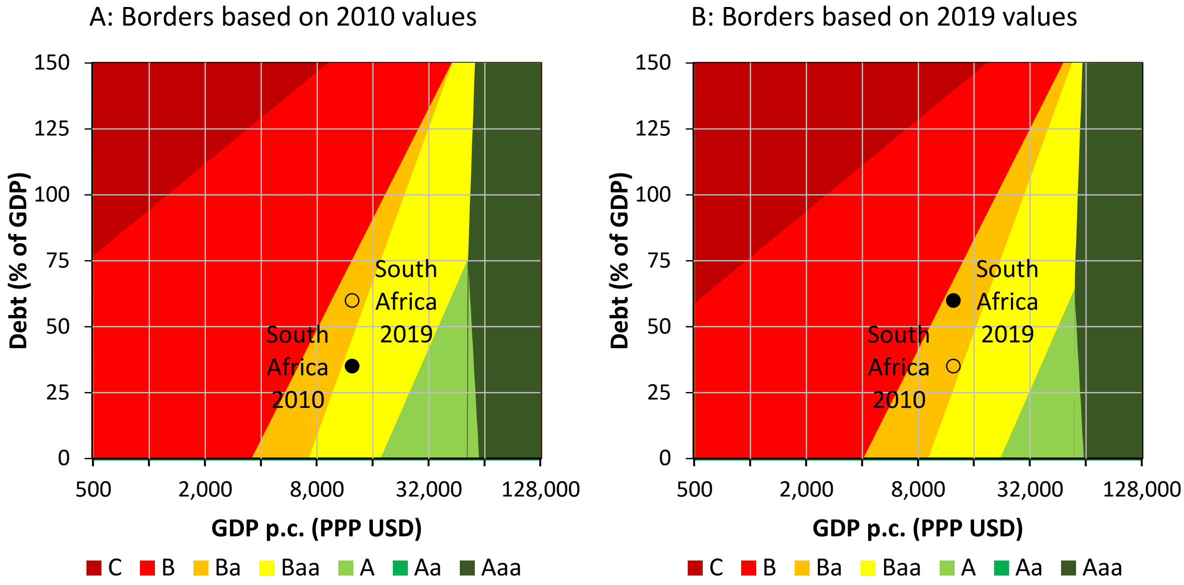


The Baa-region is also rather broad, signalling that once investment-grade status has been accomplished, debt has to fall substantially and/or GDP per capita has to rise significantly before the A-region is reached. After this relatively small region, the broad Aa-region forms a real hurdle before reaching the Aaa-region. In any case, the other economic conditions should be very strong to enter the Aaa-region, as with the typical values the required GDP per capita is almost forbiddingly high.

Importantly, the figures also show that from moderate GDP per capita levels onwards and for higher debt levels, the areas are closer to each other, implying that in particular A- and Ba-ratings can change relatively easily. Apart from the Aaa-region, the classification complies with the expectation that more debt and a lower GDP per capita lead to lower ratings, and that there is no jumping over bands for relevant values of debt and GDP per capita. Overall, the found classification is very different from the one based on an OLS regression (see Figure 8, panel F).

The impact of the control variables varies across the borders, and hence each border has its own definition of adjusted debt. It is thus not possible to gauge the performance of an estimation method by showing the observations in a figure using adjusted debt instead of debt. However, it is possible to depict the borders for a specific country and year, for example South Africa (see Figure 9; note that the backwards leaning Aaa-region overshadows the Aa-region). In 2010, South Africa was rated in the Baa-band by Fitch and S&P, but in 2019 it was rated in the Ba-band by both (the Moody's rating was one band higher in both years). The sequential local logit estimation confirms the deterioration over time and the change in bands from Ba to Baa. However, it also shows that the deterioration is not solely due to the increase in debt, but also due to changes in the other economic variables as reflected by the change in the Ba/Baa-borders: even with the 2010 debt level, South Africa would have been in the Ba-region in 2019. Similarly, due to the change in the other economic variables, South Africa is now much closer to the B-region than otherwise. Appendix A looks closer at the evolution of South Africa's creditworthiness over time and shows how the probabilistic nature of the logit regressions can be utilised relatively easily to make assessments that are more granular than the bands.

Figure 9: Impact of variables on borders: the case of South Africa (sequential local logit; regressed control variables)



5.3 Exploring the role of public finance

In the linear model, GDP per capita and the public finance variables (debt, interest and primary balance) raise the share of hits by some 25 percentage points for the sequential logits and multinomial logit, but only by 10 percentage points for OLS (namely, the difference between 26.0% of only assigning B's and the performance of the respective models shown in Appendix C). Their contribution reaches some 30-35 percentage points (20 percentage points for OLS) by including quadratic terms which allows for richer forms of the borders.

The performance can be further boosted by re-introducing the control variables (see Table 12). The quadratic terms help OLS and ordered logit, but the imposed structure is still too rigid to benefit fully. The sequential global logits and the multinomial logit see their performance increase to some 69%. However, the sequential local logit sees a drop in its performance compared to the linear specification. Clearly, the quadratic terms result in considerable overfitting of the individual borders (as seen in Section 5.1) which prevents a good overall performance. Sequential global logit and multinomial logit benefit from including all observations in the estimations for all borders. The richer specification somewhat increases the symmetry, and the sequential global logit still has the fewest far-off predictions (its detailed regression results are provided in Appendix D).

Table 12: Performance of the quadratic regressions

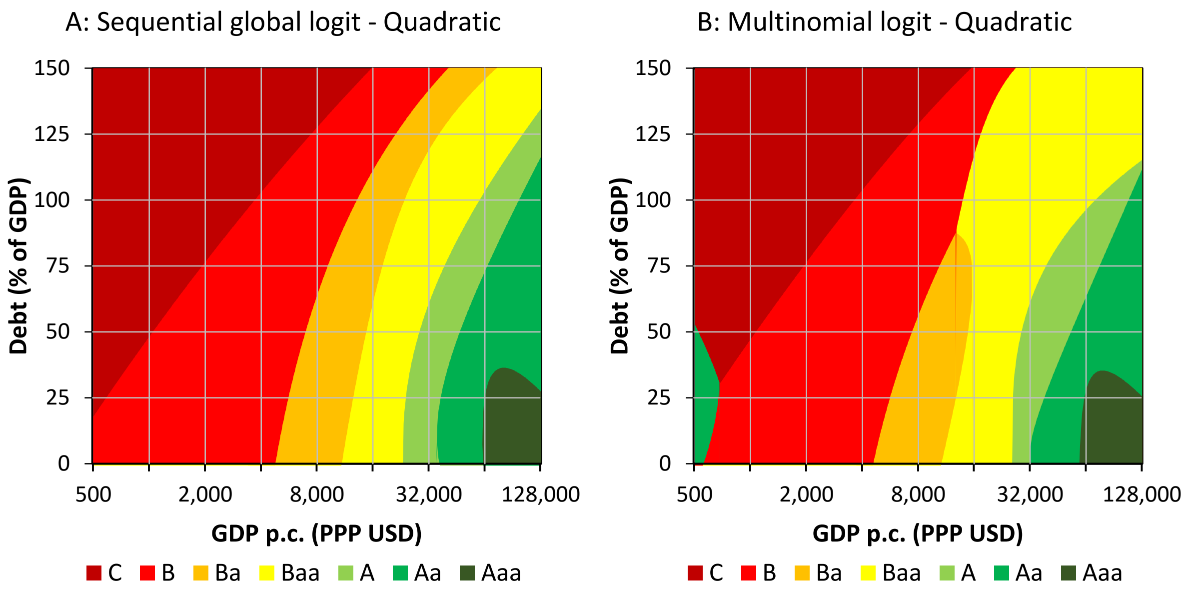
	OLS	Ordered logit	Sequential global logit	Sequential local logit	Multinomial logit
Hits	55.3%	59.2%	68.6%	61.7%	69.5%
Hits +/- 1 band	95.1%	94.3%	96.5%	85.7%	96.0%
>= 3 bands away	0.5%	0.9%	0.2%	8.5%	0.4%
Too high	21.0%	20.3%	16.5%	23.6%	16.3%
Too low	23.7%	20.5%	14.9%	14.7%	14.2%
Standard error	0.79	0.79	0.65	1.36	0.67
Hits by band					
C	37.0%	33.3%	39.8%	26.9%	41.7%
B	61.8%	75.6%	79.6%	60.2%	81.2%
Ba	61.4%	46.0%	54.0%	45.2%	46.8%

Baa	55.9%	55.9%	64.2%	63.1%	68.7%
A	56.6%	51.9%	65.9%	70.1%	69.3%
Aa	47.7%	48.4%	61.8%	61.1%	61.4%
Aaa	40.5%	72.4%	90.3%	90.0%	94.1%
Observations	3215	3215	3215	3215	3215

Note: The underlying regressions include a constant, interest, primary balance, growth, current account, free-floating dummy, reserves, ln(GDP p.c.), quadratic terms of the three public finance variables as well as ln(GDP p.c.), and their interaction terms.

The borders of the quadratic regressions provide a slightly different allocation of the region (see Figure 10). Most notably, the Aaa-region is starting at lower GDP per capita levels, and the Baa-region has tilted and become more vertical, implying higher GDP per capita levels for low debt. Also, the C-region is larger as it begins at lower debt levels. There are some signs of overfitting for multinomial logit as part of the A-region is located in the lower-left corner (in which there are no observations). Also, the Ba region stops at a debt level of around 85% of GDP, leading the B- and Baa-regions to become direct neighbours, making the sequential global logit the preferred method.

Figure 10: The relation between debt, GDP per capita and rating bands – quadratic regressions (regressed control variables)



5.4 Analysing robustness

5.4.1 Assessing the degree of overfitting

To assess the degree of overfitting in both the linear and quadratic models, cross-validation is performed with ten folds[17] (Table 13). For the linear models, the performance from the cross-validation is just slightly lower than the one on the sample, with sequential local logit and multinomial logit showing the largest effect. For the quadratic models, the difference is somewhat larger, but without changing the overall ranking for the hits while indicating that sequential global logit and the multinomial logit perform comparably well when adjacent bands are included.

[17] So, the sample is divided in ten folds, the estimation is done on nine of these folds, and the performance is assessed on the fold taken out. This process is repeated for the other folds, after which the average is taken of the ten obtained performance statistics.

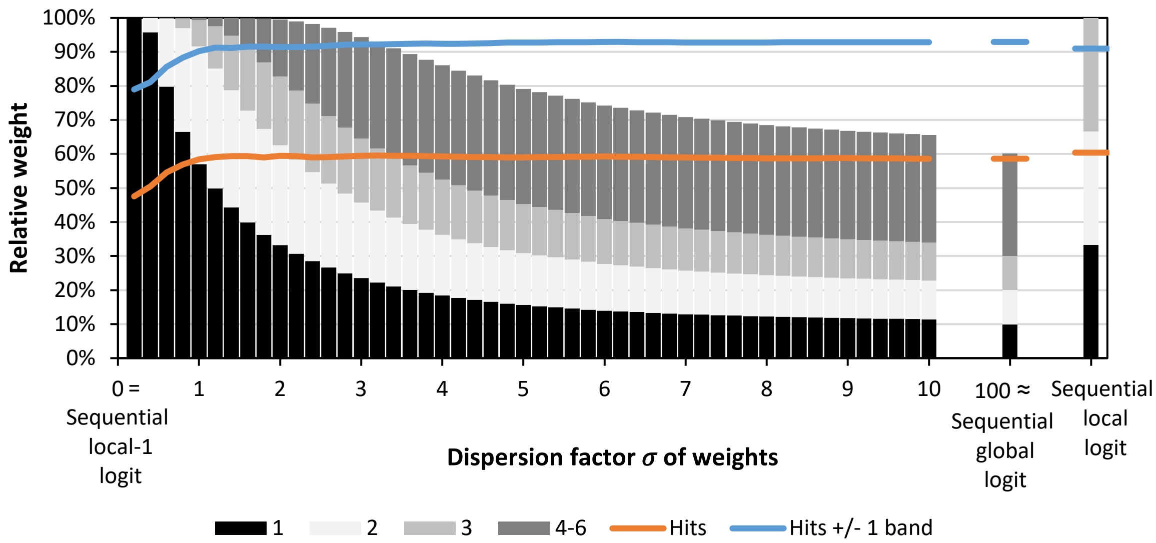
Table 13: Performance analysis of various models

		OLS	Ordered logit	Sequential global logit	Sequential local logit	Multinomial logit
Linear						
Hits	Estimation	51.9%	55.1%	62.9%	64.9%	64.0%
	Cross-validation	51.5%	55.1%	62.6%	63.9%	63.0%
Hits +/- 1 band	Estimation	93.9%	93.2%	94.4%	92.4%	94.0%
	Cross-validation	93.7%	93.2%	94.0%	92.6%	94.0%
Quadratic						
Hits	Estimation	55.3%	59.2%	68.6%	61.7%	69.5%
	Cross-validation	54.4%	58.8%	65.8%	60.4%	67.7%
Hits +/- 1 band	Estimation	95.1%	94.3%	96.5%	85.7%	96.0%
	Cross-validation	94.9%	94.2%	96.1%	85.3%	95.9%

5.4.2 Generalising the choice of included observations

Sequential local logit only includes the ratings in the neighbouring bands (usually three), while sequential global logit ratings considers all ratings equally. However, the logit regression can also be conducted with weights that are lower for observations further from the border. Specifically, for a rating s below the border between rating bands r and $r{+}1$, the weight is set as $\exp\left(-\frac{1}{2}(s\text{-}r)^2/\sigma^2\right)$ and for a rating s above the border it is set as $exp\left(-\frac{1}{2}(s\text{-}r\text{-}1)^2/\sigma^2\right)$ for a parameter σ that indicates the dispersion of the weights. When σ is close to zero, only the ratings directly neighbouring the border are relevant (see also below where this is referred to as sequential local-1 logit), while for high σ this approach coincides with sequential global logit (see Figure 11).

Figure 11: The weights of ratings and the impact on the performance in terms of (near) hits



Note: The relative weights are shown as function of the dispersion factor σ which is changing by steps of 0.2. The weights are shown for ratings at various distances from the border. For instance, when considering the Ba/Baa-border, “1” refers to the Ba1 and Baa3 ratings, “2” to the Ba2 and Baa2 ratings, “3” to the Ba3 and Baa1 ratings and “4-6” to the A and B rating bands. The relative weights will be slightly different for borders close to the extremes of the rating scale. “Hits” refers to the share of correctly predicted rating bands and “Hits +/- 1 band” to the share of predictions that are at most one band away from the true band.

When focusing almost exclusively on the neighbouring rating (σ close to zero), the model is performing rather badly as the noise in this small sample is high. This shows that a hyper-local interpretation of, *e.g.*, the Ba/Baa-border as the border between the Ba1- and Baa3-ratings is too narrow for estimations. When σ is increased, weight moves away from the neighbouring ratings to ratings a bit more distant which are still informative and thus improve performance. However, there is a trade-off as at a certain point the weight on less-informative distant ratings increases. Most benefits of including ratings further away from the borders are already obtained when the relative weight of the neighbouring ratings falls below 50% (σ = 1.2) after which performance is rather stable before declining marginally. At the point where most benefits are achieved (σ = 1.2), the combined weight of the three ratings next to the border is 97.6%, indicating that the three closest ratings contain basically all essential information. This confirms that sequential local logit, where these three ratings account for 100% of the weight, is not only intuitively appealing, but also focusses on the minimal set of relevant ratings. On the other hand, including all ratings with equal importance, as sequential global logit does, only marginal lowers the performance, but has the importance advantage of being parameter free.

5.4.3 Estimating ratings to analyse the reliance on the small number of dependent classes

The robustness of the models is further assessed by looking at individual ratings instead of bands. *Ex ante*, one would expect that with three times as many classes to assign, the performance in terms of hits would be a third or somewhat higher. The very few observations of countries with a low rating and a free-floating currency causes problems with the estimations of individual borders at the lower end of the rating scale. Hence, an analysis at such a local level is no longer able to identify or utilise economic variables that are only relevant for observations in a particular part of the rating scale. To ensure comparison across methods, both the free-floating dummy and the reserves are thus dropped.

Again, ordered logit is performing better than OLS, and the difference has widened, while global logit performs better still (Table 14). For sequential local logit, one can as before consider the three[18] ratings above and below the border (sequential local-3 logit), or run the estimations on only the ratings directly adjacent to the border (sequential local-1 logit). Having (up to) three ratings on each side of the border uses for each border fewer but more relevant observations than global logit and leads to a better performance than global logit in terms of hits. Sequential local-1 logit specifically aims to differentiate between ratings adjacent to the borders, and performs somewhat better still.[19] However, as we will see below, this method does not impose sufficient structure to obtain an economically relevant set of borders.[20] Strikingly, multinomial logit which benefits from even less structure, performs very badly in terms of hits, and is unable to outperform a random allocation over the 17 ratings. Apparently, its good performance relied crucially on the small number of bands – which is not the case for the sequential logits.[21] As before, the sequential logits and the multinomial model are skewed towards too high ratings. Closer inspection shows that this is entirely caused by large deviations, which almost necessarily are downwards. For example, several advanced countries with rather robust economic fundamentals, *e.g.* Cyprus, Iceland, Ireland and Slovenia, had a deviation of 10 notches or more during their (near) default episodes in the early 2010s. When assessed on an interval of +/- 2 ratings, the sequential logits models are broadly symmetric.

[18] Only four or five ratings can be included in the estimation of the C/B3-, B3/B2-, Aa2/Aa1- and Aa1/Aaa-borders.

[19] The better performance than reported in Figure 11 can be explained by the fact that now all 16 borders are estimated, instead of only six. Hence, the conclusion is that all observations are necessary: either through the estimation of all border or through the estimation of fewer borders but with aggregating into bands.

[20] The sequential local-1 logit uses fewer observations for each estimation. Hence, it is sometimes possible to perfectly or almost perfectly separate them. In this case, the parameters will increase in size, but this is not a problem for calculating the relative probabilities as only those bands on the "right" side of the borders have probabilities that are not close to 0. However, the border might not be optimal for observations of other ratings.

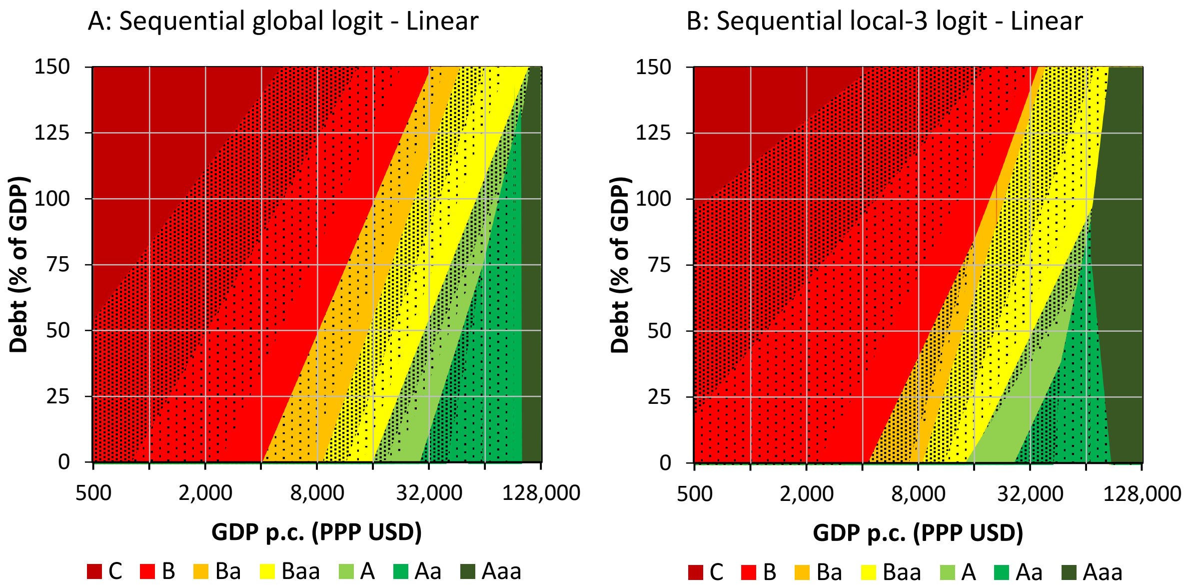
[21] Performance under cross-validation is comparable for OLS, ordered logit and sequential local-1 logit, with a minor deterioration for sequential global and local-3 logit (0.6 and 0.8 percentage points respectively in terms of hits).

Table 14: Performance of linear regressions of ratings

	OLS	Ordered logit	Sequential global logit	Sequential local-3 logit	Sequential local-1 logit	Multinomial logit
Hits	19.3%	28.0%	31.1%	35.0%	36.9%	4.4%
Hits +/- 1 rating	50.6%	57.0%	64.0%	66.2%	64.0%	8.3%
Hits +/- 2 ratings	73.3%	77.4%	80.3%	81.1%	78.7%	11.3%
Too high	38.5%	36.7%	39.8%	34.2%	32.9%	56.1%
Too low	42.2%	35.2%	29.1%	30.8%	30.2%	39.4%

The sequential global logit and the sequential local-3 logit imply very comparable borders and regions up to debt of some 100% of GDP (see Figure 12). The largest differences are for the C- and Aaa-ratings. Of course, these borders are identical to the ones found before (see Figure 8) as are the other borders that coincide with those of the bands (unless they cross any of the new borders). So, for these sequential logit estimations, the borders between individual ratings essentially carve up the regions found earlier for the rating bands. The sequential local-1 logit regression derives neighbouring borders with an overlap in observations, which provides still some structure, but not enough to ensure an economically appealing ordering (for a low-density representation, see Figure 16 in Appendix E). The multinomial logit regression obtains all borders anew, and, crucially, considers the relative likelihood of all ratings when maximising the overall likelihood. Hence, the found borders are no longer benefitting from the little bit of structure it could utilise earlier, namely the observations of three ratings, which was enough to obtain a set of coherent borders (see Figure 16 in Appendix E). As a result, the sequential local-1 logit and the multinomial models are not useful in practice.

Figure 12: The relation between debt, GDP per capita and ratings (regressed control variables)



Note: Within each rating band, darker shades refer to lower ratings. For instance, the bright yellow area without dots is Baa1, the bright yellow area with sparse dots is Baa2 and the heavily dotted bright yellow area is Baa3. Some ratings are not assigned.

5.4.4 Analysing the C/B-border to gauge the impact of having few observations in a band

The C/B-border can be assessed in more detail by using the internal ratings of the European Investment Bank (EIB), which assesses the creditworthiness of all countries to which it (potentially) has exposure in a way comparable to the CRAs. Given its mandate to support development within and outside the European Union, its ratings cover a larger number of low-rated countries than those of the CRAs (see Table 15). After removing countries without debt data, 173 remain.

Table 15: Coverage of CRAs and the EIB by rating (2010-2019)

	CRAs	EIB
Aaa	11.1%	5.6%
Aa	10.1%	5.8%
A	12.0%	7.3%
Baa	19.4%	13.9%
Ba	17.4%	16.7%
B	26.3%	36.9%
C	3.7%	13.7%
Total	100%	100%
Countries covered	140	173
Observations	3378	1572

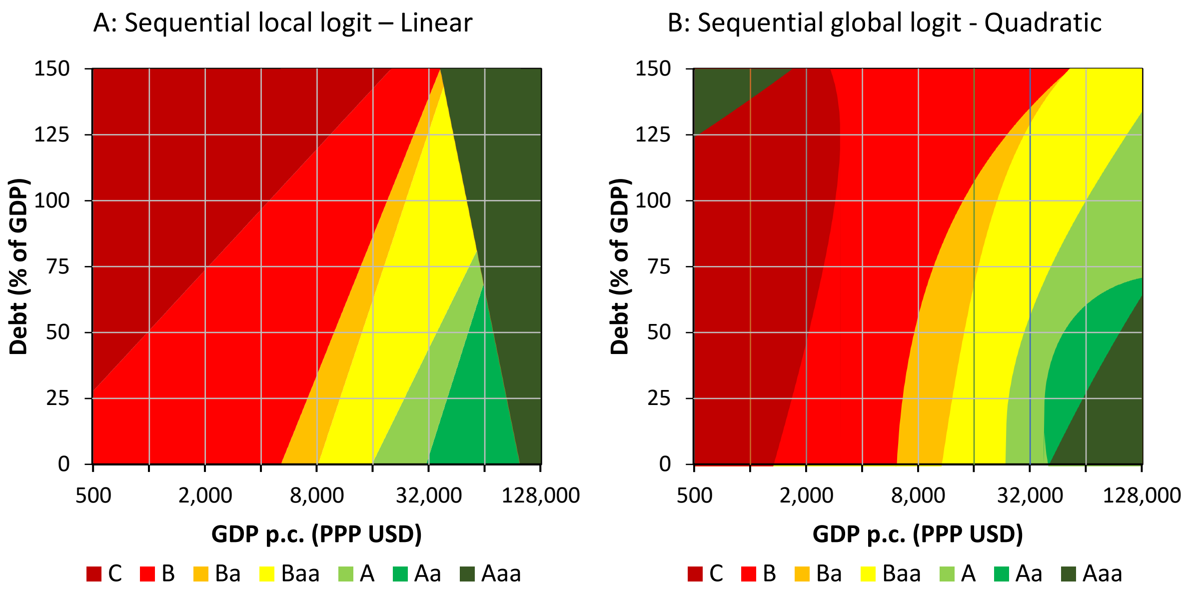
As before, sequential local logit outperforms sequential global logit for the linear specification (see Table 16), and the performance is somewhat better than for the sample of the CRAs. When including the quadratic and interaction terms of the public finance variables and ln(GDP p.c.), sequential global logit has the best performance, with sequential local logit performing rather disappointingly. This is again a direct result of the overfitting as there are now more observations in the region crossed by *e.g.* the Ba/Baa-border. For the three relevant models, the share of correct predictions in the C-band is now comparable to those of other bands, confirming that its poorer performance on the CRA sample is due to an unbalanced number of observations. Hence, the C-area is now covering lower debt levels for countries with low GDP per capita levels (see Figure 13). Conversely, sequential local logit assigns 100% of the Aaa-ratings correctly: having just 170 observations of Aa- and Aaa-observations is too few given the 10 or more parameters that have to be estimated. Indeed, the Aaa-area is quite different from those found before. Overall, these findings indicate that when the number of observations is very different for a band at the top or the bottom of the rating scale, additional efforts are needed to estimate its border correctly. To distinguish between the relevant cases at the lower end of the rating scale, it could be useful to consider economic indicators with a higher frequency (*e.g.* exchange rate developments, outcomes of domestic bond auctions, monthly tax revenues), those with a higher level of detail (*e.g.* maturity and currency composition of debt) and to qualitative factors (*e.g.* political statements and news paper articles).

Table 16: Performance for the EIB sample

	Sequential global logit		Sequential local logit	
	Linear	Quadratic	Linear	Quadratic
Hits	65.5%	67.9%	67.4%	50.5%
Hits +/- 1 band	96.2%	96.4%	94.4%	66.6%
>= 3 bands away	0.3%	0.4%	1.1%	26.4%
Too high	18.2%	18.1%	16.8%	40.7%
Too low	16.3%	14.0%	15.8%	8.8%
Standard error	0.70	0.68	0.75	2.18
Hits by ratings				
C	57.8%	61.8%	60.3%	39.7%
B	78.0%	79.2%	81.5%	48.0%
Ba	41.2%	45.5%	29.6%	24.5%
Baa	57.5%	65.3%	65.3%	53.0%
A	62.7%	71.8%	64.5%	70.0%
Aa	75.6%	46.3%	78.0%	82.9%
Aaa	87.2%	97.7%	100.0%	100.0%
Observations	1453	1453	1453	1453

Note: The underlying regressions include a constant, interest, primary balance, growth, current account, free-floating dummy, reserves and ln(GDP p.c.), and in case of the quadratic regressions quadratic terms of the three public finance variables as well as ln(GDP p.c.), and their interaction terms.

Figure 13: The relation between debt, GDP per capita and rating bands for the EIB sample (regressed control variables)



6 Conclusions

Overall, the relationship between debt and creditworthiness is found to be highly non-linear. Areas associated with rating bands differ vastly in size and shape. The band of C-ratings is associated with very low GDP per capita or very high debt levels. B-ratings span a vast region, especially compared to Ba-ratings. Once investment grade status has been achieved, traversing the broad Baa-band is a real hurdle before reaching A-ratings. The border between the Aa- and the Aaa-region starts at very high levels of GDP per capita, in effect requiring other economic fundamentals to be strong as well. Importantly, from moderate GDP per capita levels onwards and for higher debt levels, the areas are closer to each other, implying that especially A- and Ba-ratings can change relatively easily.

These findings show that any analysis of ratings needs to take non-linearities into account. OLS does not and hence exhibits the worst performance among the analysed models. Using ordered logit results in a marked improvement, especially when the analysis is done at the rating level. Even better results can be achieved by exploiting the structure of the ratings through sequential local logit and sequential global logit. The local variant performs well for the linear case, but as only a fraction of the observations is used for the estimation of each border, the borders could suffer from overfitting. For the same reason, this method cannot deal with quadratic terms. Sequential global logit uses all observations for each border and is more robust, which only for the linear case comes at a (small) cost in terms of performance. Its implied classification is visually attractive, with the borders having the correct slopes and no crossings within the economically relevant area. Both sequential logits have the appealing feature that they can be readily expanded to more classes without essentially affecting the already found borders by doing additional estimations. Multinomial logit on the other hand, only works when the number of classes is small, as it cannot capture the underlying structure otherwise. Hence, it could potentially be used for showing broader patterns but is not suitable for practical applications requiring a disaggregated analysis such as sovereign risk modelling. Overall, linear global sequential logit could be at the sweet spot between imposing structure and allowing for non-linear effects, as it allows both for robust estimations with limited risk of overfitting and for economically appealing intuitive results.

The analysed sequential logit methods account for non-linearities and can be easily adapted to a range of purposes. In general, C-ratings deserve specific attention as regards the extent to which they should be included in the analysis, as models typically have difficulties explaining these ratings due to their small number and specific characteristics. In this respect, it should also be recalled that CRAs often treat these cases differently, and that for these close-to-default countries another type of economic indicators need to be assessed – which in itself is an example of a non-linearity. The performance of all models can be improved further by exploring other variables (*e.g.* to gauge financial sector stress and the expected economic performance), which could capture some of the cases where the model is overly optimistic and make the differences between predictions and actual observations more symmetric. The sequential logit models can then be used as a starting point to understand the importance of economic variables on sovereign creditworthiness, to assign ratings to countries, or to understand the evolution of a country’s creditworthiness over time.

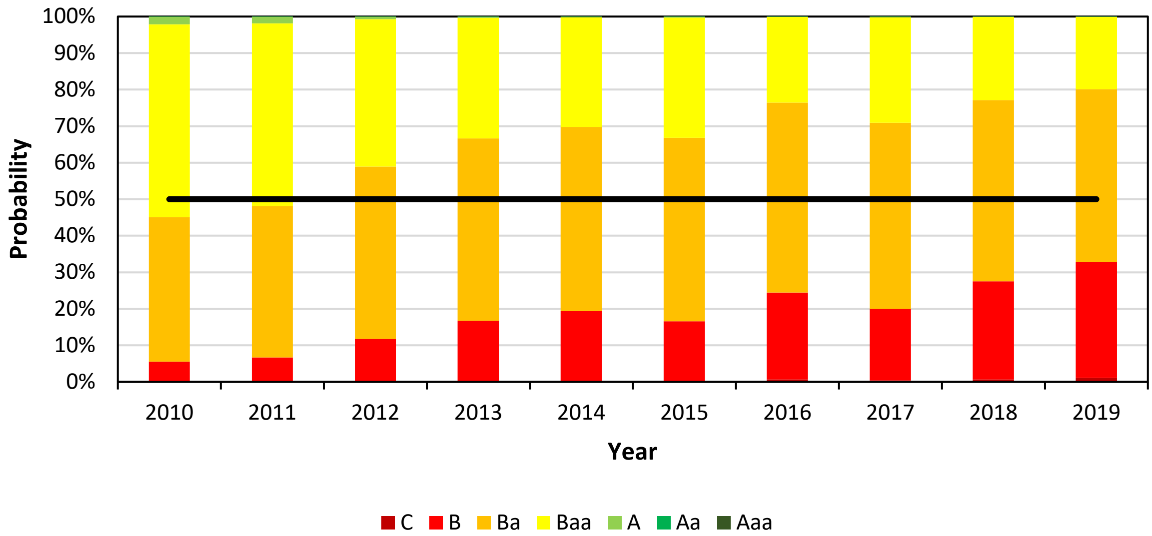
Bibliography

Afonso, A., P. Gomes & P. Rother (2009), Ordered response models for sovereign debt ratings , Applied Economics Letters, 16:8, 769-773.
Broto, C. & L. Molina (2014), Sovereign Ratings and their Assymetric Response to Fundamentals, Banco de España Documentos de Trabajo no. 1428.
D'Agostino, a. & R. A. Lennkh (2016), Euro Area Sovereign Ratings: An Analysis of Fundamental Criteria and Subjective Judgement, ESM Working Paper 2016 (14). European Stability Mechanism.
European Banking Authority (2016), Final Draft Regulatory Technical Standards on the Assessment Methodology for IRB.
Fitch Ratings (2021a), Rating Definitions.
Fitch Ratings (2021b), Sovereign Rating Criteria.
Haspolat, F.B. (2015), Analysis of Moody's Sovereign Credit Ratings: Criticisms Towards Rating Agencies Are Still Valid?, Procedia Economics and Finance 30 (2015), pp. 283-293.
Hastie, T, R. Tibshirani and J. Friedman (2017), "The Elements of Statistical Learning", Springer.
Hadzi-Vazkov, M. & L.A. Ricci (2019), The Nonlinear Relationship Between Public Debt and Sovereign Credit Ratings, IMF Working Paper 2019/19/162.
Kreamer, M. (2021), "How do you predict a sovereign default", https://medium.com/countryrisk-io/how-do-you-predict-a-sovereign-default-df7ea3f9d73.
Moody's Investor Service (2019), Sovereign Ratings Methodology.
Moody's Investor Service (2021a), Rating Symbols and Definitions.
Moody's Investor Serivce (2021b), Sovereign Default and Recovery Rates, 1983-2020.
S&P Global Ratings (2017), Sovereign Rating Methodology.
S&P Global Ratings (2021), Global Ratings Definitions.
Reusens, P. & C. Croux (2017), Sovereign credit rating determinants: a comparison before and after the European debt crisis, Journal of Banking & Finance, Vol. 77. 108-121.
Teker, D., A. Pala & O. Kent (2013), Determination of Sovereign Rating: Factor Based Ordered Probit Models for Panel Data Analysis Modelling Framework, International Journal of Economics and Financial Issues, Vol. 3, No. 1, 2013, pp. 122-132.
Tran, Y., H. Vu, P. Klusak, M. Kramer & T. Hoang (2021), Leading from behind: Sovereign credit ratings during the COVID-19 pandemic. Bennett Institute Working Paper.

Appendix A: Assigning ratings based on the cumulative distribution function

The probabilistic nature of the logic regression can be exploited to understand changes in the creditworthiness of a country over time. For example, consider South Africa (see Figure 14). In 2010, the Baa-band has the highest associated probability. However, in 2011, the probability mass starts to shift down the rating scale, and in 2012 the B-band has a higher probability than the Baa-band. Between 2013 and 2017 the picture is relatively unchanged, but in 2018 and 2019 the B-band has decisively surpassed the Baa-band in likelihood, indicating that the creditworthiness is at the lower end of the Ba-band.

Figure 14: The probabilities of all rating bands over time for South Africa (2010-2019)



Intuitively, Figure 14 suggests to look at the rating band associated with the midpoint of the probability mass, *i.e.* with the band associated with a cumulative probability of 50%.[22] This band is not necessarily the one with the highest probability, but by construction is close.[23] The cumulative probability could thus be used to assign ratings and even to further differentiate within rating bands. This can be done either through inspection, or through a more formalised approach. The simplest way, although assuming some linearity, is dividing the probability of each rating band in three equal parts. In this way, the 50% cumulative probability rule would assign a unique rating, while only six of the 16 borders have to be estimated.

The performance of assigning bands and ratings according to the cumulative distribution function is very similar (see Table 17). For the sequential logits and multinomial logit, assigning the ratings based on the cumulative probability can yield bands that are different from before, but the impact is minor. To derive individual ratings for OLS and ordered logit, the area between two adjacent borders is divided into three equal parts. The performance of these models based on estimating six borders

[22] It would be appealing to also consider the expected rating using the probabilities as weights, but this would require assigning a number to each rating - which would be contrary to the approach followed throughout.

[23] If the two approaches were equivalent, for a country in, say, the Baa-band the midpoint would indicate a Baa, and for a country in the Ba-band, the midpoint would indicate a Ba. It follows from the continuity of the probabilities that for a country on the border, the midpoint would exactly be at the Ba-Baa3 threshold as well. However, the various logit estimations are fully independent, and this condition will in general not hold. It is easy to see that it does not hold at the Aaa/Aa-border, as the Aaa- and Aa-bands have an equal probability that, due to the non-zero probabilities of the other ratings bands, is smaller than 0.5. Hence, the midpoint is not at the Aaa/Aa-threshold. However, the structure of the sample and the exponential term in the logit estimation ensure that only a few, subsequent rating bands have probabilities that are significantly different from zero, and that in case of three or more of those bands, the midpoint is likely to assign an interior one.

instead of 16 improves, especially for OLS. For sequential global logit the performance is somewhat worse, indicating that assuming linearity within a rating band is not very constraining. The considerably worse performance of sequential local logit suggest that the direct estimation of ratings is overfitting the data, as confirmed by visual inspection (see Appendix E). Finally, the performance of multinomial logit improves drastically, indicating that dividing rating bands linearly is better than estimating all borders between ratings. Apart from the multinomial logit, the performance when including one or two neighbouring bands is comparable to that of the direct estimation at the rating-level. Overall, imposing structure at the band level and assuming equally-sized ratings within bands, only has a minor cost in terms of performance.

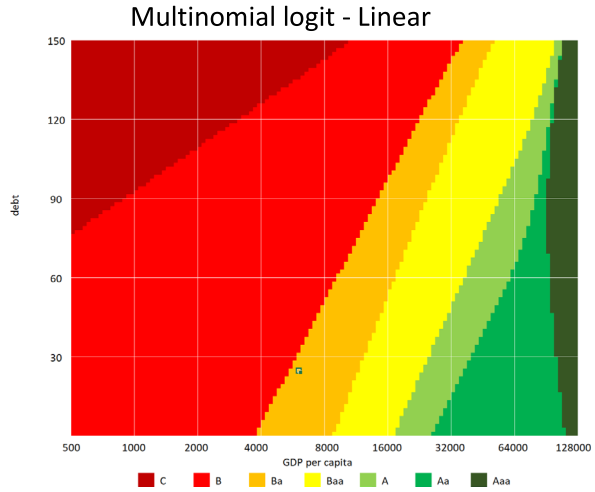
Table 17: Assigning bands and ratings according to the cumulative distribution

	Approach	OLS	Ordered logit	Sequential global logit	Sequential local logit	Multinomial logit
Bands						
Hits	Direct estimation (bands)	51.9%	55.1%	62.9%	64.9%	64.0%
	Cumulative probability (bands)			62.7%	64.2%	63.6%
Hits +/- 1 band	Direct estimation (bands)	93.9%	93.2%	94.4%	92.4%	94.0%
	Cumulative probability (bands)			94.7%	93.8%	94.7%
>=3 bands away	Direct estimation (bands)	1.0%	1.0%	0.7%	1.8%	1.0%
	Cumulative probability (bands)			0.6%	1.4%	0.7%
Ratings						
Hits	Direct estimation (ratings)	20.6%	27.2%	33.8%	38.9%	4.6%
	Direct estimation (bands) to ratings	22.5%	27.4%			
	Cumulative probability (bands) to ratings			31.5%	32.7%	33.2%
Hits +/- 1 rating	Direct estimation (ratings)	52.8%	59.1%	68.2%	71.3%	8.4%
	Direct estimation (bands) to ratings	53.6%	59.0%			
	Cumulative probability (bands) to ratings			66.7%	67.6%	67.4%
Hits +/- 2 ratings	Direct estimation (ratings)	77.2%	79.2%	84.5%	84.7%	10.6%
	Direct estimation (bands) to ratings	77.2%	79.1%			
	Cumulative probability (bands) to ratings			84.4%	83.7%	85.1%

Note: Between parentheses is indicated whether the estimation is at the band or the rating level. For OLS and ordered logit, the conversion from bands to rating is done by dividing each interior band obtained from the direct estimation into three equal parts, which then yield the borders for the individual ratings. For the other estimation methods, the rating or band associated with a cumulative probability mass of 50% is assignied, with the probability of each interior band divided into tree equal parts. For the maximum likelihood methods the performance for bands comes from Table 10. For ratings, the estimations include the free-floating dummy and reserves, and hence the performance differs from that reported in Table 14.

Allocating the bands based on the cumulative probability function has the major advantage that borders do not cross, which follows directly from the continuity of the probabilities in the economic variables. It thus follows that even if the estimation is linear, the borders are no longer linear (see Figure 15). Comparison with the earlier found allocation based on the band with the maximfum likelihood (see Figure 8 panel E), the impact is most pronounced for higher debt levels where the borders are closer to each other. Similarly, while jumps occurred frequently when borders between individual ratings were estimated individually (see Figure 12), imposing structure within bands rules out jumps completely.

Figure 15: The relation between debt, GDP per capita and ratings when using the cumulative probability to assign ratings (regressed control variables)



Appendix B: Forward looking variables

As ratings are considering the creditworthiness over a period of 3-5 years, a country's economic outlook is likely to be relevant. This is tested for the Ba/Baa-border by looking 1, 3 and 5 years ahead (see Table 18). Regressions include forecasts for stock variables and and averages for flow variables (the free-floating dummy, governance and ln(GDP p.c.) are kept unchanged as changes are difficult to anticipate or already covered by growth expectations). The signs and statistical significance are as for the contemporaneous estimation (estimations (1) and (2) of Table 9). The coefficients of debt are smaller, and those of ln(GDP p.c.) larger. The most notable difference is for the coefficient of growth, perhaps reflecting that this is the most important variable when assessing the economic outlook. When considering the performance, all forward-looking models have a higher share of correct predictions than the contemporaneous model (87.2% and 87.0% for the local and global logit respectively), although the difference for the full sample is small. The estimation for one year ahead is performing best, suggesting that the uncertainty of forecasts further out is increasing.

Table 18: Regressions for the border between speculative and investment grade (centred control variables)

	(1) Local logit centred		(2) Global logit centred		(3) Local logit centred		(4) Global logit centred		(5) Local logit centred		(6) Global logit centred	
Intercept	-0.901	***	-1.425	***	-1.763	***	-2.463	***	-2.324	***	-3.068	***
	(0.33)		(0.284)		(0.392)		(0.335)		(0.417)		(0.35)	
Debt(t+1)	-0.022	***	-0.032	***								
	(0.004)		(0.003)									
Debt(t+3),					-0.018	***	-0.028	***				
					(0.004)		(0.003)					
Debt(t+5)									-0.016	***	-0.025	***
									(0.004)		(0.003)	
Interest(t+1)	-0.002		0.002									
	(0.013)		(0.011)									
Interest(t+1:t+3)					-0.014		-0.006					
					(0.013)		(0.011)					
Interest(t+1:t+5)									-0.024	*	-0.016	
									(0.013)		(0.011)	
Primary(t+1)	-0.086	***	-0.104	***								
	(0.028)		(0.023)									
Primary(t+1:t+3)					-0.138	***	-0.164	***				
					(0.033)		(0.026)					
Primary(t+1:t+5)									-0.179	***	-0.202	***
									(0.037)		(0.027)	
Growth(t+1)	0.366	***	0.494	***								
	(0.048)		(0.044)									
Growth(t+1:t+3)					0.514	***	0.687	***				
					(0.062)		(0.056)					
Growth(t+1:t+5)									0.594	***	0.783	***
									(0.069)		(0.061)	
Current Account(t+1)	0.046	***	0.072	***								
	(0.015)		(0.013)									
Current Account(t+1:t+3)					0.064	***	0.097	***				
					(0.017)		(0.015)					
Current Account(t+1:t+5)									0.071	***	0.111	***
									(0.019)		(0.016)	

Floating	2.209	***	3.107	***	2.311	***	3.116	***	2.482	***	3.19	***
	(0.322)		(0.28)		(0.327)		(0.283)		(0.331)		(0.285)	
Reserves(t+1)	0.179	***	0.233	***								
	(0.023)		(0.02)									
Reserves(t+1:t+3)					0.223	***	0.265	***				
					(0.026)		(0.023)					
Reserves(t+1:t+5)									0.28	***	0.313	***
									(0.03)		(0.025)	
Governance	1.434	***	1.426	***	1.392	***	1.419	***	1.348	***	1.402	***
	(0.196)		(0.118)		(0.196)		(0.12)		(0.196)		(0.121)	
ln(GDP p.c.)	1.383	***	2.46	***	1.713	***	2.815	***	1.956	***	3.038	***
	(0.177)		(0.144)		(0.193)		(0.159)		(0.203)		(0.168)	
% correct adjacent bands	75.1%		74.6%		74.1%		74.2%		74.0%		73.8%	
% correct all bands	88.4%		88.5%		87.6%		88.4%		87.5%		88.4%	

Note: (t+1) indicates the one-year ahead value; (t+1:t+3) the average over the one-to-three years ahead, and (t+1,t+5) the average over the one-to-five years ahead.

Appendix C: Public finances

Table 19: Performance of regressions focussed on public finances

	OLS		Sequential global logit		Sequential local logit		Multinomial logit	
	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic
Hits	35.4%	44.7%	48.2%	58.0%	49.3%	56.2%	50.3%	59.9%
Hits +/- 1 band	84.1%	89.5%	89.4%	92.0%	86.3%	85.8%	87.4%	91.3%
>= 3 bands away	1.9%	0.9%	1.0%	1.3%	1.9%	5.3%	1.7%	2.0%
Too high	32.4%	26.6%	28.0%	21.6%	28.1%	25.5%	28.7%	22.4%
Too low	32.2%	28.7%	23.8%	20.4%	22.6%	18.3%	21.1%	17.6%
Standard error	1.11	0.96	0.94	0.85	1.01	1.14	0.98	0.88
Correctly assigned by bounds								
Adjacent bands	65.1%	70.1%	72.0%	77.3%	72.6%	76.3%	73.1%	78.4%
All bands	86.3%	88.9%	89.4%	91.5%	88.9%	89.1%	89.3%	91.5%
Hits by ratings								
C	20.4%	20.4%	11.1%	28.7%	2.8%	18.5%	2.8%	24.1%
B	35.6%	50.2%	66.2%	68.3%	73.2%	59.7%	75.0%	70.1%
Ba	40.5%	54.2%	41.0%	50.7%	24.5%	39.6%	20.1%	39.4%
Baa	36.6%	41.7%	44.8%	46.3%	53.7%	55.4%	53.1%	59.1%
A	65.6%	53.2%	39.2%	47.1%	28.3%	63.5%	38.4%	58.2%
Aa	24.2%	39.2%	40.2%	59.2%	25.5%	46.7%	20.3%	47.4%
Aaa	8.1%	25.9%	51.4%	85.4%	80.0%	86.5%	86.5%	92.4%
Observations	3215	3215	3215	3215	3215	3215	3215	3215

Note: The underlying regressions include a constant, debt, interest, primary balance and ln(GDP p.c.), and in case of quadratic regressions their quadratic terms and their interaction terms.

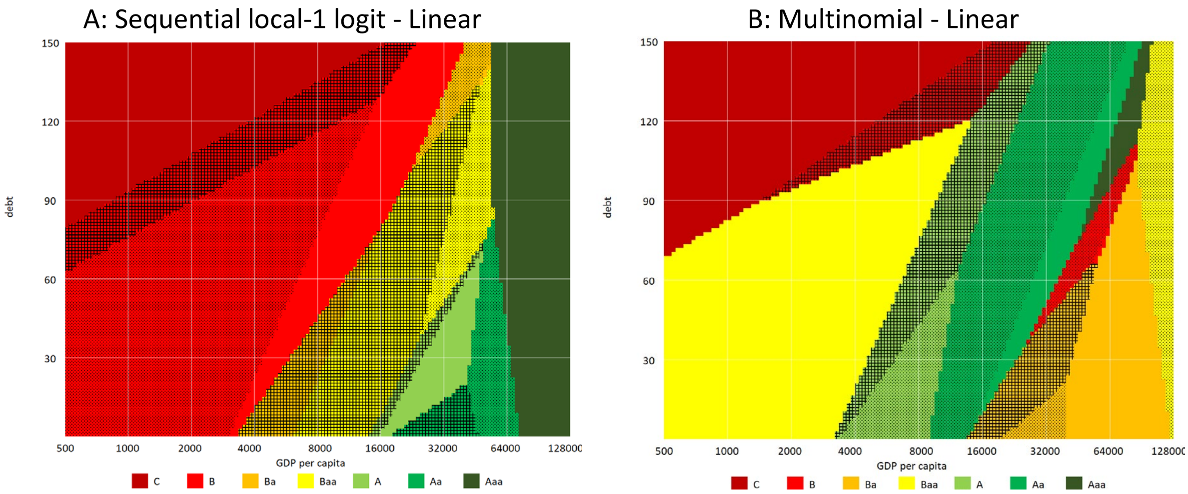
Appendix D: Quadratic regression

Table 20: Sequential global logit regression (regressed control variables)

	C/B border	B/Ba border	Ba/Baa border	Baa/A border	A/Aa border	Aa/Aaa border
Intercept	5.809 *** (0.390)	2.404 *** (0.177)	-0.246 * (0.131)	-4.46 *** (0.277)	-7.012 *** (0.508)	-48.673 *** (7.808)
Debt	-0.044 *** (0.008)	-0.044 *** (0.004)	-0.027 *** (0.004)	-0.06 *** (0.010)	-0.141 *** (0.022)	0.589 *** (0.092)
Debt^2	0.000 (0.000)	0.000 (0.000)	0.000 * (0.000)	-0.001 *** (0.000)	-0.001 *** (0.000)	-0.002 *** (0.001)
Interest	0.018 (0.042)	0.047 ** (0.020)	0.177 *** (0.025)	-0.066 (0.043)	-0.568 *** (0.145)	6.270 *** (0.787)
Interest^2	0.005 ** (0.002)	-0.007 *** (0.001)	-0.021 *** (0.003)	0.004 (0.004)	-0.053 ** (0.025)	-0.306 *** (0.045)
Primary	-0.029 (0.035)	-0.039 * (0.022)	-0.061 ** (0.029)	-0.345 *** (0.058)	-0.203 ** (0.083)	-1.537 *** (0.34)
Primary^2	-0.001 (0.003)	-0.003 (0.003)	-0.006 * (0.003)	-0.011 *** (0.002)	-0.002 (0.002)	-0.029 *** (0.010)
Growth	0.439 *** (0.053)	0.237 *** (0.032)	0.249 *** (0.032)	0.424 *** (0.047)	0.081 (0.056)	0.056 (0.162)
CA	0.011 (0.022)	0.067 *** (0.012)	0.053 *** (0.014)	0.052 *** (0.018)	-0.060 *** (0.021)	0.226 *** (0.073)
Free-floating	4.196 *** (1.043)	4.113 *** (0.619)	3.490 *** (0.332)	4.944 *** (0.347)	0.344 (0.290)	-0.283 (1.370)
Reserves	0.278 *** (0.064)	0.312 *** (0.029)	0.287 *** (0.022)	0.327 *** (0.031)	0.090 *** (0.024)	-0.196 (0.194)
Governance	1.394 *** (0.284)	1.935 *** (0.152)	1.531 *** (0.146)	0.798 *** (0.130)	0.600 *** (0.155)	8.169 *** (1.291)
ln(GDPpc)	1.446 *** (0.287)	3.490 *** (0.173)	3.924 *** (0.166)	5.179 *** (0.407)	3.887 *** (0.534)	59.707 *** (10.64)
ln(GDPpc)^2	-0.345 *** (0.130)	-0.064 (0.124)	0.643 *** (0.143)	1.235 *** (0.283)	1.407 *** (0.195)	-19.924 *** (3.848)
Debt x Interest	0.000 (0.001)	-0.001 (0.001)	-0.001 ** (0.001)	-0.001 (0.002)	-0.016 *** (0.005)	-0.048 *** (0.009)
Debt x Primary	0.002 * (0.001)	-0.001 * (0.001)	0.000 (0.001)	-0.001 (0.001)	-0.008 *** (0.001)	0.000 (0.005)
Debt x ln(GDPpc)	0.013 ** (0.005)	-0.012 ** (0.005)	-0.043 *** (0.005)	-0.019 * (0.011)	0.038 ** (0.016)	-0.522 *** (0.075)
Interest x Primary	0.014 *** (0.005)	0.003 (0.004)	0.002 (0.006)	0.013 (0.010)	-0.023 * (0.013)	0.095 ** (0.045)
Interest x ln(GDPpc)	0.064 ** (0.032)	-0.042 ** (0.018)	-0.183 *** (0.026)	-0.096 * (0.054)	-0.338 *** (0.105)	-5.73 *** (0.707)
Primary x ln(GDPpc)	0.014 (0.042)	-0.012 (0.033)	0.030 (0.04)	0.218 *** (0.049)	-0.003 (0.056)	0.984 *** (0.247)

Appendix E: Analysis with ratings and no imposed structure

Figure 16: The relation between debt, GDP per capita and ratings (regressed control variables)



ECONOMICS – WORKING PAPERS 2022/05

How much is too much?

Assessing the non-linear relationship between debt and sovereign creditworthiness

European Investment Bank
The EU bank

Economics Department
economics@eib.org
www.eib.org/economics

European Investment Bank
98 -100, boulevard Konrad Adenauer
L-2950 Luxembourg
+352 4379-22000
www.eib.org – info@eib.org

PDF: ISBN 978-92-861-5236-8